Manual d'ús de *jamovi* per **anàlisi de dades** en estudis socials

MUJADES v1.6

2017-2025

AJENJO COSP, Marc (Dep. de Sociologia, UAB)

MIGUEL QUESADA, F.J. (Dep. de Sociologia, UAB)

Amb la col·laboració de:

SÁNCHEZ GUERRERO, Laia

GRIERA ARTIGUES, Oriol

Aquest material docent es publica sota llicència *Creative Commons*, modalitat "Attribution-NonCommercial 4.0 International". Qualsevol persona és lliure de copiar i redistribuir el material en qualsevol mitjà i format, remesclar, transformar i crear a partir del material, d'acord amb les següents condicions:

Reconeixement — Heu de reconèixer l'autoria de manera apropiada, proporcionar un enllaç a la llicència i indicar si heu fet algun canvi. Podeu fer-ho de qualsevol manera raonable, però no d'una manera que suggereixi que el llicenciador us dona suport o patrocina l'ús que en feu.

NoComercial — No podeu utilitzar el material per a finalitats comercials.

No hi ha cap restricció addicional. No podeu aplicar termes legals ni mesures tecnològiques que restringeixin legalment a altres de fer qualsevol cosa que la llicència permet.

Agraiments: Al Departament de Sociologia per promoure la dedicació a aquest projecte. A Agustí Bosch per les qüestions que ens ajuden a millorar el manual, i per ser a l'origen de tot plegat.

Índex de continguts

Índex de continguts 3
1. *jamovi*: Tractament i anàlisi de dades quantitatives 7
2. Instal·lació del programari jamovi 9
Idioma de jamovi 10
Mòduls d'extensió de jamovi 11
Com afegir el mòdul per a operar directament amb codi R? 11
3. Interfície gràfica de jamovi 15
Menús de les modalitats de treball 16
1. Menú d'arxius 16
2. Modalitat de gestió de variables 18
3. Modalitat de gestió de dades 19
4. Modalitat d'anàlisi de dades 21
5. Modalitat de comentaris al text 22
6. Activador de pantalla central 23
7. Menú de configuració de jamovi 23
4. Lectura i escriptura d'arxius de dades amb jamovi 25
Introducció directa de dades 25
Introducció des de taules de dades 28
Enganxar dades des d'altres aplicacions 32
Càrrega de les dades d'arxius d'altres aplicacions 34
EXERCICIS PROPOSATS: Lectura i escriptura d'arxius de dades 37
5. Definició de dades: Tipus de variables i valors perduts 38
Correspondència entre nivell de mesura de les variables i tipus de variable/dada en jamovi 38
Les "no respostes" i la definició de valors perduts en jamovi 43
Definició de valors perduts en jamovi (I): Variables nominals i ordinals 44
Definició de valors perduts en jamovi (II): Variables quantitatives 47
EXERCICIS PROPOSATS: Definició de variables i valors perduts 49
6. Estadística descriptiva i exploratòria bàsica amb jamovi 51
Exploració bàsica descriptiva de les variables 51
Exploració de variables amb codi R 54
Generar taules de freqüències amb codi R 55
Estadística descriptiva de variables quantitatives, amb codi R 58
Resultats de l'exploració de variables: Afegir comentaris i exportar 60
Modalitat comentari textual 60
Exportació de resultats cap a altres aplicacions 61
Descripció gràfica de variables 63
Codi R per a generar representacions gràfiques 67
EXERCICIS PROPOSATS: Estadística descriptiva i exploratòria bàsica 68
7. Transformació i depuració de dades mitjançant jamovi 69
Recodificació de variables 69

Recodificació (I). De variables quantitatives a qualitatives 70
Recodificació (II). De variables qualitatives a qualitatives 77
Recodificació (III). De variables qualitatives a quantitatives 79
Recodificació (IV). De variables quantitatives a quantitatives....81
Càlculs amb variables quantitatives 81
Operacions i càlculs utilitzant una única variable 82
Operacions i càlculs amb diferents variables....84
Càlculs amb variables qualitatives.... 86
Operacions (càlculs) amb variables qualitatives multiresposta....87
Transformacions complexes 89
(1) Combinació de recodificacions i càlculs 89
(2) Construcció de tipologies amb la funció IF....93
Transformacions amb codi R 99
Filtrar casos.... 104
Detecció i depuració d'errors 107
Errors en variables individuals 107
Errors en la congruència entre variables 109
EXERCICIS PROPOSATS: Transformació de dades.... 110
8. Inferència amb una variable utilitzant *jamovi* 114
Construcció d'un interval de confiança per a una proporció (π). 114
Construcció d'un interval de confiança per a la mitjana (μ). 116
Codi R per a la construcció d'intervals de confiança (mitjana)....117
EXERCICIS PROPOSATS: Intervals de confiança 118
9. Inferència de relació entre variables 120
10. Anàlisi de taules de contingència mitjançant jamovi....122
Estadística descriptiva en les taules de contingència.... 123
Taules de contingència descriptives amb R 126
La representació gràfica de les taules de contingència.... 128
Gràfiques de taules de contingència, amb R 129
La inferència estadística amb dues variables no quantitatives: El test de la khi quadrat de Pearson. 132
A) Les condicions d'aplicació en el test de la khi quadrat de Pearson....133
B) Test de significació estadística khi quadrat de Pearson 134
C) Intensitat de l'associació entre variables nominals (V de Cramer). 135
D) Avaluació del sentit de l'associació 136
Solució als problemes d'incompliment de les condicions d'aplicació.... 141
La taula de contingència 2*2 (la correcció de Yates). 142
EXERCICIS PROPOSATS: Taules de contingència 144
11. Comparació de mitjanes amb jamovi....146
Introducció als tests de comparació de mitjanes. Quin test cal fer?....147
L'estadística descriptiva en la comparació de mitjanes i els intervals de confiança.... 149
El test de Shapiro-Wilk per a la Normalitat de la variable. 152

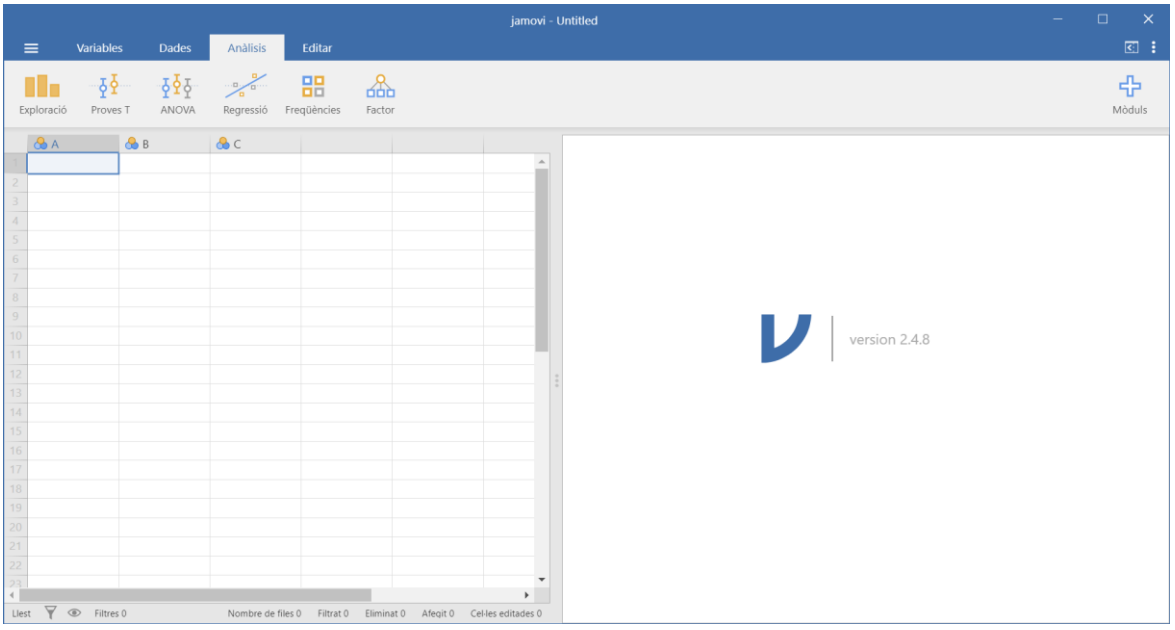
La comparació de dues mitjanes independents. 155
La comparació de més de dues mitjanes independents: l'anàlisi de la variància (ANOVA)... 161
La intensitat de la relació: El coeficient de determinació. 165
La prova no paramètrica de comparació de més de dos grups: La prova de Kruskal-Wallis . 167
EXERCICIS PROPOSATS: Comparació de mitjanes 167
12. Regressió lineal simple mitjançant jamovi 169
Descripció de la relació entre dues variables quantitatives 169
El diagrama de dispersió 170
La correlació entre dues variables 174
El model de regressió lineal simple 175
Inferència estadística en els models de regressió 178
Els supòsits del model: L'anàlisi de residuals 179
EXERCICIS PROPOSATS: Regressió lineal simple 185

Aquesta versió del tutorial utilitza imatges de pantalla de diverses versions de *jamovi*, fins a la actual 2.6.24 (2025)

1. *jamovi*: Tractament i anàlisi de dades quantitatives

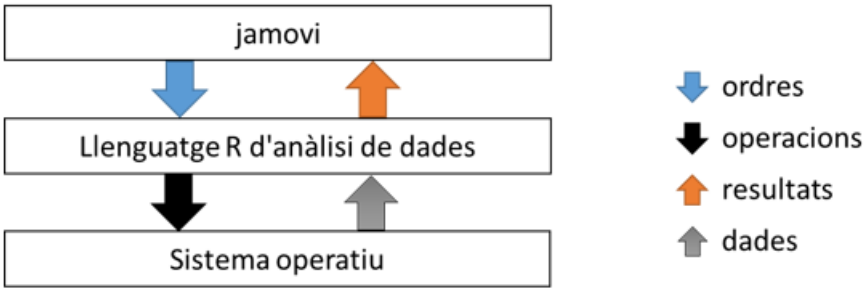
jamovi[1] és un sistema de tractament de dades i anàlisi estadística que proporciona totes les funcionalitats bàsiques requerides per a la recerca social amb mètodes quantitatius. Es tracta de software obert, lliure, multi-plataforma i és ampliable mitjançant "mòduls" que diversos programadors posen a l'abast a Internet. Està basat en el llenguatge R, i és una bona forma d'introduir-se a l'ús d'aquest potent llenguatge d'anàlisi de dades.

La instal·lació i ús de *jamovi* són gratuïts i permeten realitzar la gestió de les dades, les transformacions i els càlculs estadístics corresponents a la major part dels estudis socials típics. La seva pantalla és similar a la d'un "full de càlcul" (matriu de files-columnes, amb les dades corresponents a casos-variables). Disposa de menús i finestres per facilitar el seu ús sense necessitat de "programar codi informàtic", tot i que és simplement una interfície gràfica del llenguatge de programació R, específicament dedicat a l'anàlisi de dades.



Imatge 1. Pantalla principal de jamovi

Quan es tria una opció als seus menús, en realitat s'està executant el codi d'R associat, de forma transparent pels usuaris. Així es pot aprofitar tota la potència analítica del llenguatge R sense la necessitat de conèixer en detall el seu funcionament, programació, instal·lació i manteniment.



Imatge 2. Esquema funcional de jamovi

1 Què vol dir jamovi? "…is not an acronym and is most definitely always lowercase. (…) jamovi means what you feel when you use it. There may very well be more to the name, but sometimes mysteries are fun." (Forum de desenvolupadors de jamovi).

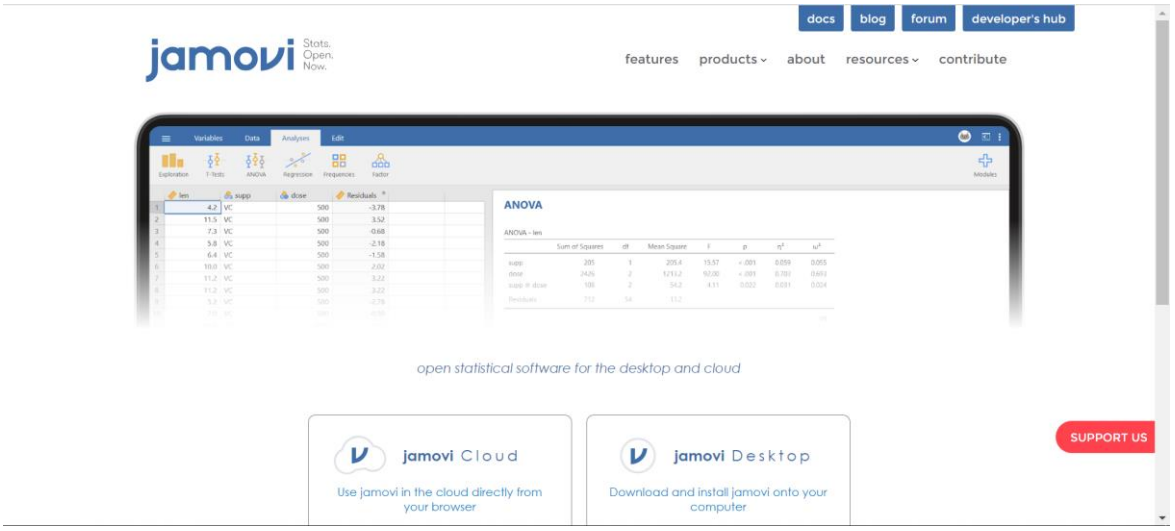
Tanmateix, *jamovi* permet també "programar" **directament** utilitzant en llenguatge R, mitjançant un dels mòduls disponibles per ampliar les capacitats bàsiques del sistema.
Als diferents capítols o tutorials de MUJADES s'introduiran, el primer lloc, les diferents operacions mitjançant el sistema de menús i, a continuació, es mostrarà el codi R per aconseguir el mateix resultat (bé sigui al mòdul Rj de *jamovi*, o directament a R). D'aquesta forma, *jamovi* esdevé una eina per a l'anàlisi de dades quantitatives tant pel seu ús directe com per la seva utilitat en l'aprenentatge de programació amb R.

L'espai web del projecte *jamovi* inclou un apartat molt interessant per a l'autoaprenentatge d'aquest sistema d'anàlisi estadística. Es pot complementar aquest tutorial amb els següents recursos *(gratuïts, disponibles en línia)*:

- el vídeo-curs de 4,5 hores a < https://blog.jamovi.org/2018/08/21/datalab.html>
- el llibre d'Estadística amb *jamovi* a < https://www.learnstatswithjamovi.com/ >

2. Instal·lació del programari jamovi

A la web principal del projecte http://www.jamovi.org, es pot descarregar *jamovi* per a diferents plataformes (versió *"Desktop"* per a Windows, Linux, Mac). També hi ha l'opció d'utilitzar una versió experimental en línia (versió *"Cloud"*), de forma que no cal descarregar-lo ni instal·lar-lo al PC local.[2]



Imatge 3. Website de jamovi (19 setembre 2023)

Per a descarregar l'aplicació, visiteu el lloc web del projecte. A l'apartat **download** hi trobareu diferents opcions de descàrrega.

1. Selecciona la descàrrega corresponent al teu sistema operatiu.[3]
2. Per a alguns sistemes existeixen tres versions disponibles:
   - "current" més actualitzada, inclou novetats i potser alguns errors,
   - "solid" més segura (recomanada), última que no ha reportat errors,
   - "legacy" més antiga, adient per a sistemes i PCs antics (>4 anys).
3. Clica la teva opció. Espera que el teu navegador conclogui el procés de descàrrega.
4. Accepta l'inici del procés d'instal·lació, o ves a la carpeta de descàrregues del teu navegador per a executar el procés d'instal·lació manualment.
5. Segueix les instruccions detallades segons el teu sistema operatiu.

Una vegada instal·lat al teu PC, busca la icona de *jamovi*, i comprova que es pot obrir sense problemes.

2 A data de febrer 2025 la versió en-línia, o "Cloud", està accessible mitjançant un pla de subscripció gratuït amb necessitat de crear un compte d'usuari GUEST, amb la versió més actualitzada, amb funcionalitats retallades i amb capacitat de servei limitat segons les connexions simultànies (es a dir, comportament imprevisible). També hi ha plans de pagament (a partir dels 7,9€ / mes) per millorar la prioritat de la connexió i les funcionalitats, tot i que en cap cas funcionarà l'editor de R en la versió en-línia. Es recomana la instal·lació de versions locals (Desktop) per treballar amb la seguretat i potencia assegurada; aquestes versions instal·lables son gratuïtes.

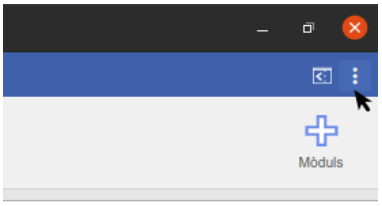
3 Al lloc web de *jamovi* hi ha indicacions detallades per a instal·lar en diferents sistemes operatius <>. La instal·lació a PCs amb Chromebook és la més complicada, i consistent en tres pasos consecutius que cal realitzar en aquest ordre: 1er) Comproba si el teu Chromebook és compatible amb aquesta instal·lació a https://www.reddit.com/r/Crostini/wiki/getstarted/crostini-enabled-devices/?rdt=49126, 2on) Instal·la FLATPACK al teu Chromebook, des de <https://flathub.org/setup/Chrome%20OS>, 3er) Instal·la JAMOVI al teu Chromebook, des de <https://flathub.org/apps/org.jamovi.jamovi>.

(cc) BY-NC

Imatge 4. Icona de *jamovi*

Idioma de jamovi

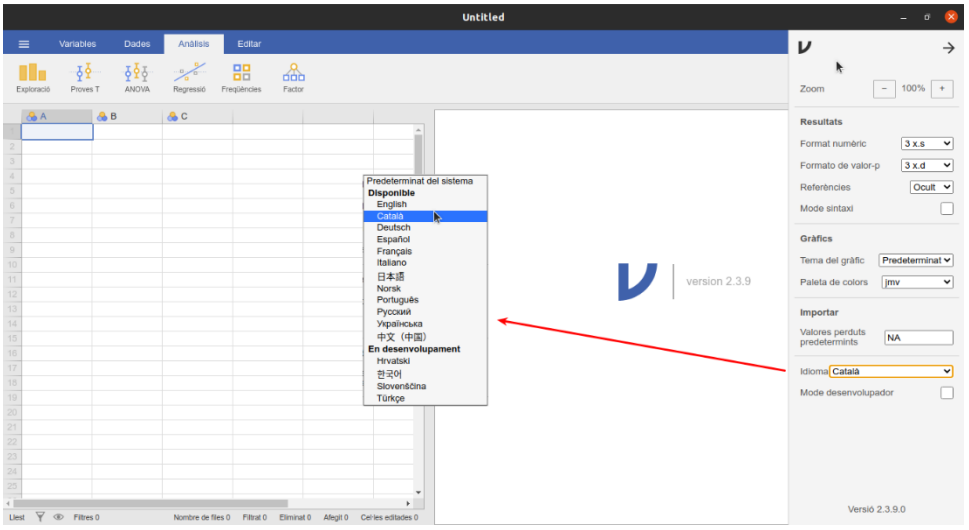
L'idioma que *jamovi* utilitza als menús i als resultats correspon, de forma predeterminada, amb l'utilitzat pel sistema operatiu instal·lat si aquest n'és un dels disponibles. També es pot canviar per altres idiomes proporcionats per la comunitat d'usuaris de la següent forma.



Imatge 5. Accés a les opcions de jamovi

Per canviar l'idioma de jamovi:

1. Localitza a la pantalla principal, a l'extrem dret de la barra blava superior, l'accés a les opcions generals de *jamovi*,
2. Al desplegable, localitza el menú de canvi d'idioma,
3. Selecciona l'idioma,
4. Cal tancar i tornar a obrir jamovi per tal que el canvi d'idioma sigui efectiu,
5. A partir d'ara l'idioma de tots els menús i resultats serà el seleccionat.



Imatge 6. Canvi d'idioma al menú d'opcions de *jamovi*

Per a qualsevol comentari o suggeriment de millora de qüestions relatives a les traduccions en "català" o "español" , podeu contactar amb els autors d'aquest manual. Gràcies als vostres comentaris podrem continuar actualitzant i millorant el manual.

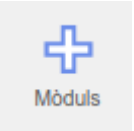
Mòduls d'extensió de jamovi

L'aplicació *jamovi* inclou un conjunt de procediments d'anàlisi estadística inicial, però es pot ampliar amb els anomenats "Mòduls". Com que és dependent del llenguatge R, *jamovi* segueix la mateixa filosofia que aquest sistema:

- incorporat directament a la instal·lació inicial, existeix un conjunt bàsic de procediments que cobreix els aspectes més habituals de l'anàlisi estadística de dades quantitatives, i
- com a mòduls especialitzats, existeixen múltiples procediments avançats desenvolupats per experts que es poden afegir (de manera optativa) al sistema base.

Aquests mòduls són molt diversos, des de petits afegits als procediments bàsics (*més* gràfics, *més* indicadors...), fins a procediments estadístics propis i especialitzats en disciplines concretes (que requereixen del seu propi manual d'instruccions i interpretació).
Els mòduls estan disponibles en línia, i son automàticament actualitzats, a un espai públic propi del projecte *jamovi*. L'accés als mòduls requereix disposar de connexió WWW per procedir a la seva descàrrega per primera vegada, però a partir de llavors, una vegada instal·lats al vostre PC local, ja funcionen sense necessitat de connexió WWW.

La pàgina web <https://www.jamovi.org/library.html> manté un llistat dels mòduls disponibles. També és possible descarregar-los directament (per a diversos sistemes operatius, i des de fonts diverses, inclosa descàrrega directa per part dels autors del mòdul)[4] i llavors fer la "càrrega auxiliar" (*library sideload*) a *jamovi*. Dins de la mateixa aplicació *jamovi* hi ha una opció de menú "Mòduls" que permet la gestió d'aquestes expansions, inclosa la seva descàrrega si es disposa de connexió WWW.



Imatge 7. Accés a la gestió d'extensions

A continuació s'exemplifica aquest procediment d'instal·lació de mòduls des de dins de *jamovi*.

Com afegir el mòdul per a operar directament amb codi R?

Al llarg d'aquest tutorial es mostrarà com realitzar determinades operacions estadístiques directament executant codi en llenguatge R. En algunes ocasions es tractarà d'una forma alternativa al sistema de menús que aconsegueix els mateixos resultats, mentre que en altres s'utilitzarà codi R per aconseguir superar determinades limitacions del sistema bàsic de menús. D'aquesta manera, el tutorial anirà introduint poc a poc l'aprenentatge d'elements bàsics del llenguatge de programació R.

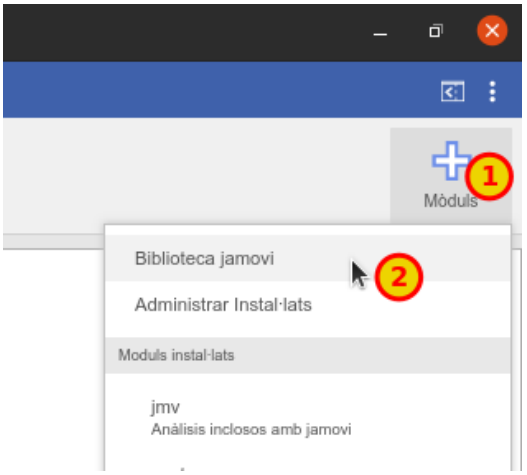
Per a poder executar codi R directament amb *jamovi* (sense utilitzar el sistema de menús) cal tenir instal·lat a l'ordinador el mòdul anomenat "**Rj – Editor to Run R code inside jamov**i".

Per fer-ho, pas a pas:

1. A la pantalla principal de *jamovi*, a la modalitat ANÀLISI DE DADES ("Anàlisis"), al menú gràfic, extrem dret, clica la icona "Mòduls", i l'opció *Biblioteca jamovi*.

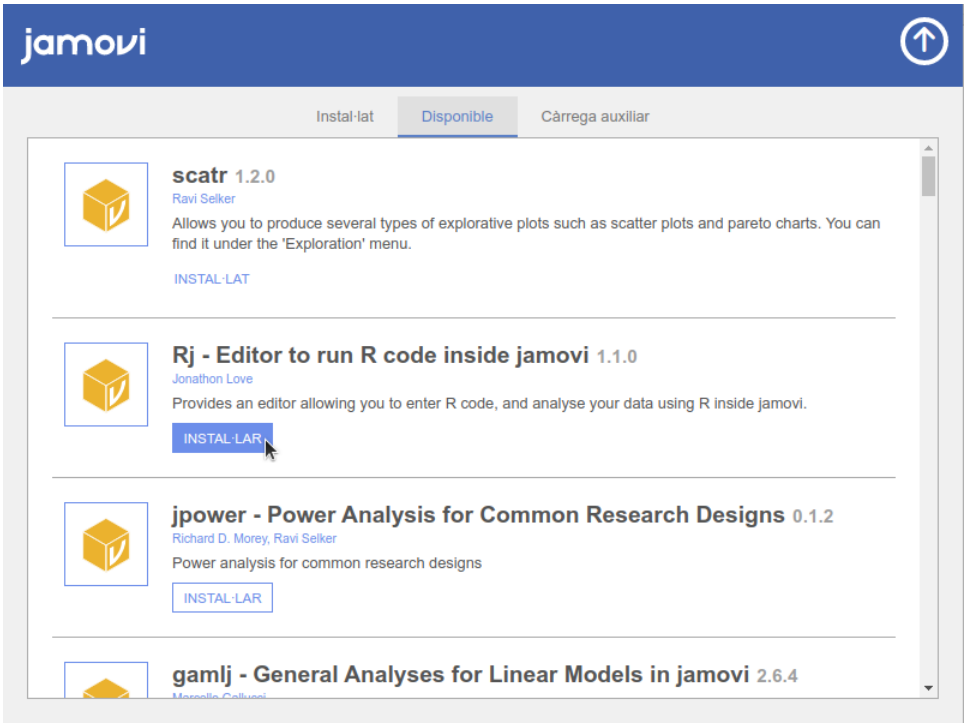
[4] Compte amb aquests mòduls, que no has estat validats pels autors de jamovi i poden generar problemes.

(cc) BY-NC



Imatge 8. Accés al sistema de gestió d'extensions

2. Si disposes de connexió a Internet, el sistema connecta amb el repositori WWW de mòduls i mostra el catàleg de les extensions actualment disponibles, juntament amb una descripció i botons per procedir a la seva instal·lació. Busca el mòdul anomenat **"Rj – Editor to ..."**.



Imatge 9. Catàleg d'extensions (*Biblioteca jamovi*)

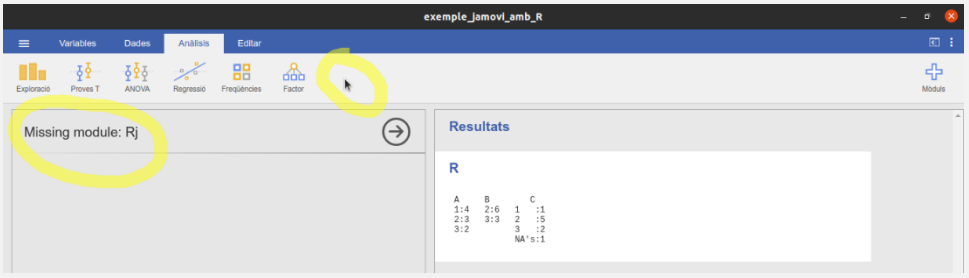
3. Clica el botó [INSTAL·LAR]. Apareix un missatge mentre s'està fent la descàrrega. Quan culmina el procés, el catàleg indica "INSTAL·LAT".
4. Tanca la finestra del catàleg de mòduls (*Biblioteca jamovi*) clicant la icona fletxa-amunt-encerclada de la part superior dreta.
5. Al menú gràfic de *jamovi* ha aparegut una **nova opció**, destacada, que dona accés a la consola per escriure i executar directament codi R amb les dades de la matriu.

Imatge 10. Menú gràfic ampliat amb mòdul *Rj-Edit*

Ara el mòdul corresponent ha quedat instal·lat al teu PC. No caldrà tornar-lo a instal·lar fins que es modifiqui, en versions posteriors de *jamovi*. (El sistema avisa quan això succeeix, i permet fer l'actualització en línia dels mòduls corresponents).

IMPORTANT: Quan s'instal·la un mòdul d'extensió de *jamovi* a un PC concret, ja es pot utilitzar per a executar els procediments inclosos al mòdul. L'arxiu propi de *jamovi* (format .OMV) que conté les dades, els procediments i els resultats d'una anàlisi es pot desar en format digital i "obrir" amb un altre PC que tingui el *jamovi* instal·lat. PERÒ aquest arxiu **generarà un ERROR** si a aquest segon PC **no hi ha instal·lats el MATEIXOS mòduls** d'extensió que s'han utilitzat en el PC original.

En anglès, en cas de missatge d'error o advertiment, cal respondre [Enable].



Imatge 11. Missatge d'error associat a l'ús de codi R directe (mòdul *Rj*)

La solució a aquest problema consisteix en identificar tots els mòduls utilitzats per a generar els resultats d'un arxiu OMV i assegurar-se que estan instal·lats al PC on es vol "obrir" l'arxiu.

Per a gestionar els diferents mòduls que hi ha instal·lats a un PC:

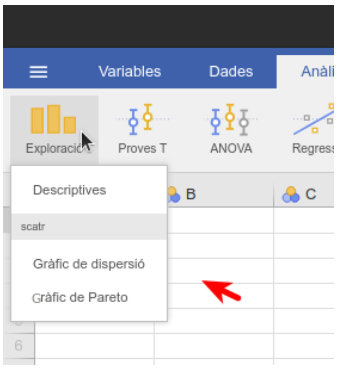
1. A la pantalla principal de *jamovi*, al menú gràfic superior, extrem dret, clica la icona "Mòduls".
2. Al desplegable selecciona l'opció "Administrar Instal·lats". Apareix un catàleg de les extensions instal·lades localment al sistema, amb possibilitat d'amagar-les del menú [OCULTAR], i de desinstal·lar-les definitivament [ELIMINAR].

Pots comprovar que hi ha una extensió impossible d'eliminar... **"jmv – Anàlisis bundled with jamovi"**. Es tracta del sistema de menús **principal** de *jamovi*, amb totes les anàlisis que formen part de la instal·lació bàsica original (es pot amagar, però continuarà sempre operativa).

EXERCICI

1. A la pantalla principal de *jamovi*, al menú gràfic superior, extrem esquerra, clica la icona "Exploració". Comprova que l'única opció disponible és "Descriptives".

2. Utilitza el procediment explicat anteriorment per a instal·lar el mòdul d'extensió anomenat **"scatr"**.
3. De quines opcions disposa ara el menú "Exploració" ?



Imatge 12. Menú de Descriptiva, ampliat amb mòdul "sctr"

ATENCIÓ: Aquestes dues noves opcions s'utilitzaran al llarg del tutorial, no continuïs fins a tenir-les disponibles.

Tingues en compte que si estàs utilitzant un PC aliè amb *jamovi* instal·lat (p. ex., aules d'informàtica, o accés remot) és possible que:

A) trobis que alguns mòduls d'extensió ja estiguin prèviament disponibles,

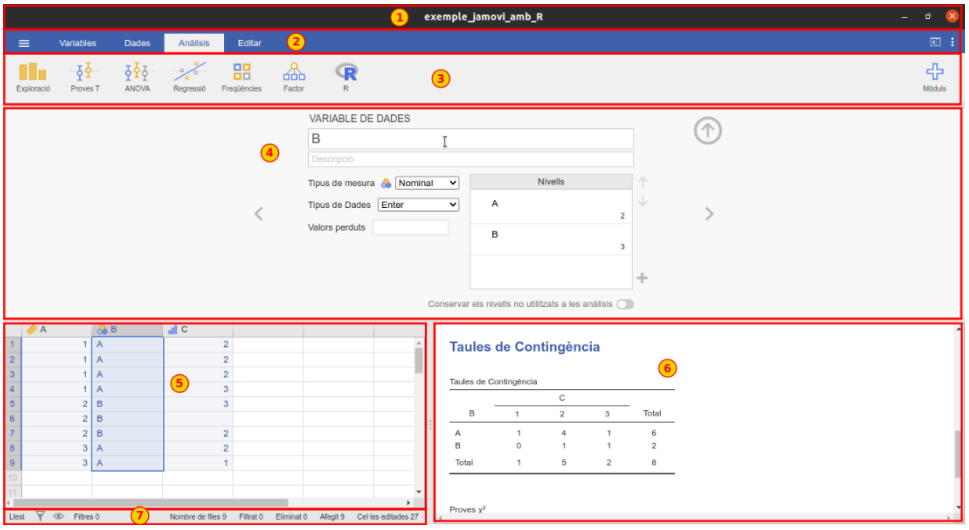
B) si instal·les nous mòduls al sistema "desapareguin" en reiniciar la màquina (per les necessàries polítiques de seguretat dels PCs públics), i

C) alguns mòduls instal·lats en versions prèvies de jamovi no estiguin operatius en versions posteriors o actuals. En accedir al menú d'instal·lació de mòduls, estaran senyalats amb una nota en vermell indicant que CAL RE-INSTAL·LAR LA NOVA VERSIÓ DEL MÒDUL.

3. Interfície gràfica de jamovi

L'aplicació *jamovi* no és més que una completa interfície gràfica d'usuari (GUI) per executar operacions en el llenguatge de programació R. Així doncs, no cal conèixer a fons aquest potent llenguatge especialitzat en l'anàlisi de dades per poder utilitzar-lo. Com a tal, el sistema de menús de *jamovi* inclou una part (mínima) de totes les potencialitats d'R, però ho fa d'una forma molt amigable i amb màxima facilitat d'ús sense haver d'aprendre programació R.
En aquesta secció es presenten sistemàticament els diferents elements d'aquesta interfície gràfica que es mostren a la pantalla de treball.

La pantalla principal està dividida en 7 àrees específiques, i les centrals mostren diversos continguts en funció de la modalitat de treball en què es troba l'aplicació.



Imatge 13. Àrees específiques a la pantalla principal

1. Finestra general de l'aplicació (amb l'identificador de l'arxiu de treball actiu i, segons el sistema operatiu, icones de gestió de l'aplicació dins del PC: tancar, minimitzar, etc..).
2. Menú textual general (amb les cinc modalitats de funcionament: arxius externs, gestió de variables, gestió de dades, operacions estadístiques i comentaris als resultats).
3. Menú gràfic específic (amb les icones corresponents a cada modalitat de funcionament).
4. Finestra de definició de variables (habitualment oculta).
5. Finestra de dades (organitzades en casos a les files i variables a les columnes) i de configuració de les anàlisis (depèn de la modalitat activa).
6. Finestra de resultats (amb les taules, gràfics i textos generats).
7. Barra d'informació i sistema de filtratge de casos.

La finestra de definició de variables (4) es mostra o s'amaga segons la modalitat de funcionament, i també fent doble-clic a sobre del nom a l'encapçalament de cada columna de la finestra de dades.

La finestra de dades (5) se substitueix per la finestra d'operacions estadístiques, a la mateixa àrea, quan s'activa aquesta modalitat de funcionament, i també fent doble-clic a sobre de qualsevol dels resultats.

(cc) BY-NC

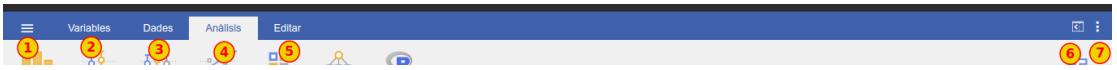
Aquestes dues finestres, com altres de "temporals" a *jamovi*, disposen a la seva part superior dreta d'icones fletxa-encerclada per a ser ocultades i així permetre la maximització de les finestres centrals de dades (esquerra) i resultats (dreta).

Menús de les modalitats de treball

L'aplicació *jamovi* funciona en cinc modalitats de treball:

- Modalitat d'IMPORTACIÓ i EXPORTACIÓ a arxius externs de dades i resultats,
- Modalitat de GESTIÓ de les VARIABLES o columnes,
- Modalitat de GESTIÓ de DADES de treball,
- Modalitat d'ANÀLISI de les DADES, i
- Modalitat per afegir COMENTARIS als RESULTATS.

A la part superior de la pantalla principal es troba l'accés als diferents menús de cada MODALITAT, que inclou també l'accés al control d'opcions de *jamovi*. Depenent de la modalitat activa es presenten conjunts diferents d'icones que apareixen a la barra inferior, així com continguts a l'àrea específica central.



Imatge 14. Menús d'opcions generals de jamovi

La barra de color blau conté els activadors d'aquest menú de modalitat i altres opcions, respectivament, d'esquerra a dreta:

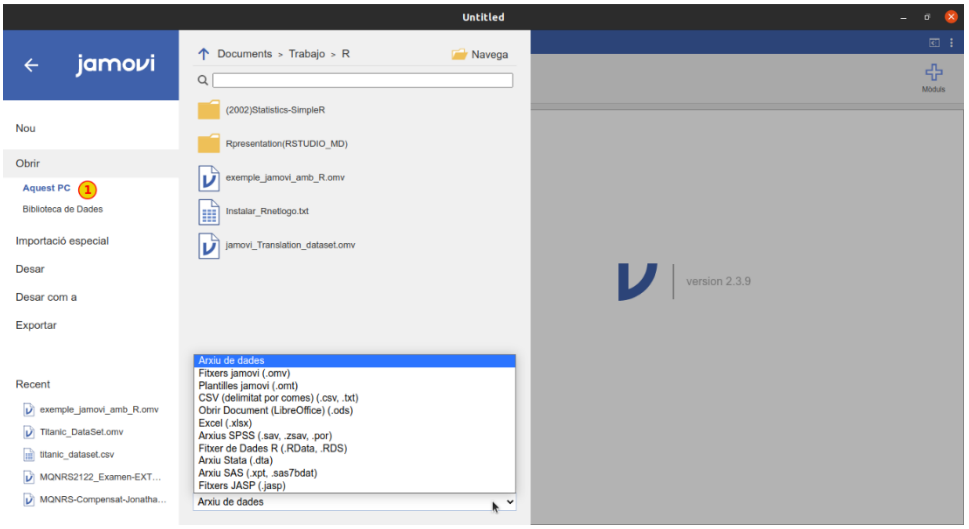
1. [Icona-tres-línies-horitzontals] Menú d'arxius: GESTIONA ARXIUS de *jamovi* i d'altres formats.
2. "Variables" Activa la modalitat de GESTIÓ DE VARIABLES: Per a treballar amb les variables.
3. "Dades" Activa la modalitat de GESTIÓ de DADES: Per a treballar amb les dades i variables.
4. "Anàlisis" Activa la modalitat d'ANÀLISI de DADES: Per a demanar resultats d'operacions.
5. "Editar" Activa la modalitat d'EDICIÓ de TEXT: Per a afegir comentaris propis als resultats.
6. [Icona-pantalla] Mostra o amaga l'àrea de resultats o de dades (treball a pantalla total).
7. [Icona-columna-tres-punts] Menú de configuració: Controla aspectes de visualització.

1. Menú d'arxius

S'utilitzarà quan es vulguin carregar dades dins de *jamovi*, desar la tasca realitzada en l'estat actual per continuar posteriorment, o descarregar dades o resultats cap a altres aplicacions.

Permet crear nous arxius de dades propis de *jamovi* (amb format anomenat OMV), obrir-ne d'existents, desar l'actual, així com importar i exportar dades i resultats (segons el format) d'arxius en altres formats per intercanvi amb d'altres aplicacions (d'anàlisi o de processament i presentació).

(cc) BY-NC



Imatge 15. Menú d'arxius

- **Nou:** Obre una nova instància de *jamovi*, sense tancar l'actual.
- **Obrir:** Mostra una finestra per seleccionar un arxiu de dades per a carregar-les dins de *jamovi*. Funciona amb diversitat d'arxius d'aplicacions habitualment utilitzades per a l'anàlisi de dades quantitatives (SPSS, Stata, SAS, JASP) i de Fulls de càlcul (LibreOffice Calc, M$-Excel), així com arxius d'R, o del format genèric per intercanvi de dades CSV.

NOTA: En tot cas, les dades han d'estar organitzades en files de casos i columnes de variables, amb la primera fila com a nom de cada variable. El procés d'obertura prova d'identificar el tipus de variable i el seu nivell de mesura a partir de la informació importada (amb desiguals resultats segons el format de l'arxiu). Si s'obre un arxiu OMV es carreguen les dades i reprodueixen totes les operacions que es van fer en l'arxiu original fins al moment dels seu "tancament".

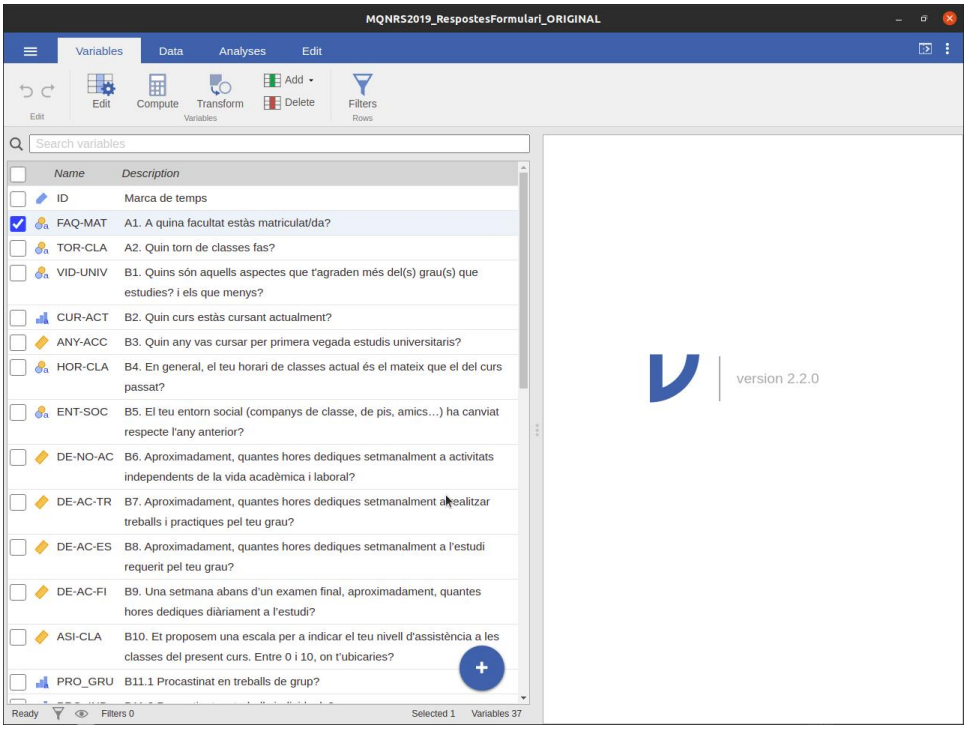
- **Importació especial:** Permet **afegir** noves dades a les ja existents, o ajuntar arxius de dades. Mostra una finestra que permet seleccionar arxius de dades de tipus propi OMV i afegir-los als que ja existeixen en el *jamovi* que estigui obert. El procés prova d'identificar si s'afegeixen nous casos/individus (si detecta que hi ha variables amb el mateix nom i tipus) o si s'afegeixen noves variables (si el nombre de casos és el mateix), amb desiguals resultats segons els continguts dels arxius originals.
- **Desar / Desar com a:** Desa en el format propi de *jamovi* (OMV) tota la informació de l'actual sessió d'anàlisi, no només la matriu de dades amb les identificacions de cada variable original i cada variable generada o transformada, sinó també tot el codi R que genera els resultats demanats, així com tot el text afegit als resultats. Una vegada hem assignat un nom a un arxiu, "Desar" el desarà amb aquest nom a la mateixa ubicació actualitzant l'anterior.
- **Exportar:** Genera diferents tipus d'arxius amb continguts específics, segons el format seleccionat per a l'exportació:
    - Resultats: En el cas d'exportacions cap a PDF, a HTML, o a LaTeX (bundle zip), es poden generar pàgines amb totes les taules i gràfics dels **resultats** i amb els textos adjunts.
    - Dades: En el cas d'exportacions cap a formats d'aplicacions d'anàlisi (SPSS, Stata, SAS, CSV o R) el què es genera són arxius únicament amb les **dades**, sense els resultats.

NOTA: Un arxiu OMV desat i obert posteriorment amb una altra instància de *jamovi* (que tingui els mateixos mòduls instal·lats) reprodueix de nou totes les operacions que es van realitzar anteriorment, per la qual cosa en ocasions pot trigar minuts en completar l'obertura, depenent de la quantitat d'operacions realitzades fins al moment de desar. El resultat, però, és una còpia exacta de la tasca tal com es va realitzar anteriorment.

2. Modalitat de gestió de variables

Disponible només a partir de la versió v2, s'utilitzarà quan es vulguin modificar les variables o columnes de dades existents, bé sigui canviant el seu nom o descripció, el seu tipus, crear-ne de noves a partir de variables existents, eliminar-ne, etc.

Mostra el submenú gràfic amb les opcions disponibles per a editar, eliminar, moure, modificar i filtrar les dades actuals, organitzades en columnes de variables.



Imatge 16. Modalitat gestió de variables

- **Editar:** Activa i desactiva la finestra d'IDENTIFICACIÓ de variables, amb la possibilitat d'ajustar-ne el nom, la descripció, el tipus de variable (nominal, ordinal, contínua), el tipus de dada (text, decimal, nombre enter), els valors perduts, i les categories o etiquetes (*"Nivells"*).
- **Calcular:** Activa i desactiva la finestra de CÀLCUL de variables, amb la possibilitat de crear NOVES variables a partir d'una o més de les existents, tot aplicant funcions estadístiques, lògiques, o matemàtiques (p. ex., el càlcul de la nova variable "densitat residencial" per a cada fila, mitjançant la divisió dels metres quadrats de l'habitatge pel nombre d'habitants).
- **Transformar:** Activa i desactiva la finestra de TRANSFORMACIÓ de variables, amb la possibilitat de crear NOVES variables a partir d'una d'existent, tot aplicant criteris

condicionals *(del tipus IF-THEN-ELSE)* fonamentats en els valors o etiquetes de la variable original (p. ex., la recodificació de la variable quantitativa "anys d'edat" per a cada fila, en una nova variable "grup d'edat" mitjançant l'assignació dels nous valors {jove/adult/gran} a cada cas/individu en funció d'un conjunt de regles d'usuari que defineix els intervals corresponents a cadascuna de les tres noves etiquetes o "nivells").

- **Afegir/ Eliminar:** Afegeix o elimina definitivament noves variables a la matriu de dades. A partir de la variable seleccionada, [clic-esquerra] desplega un menú que permet accedir a la finestra de CÀLCUL, a la de TRANSFORMACIÓ, o a crear una nova variable en blanc.
- **Filtres:** Activa i desactiva la finestra de FILTRATGE de casos, amb la possibilitat d'incloure/excloure conjunts de casos mitjançant criteris "dissenyats" per l'analista. La icona filtre (embut) de la part inferior esquerra de la pantalla té el mateix efecte. L'exclusió de casos mitjançant filtratge no és definitiva; els casos continuen existint a la matriu de dades tot i que no són tinguts en compte per a les operacions estadístiques (p. ex., filtrar la matriu de dades per obtenir resultats només per a les "dones", o pels "homes de <30 anys que viuen a Cerdanyola").

Treballant en aquesta modalitat de gestió de variables, moltes d'aquestes opcions estan també disponibles fent clic-dret a sobre de qualsevol de les variables.

La columna de caselles de selecció, a l'esquerra dels noms identificadors de les variables, permet realitzar operacions amb conjunts de variables. La casella que encapçala la columna serveix per a seleccionar / de-seleccionar totes les variables.

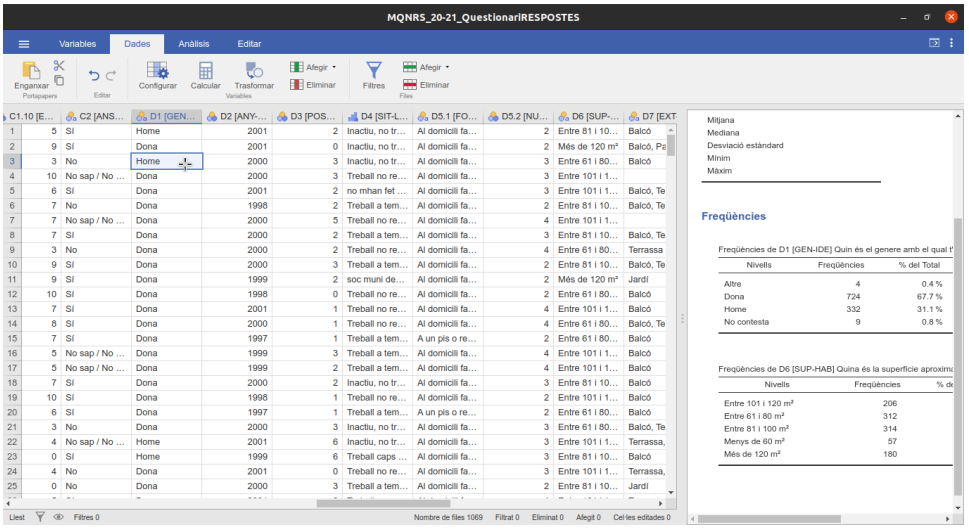
Tal i com es veurà tot seguit, les operacions sobre les variables d'aquesta modalitat de treball TAMBÉ son accessibles treballant en la modalitat de gestió de dades [Data].

3. Modalitat de gestió de dades

S'utilitza quan es volen modificar les dades existents, bé sigui en l'àmbit dels casos *(filtrant-los segons criteris, afegint-ne o eliminant-ne valors, enganxant des d'altres aplicacions...)* o en l'àmbit de les variables *(creant-ne de noves a partir de variables existents, eliminant-ne...)*. Aquesta capacitat de "gestionar" les dades directament és el principal tret diferencial de *jamovi* en comparació amb altres paquets d'anàlisi estadística.

Mostra el submenú gràfic amb les opcions disponibles per a editar, eliminar, moure, modificar i filtrar les dades actuals, que s'organitzen visualment com una matriu de files (casos) i columnes (variables).

RECORDA: Les dades, una vegada "dins de *jamovi*", adopten una organització en files i columnes. Aquesta estructura es coneix com a **MATRIU DE DADES**. Cada columna correspon a un indicador o variable, i cada fila correspon a un registre o unitat d'observació (p. ex., una persona, un municipi, un país...). En el context d'enquestes per qüestionari de caire quantitatiu, en general, cada fila inclou informació d'un qüestionari, i cada columna informació d'una pregunta (excepte en el cas de multirespostes o d'escales combinades).



Imatge 17. Modalitat gestió de dades

- **Portapapers:** Tres controls habituals en ofimàtica *(copiar, retallar, enganxar)* permeten intercanviar dades (1) entre diverses zones de la matriu, o (2) amb altres aplicacions (per exemple, intercanvi amb fulls de càlcul, mitjançant operacions de retalla/enganxa del portapapers).
- **Editar:** Registre històric per a desfer o refer accions d'edició de dades.
- Variables
    - **Configurar:** Activa i desactiva la finestra d'IDENTIFICACIÓ de variables, amb la possibilitat d'ajustar-ne el nom, la descripció, el tipus de variable (nominal, ordinal, contínua), el tipus de dada (text, decimal, nombre enter), els valors perduts, i les categories o etiquetes (*"Nivells"*).
    - **Calcular:** Activa i desactiva la finestra de CÀLCUL de variables, amb la possibilitat de crear NOVES variables a partir d'una o més de les existents, tot aplicant funcions estadístiques, lògiques, o matemàtiques (p. ex., el càlcul de la nova variable "densitat residencial" per a cada fila, mitjançant la divisió dels metres quadrats de l'habitatge pel nombre d'habitants).
    - **Transformar:** Activa i desactiva la finestra de TRANSFORMACIÓ de variables, amb la possibilitat de crear NOVES variables a partir d'una d'existent, tot aplicant criteris condicionals *(del tipus IF-THEN-ELSE)* fonamentats en els valors o etiquetes de la variable original (p. ex., la recodificació de la variable quantitativa "anys d'edat" per a cada fila, en una nova variable "grup d'edat" mitjançant l'assignació dels nous valors {jove/adult/gran} a cada cas/individu en funció d'un conjunt de regles d'usuari que defineix els intervals).
    - **Afegir/ Eliminar:** Afegeix o elimina definitivament noves variables (columnes) a la matriu de dades. A partir de la columna seleccionada (clic-esquerra) desplega un menú que permet accedir a la finestra de CÀLCUL, a la de TRANSFORMACIÓ, o a crear una nova columna (variable) en blanc.
- Files (casos)
    - **Filtres:** Activa i desactiva la finestra de FILTRATGE de casos, amb la possibilitat d'incloure/excloure conjunts de casos mitjançant criteris "dissenyats" per l'analista. La icona filtre (embut) de la part inferior esquerra de la pantalla té el mateix efecte. L'exclusió de casos mitjançant filtratge no és definitiva; els casos continuen existint a la matriu de dades tot i que no són tinguts en compte per a les operacions estadístiques (p. ex., filtrar la matriu de dades per obtenir resultats només per a les "dones", o pels "homes de <30 anys que viuen a Cerdanyola").

- **Afegir/ Eliminar:** Afegeix o elimina definitivament nous casos (files) a la matriu de dades. A partir de la fila seleccionada, [clic-esquerra] desplega un menú que permet crear-ne de noves, ja sigui a continuació o a l'extrem inferior de la matriu.

Treballant en aquesta modalitat de gestió de les dades, moltes d'aquestes opcions estan també disponibles fent clic-dret a sobre de qualsevol de les dades, o dels encapçalaments de les columnes, o dels nombres identificadors de les files de la part esquerra de la matriu.

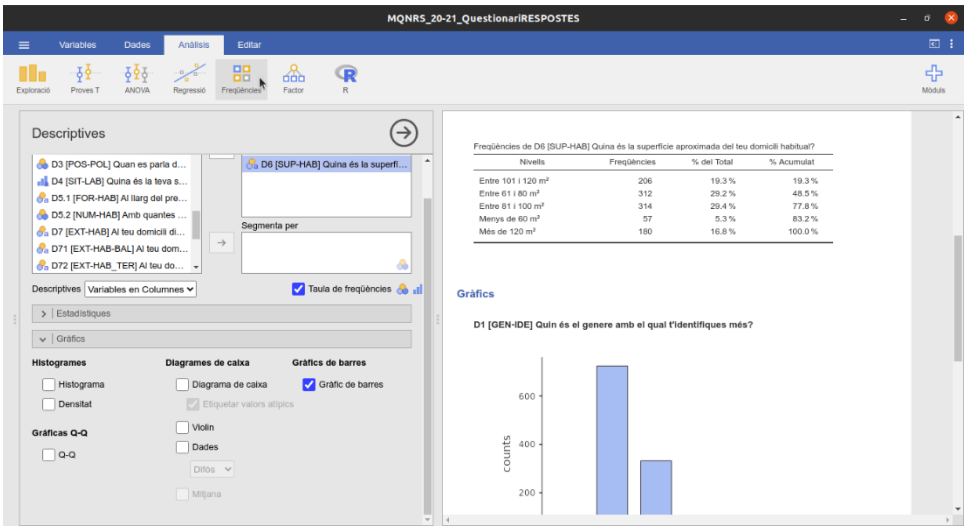
A efectes de llenguatge R, les dades de la matriu de *jamovi* són un objecte R del tipus **`data.frame`** anomenat **`data$`**

> **RECORDA**: totes les operacions sobre variables d'aquesta modalitat de treball [Dades] TAMBÉ son accessibles treballant en la modalitat de gestió de variables [Variables].

4. Modalitat d'anàlisi de dades

S'utilitzarà quan es vulguin executar operacions estadístiques a partir de dades de la matriu i generar resultats en forma de taules, d'indicadors i de representacions gràfiques.

Mostra el submenú gràfic amb els procediments disponibles per a realitzar les operacions disponibles en la instal·lació local de *jamovi*. Conté tots els procediments bàsics, més l'accés als mòduls d'expansió que han estat instal·lats.



Imatge 18. Disposició en modalitat d'operacions estadístiques

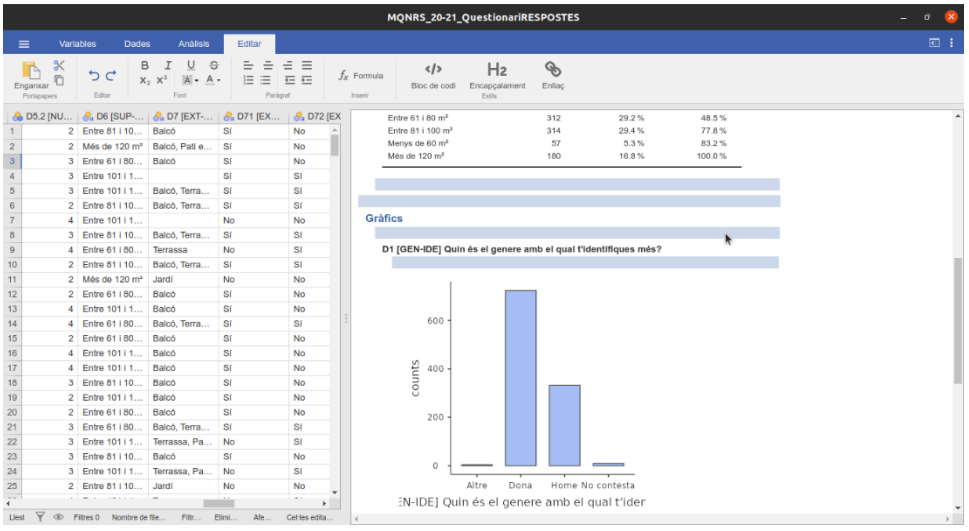
Clicant les icones del submenú gràfic desplega un menú amb els procediments estadístics disponibles, que han de ser utilitzats específicament per a condicions concretes de les dades. Correspon a qui realitza l'anàlisi conèixer les condicions d'aplicabilitat de cadascun dels procediments: *jamovi* no incorpora prou intel·ligència com per a aplicar correctament cada operació estadística, per tant es pot donar el cas de demanar un resultat impossible o no ajustat a les dades disponibles i que *jamovi* ofereixi un resultat matemàticament correcte tot i que analíticament inadequat.

RECORDA: El PC no és més que una màquina que suma molt ràpid. Tota la intel·ligència que hi ha a una anàlisi estadística l'ha d'aportar la persona que utilitza el PC com a eina. Qui ha de saber Estadística ets tu..., *jamovi* no en sap res.

5. Modalitat de comentaris al text

S'utilitzarà quan, una vegada generats els resultats, es vulguin afegir comentaris als elements de la finestra de resultats.

Mostra el submenú gràfic amb les opcions disponibles per a escriure i editar text lliure integrat als resultats de les operacions estadístiques realitzades, en forma de taules i de representacions gràfiques.



Imatge 19. Modalitat comentari de resultats

Les diferents icones d'edició de text del submenú gràfic funcionen de forma similar a les que disposa qualsevol aplicació de processament de textos *(LibreOffice Writer, M$-Word, MacOS-Pages, etc.)*. Quan s'activa aquesta modalitat, clicant a "Editar", a la finestra de resultats queden destacats (en color blau clar) les àrees on es pot afegir text, associades als diferents objectes de resultats generats per les operacions.
Clicar aquestes àrees blaves en permet l'edició, de forma que –típicament– es poden fer comentaris adjunts a cada taula, així com redactar extensions de text de presentació i de conclusions que conformen un INFORME complert dins de *jamovi*. Aquest informe es pot desar i compartir **en un mateix arxiu** (en format OMV, o bé publicar-lo directament com a PDF o pàgina web amb les corresponents opcions d'importació del menú d'arxius).

El format OMV propi de *jamovi* permet publicar i distribuir en un únic arxiu les dades originals, més totes les transformacions i càlculs realitzades, més tots els resultats generats, més tot el text que acompanya i comenta els resultats. Aquestes característiques de l'arxiu OMV són molt apreciades al mon científic: "complet", "transparent" i "compacte".
Per aquest motiu es demanarà prioritàriament com a format als lliuraments dels cursos.

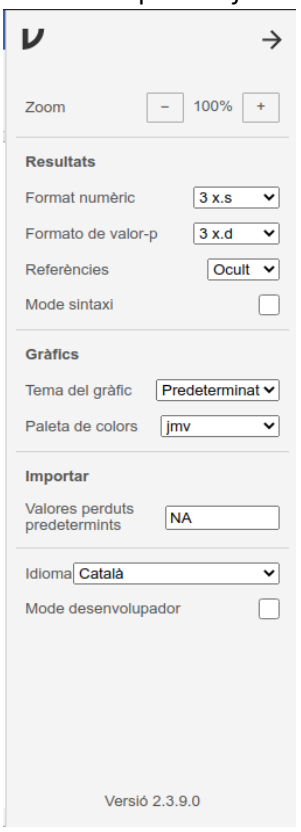
6. Activador de pantalla central

Situat a la part dreta de la barra principal superior , activa o desactiva panells addicionals a la part central de la pantalla. Serveix per a treballar més còmodament amb cadascuna de les modalitats, ocupant tota l'amplada de pantalla disponible.

En general, a l'àrea central *jamovi* mostra DOS panells, amb diversos continguts en funció de la modalitat de treball utilitzada. Aquest panells es poden modificar en amplada fent clic-dreta a la barra vertical gris que els separa i "arrossegant" a dreta o esquerra.

7. Menú de configuració de jamovi

Situat a la part dreta de la barra principal superior, desplega un llistat d'opcions que permet controlar algunes configuracions especials de l'aplicació *jamovi*.

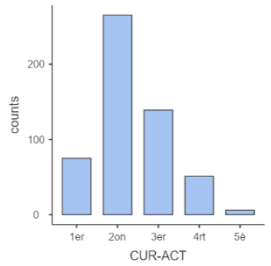
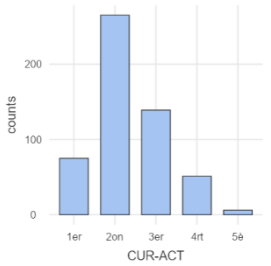
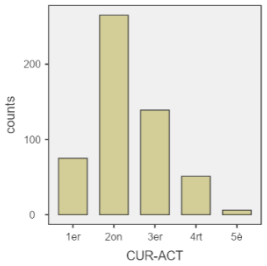
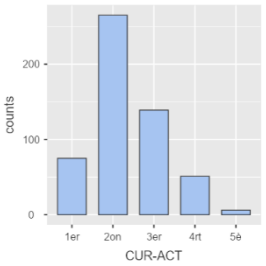


Imatge 20. Menú de configuració

- **Zoom:** Permet ajustar la mida de visualització del conjunt d'elements de l'aplicació.
- Resultats
  - **Format numèric:** Permet triar el format dels decimals a les taules, bé en forma "relativa" de dígits o xifres significatives *(sense tenir en compte els zeros a l'esquerra)* o en forma "absoluta" *(arrodonint a un nombre determinat de decimals inclosos els zeros)*, així:
    - el format 3 x.s genera 0,123 i 0,00458
    - el format 3 x.d genera 0,123 i 0,005
  - **Format de valor-p:** Permet igualment triar el format dels decimals pels càlculs de valors-p de significativitat, bé en forma "relativa" o en forma "absoluta".
  - **Referències:** Permet ocultar o visibilitzar automàticament al final dels resultats les referències bibliogràfiques pertinents al projecte *jamovi* i al llenguatge **R**, en les versions corresponents a les anàlisis realitzades.

- **Mode sintaxi:** Activa / desactiva el mode en què (1) els resultats no es formaten per a publicació sinó que apareixen tal i com els calcula el llenguatge R, i (2) s'inclou el codi R corresponent a cada resultat generat. Habitualment el mantindrem desactivat per a aconseguir resultats més presentables i directament publicables.

- Gràfics
  - **Tema del gràfic:** Diverses configuracions pels resultats gràfics generats.

			
Imatge 21. Tema "default"	Imatge 22. Tema "minimal"	Imatge 23. Tema "I . SPSS"	Imatge 24. Tema "Hadley"

  - **Paleta de colors:** Diversos conjunts de colors aplicables als resultats gràfics.

- Importar
  - **Valors perduts predeterminats:** Definició de quina etiqueta o valor serà tingut com a "valor perdut" en el moment de la importació de dades des d'arxius amb format diferent al propi de *jamovi* OMV.

- **Mode desenvolupador:** Activa un mode especial, d'utilitat per a qui programa i depura nous mòduls i extensions de *jamovi*. En general, pels analistes no té cap més efecte que disposar d'informació extra quan es produeix un error, i és preferible mantenir-lo desactivat.

EN GENERAL, moltes finestres emergents de la GUI *jamovi* disposen a la part superior dreta d'una icona [fletxa-encerclada], que pot apuntar cap a qualsevol direcció. Aquesta icona és un botó per a **TANCAR la finestra**, o per a amagar-la.

4. Lectura i escriptura d'arxius de dades amb jamovi

Introducció directa de dades

En algunes ocasions ens pot interessar analitzar amb *jamovi* dades que es troben en formats no digitals, com ara en qüestionaris, o en publicacions impreses. Caldrà introduir, teclejant manualment, aquestes dades.

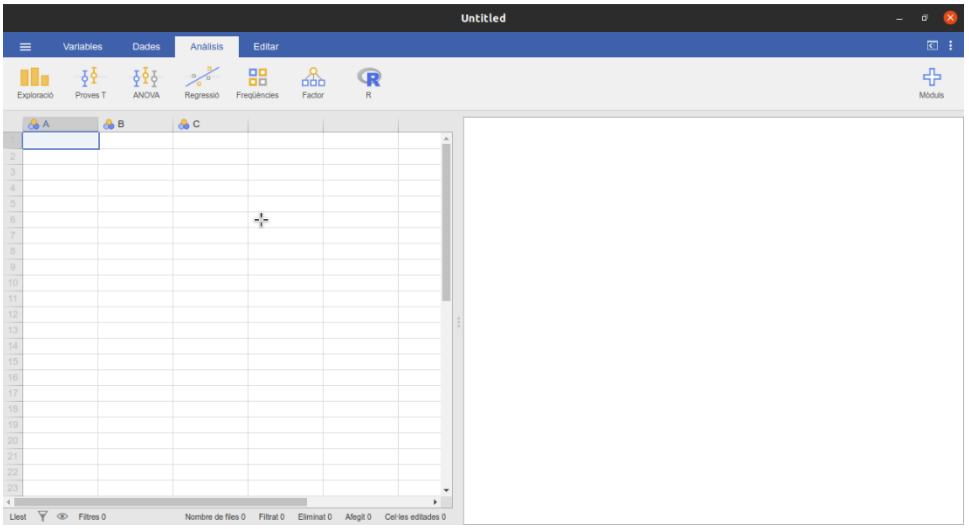
A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic:

- Hi ha unes dades publicades que es volen integrar dins d'una matriu de dades de *jamovi* per aplicar anàlisis estadístiques posteriors. Les següents dades corresponen a les diferents qualificacions, entre 0 i 10, d'una prova d'un curs:

5, 4, 8, 4, 6, 8, 5, 9, 3, 7, 2, 8, 8, 3, 4, 4, 2, 3, 3, 5, 8, 6, 4, 1, 3, 2, 4, 2, 2, 4, 2, 7, 4, 8, 5, 8, 8, 4, 2, 8, 1, 8

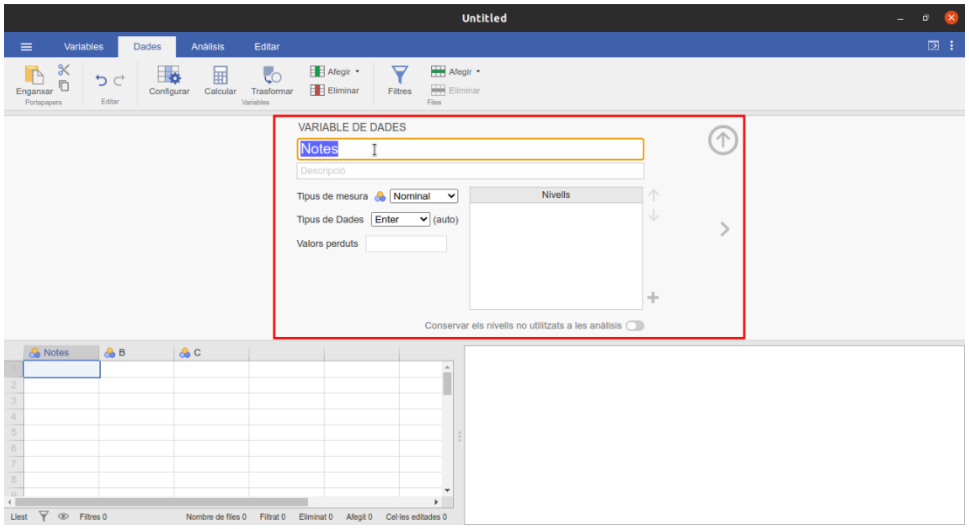
Per a introduir aquestes dades al sistema, com a matriu de dades *jamovi*:

1. Executa l'aplicació *jamovi* al teu ordinador.
   En funció del sistema operatiu utilitzat, de l'ordinador i si és la primera vegada que s'executa pot trigar fins a un minut a iniciar-se.
2. S'obrirà una nova sessió amb tres variables nominals (automàticament anomenades A, B, C), sense dades, i amb la pantalla principal en modalitat ANÀLISIS.



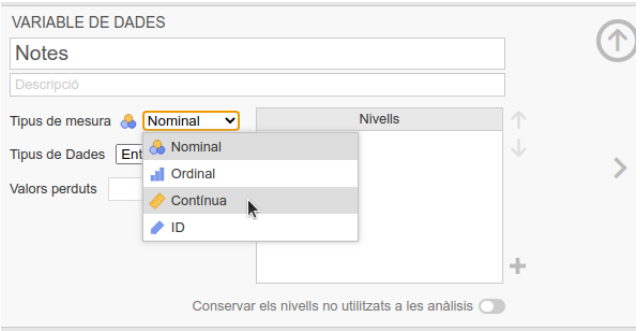
Imatge 25. Inici de jamovi

3. Primer, cal **definir la variable** que contindrà les notes de l'alumnat del curs assignant-li un nom (i opcionalment, un comentari), un tipus de variable (aquesta és numèrica, o quantitativa), i un tipus de dades (son nombres sencers i sense decimals).
   Per fer-ho, hi ha tres procediments alternatius que tenen el mateix efecte...
   OPCIO A: Clica el menú / Variables. A continuació fes doble-clic a qualsevol lloc de la primera línia o variable.
   OPCIO B: Clica el menú / Dades / Configurar
   OPCIO C: Doble-clic a l'encapçalament de la primera columna (lletra A)
4. S'obre l'àrea de definició de variables.



Imatge 26. Àrea de definició de variables

5. Tecleja el nom desitjat, substituint les lletres assignades automàticament pel sistema, per exemple, "*Notes*".
6. Pots incloure qualsevol caràcter (fins i tot accents, majúscules o espais). És recomanable utilitzar noms CURTS, abreviatures i no incloure espais especialment si es vol intercanviar l'arxiu de dades amb altres aplicacions sense que el procés d'importació doni errors.
7. A "Descripció" pots incloure una explicació o nom llarg de la variable. No s'utilitzarà per a generar els resultats estadístics, però resultarà molt aclaridor pel treball amb *jamovi*.
8. A "Tipus de mesura" (tipus de variable) clica el desplegable i selecciona "Contínua".[5]
9. A "Tipus de Dades" (nivell de mesura de les dades) selecciona "Enter" (és l'opció per defecte).
10. Pots continuant definint d'altres variables, clicant a les icones laterals > i < (alternativament, fent clic a l'encapçalament de la columna corresponent).
11. En el cas de variables de tipus nominal o ordinal, caldrà comprovar les categories de la variable ("Nivells") i la seva ordenació. Això es veurà en detall al següent capítol "5. Definició de dades amb *jamovi*: Tipus de variables i valors perduts".



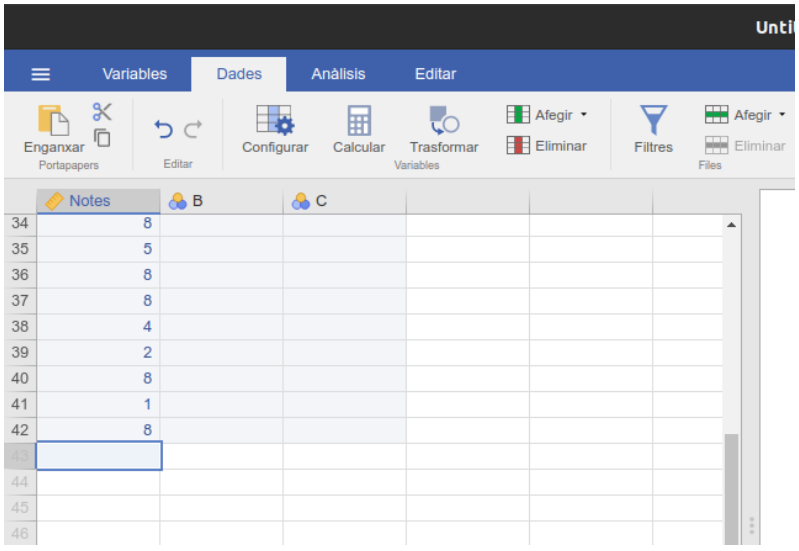
Imatge 27.Definició d'una variable numèrica sense decimals

12. Pots tancar l'àrea de definició de variable clicant la fletxa-amunt-encerclada.

5 Per a més informació sobre "Tipus de mesura" i "Tipus de Dades", consulta l'apartat posterior "5. Definició de dades amb jamovi: Tipus de variables i valors perduts".

13. Comprova que (1) la línia superior de la definició de variables, o (2) l'encapçalament de la primera columna, inclou el nom que has introduït i la petita icona que identifica variables quantitatives (una regla de mesura de color taronja).
14. Clica el primer cas de la columna/variable. La casella pren color blau. Tecleja la primera dada: 5 i prem al teclat [Intro]. La dada s'inclou al primer cas de la variable, i es destaca la casella per introduir el següent cas de la variable.
15. Continua introduint manualment totes les dades. Comprova a la numeració lateral esquerra de les files que són 42 casos.
16. En cas d'error, només cal que cliquis sobre la casella que es vulguis modificar i introduir el valor corregit a sobre, validant amb la tecla [Intro].
17. La tecla [Del] elimina dades de caselles concretes. Alternativament, es pot retallar ("Retallar") utilitzant el menú contextual que es desplega en fer clic-dreta a sobre d'una casella.

El resultat ha de ser la següent matriu de dades, amb una variable anomenada "Notes" i 42 casos.



Imatge 28.Exemple d'introducció DIRECTA de dades

PROBLEMA: La introducció manual de dades a *jamovi* està optimitzada per ser realitzada "per columnes", és a dir, introducció de les successives dades d'una mateixa variable. Això no resulta especialment útil en el cas d'introduir dades des de qüestionaris en paper; en aquest cas fora millor una introducció "per files" que no està disponible.

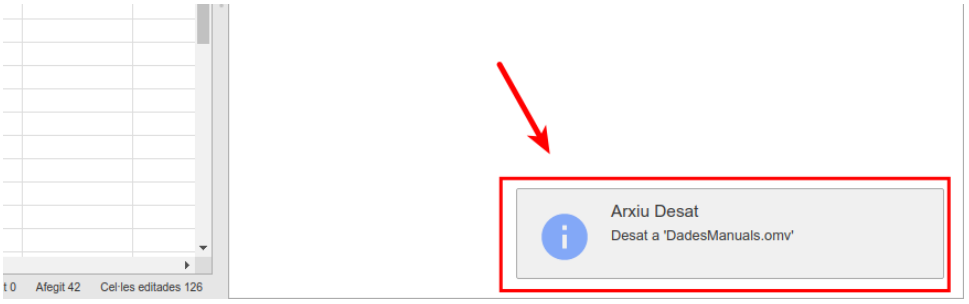
RECOMANACIÓ: Si has d'introduir manualment les dades de diferents qüestionaris en paper, és millor que (1) ho facis prèviament en un full de càlcul (no directament en *jamovi*), ja que permet la introducció contínua de totes les dades d'un mateix qüestionari **per files** i, (2) posteriorment, utilitzis qualsevol dels dos procediments que s'expliquen a continuació per a importar les dades del full de càlcul cap a *jamovi*.

Tot i que no forma part del procediment d'introducció de dades, avancem ara el procediment de DESAR les dades introduïdes en un arxiu nou en el format propi de *jamovi*. Compte!, Aquesta aplicació **NO desa automàticament** les dades, cal DESAR sovint el treball per no perdre'l degut a incidències amb l'ordinador.

Per desar un arxiu propi de *jamovi*:

1. Abans de desar les dades d'un arxiu nou, la barra superior indica "Untitled".
2. Clica al Menú / Arxius [tres línies horitzontals]

3. Al desplegable, selecciona l'opció "Desar". S'obrirà una finestra d'emmagatzematge.
4. Tecleja el nom desitjat, p. ex., "DadesManuals" (recomanable sempre sense espais)
5. Clica "Navega", per a seleccionar la ubicació on vols desar d'aquest arxiu *jamovi*.
6. S'obre una finestra d'ubicació. Selecciona la teva preferència: disc dur, usb, etc...
7. A la part inferior dreta de la pantalla principal apareix una notificació del procés d'emmagatzematge.



Imatge 29. Missatge d'indicació del procés per desar arxius

8. Comprova que la barra superior de l'aplicació ara indica "DadesManuals". L'arxiu amb les dades s'ha desat correctament.

Introducció des de taules de dades

En ocasions les dades que es volen analitzar no estan disponibles com a llistat de tots i casacun del valors o categories per a CADA cas o individu (matriu de dades), sinó com a taula de resultats on apareixen acumulats els recomptes o freqüències de cada categoria. En aquest cas les dades s'introduiran a *jamovi* en un format especial, diferent d'un valor per a cada fila: en format de "ponderació" de variables.

A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic:

- Hi ha unes dades publicades que es volen integrar dins d'una matriu de dades de *jamovi* per aplicar anàlisis estadístiques posteriors. Les següents dades corresponen a la quantitat de persones que van morir i sobreviure a l'enfonsament del trans-atlàntic Titanic, segons el tipus de passatge amb el que viatjaven.[6]

Víctimes mortals al naufragi del TITANIC				
		Van sobreviure...		
Passatge (classe)		**No**	**Si**	**Total**
Primera	(Alta)	80	136	216
Segona	(Mitjana)	97	87	184
Tercera	(Baixa)	372	119	491
Total		549	342	891

[6] Cal dir que el tipus de passatge es un indicador de la "classe social" d'aquestes persones.

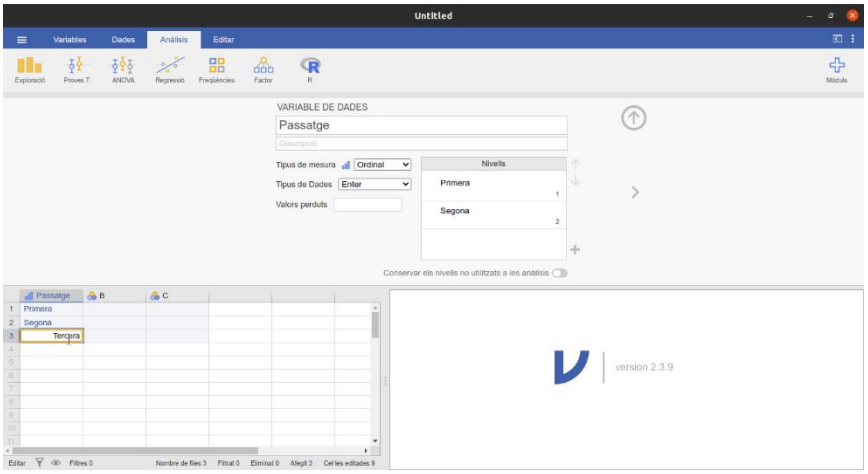
Caldrà considerar les quantitats de l'interior de la taula com a **"freqüències"** i crear a *jamovi* una columna pròpia per aquestes dades. Caldrà, igualment crear dues variables, "sobreviure" (nominal, amb dos nivells: sí, no) i "classe" (ordinal, amb tres nivells: Primera, Segona i Tercera). En total s'utilitzaran 3 columnes, dos per a les variables "reals" més una altra per a les freqüències o recomptes de cada categoria.

Per a introduir aquestes dades al sistema, com a matriu de dades *jamovi*:

1. Executa l'aplicació *jamovi* al teu ordinador.
2. S'obrirà una nova sessió amb tres variables nominals (automàticament anomenades A, B, C), sense dades, i amb la pantalla principal en modalitat ANÀLISI.
3. Primer, cal **definir les tres variables**, assignant-li a cadascuna un nom (opcionalment, un comentari), un tipus de variable i un tipus de dades. Utilitzant els menus deplegables i seguint les indicacions de l'apartat anterior:
    a. Defineix la variable per a la primera columna de dades amb el nom "Passatge", ha de ser de tipus de mesura "Ordinal" i de tipus de dades "Enter".
    b. A la primera columna de la matriu de dades, fes clic al cas 1 de la primera fila i tecleja la categoria "Primera".

Quan teclegis el que indica el manual pas a pas, recorda't de NO INCLOURE les cometes(si no és que així ho indica expressament el mateix manual)

    c. Pica la tecla [Intro], a la segona fila tecleja la categoria "Segona".
    d. A mesura que introdueixis els valors o nivells de la variable, pots comprovar com apareix a la finestra corresponent de l'àrea superior de definició de variables.



Imatge 30. Definició de variable ordinal i introducció de nivells

    e. Finalment, tecleja a la tercera fila la categoria "Tercera".
    f. Ja tens definida la variable Passatge amb les seves categories o nivells en ordre.

NOTA: Els nivells o categories de les variables Ordinals o Nominals es poden reordenar a voluntat a l'espai corresponent de l'àrea de definició de variables. Per fer-ho, fes clic al nom de

la categoria que apareix a la finestra "Nivells" per seleccionar-la, a la part dreta apareixeran unes fletxes que permeten "pujar" i "baixar" la categoria seleccionada.
Utilitza aquesta funció per a ordenar variables per tal que (1) apareguin als gràfics i taules tal com penses expressar-les, i (2) tinguin sentit determinades operacions estadístiques, com ara el càlcul de freqüències acumulades, medianes i RIQ]

4. Ara fes clic a l'encapçalament de la **segona columna**. A l'àrea superior defineix la segona variable amb el nom "Sobreviure", tipus de dades "Nominal" i tipus de mesura "Text".
5. En aquest cas, afegirem les dos categories de la variable directament a l'àrea de definició, utilitzant la finestra de "Nivells".
    a. A la part inferior dreta d'aquesta finestra, fes clic al signe + per afegir un nou nivell.
    b. Tecleja la nova categoria corresponent, per exemple "Sí".
    c. Fes igual per introduir la nova categoria "No".
6. Finalment, fes clic a l'encapçalament de la **tercera columna**. A l'àrea superior defineix la variable que contindrà les ponderacions o freqüències de cada combinació de nivells, amb el nom "Casos", tipus de dades "Contínua" i tipus de mesura "Enter". Aquesta variable NO té nivells perquè és contínua (només en tenen les nominals o ordinals)

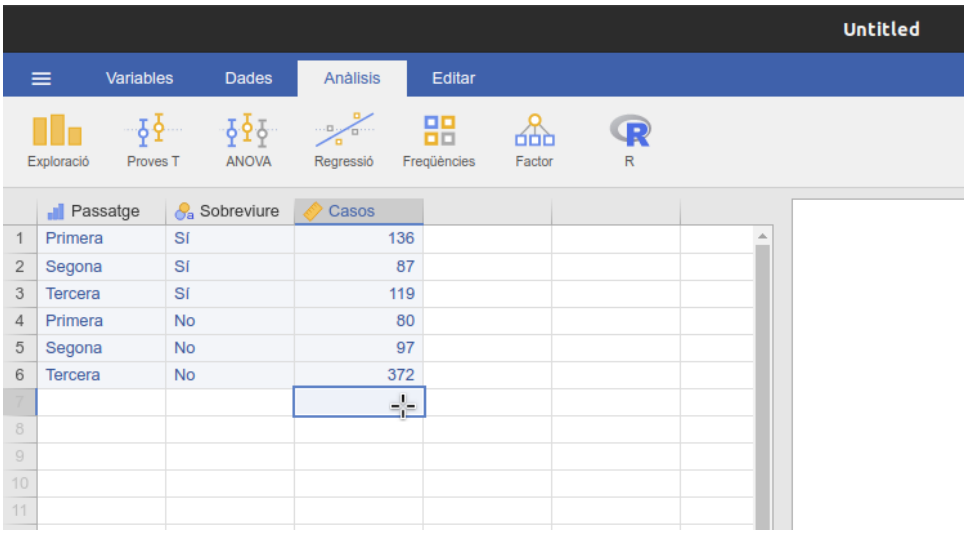
Ja estan creades les tres variables a les tres columnes que es necessiten per a introduir les dades ponderades de la taula. A continuació es tracta d'introduir les dades a la matriu de dades inferior. No tanquis de moment l'àrea de definició.

Per a introduir les dades corresponents a la taula superior a *jamovi*:

1. A la primera columna ja hi ha les tres possibles categories de la variable "Passatge".
2. A la segona columna, clica a la primera fila de la variable "Sobreviure" i tecleja la categoria "Sí" seguida de la tecla [Intro]
3. **ATENCIÓ**, ara farem un error a propòsit per a mostrar una qüestió important... Tecleja "si", sense accent ni majúscula. A la finestra de "Nivells" de l'àrea de definició ha aparegut la categoria "si" i *jamovi* la considerarà com a DIFERENT de la categoria "Sí". Això és un error d'introducció de dades, com si p.e. algunes persones d'un qüestionari visquessin a "Catalunya" d'altres a "Cataluña" i l'ordinador considerés que son dos territoris diferents. **CORREGEIX l'error**: clica a la cel·la on indica "si" i tecleja "Sí". Recorda que és responsabilitat teva la introducció de les dades sense errors com aquest.
4. A les tres següents files de la columna "Sobreviure" tecleja tres "No"
5. Ara cal que a la primera columna s'aporti informació per aquestes darreres 3 files que tenen nivell "No" per a la variable "Sobreviure". Ho farem exemplificant la possibilitat de retallar/enganxar dades mitjançant el porta-papers:
    a. Primer selecciona les cel·les que es volen copiar. Fes clic a sobre de "Primera", manté el dit i arrossega fins a seleccionar "Primera" "Segona" i "Tercera". Les tres cel·les queden destacades en blau.
    b. Ara fes clic-dret, i al desplegable selecciona "Copiar".
    c. Fes clic per seleccionar la fila 4 de la primera columna, on vols que aparegui el que has copiat.
    d. Ara fes clic-dret, i al desplegable selecciona "Enganxar".

Aquesta és la forma de copiar mitjançant el porta-papers. Es pot utilitzar (1) a la mateixa matriu de dades de *jamovi*, o (2) entre dos matrius de *jamovi* d'arxius oberts simultàniament, o fins i tot (3) entre altres aplicacions (per exemple, fulls de càlcul o navegadors web) i *jamovi*.

6. Finalment, a la tercera columna, cal introduir a la cada fila la quantitat de casos que correspon a la combinació de nivells definida per les columnes primera i segona. Així:
7. A la columna de "Casos", a la primera fila cal introduir 136 que son els casos de passatge "Primera" que "Sí" van sobreviure.
8. Consulta la taula superior amb les dades i Introdueix la resta de casos.



Imatge 31. Exemple d'introducció de dades ponderades

Aquesta forma d'introduir les dades, coneguda com a "ponderació de dades", requereix de procediments específics al moment d'executar les anàlisis estadístiques, com ja s'indicarà més endavant. Caldrà indicar quina columna és la que conté informació sobre el "pes" o la "ponderació" de cadascun dels valors de la resta de columnes de dades.

Fixat en que cada fila conté cadascuna de les possibles combinacions de les variables més una columna amb els casos de la combinació. S'utilitza un total de files x columnes de la taula original (en aquest cas, 3x2=6). A la forma "detallada" d'introduir la informació, contràriament, cada fila és un cas, o persona, o qüestionari, o població, o el que sigui la unitat mostral, de forma que per introduir les dades del Titanic caldria un total de 691 files.

EXERCICI

1. Una vegada introduïdes les dades del Titanic, pot desar aquest nou arxiu jamovi, seguint les explicacions de l'apartat anterior?
2. Deixa'l en un lloc on tinguis accés per si cal que obris aquest arxiu de dades més endavant al manual MUJADES.

(cc) BY-NC

Enganxar dades des d'altres aplicacions

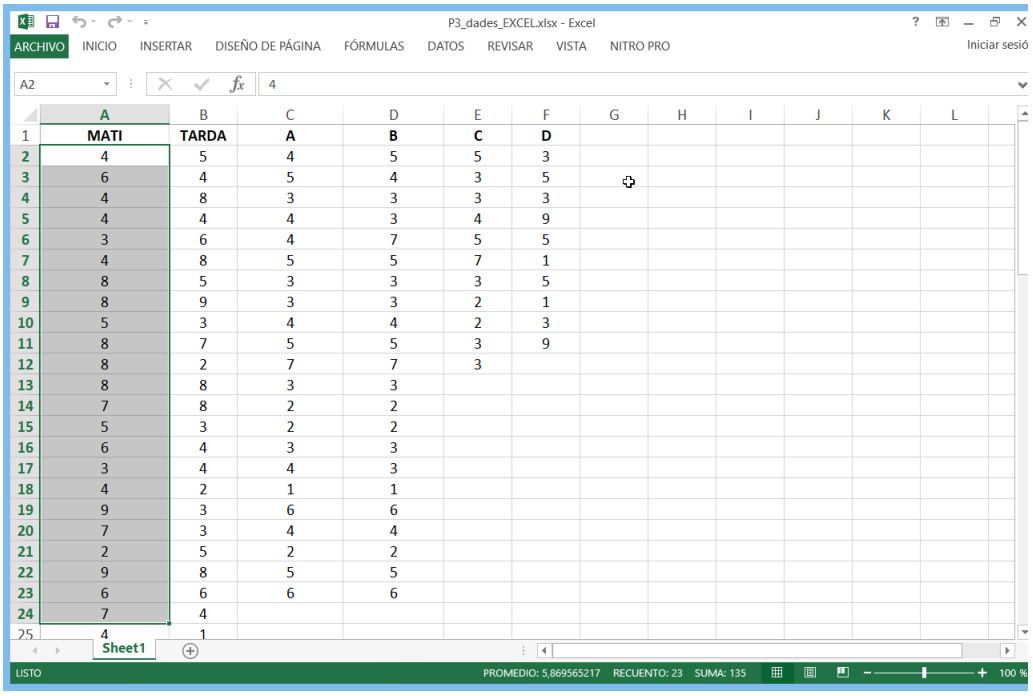
En ocasions interessa analitzar amb *jamovi* dades que es troben disponibles directament en arxius digitals desats amb altres aplicacions, o publicades en format electrònic o en línia. Per exemple, les dades poden ser a un arxiu de full de càlcul (*.ods, *.xlsx), o publicades en formats propis de diverses aplicacions d'anàlisi estadística. Una opció és retallar aquestes dades de l'aplicció original i enganxar-les a *jamovi*.

A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic:

- Hi ha unes dades a un full de càlcul, que es volen carregar dins d'una matriu de *jamovi*.
- Aquestes dades complementen unes dades que anteriorment ja s'han introduït manualment a *jamovi*.

Adjunt a l'espai de descarrega d'aquest manual (o l'espai docent Campus Virtual CV de l'assignatura a la UAB) hi pots trobar l'arxiu *MJDS_Dades_EXCEL.xlsx*, un full de càlcul on hi ha dades corresponents a diverses distribucions, en concret son notes d'exercicis.

1. Descarrega aquest arxiu format .XLSX a una carpeta local del teu PC.
2. Fes doble-clic, per a obrir l'arxiu en la teva aplicació de full de càlcul.



	A	B	C	D	E	F
1	MATI	TARDA	A	B	C	D
2	4	5	4	5	5	3
3	6	4	5	4	3	5
4	4	8	3	3	3	3
5	4	4	4	3	4	9
6	3	6	4	7	5	5
7	4	8	5	5	7	1
8	8	5	3	3	3	5
9	8	9	3	3	2	1
10	5	3	4	4	2	3
11	8	7	5	5	3	9
12	8	2	7	7	3	
13	8	8	3	3		
14	7	8	2	2		
15	5	3	2	2		
16	6	4	3	3		
17	3	4	4	3		
18	4	2	1	1		
19	9	3	6	6		
20	7	3	4	4		
21	2	5	2	2		
22	9	8	5	5		
23	6	6	6	6		
24	7	4				
25	4	1				

Imatge 32. Exemple de dades disponibles a un full de càlcul

Observa que la columna B del full (variable "TARDA") conté les mateixes dades de qualificació del curs pel grup de tarda que anteriorment has introduït manualment. A la columna A del full (variable "MATI") hi ha dades corresponents a la mateixa prova pel grup de matí de la mateixa assignatura.

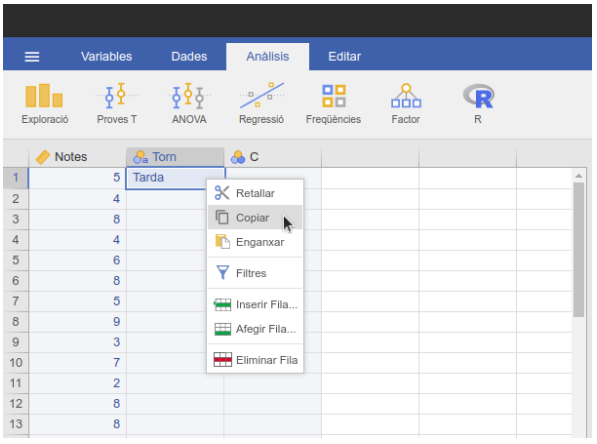
A continuació es mostrarà com ho hem de fer per AFEGIR aquestes dades del MATI a l'arxiu de dades anterior. A més, primer s'afegirà una nova variable a la matriu de dades per a identificar el torn d'assistència de cada cas (individu) al mati o a la tarda. Tot plegat s'utilitzarà per a practicar les operacions disponibles amb *jamovi* i el porta-papers del sistema operatiu.

EXERCICI

1. Si està tancat, obre l'arxiu de *jamovi* que has desat anteriorment amb les dades introduïdes manualment. Si ja el tens obert només cal que continuïs treballant amb aquell arxiu. (Si no has fet cas i no ho has desat, o no saps on ho has desat... hauràs de tornar a fer l'apartat ***"Introducció directa de dades"*** per disposar ara de les dades).
2. Afegeix una nova variable a la matriu de dades de *jamovi*, a la segona columna (que en crear la matriu de nou, automàticament s'ha anomenat B). Per fer-ho, pots consultar la secció anterior.
3. Anomena aquesta nova variable "Torn", i configura-la de tipus de variable "Nominal", i amb tipus de dades "Text".
4. Pots tancar la finestra d'identificació de variables clicant a la icona de tancament de menús .

A continuació introduirem les dades restants fins aconseguir una matriu de dades amb dues variables per a cadascun del 42 casos.

1. A la finestra de dades, clica la casella (buida) del primer individu per a aquesta segona variable "Torn" que has creat.
2. Introdueix manualment el valor de la categoria corresponent: Tarda.
3. Podries fer-ho manualment pels 41 individus restants, però millor utilitzar el procediment de retallar/enganxar: Fes clic-dreta a sobre de la casella que conté el valor "Tarda". Es desplega el menú contextual.
4. Selecciona l'opció "Copiar" (alternativament, combinació de tecles [Ctrl]+[C]).

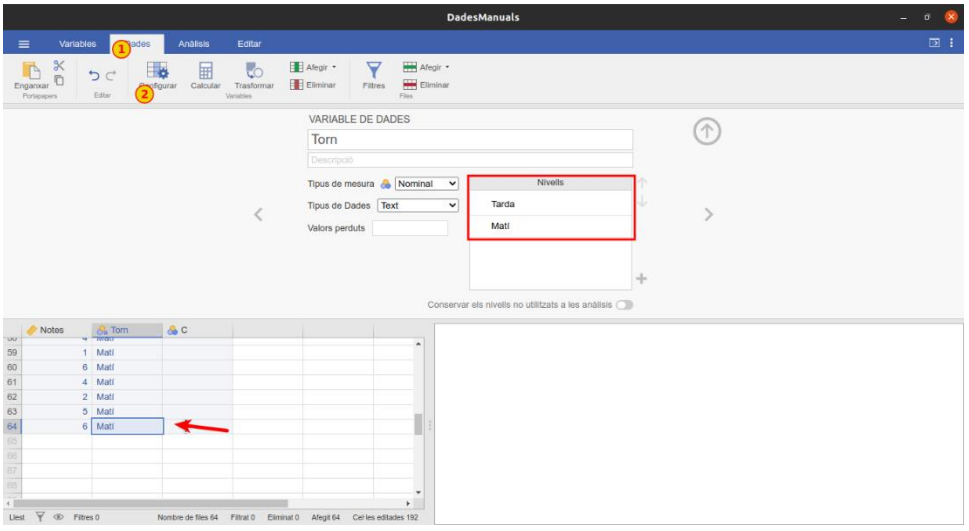


Imatge 33. Menú contextual de dades

5. Amb el ratolí, clica la casella corresponent al Torn del segon individu, i MANTENINT EL DIT "arrossega" cap avall fins a seleccionar totes les caselles buides fins a la fila 42. Aquest és el destí on vols copiar el contingut del porta-papers ("Tarda").
6. Amb totes aquestes caselles de la segona columna seleccionades (en color blau) fes clic-dreta per accedir de nou al menú contextual. Clica l'opció "Enganxar" per a enganxar-hi el porta-papers. Ara tots els 42 primers individus consta que són del torn de tarda.
7. Ves al full de càlcul prèviament obert amb les dades de l'arxiu extern *MJDS_dades_EXCEL.xlsx*. Utilitza el mateix mètode, de clicar i arrossegar mantenint el dit, per a seleccionar totes les notes dels individus del torn de matí (A2 fins a A57).

(cc) BY-NC

8. Clic-dreta, al menú contextual opció "Copiar" (alternativament, combinació de teclat [Ctrl]+[C]). Ara totes les dades del full de càlcul són al porta-papers.
9. Ves a *jamovi*. Clica la casella on vols "començar a copiar" les dades, és a dir, la primera buida de la primera columna, corresponent a la variable "Notes".
10. Clic-dreta, al menú contextual opció "Enganxar" (alternativament, combinació de teclat [Ctrl]+[V]). Ara les dades seleccionades del full de càlcul són a *jamovi*.
11. Per acabar, cal omplir amb el valor "Matí" totes les caselles de la segona columna corresponents a aquests individus (de la fila 43 a la 98). Comença introduint manualment el valor "Matí" a la casella del primer individu d'aquest torn.
12. Utilitza el mateix procés anterior (clica casella + "Copiar" + seleccionar múltiples cel·les + "Enganxar") per acabar d'introduir ràpidament les dades dels 98 individus.



Imatge 34. Retalla/Enganxa dades a la matriu

Si ara obres la finestra d'identificació de variables, amb Menú / Variables, o amb Menú / Data / Configurar (o doble-clic a encapçalament d'una columna) pots comprovar que la variable nominal té dues categories: Matí i Tarda.

Com has pogut veure, és possible (i molt útil) transportar i copiar dades entre diferents aplicacions, i també dins de la matriu de dades de *jamovi*.

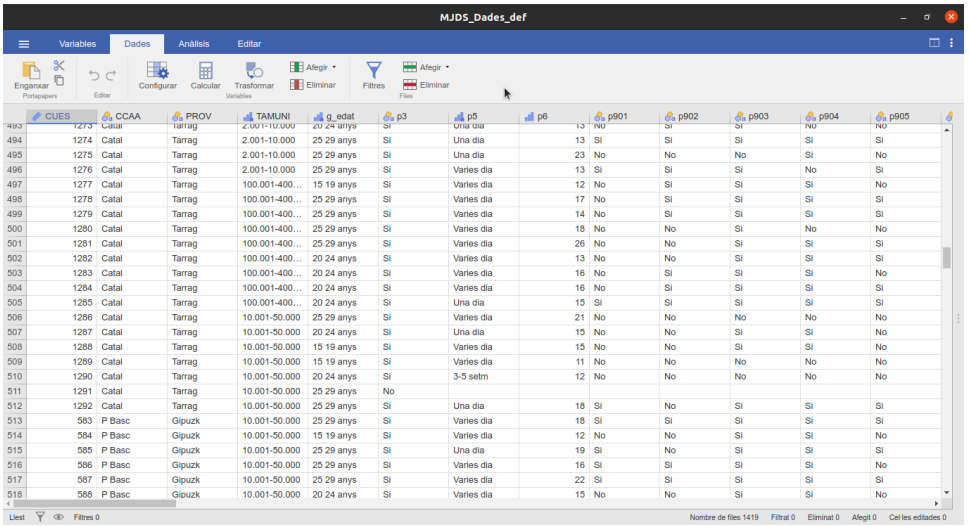
RECOMANACIÓ: Sempre és millor fer retalla/enganxa que no introduir manualment els valors, especialment els "textuals", per estalviar-se problemes d'errors de teclat.

Recorda! A *jamovi* "Matí" és una categoria ("Nivell") totalment diferent de "matí" o de "Mati".

Càrrega de les dades d'arxius d'altres aplicacions

En ocasions ens interessa analitzar amb *jamovi* dades que es troben emmagatzemades en format digital a arxius propis d'altres aplicacions de gestió i anàlisi estadística. El procés d'obertura i importació de dades a *jamovi* funciona amb arxius del propi format (*.OMV) i amb d'altres formats propis de diferents sistemes d'anàlisis de dades, bé siguin fulls de càlcul genèrics o aplicacions d'anàlisi (p. ex., ODS, XLSX, R, SPSS, STATA,...)

Quan s'obre un arxiu de dades en *jamovi* es visualitza com una matriu de dades amb files i columnes on cada fila és un cas (individu de la mostra) i cada columna correspon a una variable (atribut o resposta dels individus).



Imatge 35. Matriu de dades a jamovi

A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic:

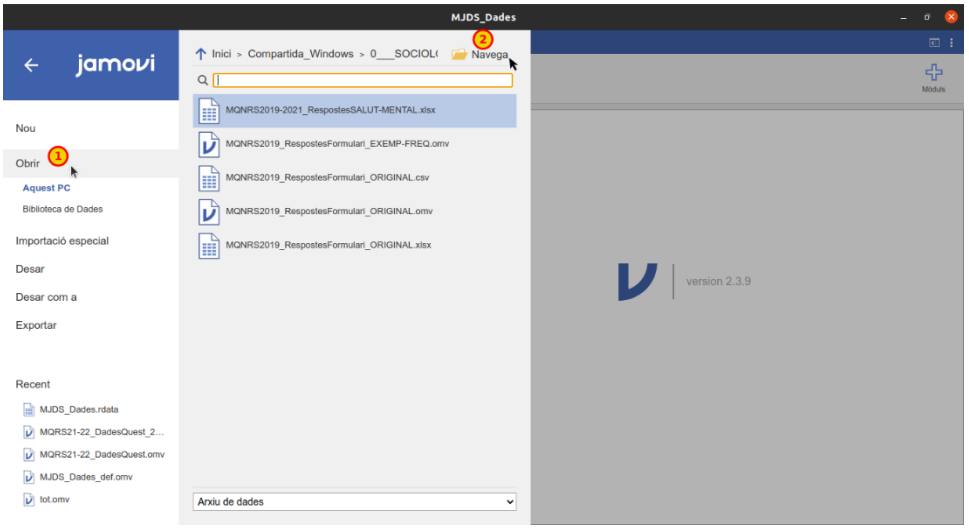
- Existeix una matriu de dades en format R que es vol importar i analitzar amb *jamovi*.

A l'espai de descarrega d'aquest manual (o l'espai docent Campus Virtual CV de l'assignatura a la UAB) hi pots trobar l'arxiu *MJDS_Dades.rdata*, un arxiu de dades corresponents a una enquesta del CIS en format propi de R.

1. Descarrega aquest arxiu format .rdata a una carpeta local del teu PC.
2. Executa l'aplicació *jamovi* al teu ordinador.
3. Cal obrir una nova sessió, sense dades prèvies. Si tens una sessió oberta, considera desar-la en un arxiu *jamovi* (.OMV) per si vols continuar posteriorment a partir d'aquí.
4. Per a obrir el fitxer *MJDS_Dades.rdata* amb *jamovi*, clica el menú d'arxius (dalt, esquerra).

Imatge 36. Ubicació menú d'arxius

5. Selecciona l'opció Obrir / Aquest PC. Al panell central, clica [Navega] (dalt, dreta)



Imatge 37. Obrir un nou arxiu de dades

6. S'obre una finestra on hauràs de seleccionar l'arxiu de dades a obrir. Tria l'arxiu *MJDS_Dades.rdata* que has descarregat al teu PC local.
7. A la part inferior dreta de la pantalla principal apareix una notificació del procés d'obertura d'arxiu i càrrega de dades.
8. Una vegada completat el procés, *jamovi* disposa de la matriu de dades, que conté 1.419 individus entrevistats (files) i 70 variables (columnes).

L'aplicació *jamovi* pot obrir **directament** arxius de dades en diferents formats, corresponents a altres aplicacions abastament utilitzades en l'anàlisi estadística de dades.

- Arxius propis de *jamovi* (.omv)
- Arxius de text amb columnes de dades delimitades per comes (.csv, .txt)
- Arxius propis de l'aplicació SPSS (.sav, .zsav, .por)
- Arxius propis d'R (.RData, .RDS)
- Arxius propis de l'aplicació STATA (.dta)
- Arxius propis de l'aplicació SAS (.xpt, .sas7bdat)
- Arxius propis de l'aplicació JASP (.jasp) [7]

L'opció del menú d'arxius "Importació especial" permet **AFEGIR dades noves a les ja existents** a la matriu de dades de l'espai de treball. D'aquesta forma es poden afegir casos o variables per a poder AJUNTAR diferents arxius "parcials" de dades fent servir l'opció d'importació i seleccionant (amb Ctrl+Clic) els diferents arxius que es vulguin ajuntar en un mateix espai de treball de *jamovi*.

CAS: Existeix un conjunt de qüestionaris en paper, i en el moment d'introduir les dades en format digital es va repartir la tasca entre dues persones, o dues empreses de servei, per accelerar el procediment *("gravació" de dades en paral·lel, per estalviar temps)*. Ara hi ha DOS arxius, i cada arxiu conté les MATEIXES variables i UNA PART dels casos d'enquestes contestades.

- A l'espai Moodle corresponent (PRÀCTICA 3) de l'assignatura, o als arxius adjunts a aquest manual MUJADES, pots trobar dos arxius en format *jamovi* que contenen dades de la mateixa enquesta del CIS: *MJDS_Dades0_A.omv* i *MJDS_Dades0_B.omv*.

7 *JASP* és una aplicació gratuïta per realitzar càlculs estadístics que va ser desenvolupada anteriorment pels mateixos creadors de *jamovi* i actualment s'orienta més cap a l'estadística bayesiana. Es pot descarregar, per a diferents plataformes, a <http://jasp-stats.org>.

- Descarrega els dos arxius al teu PC local.
- Obre cadascun dels arxius per separat. En quin dels dos arxius es troben les dades de les enquestes realitzades a Barcelona? Quantes entrevistes en total conté aquest arxiu?
- Prova d'unificar totes les dades en un nou i únic arxiu de *jamovi*. Obre de nou jamovi, veuràs que la matriu de dades (files i columnes) ha d'estar buida. Utilitza l'opció "Import" del menú d'arxius, tot seleccionant **alhora** els dos arxius en el moment d'indicar QUINS arxius cal carregar a *jamovi*.
- Comprova que el total sigui de 1.419 enquestes, i desa l'arxiu unificat amb el nom que vulguis.

NOTA: En la versió actual (v2.4.8) aquest procediment d'IMPORTACIÓ especial només funciona correctament **si tots** els arxius que s'ajunten són en format propi de *jamovi* (.omv). Amb altres formats els resultats son "desiguals" i, en general, poc fiables.

EXERCICIS PROPOSATS: Lectura i escriptura d'arxius de dades

Després de seguir completament l'apartat del tutorial sobre com gestionar arxius de dades i com introduir i importar dades a *jamovi*, pots respondre a les següents qüestions.

(A) A la carpeta del CV corresponent a la pràctica P3 (o al repositori DDD-UAB on és publicat aquest Manual) hi trobaràs l'arxiu "*MJDS_Dades_EXCEL.xlsx*". A més dels registres de les notes del MATI i la TARDA, a la seva dreta hi trobaràs 4 columnes que corresponen a les notes de 4 grups d'alumnat identificats amb les lletres A, B, C i D.

1. Crea un nou arxiu de *jamovi* i introdueix les dades d'aquests 4 grups de matricula, de forma que es puguin analitzar posteriorment segons la nota i el grup de cada alumne.
2. Lliura al CV l'arxiu *jamovi* (*.omv) amb les dades aquestes dades.

(B) Al capítol 4 del manual MUJADES apareix una taula amb dades sobre la "població" de passatgers del Titanic. Aquesta taula mostra un recompte de persones amb determinades característiques, combinant dos criteris, un a les files i altre a les columnes: un criteri (o variable) és en quina classe viatjaven i l'altre és si van sobreviure o no.

1. Cerca a Internet una taula similar a aquesta, sobre qualsevol població. Pots trobar d'altres exemples <aquí>, o a qualsevol altre Institut d'Estadística oficial.
2. Retalla i enganxa la taula que hagis seleccionat a un document de text (ODT o DOCX), i completa aquest document afegint també la referencia de l'adreça URL on es pot consultar la taula original. Genera un PDF a partir d'aquest document.
3. Crea un nou arxiu de *jamovi* i introdueix aquestes dades de forma que es puguin analitzar posteriorment.
4. Lliura al CV els dos arxius: (1) el document amb la adreça i la taula (*.pdf) i l'arxiu de jamovi amb les dades (*.omv).

(cc) BY-NC

5. Definició de dades: Tipus de variables i valors perduts

Correspondència entre nivell de mesura de les variables i tipus de variable/dada en jamovi

En tota anàlisi estadística el primer que cal és demanar-se amb quin tipus de variable ens trobem, des del punt de vista del nivell de mesura. A efectes estadístics hi ha dos grans tipus de variables, d'una banda les qualitatives o categòriques (on sovint interessa distingir entre nominals i ordinals) i, d'altra les quantitatives o numèriques.

A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic:

- Tenim un qüestionari realitzat pel CIS, i es vol preparar una matriu de dades per a treballar amb *jamovi*.
- Disposem d'una versió en format R d'aquestes dades (descarregada de la Web del CIS), que pot ser importada dins de *jamovi* com a base per a la identificació de les variables.

A l'espai docent Campus Virtual (CV) de l'assignatura, o als arxius adjunts a aquest manual MUJADES, hi pots trobar els arxius:

- *MJDS_QüestionariCIS.pdf*, que reprodueix una mostra des qüestionari del CIS amb diverses preguntes, i
- *MJDS_Dades.rdata*, que inclou les respostes de les 1.419 persones entrevistades a 70 variables.

1. Descarrega tots dos arxius a una carpeta local del teu PC.
2. Obre l'arxiu PDF, per tenir-lo disponible al llarg del següent tutorial, i poder identificar les variables i les possibilitats de resposta ("nivells" a *jamovi*).
3. Les variables tenen com a nom un codi format per la lletra "p" i la xifra que expressa l'ordre de la pregunta al qüestionari.
4. Comprova en el qüestionari del CIS que hi ha exemples de variables dels diferents tipus o de mesura. Per exemple:
   - Variables **nominals**: Utilització o no d'Internet (p3), sexe (p40), estat civil (p41).
   - Variables **ordinals**. Freqüència d'utilització d'Internet (p5), probabilitat de quedar amb algú que hagi conegut per Internet (p16), nivell d'estudis del pare (p54).
   - Variables **quantitatives**. Edat que tenia en començar a utilitzar Internet (P6), nombre de contactes a les xarxes socials (p15), escala ideològica (p33).[8]

Encara que aquests són els 3 tipus de variables més habituals en les anàlisi de dades socials, *jamovi* utilitza una altra terminologia. La combinació de TIPUS de VARIABLE amb de TIPUS de DADES permet distingir entre dotze modalitats de variables:

8 Les escales d'opinió o d'auto-posicionament són, en realitat, variables de tipus ordinal, però hi ha un ampli consens (poc fonamentat) de considerar-les com a numèriques o quantitatives a efectes estadístics si tenen més de 7 opcions de resposta.

(cc) BY-NC

		Tipus de variable *(jamovi)*			
		"Contínua"	"Ordinal"	"Nominal"	"ID"
Tipus de les dades *(jamovi)*	"Enter"	1, 2, 99,...	1, 2, 99,...	1, 2, 99,...	1, 2, 1001369
	"Decimal"	0.5, 12.342,...	impossible	impossible	Impossible
	"Text"	impossible	Jove (<30 anys)	Sí, No, Home,...	ABx76yD

Seguint la terminologia de *jamovi*, segons el tipus **de la variable** hi ha:

- **"Contínua":** Per a variables quantitatives (numèriques) que disposen d'una unitat d'escala, com ara el nombre de germans, o quants dies a la setmana es fa esport.
- **"Ordinal":** Per a variables ordinals que permeten establir una relació de més o menys entre cada parell de categories en relació al concepte mesurat, com ara el nivell educatiu assolit, o l'opinió respecte d'un tema *(D'acord, Neutral, En desacord)*.
- **"Nominal":** Per a variables nominals que permeten classificar tots els casos en grups mútuament excloents, com ara la religió professada, o la participació en un esdeveniment *(Sí, No)*.
- **"ID":** Per a identificar de forma única cada cas, persona o element de la mostra, com ara el número de identificació fiscal, una adreça de correu electrònic, o el teu NIU. No mostra les categories a la finestra d'identificació, estalviant recursos del PC.

NOTA: Resulta molt convenient utilitzar aquest tipus ID en el cas de dades textuals corresponents a les respostes literals a preguntes **obertes** d'un qüestionari.

En funció del tipus de variable *jamovi* permet la realització o no de determinats procediments estadístics. Per aquesta raó és IMPRESCINDIBLE tenir definit de forma adient el tipus de variable per poder realitzar els procediments pertinents.

A més, segons el tipus **de les dades**, *jamovi* distingeix entre:

- **"Enter"**: Dades numèriques sense decimals, o variables ordinals identificades "internament" mitjançant un codi numèric (factor) però amb una "etiqueta" associada a cada codi (p. e., intervals).
- **"Decimal"**: Només per a variables contínues; tractades com a números amb decimals.[9]
- **"Text"**: Només per a variables nominals o ordinals, tractades com a text, cadenes de caràcters alfabètics i/o numèrics.

En funció del tipus de dades *jamovi* utilitza determinats formats per a l'emmagatzematge intern de les dades, la qual cosa té conseqüències en la generació de resultats estadístics. Resulta rellevant en algunes anàlisis, perquè permet el tractament "dual" d'una mateixa variable (nominal o ordinals) bé com a factor d'agrupació o bé com a quantitatives.

9 Els nombres decimals a jamovi utilitzen un **PUNT** (i no una coma) per a separar la part sencera de la decimal, seguint l'estàndard anglosaxó de notació .

(cc) BY-NC

En quant a la correspondència entre el nivell de mesura i les classes de variables/dades en *jamovi*, és RECOMANA mantenir la següent lògica:

- **Nivell de mesura nominal**. Exemples al qüestionari CIS: utilització o no d'Internet (p3), sexe (p40) o estat civil (p41).
  => Cal que estiguin definides com a mesura="Nominal" i dada="Text".
- **Nivell de mesura ordinal**. Exemples al qüestionari CIS: freqüència utilització d'Internet (p5), probabilitat de quedar amb algú que hagi conegut per Internet (p16) o nivell d'estudis del pare (p54).
  => Cal que estiguin definides com a mesura="Ordinal" i dada="Text".
- **Nivell de mesura quantitatiu**. Exemples al qüestionari CIS: edat que tenia en començar a utilitzar Internet (P6), nombre de contactes a les xarxes socials (p15) o escala ideològica (p33).
  => Tot i que el més lògic és que la seva definició fos mesura="Contínua" i data="Enter", aquesta definició en pot portar algun problema, ja que si la mesura és *Continua* no permet, per exemple, fer una taula de freqüències. Per aquest motiu, en un moment inicial les definirem com a mesura="Ordinal" i data="Enter" i avaluarem, en alguns casos, la possibilitat de canviar la definició al llarg de l'anàlisi.

RECORDA: A *jamovi*, les variables nominals son de mesura Nominal amb dades tipus Text, les variables ordinals son de mesura Ordinal amb dades tipus Text, i les variables quantitatives son de mesura Ordinal amb dades Enter (o Decimal si tenen decimals).

Tal i com veurem a continuació, en el moment de processar una ordre d'obertura d'un arxiu de dades que no sigui del seu propi format, l'aplicació *jamovi* **prova d'estimar el tipus** de variable a partir de les dades contingudes a l'arxiu. Aquesta identificació automàtica **no sempre** és la forma més adient per definir les variables.

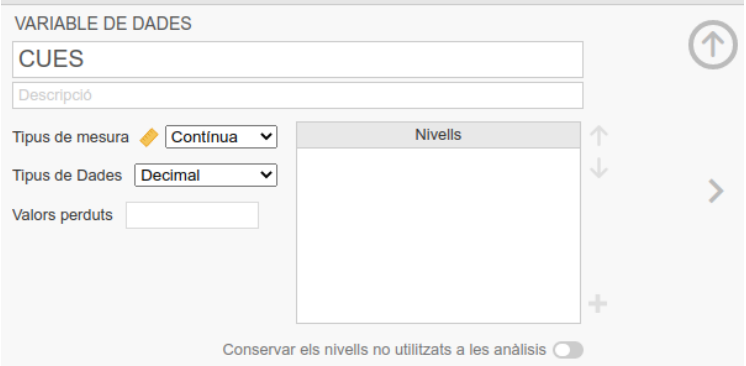
Sempre cal revisar la definició de cadascuna de les variables quan s'obre (o s'importa) amb *jamovi* un arxiu de dades amb variables que no hem definit personalment.

A continuació el tutorial t'acompanyarà en la identificació de variables a partir de l'arxiu en format R (*.rdata).

1. Obre des de dins de *jamovi* l'arxiu de dades R anomenat *MJDS_Dades.rdata*.
2. Per accedir al panell de definició de les variables, clica la pestanya "Dades" del menú superior, i a continuació clica la icona "Configurar". Comprova com clicar aquesta icona serveix per a obrir o tancar aquest panell.
   (Una drecera per aconseguir el mateix efecte és fer doble-clic al nom de la variable que encapçala cadascuna de les columnes de la matriu de dades).

Imatge 38. Accés a l'àrea d'identificació de variables

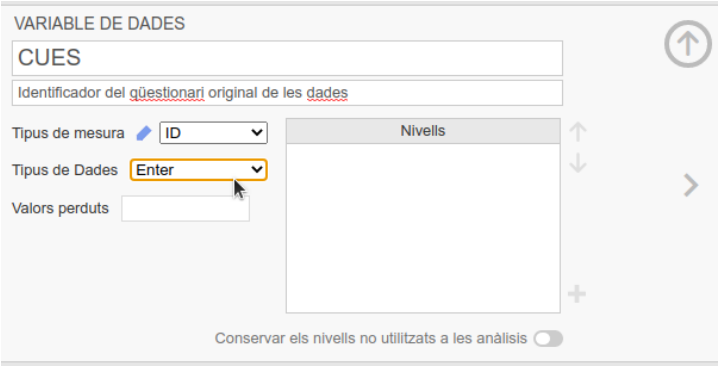
3. El panell d'identificació i definició de variables es pot amagar fent clic a la fletxa encerclada que es mostra a la seva part superior dreta.
4. Es pot passar seqüencialment d'una variable a la següent o anterior clicant les fletxes > < que es mostren a ambdós laterals del panell.
5. En aquest cas, obre el panell de definició per a la variable anomenada "CUES".



Imatge 39. Identificació automàtica (Obrir) de la variable CUES

Tal i com pots veure, *jamovi* ha identificat automàticament la "CUES" com a variable numèrica (*Contínua + Decimal*) a partir de les dades originals (nombres enters) i de la seva definició anterior a l'arxiu R. Però aquesta definició ***és incorrecta***, perquè CUES és un simple identificador (ID) del qüestionari d'on es van extreure les dades, per tant, per a definir correctament d'aquesta variable fes el següent:

1. Clica al menú desplegable del tipus de variable, per canviar a ID,
2. Canvia també el tipus de dada, i identifica-la com a "Enter" (nombre sencer),
3. A sota del nom de la variable hi ha un espai per incloure la seva descripció, i pot ser molt útil teclejar aquí, per exemple, "Identificador del qüestionari original de les dades".



Imatge 40. Correcta identificació de la variable CUES

Les dues variables següents, CCAA i PROV, són variables que tenen un nivell de mesura nominal. La definició automàtica que n'ha fet *jamovi* és "Nominal", en quant a la mesura, cosa que és correcta; i "Enter" en quant al tipus de dada. Cosa que **és incorrecta**, ja que son els noms de comunitats i de províncies, i haurien d'estar definides com a "Text".
Per a definir-les correctament fes el següent:

1. No canviïs el tipus de variable,
2. Clica sobre el selector desplegable de "Tipus de Dades", i canvia el tipus a "Text",
3. A la descripció (sota del nom de la variable) pot ser útil afegir respectivament "Comunitat Autònoma" i "Província".

Pel què fa a les dues següents variables, anomenades TAMUNI i g_edat, observa que han estat definides automàticament com a "Nominal" i "Enter". En ambdós casos, ambdues coses són **incorrectes**, ja que es tracta de variables tipus "Ordinal" (la primera, amb 7 categories que representen intervals de població del municipi i, la segona, amb 3 grups quinquennals d'edat, ambdues perfectament ordenables de menor a major), i amb dades de tipus "Text". Per tant, cal canviar ambdues característiques. En cadascuna d'aquestes dos variables, fes:

1. Clica sobre "Nominal", i canvia el tipus de variable a "Ordinal",
2. Clica sobre el selector desplegable de tipus de Dades, i canvia el tipus a "Text",[10]
3. A la descripció (sota del nom de la variable) pot ser útil afegir, respectivament, "Mida del municipi" i "Grup d'edat".

IMPORTANT: En el cas de les variables de tipus "Ordinal" és possible que *jamovi* les defineixi automàticament de forma correcta, però que l'ordenació de les diferents categories sigui incorrecta. Sovint succeeix que quan es fa una importació de dades *jamovi* ordena alfabèticament les categories, i no segons la ordenació "real" dels diferents intervals o "nivells" de la variable.

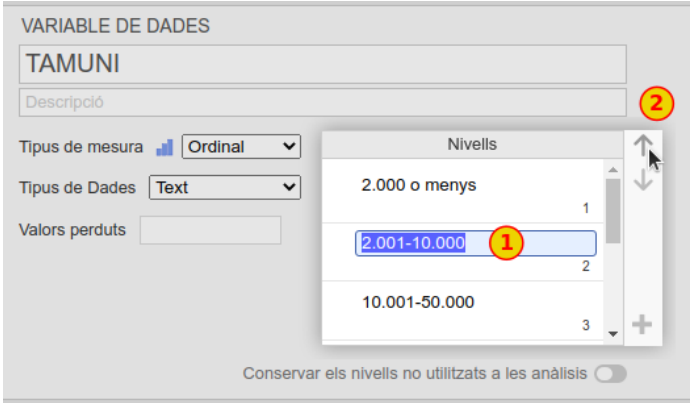
Comprova ara si les categories o "nivells" de les variables TAMUNI i g_edat han quedat correctament ordenades durant el procés d'importació de les dades.
En aquest cas **no cal canviar l'ordre** dels nivells: s'han importat correctament a la matriu de dades de *jamovi*.

Si calgués **modificar l'ordre** de les categories d'una variable "Ordinal":

- Clica a sobre d'una de les categories *("Nivells")* i, quan quedi seleccionada, fes servir les fletxes que apareixen a la part dreta de la finestra de categories per corregir l'ordenació "movent" el nivell seleccionat cap amunt o avall.

10 Tot i que representen intervals numèrics, les "etiquetes" de cada categoria són en realitat textuals.

(cc) BY-NC



Imatge 41. Procediment d'ordenació manual de categories

NOTA: Aquesta ordenació manual és especialment rellevant en les variables noves que generis a partir d'alguna transformació mitjançant *jamovi*. També es pot fer amb les variables de tipus "Nominal" per motius purament estètics (per exemple, mantenir per a totes les variables "No" com a primera opció i "Sí" com a segona). NOMÉS és possible si la variable ordinal te les dades definides com tipus "Text"; no funciona per a dades definides com a tipus "Enter" o "Decimal".

Les "no respostes" i la definició de valors perduts en jamovi

Hauràs pogut observar que en la part esquerra inferior de la definició de la variable hi ha una pestanya que diu "Valors perduts" (*Missing values*, en anglès). Sovint algunes persones entrevistades deixen de contestar alguna de les preguntes del qüestionari. Són els *no ho sap*, *no contesta* o *no procedeix*, que tècnicament es coneixen amb el nom de "valors perduts".

És pràctica habitual, fer el següent:

- Si la pregunta *no procedeix*, és a dir, no cal fer la pregunta degut a la seva resposta a preguntes anteriors, se sol deixar **en blanc** sense anotar res. Per exemple, si abans ja s'ha respòs que no disposa de xarxes socials, no procedeix preguntar quants amics té a les xarxes socials. Per aquesta persona, doncs, es deixaria la informació sobre els amics en blanc.
  La presència, en una variable, de molts *valors que no procedeixen* pot comportar un problema en la representativitat de les dades. Així, es pot donar el cas que la mostra sigui prou gran per assegurar representativitat estadística (assegurar que els resultats i conclusions fonamentats en la mostra es poden "aplicar" al conjunt de la població), però que la presència de molts valors perduts en una variable concreta redueixi, en aquest cas, la mida de la mostra i que, per tant, no se'n puguin treure conclusions extrapolables a la població per aquella variable o pregunta en concret.
- Si la persona ha decidit no contestar, ja sigui perquè no ho sap o perquè no vol respondre, se sol assignar un **codi especial**: en variables de text sol ser "N.S." o "N.C.", mentre que en variables numèriques se solen deixar codis que no tinguin un significat substantiu, per exemple 98 i 99, si aquests valors no existeixen, o 998 i 999, en cas contrari.
  A més dels possibles efectes sobre la representativitat, cal ser molt curós amb la utilització de codis o valors numèrics per a expressar valors perduts de la variable, ja que si no s'eliminen d'alguna manera, s'estarà "deformant" la realitat social mesurada.

Quan *jamovi* troba un **espai en blanc** a les dades li assigna, automàticament, el significat de "Valor perdut" en la matriu de dades, i el considera com a "*Perdut*" en les anàlisis. Aquest codi (en blanc)

es pot considerar com el valor perdut per defecte o **del sistema**. En canvi, els altres, els que no es corresponen a blancs, cal que els definim nosaltres expressament com a valors perduts **d'usuari**, de manera que *jamovi* tampoc els tingui en compte en els procediments de càlcul corresponents.

Definició de valors perduts en jamovi (I): Variables nominals i ordinals

Per poder entendre millor el què estem fent quan definim valors perduts, cal explicar abans una funció que és més pròpia de l'estadística descriptiva (següent capítol): la taula de freqüències.

Anem a fer una taula de freqüències bàsica de tres variables, TAMUNI que ja tenim ben definida, i p3 i p5 que prèviament caldrà revisar i definir correctament:

1. Entra a "Configurar" de p3 i p5, defineix-les adequadament i anomena-les, respectivament, "Ús d'Internet en els darrers tres mesos" i "Freqüència d'ús d'Internet".
2. Son variables, respectivament, de tipus Nominal (p3) i Ordinal (p5), amb dades de tipus "Text" (p3) i "Enter" (p5).
3. A la pestanya "Anàlisis" selecciona l'opció "Exploració", a l'extrem esquerra.
4. Al menú desplegable, selecciona l'opció "Descriptives". S'obre un panell de definició del procediment.

Imatge 42. Panell típic de definició de procediments estadístics

5. A l'esquerra del panell hi ha un llistat amb totes les variables disponibles (identificades amb la corresponent icona de nivell de mesura). Busca i selecciona amb un clic TAMUNI, p3 i p5. A la part central superior hi ha un botó QUADRAT FLETXA-DRETA (no confondre amb la fletxa-encerclada, que serveix per a tancar aquest panell). Clica per "passar" les variables a l'espai dret del panell, on hi haurà les variables que s'utilitzaran per a generar l'anàlisi.
6. Com pots veure, a la finestra de resultats de la dreta van apareixen, immediatament i de forma actualitzada, els resultats (segons el tipus de variable).
7. Per defecte aquest procediment només genera 7 resultats: recompte de casos totals (N), recompte de casos perduts *(Missing)*, mitjana aritmètica *(Mean)*, desviació típica *(Standard deviation)*, mediana *(Median)*, valor mínim *(Minimum)* i valor màxim *(Maximum)* de la distribució. Com que en aquest cas estem analitzant tres variables qualitatives només ens interessa el nombre de casos i el recompte dels perduts.
8. Per a generar taules de freqüències cal activar específicament aquesta opció: clica el botó a sota del llistat dret de variables disponibles "[ ] Taula de freqüències". Els resultats s'actualitzen automàticament per a variables definides com tipus NOMINAL o ORDINAL (tal com indiquen les icones al costat del botó que acabes d'activar). El mateix botó serveix per a desactivar les taules de freqüències.
9. Per amagar el panell central pots clicar a la FLETXA-EN-CERCLE de la part superior.

10. Si vols visualitzar el panell central de definició d'un procediment, fes doble-clic a sobre de qualsevol resultat (taules) al panell dret de resultats.

Si et fixes en els resultats de "Descriptives", i referint-nos a les variables p3 i p5, veuràs que p3 no conté cap valor que es correspongui a un perdut mentre que a p5 n'hi ha 118. Són els que no han fet servir Internet en els darrers tres mesos. En aquest cas, la persona encarregada de generar la base de dades ha decidit deixar en blanc la pregunta quan aquesta no era procedent (degut a la resposta negativa a p3), mentre que, com veiem a la taula de freqüències de més avall, si no contestava a p5 li ha assignat el codi "N.C.":

Descriptives

	TAMUNI	p3	p5
N	1419	1419	1301
Perduts	0	0	118
Mitjana			1.68
Mediana			1
Desviació estàndard			1.10
Mínim			1
Màxim			7

Freqüències de p3

Nivells	Freqüències	% del Total	% Acumulat
Si	1301	91.7 %	91.7 %
No	118	8.3 %	100.0 %

Freqüències de p5

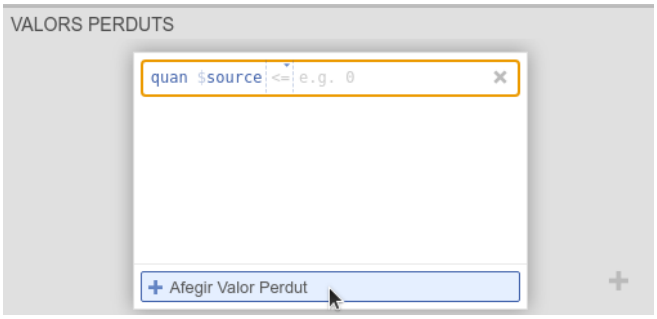
Nivells	Freqüències	% del Total	% Acumulat
Varies dia	822	63.2 %	63.2 %
Una dia	244	18.8 %	81.9 %
3-5 setm	123	9.5 %	91.4 %
1-2 setm	69	5.3 %	96.7 %
Menys 1 setm	30	2.3 %	99.0 %
Gaireb mai	12	0.9 %	99.9 %
N.C.	1	0.1 %	100.0 %

Imatge 43. Resultat típic d'anàlisi descriptiva, amb valors perduts

ATENCIÓ: Recorda que jamovi només "executa" ordres del llenguathe R. Com que R "escriu" els resultats *"en anglès"*... els punts (.) que apareixen a les xifres dels resultats son **separacions decimals** *(que nosaltres denotem amb comes)*.

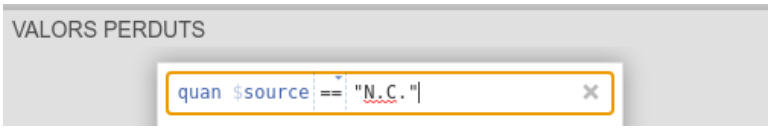
De fet, aquest codi "N.C.", assignat als que no han contestat, no té cap significat substantiu a la variable "Freqüència d'ús d'Internet", de manera que cal descartar-ho; el definirem com a "valor perdut". L'actual versió de *jamovi* permet definir valors perduts en la mateixa definició de la variable, en la pestanya anomenada "Valors perduts".
Per fer-ho:

1. Fes clic al menú Dades, i a Configurar. Selecciona la variable p5, per tornar a obrir el panell de definició d'aquesta variable. (Alternativament, a la matriu de dades, fes doble clic al nom de la variable.)
2. Clica sobre l'espai que hi ha a la dreta de "Valors perduts", i s'obrirà una nova finestra.
3. Clica al botó inferior [+ Afegir Valor Perdut].



Imatge 44. Panell per definir valors perduts

4. T'apareix el text "quan $source..." i un lloc per omplir. Escriu-hi =="N.C.", tal i com es mostra a la imatge:



Imatge 45. Exemple d'escriptura al panell per definir valors perduts

El què has fet és definir, com a perduts, els valors que en la variable original (*$source*) equivalguin a (==) "N.C.".
Observa que hem fet servir una doble igualtat (==), i no una igualtat simple(=), i que *N.C.* en tractar-se d'un text, l'hem d'escriure "entre cometes".

5. Si volguessis afegir una altra categoria a la de valors perduts, tornaries a clicar sobre el botó [+ Afegir Valor Perdut], i escriuries una altra categoria (per exemple, "N.S.", si fos cas).

6. Per tornar a la pantalla principal, clica a la FLETXA-EN-CERCLE-AVALL de la part superior dreta.

Observa que aquest procediment ha tingut un efecte directe sobre els resultats que havies demanat: la variable p5 ara té 1.300 casos vàlids (un menys que abans) i 119 de perduts (un més). A més, en fer freqüències ha desaparegut la categoria "N.C.", de manera que ara la freqüència relativa, o percentatge, només es fa en relació a 1.300 casos vàlids:

Descriptives

	TAMUNI	p3	p5
N	1419	1419	1300
Perduts	0	0	119
Mitjana			1.67
Mediana			1.00
Desviació estàndard			1.09
Mínim			1
Màxim			6

Freqüències de p5

Nivells	Freqüències	% del Total	% Acumulat
Varies dia	822	63.2 %	63.2 %
Una dia	244	18.8 %	82.0 %
3-5 setm	123	9.5 %	91.5 %
1-2 setm	69	5.3 %	96.8 %
Menys 1 setm	30	2.3 %	99.1 %
Gaireb mai	12	0.9 %	100.0 %

Imatge 46. Resultat d'anàlisi descriptiu, eliminats els valors perduts

OBSERVACIÓ IMPORTANT! Aquesta definició de perduts no altera la variable, de manera que en qualsevol moment la pots desactivar i tornarà aparèixer el valor *N.C.* a les taules. Per desactivar-la has de tornar a entrar a la finestra de "Valors perduts" i clicar sobre la creueta que hi ha a la dreta de cada definició de valor perdut.

(cc) BY-NC

ATENCIÓ! Fins a la versió 2.3.9, i només en el cas de variables Nominals o Ordinals, *jamovi* només accepta **fins a DOS** valors o nivells definits com a "perduts" *(degut a un error intern, corregit a versions posteriors)*. En aquest cas, si cal definir MES DE DOS nivells d'una variable com a perduts la estratègia correcta és crear una nova variable, mitjançant una transformació, de forma que contingui <u>només</u> els nivells correctes. Els casos que quedin en blanc seran tinguts com a "perduts".

Definició de valors perduts en jamovi (II): Variables quantitatives

La lògica dels valors perduts quan la variable és quantitativa és pràcticament idèntica a l'anterior. A més, la necessitat de tenir ben definides les variables és encara més important, ja que la presència de valors no substantius comporta errors molt importants en els càlculs estadístics. Ho exemplificarem amb la variable p33, que conté informació sobre el posicionament ideològic de la persona entrevistada (en una escala on 1 és extrema esquerra i 10 extrema dreta).

Per veure-ho anem, un cop més a utilitzar els descriptius, on veurem un resum de la variable, així com una taula de freqüències. Abans, però definim correctament aquesta variable quantitativa, que tal i com s'explica a l'inici del capítol és: mesura="Ordinal" i data="Enter".

1. Entra a p33 (doble clic) i veuràs que, per defecte, *jamovi* l'ha definida, **incorrectament**, com mesura="Contínua" i data="Decimal".
2. Canvia-ho i anomena-la *Autoposicionament polític (1-10),* com a mesura="Ordinal" i data="Enter".
3. A la pestanya "Anàlisis" selecciona l'opció "Exploració", a l'extrem esquerra. I al menú desplegable l'opció "Descriptives".
4. Selecciona, del panell de l'esquerra, la variable "*Autoposicionament polític (1-10)"*, i amb el botó central **[->]** FLETXA-DRETA envia-la a la dreta (finestra "Variables").
5. Activa també l'opció "[ ] Taules de freqüències".

Els resultats obtinguts són els següents:

Descriptives

	Autoposicionament polític (1-10)
N	1419
Perduts	0
Mitjana	38.2
Mediana	6
Desviació estàndard	44.7
Mínim	1
Màxim	99

Freqüències de Autoposicionament polític (1-10)

Nivells	Freqüències	% del Total	% Acumulat
1	55	3.9 %	3.9 %
2	60	4.2 %	8.1 %
3	157	11.1 %	19.2 %
4	131	9.2 %	28.4 %
5	264	18.6 %	47.0 %
6	98	6.9 %	53.9 %
7	83	5.8 %	59.8 %
8	38	2.7 %	62.4 %
9	6	0.4 %	62.9 %
10	15	1.1 %	63.9 %
97	261	18.4 %	82.3 %
98	178	12.5 %	94.9 %
99	73	5.1 %	100.0 %

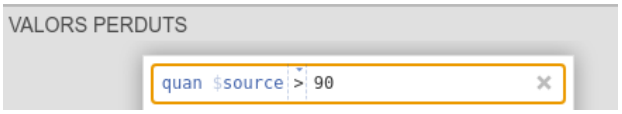
Imatge 47. Resultat típic d'anàlisi descriptiva, amb valors perduts

En aquest cas, on encara no hem definit cap valor perdut, observem que la mitjana d'autoposicionament polític és de 38,2, un valor <u>impossible</u> en una variable que només pot tenir valors entre 1 i 10. El motiu el veiem a la dreta, en la taula de freqüències, on observem que hi ha tres valors, 97, 98 i 99, que no són pròpiament un posicionament polític sinó que es corresponen (vegeu qüestionari) a "no té cap ideologia", "no ho sap" i "no contesta", respectivament. Aquests

valors, que poden tenir el seu interès en algun moment, hauríem de descartar-los de moment si volem saber és la mitjana de la variable o altres indicadors estadístics que no els tinguin en compte.

Per fer-ho seguirem la mateixa lògica que abans:

1. Obre el panell de definició de la variable "Autoposicionament polític (1-10)".
2. Clica sobre l'espai que hi ha just a la dreta de "Valors perduts", i s'obrirà una nova finestra de diàleg; i clica al botó inferior [+ Afegir Valor Perdut]
3. En el text "quan $source...", omple'l de la següent manera:



Imatge 48. Exemple d'escriptura al panell per definir valors perduts

El què has fet és definir, com a perduts, els valors que en la variable p33 (*$source*) siguin més grans (>) a 90. Observa que haguessis pogut escriure >=97 o >10; i que el resultat fora el mateix.
En aquest cas com que es tracta d'un valor numèric, no l'has descriure entre cometes.

4. Per tornar a la pantalla principal, clica a la FLETXA-EN-CERCLE-AVALL de la part superior dreta.

El procediment ha tornat a tenir un efecte directe sobre els resultats que havies demanat: la variable p33 ara té 907 casos vàlids, 512 de perduts i tots els indicadors estadístics han canviat, de manera que ara la mitjana és de 4,59 punts. A més, en fer freqüències ja només veiem els valors compresos entre 1 i 10:

Descriptives

	Autoposicionament polític (1-10)
N	907
Perduts	512
Mitjana	4.59
Mediana	5
Desviació estàndard	1.90
Mínim	1
Màxim	10

Freqüències de Autoposicionament polític (1-10)

Nivells	Freqüències	% del Total	% Acumulat
1	55	6.1 %	6.1 %
2	60	6.6 %	12.7 %
3	157	17.3 %	30.0 %
4	131	14.4 %	44.4 %
5	264	29.1 %	73.5 %
6	98	10.8 %	84.3 %
7	83	9.2 %	93.5 %
8	38	4.2 %	97.7 %
9	6	0.7 %	98.3 %
10	15	1.7 %	100.0 %

Imatge 49. Resultat d'anàlisi descriptiu, eliminats els valors perduts

OBSERVACIÓ IMPORTANT! Tal i com passava abans, aquesta definició de perduts no altera la variable. Si observes la columna de la variable "Autoposicionament polític (1-10)", segueix havent-hi aquests valors que hem definit com a perduts. Ara apareixen en un color clar a la matriu de dades. L'important és que ara consten com a *Valors perduts*; no es tenen en compte en els càlculs però sempre els podem recuperar, si ho necessitem.

EXERCICI

1. Repassa **TOTES** les variables de l'arxiu *MJDS_Dades.rdata* i defineix-les correctament en quant al nivell de mesura.
2. És habitual cometre algun error en fer-ho. Si et passa veuràs que **en alguns casos no és possible corregir aquest error** (sobretot si, per error, has definit una variable com a decimal). Si tens aquest problema es recomana tancar immediatament el fitxer **sense** gravar-lo i tornar-lo a obrir. D'aquesta forma recuperaràs totes les modificacions fetes fins a l'última vegada que vas desar l'arxiu.
3. Al mateix temps, ves introduint descripcions per a cada variable, basades en les preguntes literals del qüestionari. Per fer-ho, ajuda't de l'arxiu *MJDS_QüestionariCIS.pdf*.
4. Controla amb la funció estadística freqüències si hi ha valors perduts en les variables que s'escaigui. Defineix aquests valors perduts.
5. Recorda **desar** sovint els resultats de les teves modificacions. Utilitza el menú d'Arxiu / Desar com a ... i desa un arxiu en format propi de *jamovi* (per exemple, *MJDS_Dades_def.omv*) en una ubicació que puguis recuperar un altre dia (USB, núvol,...).

ATENCIÓ ! Si no fas l'exercici fins tenir revisades i ben identificades TOTES les variables de l'arxiu, i no has desat les dades en format propi de *jamovi* (*MJDS_Dades_def.omv*) no podràs continuar amb el tutorial a partir d'aquest punt.

EXERCICIS PROPOSATS: Definició de variables i valors perduts

(A) A la carpeta del CV corresponent (o al repositori DDD-UAB on és publicat aquest Manual) hi trobaràs l'arxiu anomenat "*MJDS_DadesENQUESTA_22nov.xlsx*". Es tracta d'un extracte de les dades recopilades mitjançant el qüestionari realitzat al curs 2021-2022. Hi trobaràs 6 columnes que corresponen a les respostes d'un conjunt de persones entrevistades a 6 de les preguntes del qüestionari, amb informació sobre la seva Universitat i Estudis, gènere, despesa en oci, edat i motivació.

1. Crea un nou arxiu de *jamovi* i introdueix les dades de resposta del qüestionari, de forma que posteriorment es puguin processar i analitzar.
2. Ajusta correctament les definicions del tipus de variable en funció del tipus de dades.
3. Realitza la depuració de les dades: Revisa les distribucions dels valors de les variables fins a detectar errades, incoherències i "valors perduts". Utilitza las capacitats de l'editor de *jamovi* per a fer comentaris de text a les taules que hagis generat, per indicar quin son el valors o casos problemàtics i com els tractaràs.
4. Modifica a *jamovi* les dades de forma corresponent (valors perduts, canviar valors, eliminar, etc...) per tal de solucionar els problemes de l'arxiu i deixar-ho preparat per a ser analitzat.
5. Finalment, desa el nou arxiu de *jamovi* amb les dades depurades.

Fins ara el manual MUJADES t'ha acompanyat en la introducció a *jamovi*. Ja has vist què és, com descarregar-la i instal·lar-la al teu propi PC. (També existeix una versió remota, per executar per Internet, accessible des de la pàgina web de jamovi, tot i que encara està en desenvolupament i no és tan fiable com treballar amb la instal·lació local).
Igualment, has vist com introduir "manualment" dades de les que disposis, i com desar-les en format *jamovi* (*.OMV) i com importar informació dins la matriu de dades: des de jamovi, o des d'altres aplicacions.

Has pogut comprovar la importància de fer una definició **correcta** de les variables, així com la rellevància del tractament dels casos "perduts".

Fins aquí has seguit els primers passos amb l'aplicació, fins a disposar d'una matriu de dades amb informació, per exemple, pròpia de les respostes a una enquesta social per qüestionari.

En els apartats següents el Manual MUJADES t'acompanyarà en l'aprenentatge de com gestionar i millorar la informació de que disposes, així com en una introducció als procediments bàsics d'estadística descriptiva aplicats a les variables disponibles.

6. Estadística descriptiva i exploratòria bàsica amb jamovi

Exploració bàsica descriptiva de les variables

Es fonamental poder "entendre" de forma simple i sintètica com es distribueixen els valors de les variables, per comprendre l'extensió de l'atribut o resposta que representen. El tipus de la variable determina quines operacions estadístiques se'n poden realitzar i, per tant, quins estadístics es poden calcular per a cada variable.

Així, una forma simple i adient de descriure o explorar la distribució d'una variable de tipus nominal o ordinal és generar la seva **taula de freqüències**. Contràriament, en el cas d'una variable de tipus quantitatiu (o continu) la seva distribució es pot copsar millor a partir d'estadístics com la **mitjana,** la **mediana** i la **desviació típica**. Malgrat això, per a les variables quantitatives també és interessant, tal i com hem vist en l'anterior capítol, poder VEURE-LES, ja que això ens permet copsar amb major exactitud qüestions com els valors sense un significat substantiu, que haurem de transformar en valors perduts.

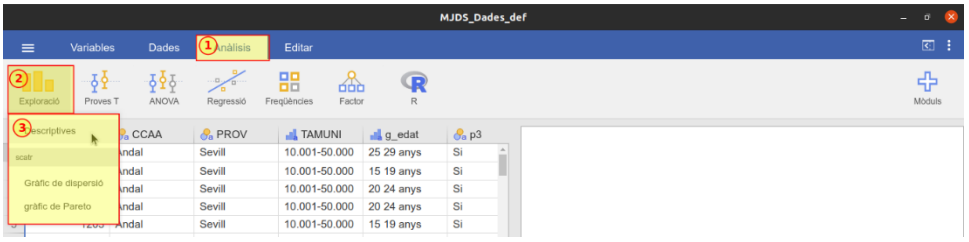
A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic que és una continuació de la tasca realitzada en l'anterior capítol. Així, disposem de:

- Un qüestionari realitzat pel CIS (fitxer *MJDS_QüestionariCIS.pdf*).
- La matriu de dades corresponent, on hi ha les variables correctament identificades, amb el tipus de variable i els valors perduts definits, i que està gravada en format de *jamovi* (*MJDS_Dades_def.omv*). Aquesta matriu de dades prové de la correcta definició de la matriu original, que era en format R (*MJDS_Dades.rdata*) i que has transformat en el apartat immediat anterior.
- Sobre aquesta matriu es vol elaborar un **informe amb una descripció exploratòria** de totes les variables.

Si no has realitzat encara la pràctica anterior (identificació i definicició de variables) però vols continuar amb el tutorial... pots obrir l'arxiu *MJDS_Dades_def.omv* disponible a l'espai CV de l'assignatura, o al conjunt d'arxius adjunts a MUJADES.

Obre el teu arxiu *MJDS_Dades_def.omv* (o el disponible al CV) per poder seguir amb el tutorial a partir d'aquest punt.

Tal i com es va avançar en l'anterior apartat, a *jamovi* disposem d'un procediment per a seleccionar un conjunt de variables i demanar la generació d'un conjunt d'indicadors propis de l'estadística descriptiva, i imprescindibles per a una anàlisi exploratòria de les dades.



Imatge 50. Ubicació del menú d'anàlisi estadístiques DESCRIPTIVES

A continuació ho exemplificarem amb les següents variables, cadascuna d'un nivell de mesura diferent, en les quals, a més, hem definit alguns valors perduts (vegeu la pràctica anterior):

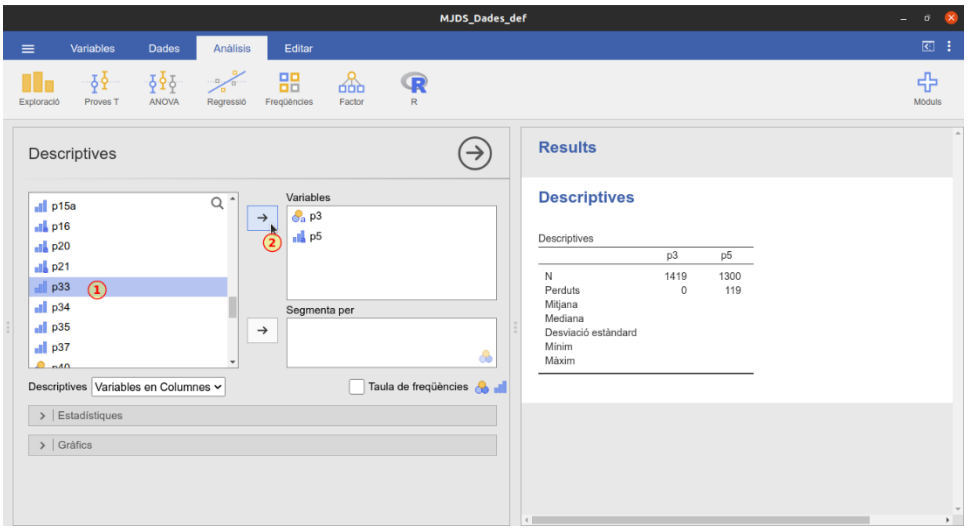
- Variable **nominal**, definida com a Nominal/Text*: Ús d'internet en els darrers tres mesos*, (p3). No té valors perduts definits.
- Variable **ordinal**, definida com a Ordinal/Text*: Freqüència d'us d'Internet*, (p5). El codi "N.C." consta com a valor perdut.
- Variable **quantitativa**, definida com Ordinal/Enter: *Autoposicionament polític (1-10)*, (p33). Consten com a valors perduts tots el valors superiors a 10 (o superiors a 96).

Segueix les passes que s'indiquen a continuació:

1. Comprova que les tres variables estan ben definides en el fitxer de *jamovi*. Per fer-ho, fes doble-clic a sobre de cadascuna i comprova que el tipus (mesura i dada) es corresponen amb els anteriors.
2. Comprova la definició dels valors perduts. Si cal, defineix els nivells o categories corresponents (p.e., "No sap que es", o "N.C.", o 97, o 98...) com a Valors perduts.

> **RECORDA!** Es molt important que totes les variables estiguin ben definides, perquè *jamovi* permet fer uns procediments o uns altres **en funció del nivell** de la variable. Veuràs que les icones de tipus apareixen en diversos llocs tot indicant quines variables es poden utilitzar en el procediment concret.

3. Clica l'etiqueta "Anàlisis" del menú superior i selecciona l'opció "Exploració", a l'extrem esquerra del submenú de procediments estadístics i, al menú desplegable, selecciona l'opció "Descriptives". S'obre un panell central de definició del procediment:



Imatge 51. Panell típic de selecció de variables per procediments estadístics

4. A l'esquerra del panell hi ha un llistat amb totes les variables disponibles (identificades amb la corresponent icona de nivell de mesura). Busca i selecciona amb un clic la p3. A la part central (2) hi ha un botó QUADRAT FLETXA-DRETA (no confondre amb la fletxa-encerclada, que serveix per a tancar aquest panel...). Fes clic per a "passar" la variable a l'espai dret del panel, on hi haurà les variables que s'utilitzaran per a generar l'anàlisi.
5. Fes el mateix amb les variables p5 i p33.
6. Com pots veure, a la finestra de resultats de la dreta van apareixen, immediatament i de forma actualitzada, els resultats (segons el tipus de variable).

7. Tal i com hem comentat en el capítol anterior, per defecte aquest procediment només genera set resultats: recompte de casos totals (*N*), recompte de casos perduts (*Missing*), mitjana aritmètica (*Mean*), mediana (*Median*), desviació típica (*Standard deviation*) valor mínim (*Minimum*) i valor màxim (*Maximum*) de la distribució. Més endavant veuràs com afegir o treure altres estadístics.
8. Per a generar taules de freqüències cal activar específicament aquesta opció: Fes clic al botó a sota del llistat dret de variables disponibles "**[ ] Taules de freqüències**".
   Els resultats s'actualitzen automàticament per a variables definides com tipus NOMINAL o ORDINAL (tal com indiquen les icones al costat del botó que acabes d'activar).
   El mateix botó serveix per a desactivar les taules de freqüències.
9. Si vols amagar el panell central, pots clicar a la FLETXA-EN-CERCLE de la part superior dreta.
10. Si vols visualitzar el panell de definició d'un procediment, fes doble-clic a sobre de qualsevol resultat al panell dret de resultats.

Per què creus que, en la primera taula, de les variables p3 i p5 no obtenim cap estadístic de resum? Tindria sentit que ens donés, per exemple, la mitjana?

Perquè, en canvi, sí que observem mesures de resum per a la p33? Quant val la mitjana? Quina interpretació li dones, tenint en compte com s'ha mesurat la variable?

EXERCICI

1. Ajudant-te de l'arxiu on hi ha el qüestionari que es va utilitzar per recollir les dades (*MJDS_QüestionariCIS.pdf*), repassa les variables fins que en trobis algunes (4 o 5) que siguin descriptives de la població espanyola i que facin referència a aspectes que t'interessin.
2. Genera taules de freqüències i estadístics descriptius per a visualitzar la distribució de cadascuna de les variables que has seleccionat.
3. Coincideix amb la idea prèvia que intuïtivament tens sobre les característiques de la població espanyola objecte de l'estudi?
4. Desa l'arxiu actualitzat en format propi de *jamovi* (.omv), posant-li un nom que t'ajudi a identificar-lo més endavant (p. ex., *Dades_CIS1.omv*).
5. Tanca **totalment** *jamovi*.
6. Torna a obrir amb *jamovi* l'arxiu que acabes de desar.
7. Comprova com s'han conservat les dades JUNTAMENT amb les taules que has generat.

(cc) BY-NC

En resum:

- Per obtenir una taula de **freqüències** on apareixen els valors absoluts, els percentatges sobre el total i el percentatge acumulat (els solem notar, respectivament, com a f_i p_i i P_i), cal clicar Menú / Anàlisis / Exploració / Descriptives, i activar l'opció "[ ] Taules de freqüències",
- Els estadístics descriptius QUANTITATIUS per defecte (mitjana aritmètica, mediana, desviació, mínim i màxim) només es calculen si la variable està definida com a "Tipus de Dades=Enter" o com a "Tipus de Dades=Decimal"; mai si està definida com a "Tipus de Dades=Text".
- Pots veure tots els estadístics que ofereix *jamovi* clicant la secció desplegable "Estadístiques", a la part central de la finestra de definició del procediment d'estadístiques descriptives.

Exploració de variables amb codi R

Si està instal·lat el mòdul corresponent a *jamovi* es pot fer servir directament el llenguatge de programació "R", orientat a l'anàlisi de dades, per a explorar la distribució de variables de la matriu de dades. Es tracta, en aquest cas, d'introduir una ordre R molt simple, **`summary()`**, dins del sistema perquè *jamovi* l'executi.

Per fer això cal:

1. Afegir a *jamovi* la capacitat addicional *(que no està implementada per defecte al moment d'instal·lar-ho)* d'executar directament ordres d'R mitjançant l'anomenada "consola",
2. Activar la "consola" fent clic a Menú / R / Rj-Editor.
3. Escriure l'ordre en el llenguatge R, i
4. Executar la consola d'R per a generar el resultats.

Per a instal·lar el mòdul o extensió (Rj) que permet l'execució, des de *jamovi*, de codi R, cal disposar de connexió a Internet i seguir les passes indicades a la secció anterior *"*Com afegir el mòdul per a operar directament amb codi R?*"*. Una vegada instal·lat un mòdul aquest quedarà integrat al teu sistema local *jamovi*, fins que el desinstal·lis (utilitzant, igualment, la finestra de "Mòduls").

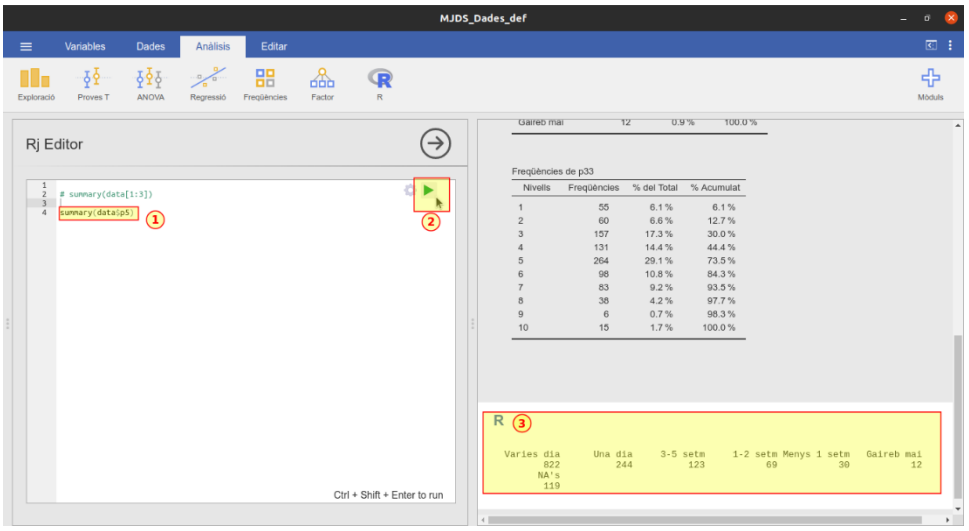
Per a escriure codi R i executar-lo amb una matriu de dades a *jamovi*, cal seguir les següents passes:

1. A l'apartat "Anàlisis", clica la icona "R". Al desplegable, selecciona l'opció "Rj Editor",
2. Apareix la "consola" d'ordres R, i predeterminadament, una ordre inactiva d'exemple:
   `# summary(data[1:3])`
3. Qualsevol línia que comenci per # es considera inactiva (està de color verd) i no s'executarà; s'utilitza molt per afegir comentaris al codi R,
4. Situa el cursor a qualsevol línia de la consola diferent de la que comença per #, i tecleja (NO facis retalla i enganxa d'aquí...), poc a poc, la teva pròpia ordre R:
   `summary(data$p5)`

Com has pogut veure, *jamovi* t'ajuda proposant ordres i variables disponibles, que pots seleccionar fent clic als menús que hi apareixen.

(cc) BY-NC

5. Aquesta simple ordre que has escrit en llenguatge R significa: "mostra un resum descriptiu de la variable que està ($) a la matriu de dades (data) i que es diu p5".
6. Per a **executar el codi** que hi ha escrit a la consola cal clicar al petit triangle verd de la part superior dreta de la consola (al costat d'una roda dentada). Alternativament, també es pot executar picant alhora la combinació de teclat [Ctrl] + [Majúscules] + [Intro]. Fes-ho.
7. Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.



Imatge 52. Exemple resultat d'execució codi R

8. Per tancar la consola R cal clicar a la fletxa-encerclada de la part superior dreta (igual que en el cas de qualsevol altre procediment d'anàlisi de *jamovi*).
9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Si vols explorar, d'una forma molt simple, **tot el conjunt** de variables d'una base de dades només cal que entris en el mode consola R i executis l'ordre:

```
summary(data)
```

Pots provar-ho ara per veure com són TOTES les variables de l'arxiu. Si vols explorar un subconjunt petit de variables és preferible utilitzar el sistema de menús propi de *jamovi*.

Recorda que aquesta consola dona accés a tota la potència del llenguatge R, que supera en molt les funcions que actualment pot desenvolupar *jamovi* per la via dels seus menús i les seves extensions. Per exemple, utilitzant l'ordre d'R "ggplot2" es pot produir i modificar gairebé qualsevol tipus de representació gràfica a partir de les dades, mentre que amb els menús de *jamovi* aquesta producció és molt més restrictiva.

Generar taules de freqüències amb codi R

Hi ha diverses ordres del codi R que permeten generar taules de freqüències bàsiques, o recompte dels casos per a cada categoria de la variable. Algunes d'aquestes ordres o funcions depenen de la

prèvia instal·lació de paquets específics (al sistema R), d'altres depenen de la prèvia instal·lació de mòduls d'extensió (al *jamovi*), i d'altres estan directament accessibles perquè formen part del conjunt bàsic d'operacions R acceptades per *jamovi*.

IMPORTANT: Les ordres R cal "escriure-les" a la consola de jamovi. Si el que fas és copiar, p.e. d'aquest documents PDF, i enganxar la ordre es MOLT PROBABLE que doni errors. Això es degut a que has enganxat caràcters de PDF que semblen adients per a R quan realment no ho son. Els problemes més habituals tenen relació amb guions (-), espais blancs ( ) o cometes (").

La solució es, simplement, teclejar a la consola R de jamovi completament la ordre.

El següent apartat del tutorial t'acompanyarà en la utilització pràctica d'ordres d'R que són directament accessibles amb la versió de R que hi ha incorporada a la instal·lació de *jamovi*.

1. Accedeix a la consola d'R, tal i com hem fet abans, del menú: Anàlisis / R / Rj Editor...
2. Per a generar una taula de freqüències simple de la variable tipus Nominal/Text p3 (Ús d'Internet en els darrers tres mesos), escriu el següent codi a la consola:
   `table(data$p3)`
3. Aquesta ordre en llenguatge R significa: "mostra una taula de recompte de casos, o taula de freqüències, de la variable que està ($) a la matriu de dades (data) i es diu p3".
4. Executa aquest codi clicant el petit triangle verd de la part superior dreta de la consola, o amb la combinació de teclat [Ctrl] + [Majúscules] + [Intro].
5. Comprova que a la finestra de resultats s'ha generat un recompte de les categories de la variable:
```
   Si   No
 1301  118
```
6. Aquesta taula de freqüències és mol bàsica i no disposa ni de títol ni d'altres càlculs com ara freqüències relatives (%) o acumulades. Pots veure com les taules generades amb R son diferents de les que genera el sistema de menús de *jamovi*.

EXERCICI

- Genera, amb codi R, una taula de freqüències simple de la variable tipus ORDINAL/Text p5 (Freqüència d'ús d'Internet) i de la variable Ordinal/Enter p33 (Autoposicionament polític (1-10)).
- Quines mancances trobes en aquestes taules generades en codi R?

Per generar una taula completa d'alguna de les variables, farem el següent (vegem-ho amb la p33 (*Autoposicionament polític (1-10)*):

1. Per poder generar les taules **en vertical,** les quals permeten una lectura més comprensible, escriurem el següent codi:
   `cbind(table(data$p33))`
2. Com pots veure, la ordre o funció R `cbind()` serveix per a "girar" els resultats.
   En els següents passos veuràs com utilitzar altres ordres R per modificar d'altres maneres els resultats de la taula bàsica de recompte o freqüències.

3. Per a generar la taula vertical, que contingui les **freqüències relatives** de la variable P33, escriu i executa el següent codi a la consola:
```
cbind(prop.table(table(data$p33)))
```
4. I si volem que ens doni els **percentatges**, només hem de multiplicar la taula per 100:
```
cbind(prop.table(table(data$p33))*100)
```
5. Podem obtenir en **una mateixa taula** diferents elements. Per exemple, si volem que ens mostri en una única taula, les freqüències absolutes i els percentatges, li podem dir:
```
cbind(table(data$p33), prop.table(table(data$p33))*100)
```
6. El què li hem dit, en llenguatge R, és: Fes una taula vertical (cbind) que tingui dues columnes (,) . La primera ha de contenir una taula de freqüències absolutes (table) de la variable p33 i la segona que contingui les freqüències relatives (prop.table) i multiplicades per 100 (*100), de la taula de freqüències absolutes de la variable p33.

AVANÇAT

Una de les més útils potencialitats del llenguatge R és poder "posar nom" a qualsevol dels resultats obtinguts, de forma que es poden reutilitzar per a posteriors operacions. Per exemple, el codi **7+2** genera com a resultat **9**, mentre que el codi **A<-7+2** genera un objecte (una entitat) que hem anomenat **A**, i que té emmagatzemat a dins el nombre **9**. Així, si li després diem **A+1** donarà com a resultat **10**.
Per anomenar qualsevol resultat a R cal: (1) indicar el nom desitjat, (2) seguit de **<–** i (3) seguit de l'operació corresponent.

Utilitzant diferents funcions, i la possibilitat d'anomenar i re-utilitzar resultats, es pot combinar una sèrie consecutiva d'ordres d'R per a generar una taula vertical amb diversos resultats a les columnes. El procediment que es veurà a continuació **és el mateix** que s'ha fet abans, però ara ALTERNATIVAMENT s'aniran "creant" objectes i treballant amb ells per a modificar-los i aconseguir els següents objectes:

1. Per a generar una taula de freqüències de la variable P33 més completa, escriu el següent codi a la consola, a línies diferents i sense executar-lo fins que estiguin totes les línies:
```
dades <- data$p33
n_i <- table(dades)
f_i <- prop.table(n_i)*100
F_i <- cumsum(f_i)
cat("Taula de freqüències: variable p33")
cbind(n_i, f_i, F_i)
```
2. Observa que a les primeres línies de codi hem generat i anomenant objectes, amb **<-**, que utilitzem en operacions posteriors.
3. Aquesta seqüencia d'ordres significa, en llenguatge R:
   Anomena **dades** als valors de la variable p33,
   Anomena **n_i** a la taula de freqüències absolutes d'allò anomenat **dades**,
   Anomena **f_i** a la taula de percentatges d'allò anomenat **n_i** (és a dir, l'anterior taula de freqüències absolutes); les proporcions (prop) es multipliquen per 100,
   Anomena **F_i** a la taula de freqüències acumulades d'allò anomenat **f_i** (és a dir, l'anterior taula de freqüències relatives en %),
   Escriu a la finestra de resultats el text següent: "Taula de freqüències: variable p33",
   Mostra verticalment les columnes de les taules que s'anomenen **n_i, f_i, F_i**
4. Observa la nova funció **`cat()`** que genera un text a la finestra de resultats. És una forma habitual d'utilitzar codi R per a identificar resultats.

5. Si executes tot el codi generaràs la següent taula a la finestra de resultats:

```
Taula de freqüències: variable p33

   n_i      f_i     F_i
1   55   6.0639   6.064
2   60   6.6152  12.679
3  157  17.3098  29.989
4  131  14.4432  44.432
5  264  29.1069  73.539
6   98  10.8049  84.344
7   83   9.1510  93.495
8   38   4.1896  97.685
9    6   0.6615  98.346
10  15   1.6538 100.000
```

RECORDA que *jamovi* només "executa" ordres del llenguathe R. Com que R "escriu" els resultats "en anglès" els punts (.) que apareixen a les xifres dels resultats son **separacions decimals** *(que nosaltres denotem amb comes)*.

EXERCICI

- Genera el mateix tipus de taula de freqüències per a les variables p3 i p5. Pots utilitzar qualsevol dels dos procediments alternatius que s'han vist anteriorment per aconseguir el mateix resultat.
- Recorda que si tens la instrucció complerta escrita en codi R, en realitat només hauràs de canviar, en cada cas, el nom de la variable.

TOTHOM que treballa amb codi R tenim diversos arxius de text (*.txt) amb diferents ordres de R escrites per a generar processos o resultats. Es fan servir com si fossin "llibreries" de codi R reutilitzable. Quan necessitem fer el mateix per altres variables només cal trobar el procediment en concret, retallar i enganxar el codi anterior a dins de la consola del Rj Editor de *jamovi*, canviar el nom de les variables, i executar.

Estadística descriptiva de variables quantitatives, amb codi R

Tots els estadístics bàsics i propis de les variables quantitatives es poden demanar utilitzant ordres de codi R. Ara bé, has de tenir en compte que:

- Les variables quantitatives amb les que treballem amb *jamovi* no són, pròpiament, numèriques, ja que les hem definides com a Ordinal/Enter. Per tant, el primer que haurem de fer és canviar-les, utilitzant directament el codi R,
- Per a generar una còpia exacta, però numèrica d'una variable definida Ordinal/Enter: `data$p33num<-as.numeric(as.character(data$p33))`

Li estem dient que generi una nova variable anomenada "p33num" que sigui la versió numèrica de la p33, passada prèviament a tipus "character".

- Si demanes qualsevol estadístic descriptiu d'una variable que contingui valors perduts, donarà error. Cal indicar expressament a R que no els tingui en compte, afegint a la instrucció corresponent:
  `na.rm=T`

> **IMPORTANT**: A les ordres R cal incorporar EXPRESSAMENT l'opció per NO tenir en compte valors perduts (`na.rm`=T), altrament *jamovi* donarà un missatge d'error i no calcularà l'estadístic demanat.

A continuació veuràs el procediment per demanar els diferents estadístics. Treballarem amb la variable *Nombre d'amics i agregats a les xarxes socials* (p15), què prèviament haurem de passar a numèrica seguint la lògica d'abans:

```
data$p15num<-as.numeric(as.character(data$p15))
```

Alguns dels estadístics que podem demanar utilitzant codi d'R són:

- La **mitjana** ($\overline{x}$). Escriu, a la consola d'R la funció (i executa-la amb el petit triangle verd, superior esquerra):
  `mean(data$p15num, na.rm = T)`
  La mitjana d'amics a les xarxes socials, és de 248,7 amics.

> Observa quin resultat produeix R amb la següent variant de la mateixa ordre:
> **`mean(data$p15num)`**
> R prova de calcular una mitjana tenint en compte <u>tots</u> els valors, però com que alguns del quals son valors perduts (NA) el resultat també és valor perdut (NA). Per aquesta raó cal incloure, com s'ha vist anteriorment, l'opció ,na.rm=T que indica *"estableix com a veritat (T, true) l'opció d'eliminar (remove) els valors perduts (na)"*.

- Els **quartils** (Q_1, Q_2, Q_3). Escriu a la consola, i executa:
  `quantile(data$p15num, na.rm = T)`
  La persona que té menys amics i la que en té més, tenen, respectivament 1 amic i 4.600 amics. A més, el 25% que menys amics té, en té 77 o menys. Hi ha un 50% d'individus que tenen 175 amics o menys i un 50% que en tenen 175 o més (mediana o quartil segon). El 75% que menys amics té, en té 300 o menys (anàlogament, el 25% que més en té, té 300 amics o més).
- El **coeficient de variació** (*CV*), que es correspon a la desviació típica dividida per la mitjana. Caldrà fer càlculs amb R, concretament una divisió. Escriu a la consola, i executa:
  `sd(data$p15num, na.rm = T) / mean (data$p15num, na.rm = T)`
  El nombre d'amics a les xarxes té un coeficient de variació de l'ordre del 133,7%.
- El **rang interquartíl·lic** (*IQR*). Escriu i executa:
  `IQR(data$p15num, na.rm = T)`
  Per encabir el 50% de les observacions "centrals" d'aquesta distribució de valors necessitem un rang total de 223 amics. Comprova que es tracta de la diferència entre el tercer i el primer quartil, calculats anteriorment.

Algunes persones prefereixen treballar amb el sistema de menús (restringit, però amb resultats més ben dissenyats), mentre que d'altres prefereixen escriure codi i aprofitar tota la potència d'R. En tot cas *jamovi* permet utilitzar totes dues estratègies, combinar-les, i introduir-se de forma simple i poc a poc a la programació en R, que és la estratègia que recomanem.

RECORDATORI: Si ara deses un arxiu .omv que conté resultats generats directament amb codi R, i després es vol obrir aquest arxiu des d'un altre PC o sistema *jamovi* que NO TINGUI instal·lat el mateix mòdul (Rj) **es generarà un error** per incompatibilitat de la configuració dels dos sistemes. Aquest problema es pot resoldre –simplement– instal·lant el mòdul "Rj" al sistema "receptor" i tornant a obrir l'arxiu. En el cas de contenir codi directe R, *jamovi* sempre demana –per motius de seguretat– la confirmació per a obrir aquests arxius.

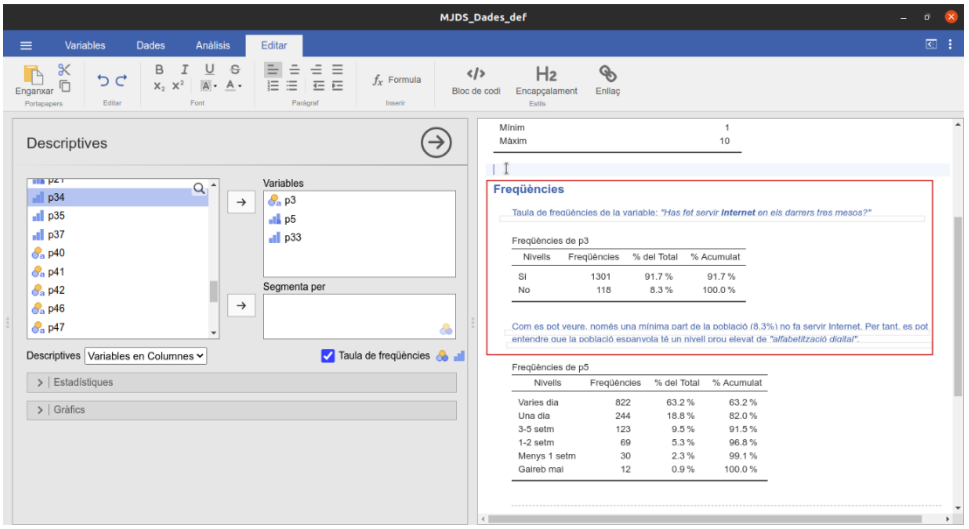
Resultats de l'exploració de variables: Afegir comentaris i exportar

El sistema *jamovi* genera resultats a la finestra de la part dreta de la pantalla, en forma de taules i de representacions gràfiques. Aquests resultats són "objectes" que es van afegint a un espai continu de dalt a baix, i que poden ser recuperats en tornar a obrir l'arxiu en format OMV propi de *jamovi*. En aquesta secció es mostraran dues interessants possibilitats de treball a partir dels resultats generats en *jamovi*: afegir comentaris personals, i exportar cap a altres aplicacions.

Modalitat comentari textual

El contingut de les taules generades per *jamovi* NO es pot editar amb *jamovi*. Però és possible incorporar text propi, notes i títols a aquest resultats generats. Per fer-ho cal entrar en la modalitat COMENTARI de resultats mitjançat el menú corresponent (menú "Editar").
Alternativament, i si et trobes a la modalitat ANÀLISI, pots fer clic-esquerra a les zones de la finestra de resultats on és possible la introducció o edició de text (identificades per l'aparició d'un petit signe > blau, quan es passa el punter per sobre).
Alternativament, en qualsevol modalitat es pot fer clic-dreta a sobre d'una taula o gràfica i al menú contextual desplegat seleccionar l'opció "Afegir Nota".

1. Clica el menú "Editar".
2. A la finestra de resultats, queden destacats en blau (molt clar, però més destacat si es passa el punter per sobre...) els espais on es pot incloure text propi: a l'inici i final de cada procediment estadístic demanat, i al peu de cada taula generada. També inclou el títols dels procediments, que poden ser modificats lliurement.
3. Clica sobre qualsevol d'aquestes àrees editables. Escriu el text desitjat i dona-li format bàsic amb les diferents opcions del submenú d'icones (negreta, cursiva, subratllat, sub/super-índex, color de lletra o de fons, justificació, llistes...). L'opció "Encapçalament" (H_2) serveix per a generar subtítols de forma automàtica.
4. Busca la taula de freqüències de la variable p3. Comprova que no és massa clara ni explicativa. Entra en mode comentari i canvia l'encapçalament de la taula perquè inclogui la expressió explícita de la pregunta del qüestionari "*Has fet servir Internet en els darrers tres mesos?*".
5. Afegeix algun comentari al peu de la taula en relació a conclusions derivades dels resultats.



Imatge 53. Exemple de comentaris afegits a resultats estadístics

Exportació de resultats cap a altres aplicacions

El contingut de la finestra dreta de resultats es composa d'un conjunt d'objectes (els resultats estadístics o gràfics), generats per *jamovi* a partir de les "comandes" de l'anàlisi en concret. Cada anàlisi demanat és un objecte-Anàlisi, que conté diversos objectes-Taula.

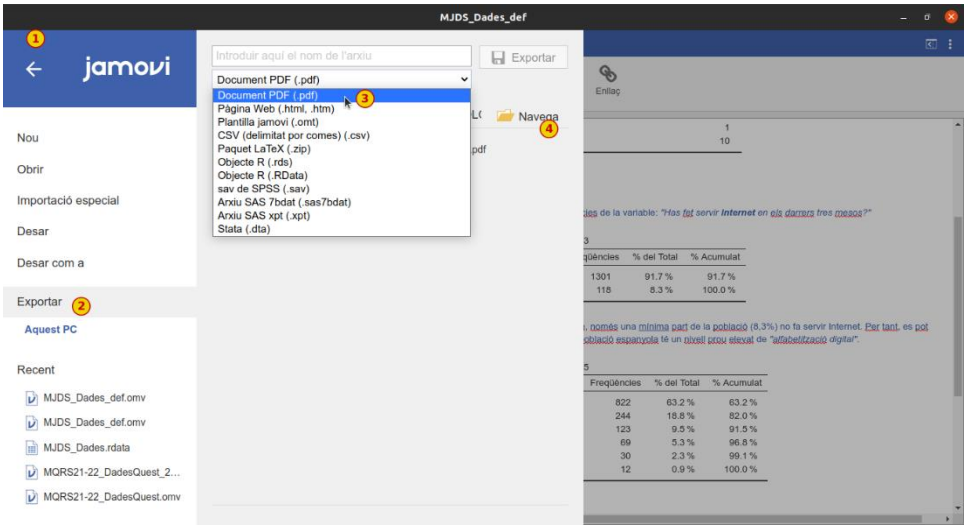
Mitjançant diversos procediments és possible exportar els resultats, taules i gràfics, inclosos els comentaris afegits:

(A) del conjunt de tots resultats de l'arxiu .omv actual (TOT),
(B) de cada anàlisis per separat, incloent totes les seves taules i gràfics (ANÀLISI), o
(C) d'una taula o gràfic en particular (TAULA).

En funció del procediment d'exportació utilitzat serà possible modificar més o menys fàcilment els resultats obtinguts, amb l'objectiu final d'integrar les evidències estadístiques dins d'un treball, informe o una publicació sigui acadèmica, científica o de divulgació.

(A) Per a exportar el conjunt de **tots els resultats**, inclosos taules, gràfics i comentaris, de forma ràpida i simple, **en format PDF**:

1. Clica sobre el menú d'Arxius,
2. Selecciona "Exportar",
3. Selecciona, al desplegable corresponent, com a format d'exportació "PDF Document (.pdf))",
4. Clica "Navega" per obrir la finestra que permet triar el nom i la ubicació de l'arxiu exportat. Procura desar el PDF a un lloc d'on sigui senzill recuperar posteriorment. **IMPORTANT**: No oblidis d'incloure al nom de l'arxiu la "extensió" corresponent al format PDF, per exemple, "elsmeusresultats**.pdf**".



Imatge 54. Exportació de resultats a PDF

5. Ara busca, al teu sistema local d'arxius, l'arxiu PDF que acabes d'exportar i fes-hi doble-clic per a obrir-lo amb la teva aplicació de lectura de PDF *(si és que en tens cap d'instal·lada; recorda que hi ha moltes disponibles i gratuïtes, com ara SumatraPDF o Acrobat® Reader)*.
6. Una vegada generat, aquest arxiu PDF no és fàcil de modificar. És com si fessis una "impressió" dels resultats, però no en paper sinó en format digital.

ALTERNATIVA: Menú Arxiu / Exportar / ... a format HTML, que es pot obrir fàcilment amb editors de text, i permet modificar, editar i complementar amb comentaris.

(B) Per a exportar **cada anàlisis per separat, incloent totes les seves taules, gràfics i comentaris** de text, cap a una aplicació de processament de documents amb la qual es redactarà l'informe:

1. Obre un nou document en blanc al teu processador de text favorit (LibreOffice Writer, MacOS Pages, M$-Word...). Comença a redactar el teu informe, al qual hi incorporaràs els resultats estadístics.
2. A *jamovi*, fes clic-esquerra a sobre de qualsevol de les taules de resultats de l'anàlisi en concret que vulguis exportar, al menú contextual selecciona l'opció "Analitza". Comprova que s'ha seleccionat (color blau) el que t'interessa exportar. Selecciona "Copiar" per transferir l'objecte seleccionat cap al porta-papers.
3. De nou, al teu processador de text, situa el cursor on vulguis "enganxar" els resultats exportats de *jamovi* i fes "Enganxa" (amb la combinació de teclat [Ctrl]+[V]).
4. En aquest cas, cadascun dels objectes s'han enganxat al processador de text com a taules i com a imatges, amb àmplies possibilitats per a modificar, editar, seleccionar, eliminar o moure qualsevol part, i integrar-ho dins del redactat general.

(C) Per a exportar **només una taula** en particular:

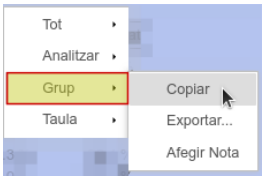
1. Obre un document al teu processador de text amb el teu informe, on hi vols incorporar els resultats estadístics.
2. A *jamovi*, fes clic-esquerra a sobre de qualsevol de les taules de resultats que vulguis exportar, al menú contextual selecciona l'opció "Taula" / "Copiar" per transferir la taula cap al porta-papers.

3. De nou, al teu processador de text situa el cursor on vulguis "enganxar" els resultats exportats de *jamovi* i fes "Enganxa" (amb la combinació de teclat [Ctrl]+[V]).
4. En aquest cas, la taula s'ha enganxat al processador de text com a taula, amb àmplies possibilitats per a la seva edició i integració dins del redactat general.
5. Aquest procediment exporta taules, PERÒ SENSE les notes o comentaris afegits a *jamovi*.

(C) Per a exportar **només un gràfic** en particular:

1. Obre un document al teu processador de text amb el teu informe (o a la teva aplicació de presentacions, o de disseny gràfic...) al qual hi incorporaràs la imatge gràfica.
2. A *jamovi*, fes clic-esquerra a sobre de qualsevol de les representacions gràfiques en concret que vulguis exportar, al menú contextual selecciona l'opció "Imatge" / "Copiar" per transferir la taula cap al porta-papers.
3. De nou, a la aplicació on s'exportarà situa el cursor on vulguis "enganxar" els resultats exportats de *jamovi* i fes "Enganxa" (amb la combinació de teclat [Ctrl]+[V]).
4. En aquest cas, la imatge s'ha enganxat al processador de text (o presentació, o disseny gràfic) com a imatge, amb les possibilitats d'edició i integració pròpies de l'aplicació "receptora".
5. Aquest procediment exporta imatges gràfiques, PERÒ SENSE les notes o comentaris afegits a *jamovi*.

Si vols exportar una taula o una gràfica **conjuntament amb el seu títol i comentaris** de text, al menú contextual cal seleccionar l'opció "Grup" / "Copiar".



Imatge 55. Menú contextual (clic-dreta) de resultats

ALTERNATIVA: Es pot desar una imatge particular *(no un "grup", i no s'inclou text)* a un arxiu amb diverses opcions de format gràfic. Posteriorment, aquest arxiu es pot importar a aplicacions de processament de documents, presentació o disseny gràfic. Per fer-ho, clic-dreta a sobre de qualsevol imatge resultat a *jamovi*, al menú contextual selecciona l'opció "Exportar..." i al desplegable de format d'arxiu no triar PDF sinó SVG *(recomanat)*, o PNG.

Descripció gràfica de variables

De manera anàloga al càlcul d'estadístics, disposem de representacions gràfiques de dades que permeten una exploració de les variables i són, també, una expressió resumida de les dades. El sistema de menús *jamovi* ofereix un conjunt de representacions gràfiques, que per defecte no estan activades, i que es reparteixen als diferents menús de configuració de les opcions d'anàlisi estadística corresponent.

Respecte de l'estadística descriptiva bàsica, l'opció per a activar representacions gràfiques està disponible al menú Anàlisis / Exploració / Descriptives (excepte el "Gràfic de Pareto", que forma part d'un mòdul específic que cal instal·lar expressament, anomenat "scatr").

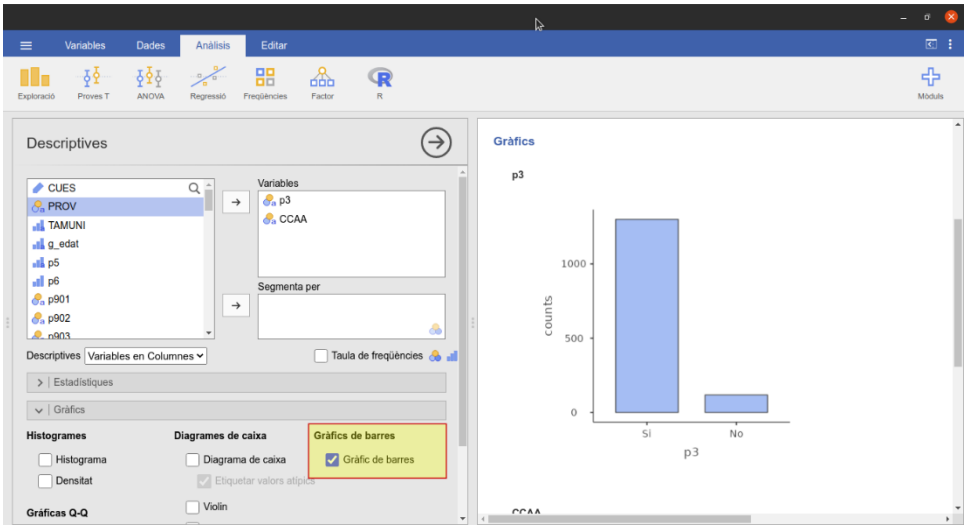
A continuació, i de manera molt sintètica, recordem quin tipus de gràfic és el més adequat en funció del tipus de mesura de la variable:

- Variables nominals (en *jamovi,* Nominal/Text). El més correcte és fer gràfics de sectors i gràfics de barres. Si el nombre de categories és bastant gran, però, el diagrama de sectors és poc clar. El sistema de menús de *jamovi* no té incorporat el diagrama de sectors, que haurem de generar executant codi amb la consola d'R.
- Variables ordinals (en *jamovi,* Ordinal/Text). Atès que la variable està ordenada, el més correcte és fer gràfics de barres, que donen, precisament, aquesta sensació d'odre.
- Variables quantitatives (en *jamovi*, Ordinal/Enter). Són adequats tant els diagrames de caixa com els histogrames.

Com a exemple de variables nominals farem servir les variable *Ús d'Internet en els darrers tres mesos* i *Comunitat Autònoma de residència* (respectivament, p3 i CCAA); com a exemple de variable ordinal *Freqüència d'ús d'Internet* (p5), i com a variables quantitatives *Autoposicionament polític (1-10)* i *Nombre d'amics a les xarxes socials* (p33 i p15, respectivament).

Per a graficar les variables p3 i CCAA (nominals), farem el següent:

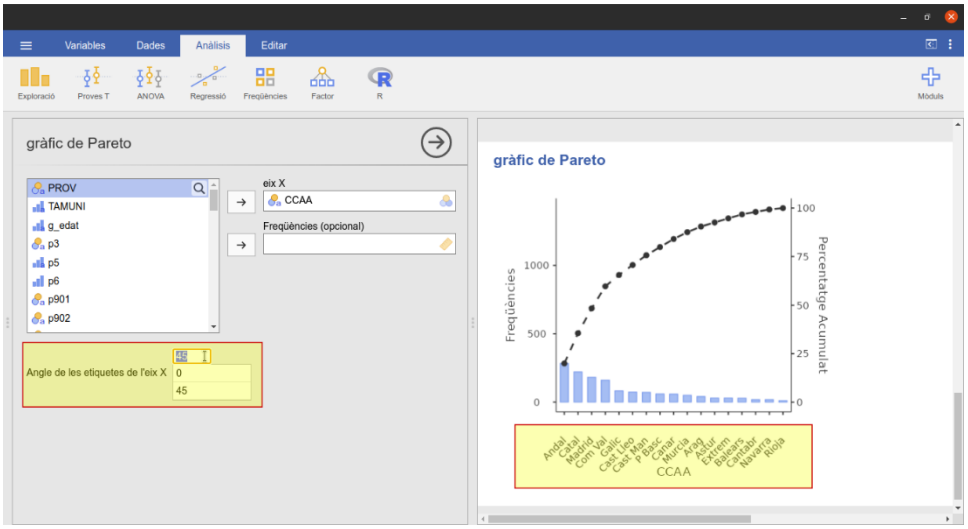
1. Com que es tracta de variables nominals el més adequat, entre les opcions que ofereix *jamovi*, és un diagrama de barres.
2. Clica al menú: Anàlisis / Exploració / Descriptives. A la configuració del procediment (pantalla esquerra) busca i selecciona les variable p3 i CCAA, i envia-les a la finestra "Variables". A la part inferior clica la secció "Gràfics" per a desplegar les opcions gràfiques. A l'apartat "Gràfics de Barres" activa l'opció "Gràfic de Barres".



Imatge 56. Generar diagrames de barres amb jamovi

3. A la finestra dreta de resultats, baixa fins a visulitzar la gràfica de la variable CCAA.
4. Observa que algunes etiquetes d'aquesta variable queden encavalcades i illegibles, un problema que tindrem sovint quan el nombre de categories sigui gran, o el text d'aquestes molt llarg. Més endavant veuràs com resoldre això amb codi R.
5. Una altra opció, si tens instal·lat el mòdul d'extensió anomenat "scatr", és generar un Gràfic de Pareto, que es pot utilitzar per variables de qualsevol tipus.
6. Clica al menú Anàlisis / Exploració / Gràfic de Pareto, a la configuració del procediment (pantalla esquerra) busca i selecciona, per exemple, la variable CCAA, i clica el botó central per a incloure la variable al conjunt d'anàlisi "eix X".

7. Pots comprovar com les etiquetes també s'encavalquen. Ara bé, en aquest cas pots seleccionar, a la part inferior, la secció "Angle de les etiquetes de l'eix X" i indicar 45 graus per "girar" les etiquetes de l'eix horitzontal i millorar la llegibilitat del gràfic.



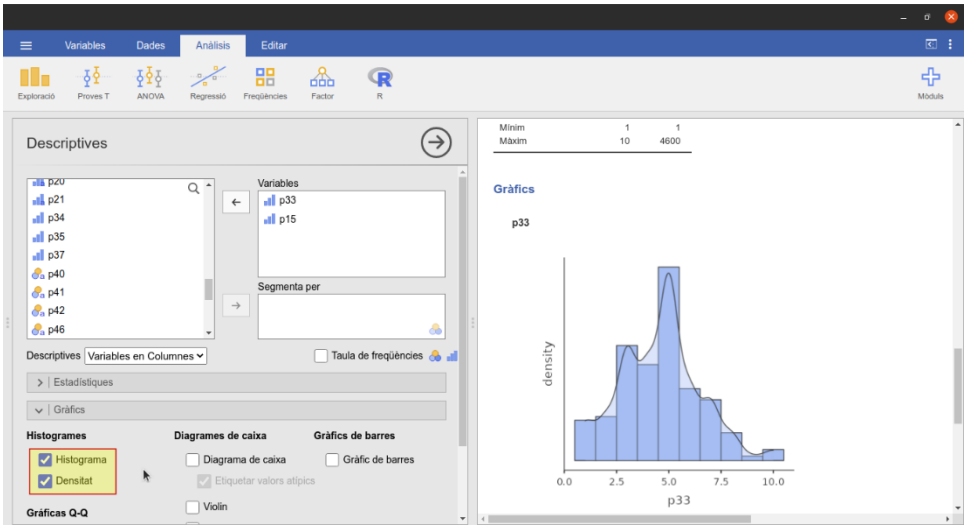
Imatge 57. Exemple de Gràfic de Pareto generat amb jamovi

EXERCICI

- Tal i com s'ha comentat més amunt, en variables ordinals (en *jamovi* Ordinal/Text) fem, en jamovi, els mateixos de gràfic que en variables nominals.
- Fes els dos mateixos gràfics que acabes de fer, però per a la variable *Freqüència d'ús d'Internet* (p5).

Per representar gràficament les variables **quantitatives** *Autoposicionament polític (1-10)* i *Nombre d'amics a les xarxes socials* (p33 i p15, respectivament), fes el següent:

1. Clica el menú: Anàlisis / Exploració / Descriptives. A la configuració del procediment (pantalla esquerra) busca i selecciona les dues variables, p33 i p15, clica el botó central per a incloure la variable al conjunt d'anàlisi i, a la part inferior, clica a la secció "Gràfics" per a desplegar les opcions de gràfiques. A l'apartat de l'esquerra "Histogrames" activa les opcions "Histograma" i "Densitat".



Imatge 58. Generar un histograma (i la corba de densitat) amb jamovi

2. Quan un histograma s'acompanya de la gràfica de densitat, com hem fet ara, és més senzill interpretar la "forma" de la distribució dels valors, en quant a biaix o asimetria lateral, el punt central (o "pic" de la distribució), i la dispersió respecte del punt central.
3. Anem, ara a dibuixar un diagrama de caixa per ambdues variables. A la mateixa finestra on ets, desmarca les opcions "Histograma" i "Densitat" i activa les opcions "Diagrama de caixa" i, si vols que ens dibuixi també la mitjana, la pestanya "Mitjana".
4. Pots comprovar que el gràfic per a la ideologia és molt clar, mentre que no ho és tant per al nombre d'amics, degut, precisament, al rang que té aquesta variable (d'1 amic a 4.600).
5. En el cas de la ideologia, per exemple, pots comprovar a la taula de descriptives que la mediana, o segon quartil, (Md, Q_2) correspon a un cas amb valor 5, i com la part central de la població (el 50% central) es situa entre els valors 3 i 7, amb un biaix (o asimetria) que sobre-carrega les posicions cap a l'esquerra de la escala ideològica. Igualment pots veure com hi ha menys individus a l'extrem dret de la escala. Així com que la mitjana se situa lleugerament per sota del 5.

RECORDA: Els diagrames o gràfics de caixes són molt útils i representen gràficament la distribució d'una variable quantitativa mitjançant la visualització de cinc punts de localització claus (mínim, primer quartil, mediana, tercer quartil i màxim) així com a qualsevol observació puntual que es pugui classificar com a sospitosa de ser un valor atípics (*"outlier"*). En aquest cas, a més, hem demanat que ens ensenyi on se situaria la mitjana. Disposar de tota aquesta informació sintetitzada al mateix diagrama proporciona una excel·lent primera visió general de la centralitat, la dispersió, la forma, i la ubicació de les dades de la variable.

La part central de la distribució, que correspon al 50% central de la població ordenada segons la variable, queda representada per la CAIXA (el rectangle central de la gràfica) que coincideix amb el rang interquartíl·lic RIQ (Q_3-Q_1) i que conté la Mediana, representada per una LÍNIA horitzontal. Els valors atípics, quan n'hi ha, s'identifiquen per PUNTS, i s'utilitza com a criteri per a la seva identificació la situació "lluny" de la part central de la distribució. Són aquells casos allunyats una vegada i mitja el RIQ fora de la caixa, és a dir: superiors a Q_3+(1,5*IQR), o inferiors a Q_1-(1,5*IQR). Aquests límits d'1,5 IQR queden representats per les "PATILLES" en forma de línies verticals a dalt i a baix de la caixa. *(Aquesta representació es coneix com "box and whiskers plot", o diagramas de caixa i patilles)*

Codi R per a generar representacions gràfiques

A continuació el tutorial MUJADES t'indicarà alguns exemples de codi R que pots executar per a generar diversos tipus de gràfics, alguns d'ells impossibles de generar directament amb els menús de *jamovi*. En tots els gràfics es genera primer una versió senzilla i, a continuació, una (o unes) de més completa, on s'incorporen elements al gràfic. Es tracta que juguis, que identifiquis què fa cada element i, per exemple, que pensis si es podria millorar.

Un dels gràfics que no es pot fer amb *jamovi* és el DIAGRAMA DE SECTORS (útil <u>només</u> en cas de variables nominals de poques categories). Anem a fer-lo amb la variable *Ús d'Internet en els darrers tres mesos* (p3).

1. Per a generar la versió simple només cal escriure i executar:
```
pie(table(data$p3))
```
2. Una versió més completa podria contenir **un títol** i controlar els **colors**, mitjançant el codi R següent:
```
pie(table(data$p3), main="Has utilitzat Internet en els darrers mesos?",
col=c("red","blue"))
```
3. Prova de fer un diagrama de sectors amb la variable *Comunitat Autònoma de residència* (CCAA) i veuràs perquè **NO ÉS** un gràfic aconsellable quan la variable té MOLTES categories.

En quant al codi per fer un DIAGRAMA DE BARRES, anem a generar-lo amb la variable *Freqüència d'ús d'Internet* (p5).

1. Per a generar la versió simple només cal escriure i executar:
```
barplot(table(data$p5))
```
2. Si, per exemple, volem incloure un **títol** al gràfic, canviar els límits de l'**eix de les Y**, i modificar l'orientació de les **etiquetes** de l'eix de les X, caldrà executar:
```
barplot(table(data$p5), main="Freqüència d'ús d'Internet",
ylim=c(0,1000), las=2)
```
3. Per fer-lo en **percentatges** en comptes de valors absoluts, afegint una **etiqueta a l'eix de les Y** i incloure **colors** diferents a cada columna:
```
barplot(
prop.table(table(data$p5))*100,
main="Freqüència d'ús d'Internet",
ylim=c(0,70), ylab="Percentatge d'ús",
col=palette("Dark 2"),las=2
)
```

Com pots veure, el codi R es pot escriure en diferents líneas. Sempre que es respecti la estructura bàsica de **`ordre(dades, opcions)`** en jamovi ho "llegirà" i executarà correctament. No oblidis la correcta ubicació dels parèntesis i que cal teclejar-lo directament i no copiar/enganxar.

En quant a representar un HISTOGRAMA, anem a fer-ho amb la variable *Nombre d'amics a les xarxes socials* (p15).

1. Per fer l'histograma cal, en primer lloc, que la variable sigui numèrica. Per això, i tal i com hem fet en altres ocasions, prèviament generem la <u>versió numèrica de p15</u> amb la següent instrucció:
```
data$p15num<-as.numeric (as.character (data$p15))
```

2. La versió més senzilla de l'histograma és:
```
hist(data$p15num)
```
3. Si volem canviar els **títols** (del gràfic i dels dos eixos), volem donar-li **color**, i volem que ens generi, per exemple, **20 barres** diferents, li direm:
```
hist(data$p15num, breaks = 20,
main = "Histograma de Contactes a Xarxes Socials",
xlab="Nombre de contactes", ylab="Freqüència absoluta",
col="green")
```
4. Compara la versió anterior senzilla amb aquesta versió millorada, i prova d'identificar quines parts del codi R produeixen cadascuna de les millores .

Per generar amb codi R un DIAGRAMA DE CAIXA ho farem amb les variables *Nombre d'amics a les xarxes socials* (p15) i *Autoposicionament polític* (p33).

1. Tant en un cas com en l'altre, si vols dibuixar l'histograma has de convertir prèviament les variables a numèriques. Per tant, executa:
```
data$p15num<-as.numeric (as.character (data$p15))
data$p33num<-as.numeric (as.character (data$p33))
```
2. En ambdós casos la versió simple seria dir-li i executar:
```
boxplot(data$p33num)
boxplot(data$p15num)
```
3. Anem, en primer lloc, a millorar el gràfic d'ideologia posant un **títol** i mostrant què volen dir alguns dels codis de p33 (observa que, en el fons, això són dues instruccions diferents, de manera que obtindràs dos gràfics):
```
boxplot(data$p33num,yaxt="n", main="Autoposic. pol",col=c("pink"))
axis(2,at=c(1,5,10),labels=c("Ext.Esq","Centre","Ext.Dr"),las=1)
```
4. Si observes el gràfic fet amb p15num veuràs que el principal problema és el rang, de manera que anem a reduir (per exemple fins el 900) el què volem que ens mostri. És com fer ampliació d'una part del gràfic, excloent els valors extrems que ho distorsionen:
```
boxplot(data$p15num, ylim=c(0,900),
main="Nombre d'amics a les xarxes socials",
col=c("magenta"))
```

EXERCICIS PROPOSATS: Estadística descriptiva i exploratòria bàsica

(A) De la matriu de dades "MJDS_Dades_def.omv", **selecciona tres** variables (diferents de les ja utilitzades), una NOMINAL, una altra ORDINAL, i una tercera QUANTITATIVA. Ajuda't amb el qüestionari del CIS "MJDS_QüestionariCIS.pdf".
Genera amb *jamovi*, la representació gràfica **que millor** representi cadascuna de les tres variables.
Redacta un document en format DOCX amb les tres gràfiques importades al processador de text, cadascuna acompanyada d'un **breu comentari** respecte de la variable que representa.

7. Transformació i depuració de dades mitjançant jamovi

L'objectiu d'aquest apartat del tutorial MUJADES és guiar en la realització d'algunes de les tasques més habituals en el tractament de les enquestes quantitatives: la **depuració**, detecció i correcció de possibles errors en les dades, així com la **transformació** de variables.

Aquestes tasques, que impliquen típicament la generació de "nova" informació a partir de les dades originals, no formen part pròpiament de les operacions estadístiques. Per aquest motiu algunes de les aplicacions informàtiques disponibles per a l'anàlisi de dades quantitatives no inclouen eines prou eficients per desenvolupar-les. Contràriament, *jamovi* està especialment dissenyat per a donar suport a totes aquestes tasques de gestió de dades, tal com es veurà a continuació.[11]

Tot i que, en la seqüencia de treball habitual a l'anàlisi de dades en estudis socials, primer es depura el fitxer i després es fan les transformacions de variables pertinents per l'anàlisi, aquí -per una qüestió didàctica- s'invertirà l'ordre: es començarà tractant la transformació de les variables i, tot seguit, una vegada que disposis de les eines, es treballarà en la depuració de les dades.

Utilitzaras, en tot moment, l'arxiu de dades *MJDS_Dades_def.omv*. Es tracta d'una matriu creada a partir de l'estudi del CIS número 2889, titulat "*Actitudes hacia las tecnologías de la información y la comunicación"* de 12 de maig de 2011.[12] De totes les variables de l'estudi només se n'han conservat algunes, les que podeu consultar en el document *MJDS_QüestionariCIS.pdf*, que necessitaras per anar seguint el tutorial pràctic pas a pas.

A l'espai docent Campus Virtual (CV) de l'assignatura, o als arxius adjunts a MUJADES, hi pots trobar els arxius requerits: *MJDS_Dades_def.omv*, i *MJDS_QüestionariCIS.pdf* corresponents a l'enquesta del CIS.

1. Descarrega aquests dos arxius a una carpeta local del teu PC.
2. Fes doble-clic sobre el fitxer *MJDS_Dades_def.omv*
3. Una vegada obert a *jamovi* disposes d'una matriu de dades, amb 1.419 individus entrevistats i 70 variables ben definides (tipus de mesura i de data correctes i valors perduts definits).

Recodificació de variables

Tot i que l'ús més habitual de la "recodificació" de variables és l'agrupació d'un rang de valors existents a nous valors *(tal i com es veu en el primer exemple que es mostra a continuació)*, aquesta funció també es fa servir per assignar una nova codificació als valors originals.

En qualsevol cas, la "recodificació" comporta sempre un canvi en els valors de variables *(anomenats "nivells" a jamovi)*, canvi que pot implicar o no una modificació en la mètrica de la variable. En aquest sentit es poden donar quatre situacions diferents, en funció del tipus les variables original i final:

11 Addicionalment es guiarà també amb la pràctica d'alguns procediments que són més senzills, o possibles només, amb programació directa de codi en llenguatge R.

12 Es pot accedir a les dades originals a <http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=12624>. Tot i que originàriament les dades es van descarregar del servidor del CIS en un format compatible amb l'aplicació comercial d'anàlisi SPSS, van ser convertides en una base de dades de R (amb extensió *.rdata) i modificades convenientment per facilitar el seu ús a MUJADES.

(cc) BY-NC

		Variable FINAL *(transformada)*	
		Qualitativa (Nominal o Ordinal)	**Quantitativa** (Entera o Decimal)
Variable ORIGEN *(inicial)*	**Quantitativa** (Entera o Decimal)	[Edat: 0 - 90] *es transforma en* **[Joves / Adults / Vells]**	[Ideologia escala 1 a 10] *es transforma en* **[Ideologia escala 1 a 5]**
	Qualitativa (Nominal o Ordinal)	p.e., "nivell d'estudis" [SE / Prim. / Sec. / FP / Grau / Doctorat] *es transforma en* **[Prim o menys / Sec. / Univ.]**	p.e., "Ús d'Internet" [Si / No] *es transforma en* **[1 / 0]**

En els quatre casos, per "recodificar" una variable amb *jamovi*, cal utilitzar el mateix procediment d'aplicació d'una **regla de transformació**. Es tracta de generar una NOVA columna, o variable, a la base de dades, en funció dels valors d'una columna ja existent:

Variable ORIGEN + ***Regla de Transformació*** = Variable FINAL

Com es transforma amb codi R? En principi, totes les recodificacions requerides en aquest curs es poden realitzar amb el sistema de menús i transformacions de variables actualment implementat per *jamovi*. Però també es pot fer servir la consola Rj per a executar operacions directament amb codi R, i a l'apartat posterior "Transformacions amb codi R" s'explicarà com utilitzar R per a realitzar les transformacions que es veuran a continuació.

Recodificació (I). De variables quantitatives a qualitatives

En el següent exemple (recodificació de quantitativa a qualitativa), a partir de les variables p33, p34 i p35 (respectivament, ideologia de l'individu, del seu pare i de la seva mare), se'n generen tres de noves, que s'anomenaran ideol-entrev, ideol-pare i ideol-mare, i que tenen, a més dels valors perduts, les categories esquerra, centre i dreta.

Tal i com pots veure al *MJDS_QüestionariCIS.pdf* aquestes variables són quantitatives, de manera que estaran correctament definides en *jamovi* com a ORDINAL/ENTER, tenint com a valors perduts 96, 97, 98 i 99 (o >90).

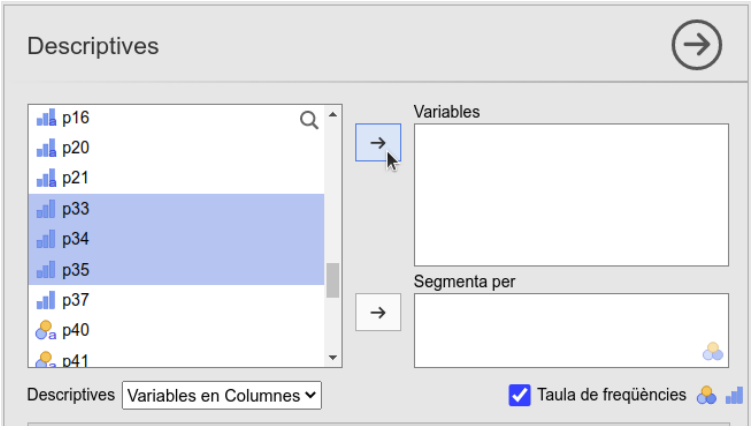
Comprova que aquesta definició és correcta:

1. Fes doble-clic al nom de la variable p33, que encapçala la columna de dades i s'obrirà el panell superior de definició de la variable.
2. Comprova que és una variable de tipus ORDINAL, que les dades són de tipus ENTER i que té definits valors perduts tots els superiors a 90.
3. Amb la fletxa de la dreta passa a la p34 i comprova el mateix. Així com a la p35.

A continuació, i per assegurar-nos que no hi ha cap problema, és útil obtenir un resum descriptiu de cada variable implicada.

1. Clica el menú superior "Anàlisis", i a la icona "Exploració" del submenú general de procediments analítics,

2. Al desplegable, clica l'opció "Descriptives" per obrir la finestra de definició del procediment.
3. Al llistat de la esquerra, pots seleccionar diverses variables, fent CTRL+Clic. D'aquesta forma pots seleccionar p33, p34 i p35, abans de clicar al botó FLETXA superior i incorporar aquestes tres variables a l'espai de treball de la esquerra del panell.
4. Clica el quadre corresponent per activar "Taules de freqüències":



Imatge 59. Selecció de diverses variables amb [Ctrl] + Clic

5. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general / Desar.

RECORDA: La "taula de freqüències" només es pot demanar per a variables de tipus NOMINAL o ORDINAL. Per aquesta raó és tan rellevant haver definit correctament el tipus de variable prèviament com a ORDINAL (originàriament havien estat importades com a CONTÍNUES).

El resultat obtingut és el següent:

Freqüències de p33

Nivells	Freqüències	% del Total	% Acumulat
1	55	6.1 %	6.1 %
2	60	6.6 %	12.7 %
3	157	17.3 %	30.0 %
4	131	14.4 %	44.4 %
5	264	29.1 %	73.5 %
6	98	10.8 %	84.3 %
7	83	9.2 %	93.5 %
8	38	4.2 %	97.7 %
9	6	0.7 %	98.3 %
10	15	1.7 %	100.0 %

Freqüències de p34

Nivells	Freqüències	% del Total	% Acumulat
1	42	5.3 %	5.3 %
2	49	6.1 %	11.4 %
3	128	16.0 %	27.4 %
4	102	12.8 %	40.2 %
5	172	21.6 %	61.8 %
6	91	11.4 %	73.2 %
7	101	12.7 %	85.8 %
8	72	9.0 %	94.9 %
9	13	1.6 %	96.5 %
10	28	3.5 %	100.0 %

Freqüències de p35

Nivells	Freqüències	% del Total	% Acumulat
1	44	5.5 %	5.5 %
2	44	5.5 %	10.9 %
3	117	14.5 %	25.4 %
4	127	15.8 %	41.2 %
5	204	25.3 %	66.5 %
6	99	12.3 %	78.8 %
7	87	10.8 %	89.6 %
8	52	6.5 %	96.0 %
9	17	2.1 %	98.1 %
10	15	1.9 %	100.0 %

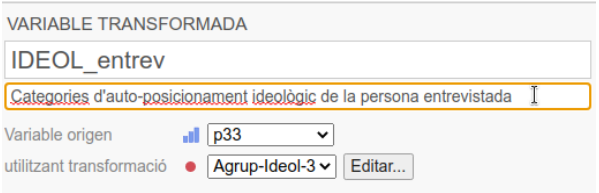
Imatge 60. Exemples de resultat de taules de freqüències

Observa com les variables es troben ben definides (sense perduts), de manera que podem aplicar una regla de transformació que assigni, per exemple, el valor *Esquerra* als valors originals 1 a 4, el valor *Centre* al valor original 5, i el valor *Dreta*, als originals 6 a 10.

Aquesta regla caldrà aplicar-la a les tres variables. Començarem per crear-la per a la p33:

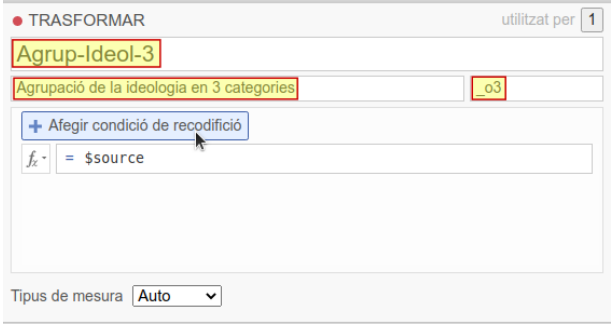
1. A la pantala de Dades, situa't sobre la variable p33 i clica la pestanya superior "Transformar".

2. Es crea una nova columna a la dreta de la columna p33, per defecte anomenada "p33 (2)" i basada en la variable d'origen p33 *(observa que, de moment, és una simple còpia sense els valors originals perduts)*.
3. Canvia el nom per IDEOL_entrev (amb guió-baix). Pots incloure una descripció més expressiva.



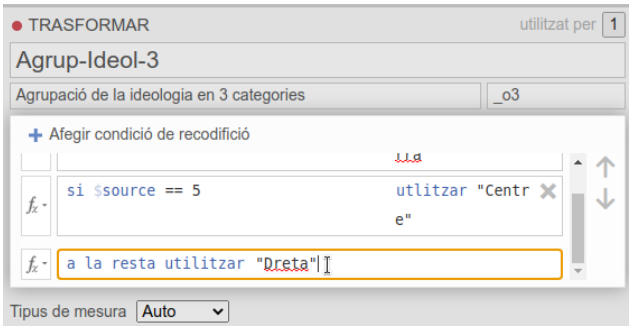
Imatge 61. Exemple de nova variable transformada

4. Fins aquí has creat una nova variable (o columna) que és una còpia exacta de la variable original p33. A continuació crearem la transformació necessària per fer la recodificació de variable quantitativa (1-10) a variable ordinal (Esquerra-Centre-Dreta), utilitzant els criteris esmentats més amunt.
5. Fes clic al desplegable que actualment indica [Cap] per a "utilitzant transformació", i selecciona l'opció [Crear Nova Trasformació...].
6. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte, anomenada "Transformar 1", que no té cap descripció, i amb una regla molt simple, *=$source*, que ve a significar *"fes que la nova variable tingui els mateixos valors que la variable original"*, i un tipus de mesura automàticament dependent de la variable original (Auto).
7. Primer **modifica el nom** perquè sigui més significatiu, canvia "Transform 1" per "Agrup-Ideol-3". Si vols pots introduir a sota una descripció més detallada, com ara "Agrupació de la ideologia en 3 categories".
8. A continuació, a la dreta de la descripció pots **introduir un sufix** amb el qual s'identificaran les noves variables a les que s'apliqui aquesta transformació: es pot utilitzar "_o3" (sense les cometes) per destacar que la variable transformada serà la versió ordinal amb categories.
9. A l'espai corresponent a la condició de recodificació cal **introduir la regla** de transformació (o regles) que canvia els valors segons el criteri corresponent, especificades en un llenguatge de programació comprensible per *jamovi*.
10. En aquest cas es tracta d'un conjunt de 3 regles, o "condicions de recodificació": Assignarem la etiqueta "Esquerra" als autoposicionaments ubicats entre 1 i 4, i "Centre" als ubicats a 5, de forma que la resta seran etiquetats com "Dreta" (de 6 a 10). Comença clicant el botó [Afegir condició de recodificació] per afegir la primera regla de recodificació.



Imatge 62. Botó per afegir condicions de recodificació

11. La primera condició pot ser: *Si el valor de la variable original és menor de 5, el valor a la variable transformada serà "Esquerra"*. En el llenguatge de programació que *jamovi* utilitza aquesta regla es pot expressar com:
    `si $source < 5 utilitza "Esquerra"`
12. Un cop la tinguis escrita, torna a clicar [Afegir condició de recodificació] per afegir la segona condició. Aquesta, que fa referència al posicionament de centre, es pot expressar en llenguatge de programació com:
    `si $source == 5 utilitza "Centre"`
13. Afegeix la condició final, que serà pels que encara no has canviat (els majors de 5), de manera que es pot expressar programàticament com:
    `A la resta utilitza "Dreta"`



Imatge 63. Combinació de les tres regles de recodificació

> **ATENCIÓ**: *jamovi* escriu una part del codi informàtic per tu *(la part blava)*, però cal que introdueixis personalment els criteris per a recodificar la variable *(la part negra)*. Es molt important que es respectin estrictament les convencions de codificació, per exemple, la **doble igualtat** (==), i no el signe simple d'igualtat (=), o les **cometes** per indicar que es tracta d'un text (**"xxxxx"**).

14. Deixa el tipus de mesura com a AUTO per a tenir una variable que sigui igual que l'original (ordinal). Tot i així, el tipus de dada te l'haurà passada a Text, de manera que ara tindràs una variable ORDINAL/TEXT.
15. Per tancar aquesta finestra, cal que cliquis la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra.
16. Com pots veure ara al panell de transformació, la nova variable està basada en els valors originals de la p33 i, a més, utilitza la transformació anomenada "Agrup-Ideol-3" que acabes de crear.
17. Recorda de desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Quan estàs recodificant amb transformació de variables, els operadors que pots utilitzar són:

> més gran	>= més gran o igual	== igual
< més petit	<= més petit o igual	!= no igual

Ara pots visualitzar els canvis fets a la matriu de dades. Comprova, fent freqüències de la variable antiga i nova, que s'han aplicat correctament les tres regles de transformació.

Freqüències de p33

Nivells	Freqüències	% del Total	% Acumulat
1	55	6.1 %	6.1 %
2	60	6.6 %	12.7 %
3	157	17.3 %	30.0 %
4	131	14.4 %	44.4 %
5	264	29.1 %	73.5 %
6	98	10.8 %	84.3 %
7	83	9.2 %	93.5 %
8	38	4.2 %	97.7 %
9	6	0.7 %	98.3 %
10	15	1.7 %	100.0 %

Freqüències de IDEOL-entrev

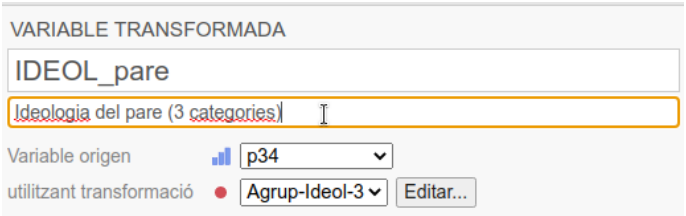
Nivells	Freqüències	% del Total	% Acumulat
Esquerra	403	44.4 %	44.4 %
Centre	264	29.1 %	73.5 %
Dreta	240	26.5 %	100.0 %

Imatge 64. Resutat d'una recodificació (freqüències)

Si tot ha funcionat correctament amb aquesta primera variable, ara ja pots procedir a aplicar LA MATEIXA transformació a les altres dues variables, per a recodificar-les en les noves variables ordinals "IDEOL_pare" i "IDEOL_mare" (respectivament, basades en p34 i p35).

Per a recodificar les variables origen p34 i p35 amb la transformació prèvia, fes:

1. A *jamovi*, clica al menú general "Dades" i fes clic al nom de la variable p34, que encapçala la columna de dades. Prem la tecla [Ctrl] i fes clic al nom de la variable p35. Has seleccionat dues columnes de valors.
2. Clica al menú superior la icona "Transformar" del submenú de procediments amb dades. Es creen dues noves columnes a la dreta de la p35, per defecte anomenades "p34 (2)" i "p35 (2)" , basades en les variables d'origen.
3. Clica el desplegable corresponent a "utilitzant transformació" (que per defecte indica "Cap").
4. Selecciona la transformació **que recodifica en tres intervals** ordinals, i que hem anomenat "Agrup-Ideol-3".
5. Comprova com s'han creat les variables "p34_o3" i "p35_o3" tal i com volíem.
6. Per a tancar la finestra de regles de transformació, clica la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra.
7. Canvia els noms de les noves variables per IDEOL_pare i IDEOL_mare (amb guió-baix) i inclou una descripció més expressiva, per exemple, *Ideologia del pare (3 categories), etc...*



Imatge 65. Exemple de transformació

8. Tal i com pots veure ara al panell de transformació, la nova variable està basada en els valors originals de la p34, i utilitza la transformació anomenada "Agrup-Ideol-3" que has creat anteriorment per a una altra variable.

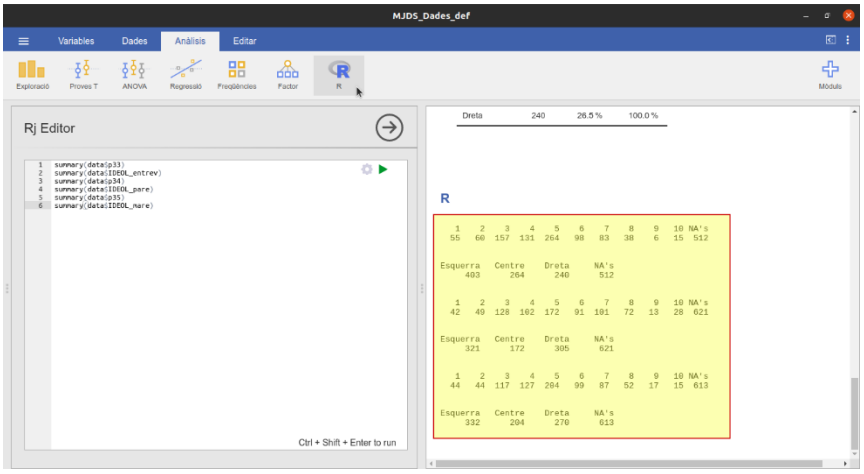
9. Per a tancar la finestra de definició de variables, clica a la icona circular FLETXA-AMUNT de la part superior dreta d'aquesta finestra.
10. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Fes el mateix que en el cas anterior anterior. Comprova, fent freqüències de les variables antiga i nova, que s'han aplicat correctament les regles de transformació.

Una de les formes de comprovar que les transformacions han estat correctes és fent servir la "consola R". En aquest cas, si vols una descripció sintètica de les sis variables que a hores d'ara tens (dues per la ideologia de la persona entrevistada, dues per la del pare i dues per la de la mare), pots fer:

Prerequisit: Tenir instal·lat a *jamovi* el mòdul d'extensió anomenat Rj.[13]

1. A l'apartat "Anàlisis", clicar a la icona "R". Al desplegable, seleccionar l'opció "Rj Editor",
2. Apareix la "consola" d'ordres R, i per defecte, una ordre inactiva d'exemple:
   ```
   # summary(data[1:3])
   ```
   Qualsevol línia que comenci per # es considera inactiva (està de color verd).
3. Escriu les següents ordres a la consola, en sis línies diferents:
   ```
   summary(data$p33)
   summary(data$IDEOL_entrev)
   summary(data$p34)
   summary(data$IDEOL_pare)
   summary(data$p35)
   summary(data$IDEOL_mare)
   ```
   Com has pogut veure, *jamovi* t'ajuda proposant ordres i variables disponibles.
4. Cadascuns d'aquestes ordres que has escrit en llenguatge R signifiquen: "mostra un resum descriptiu de la variable de la matriu de dades que es diu XXXX".
5. Per a executar el codi que hi ha escrit a la consola cal clicar al petit triangle verd de la part superior dreta de la consola. Alternativament, també es pot executar picant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho.
6. Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.



Imatge 66. Exemple de resultats d'exploració mitjançant codi R

13 A l'anterior apartat "Exploració de variables amb codi R" es descriu en detall com instal·lar aquesta funcionalitat de jamovi.

7. Per tancar el panell central, cal fer-ho com en tots els procediments d'anàlisi de *jamovi*, clicant la fletxa encerclada de la part superior dreta.
8. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

EXERCICI

1. A partir de la variable p6, genera una nova variable que ens digui quant va començar a fer servir Internet, si "de nen" (abans dels 10 anys), si "de jove" (entre els 10 i els 17) o "de gran" (després dels 17 anys). Per fer aquesta recodificació has de tenir en compte l'ordre en què escrius les condicions. Comprova que el resultat és correcte.

ATENCIÓ! Si feu servir regles de transformació creades prèviament, **tingueu cura** de no modificar-les, ja que això afectaria totes les variables sobre les que s'ha aplicat. És per això que a la finestra corresponent de l'especificació de les regles o funcions de transformació hi ha un indicador del nombre de variables a les quals s'ha aplicat la regla: utilitzat per 3, a la part superior dreta. Serveix per tenir cura i no "tocar" transformacions que afecten a diverses variables.

Alternativament, també es poden generar intervals qualitatius ordinals a partir d'una variable contínua numérica però utilitzant el procediment CALCULAR del menú Dades de *jamovi*.

Amb una única fórmula, complexa, amb múltiples condicions IF() organitzades en una "cascada lógica" es pot aconseguir d'una forma simple i relativament comprensible el mateix resultat directament i sense fer transformacions.
Per exemple, si tenim la variable contínua *"Edat en el moment d'abandonar els estudis"* (p48) i volem generar una altra d'intervals d'edat d'abandonament ("fins als 16 anys", "entre els 17 i els 19", "entre 20 i 25", "deprés dels 25 anys"):

1. Seleccionem p48, demanem per fer un CALCUL d'aquesta variable,
2. Utilitzarem la funció IF() per anar definint els diferents intervals, en ordre creixent, i amb IF() dins d'altres IF() per a definir els intervals successius. Així,
3. Es comença per definir el interval inferior, assignannt el seu valor o etiqueta, amb IF(p48<17,"fins als 16 anys",
4. A continuación de la coma, i com a segon terme d'aquest IF(), es defineix el segon interval, i la fórmula queda així
   IF(p48<17,"fins als 16 anys",IF(p48<20,"entre els 17 i els 19",
5. A continuación es continua, i com a segon terme d'aquest segon IF(), es defineix el tercer interval, i la formula queda així
   IF(p48<17,"fins als 16 anys",IF(p48<20,"entre els 17 i els 19",IF(p48<25,"entre 20 i 25"
6. I quan s'ha arribat a l'interval superior, ja es pot completar directament amb el valor corresponent, queda així
   IF(p48<17,"fins als 16 anys",IF(p48<20,"entre els 17 i els 19",IF(p48<25,"entre 20 i 25","després dels 25 anys"
7. Finalment per complatar aquesta fórmula cal tancar TOTS els paréntesis que s'han obert. Son un menys que el nombre d'intervals generats (k-1 parèntesis).

8. El resultat final, en aquest cas, sería la fórmula única que genera una nova variable d'intervals:
   IF(p48<17,"fins als 16 anys",IF(p48<20,"entre els 17 i els 19",IF(p48<25,"entre 20 i 25","després dels 25 anys")))

Igualment es pot fer amb qualsevol nombre d'intervals, ajustant la seqüencia de IF(), inclosos dins d'altres IF(). Els casos amb valors perduts a la variable original continuaran tenint valor perdut a la calculada.

EXERCICI

1. Utilitza aquest mètode de calcular per, a partir de la variable p6, genera una nova variable que ens digui quant va començar a fer servir Internet, si "de nen" (abans dels 10 anys), si "de jove" (entre els 10 i els 17) o "de gran" (després dels 17 anys). Comprova que el resultat és el mateix que amb el mètode de transformació.
2. Quin mètode et sembla més senzill? O més eficient? O més ràpid?

Recodificació (II). De variables qualitatives a qualitatives

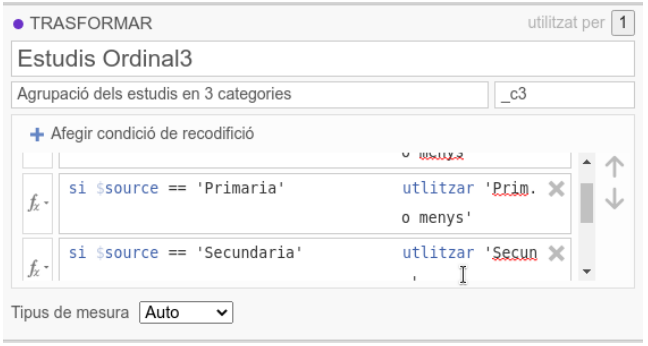
En el següent exemple, a partir de les variable ESTUDIS (ordinal de 6 categories) se'n generarà una de nova que s'anomenarà "estudis3"; amb només tres categories: *Primaria o menys*, *Secundaria* i *Universitaris*.

La variable ESTUDIS hauria d'estar definida com a ORDINAL/TEXT i amb el valor perdut 'N.C.'. Comprovarem, en primer lloc, si això és correcte i, posteriorment, farem l'agrupació.

1. Fes doble-clic sobre el nom de la variable ESTUDIS, que encapçala la columna de dades. S'obre el panell superior de definició de la variable.
2. Comprova que efectivament la variable ESTUDIS està definida com a tipus ORDINAL, amb dades de tipus TEXT, i que el codi N.C. ha estat definit com a perdut. En cas que no sigui així, modifica-la, fins que s'ajusti a aquests paràmetres.[14]
3. Clica a sobre de l'etiqueta "Dades" del menú superior i a la icona "Transformar" del submenú de procediments amb dades. Es crea una nova columna a la dreta de la variable ESTUDIS que, per defecte, anomena "ESTUDIS (2)" i que es basa en la variable d'origen.
4. Canvia el nom a "estudis3". Pots incloure una descripció més expressiva (p. ex. *Nivell d'estudis en 3 categories*).
5. A continuació cal crear una nova transformació per poder fer la recodificació. Clica el desplegable que actualment indica [Cap] per a "utilitzant transformació", i selecciona l'opció [Crear Nova Trasformació...].
6. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada "Transformar X", amb una regla molt simple *"els mateixos valors que la variable origen"* (=$source), i un tipus de variable automàticament dependent de la variable original.
7. Primer **modifica el nom** de la transformació per un de més significatiu: canvia "Transform X" per "Estudis Ordinal3". Si vols pots introduir una descripció més detallada, com ara "Agrupació dels estudis en 3 categories".

14 Pots consultar apartats anteriors d'aquest manual pràctic, on s'ha vist com fer-ho.

8. A continuació, a la dreta de la descripció, pots **introduir un sufix** amb el que s'identificaran les noves variables que siguin transformades per aquesta regla.
9. A l'espai corresponent a la condició de recodificació cal **introduir la regla** de transformació que canvia els valors segons el criteri corresponent, especificades en un llenguatge de programació comprensible per *jamovi*. En aquest cas assignarem les etiquetes corresponents a la recodificació desitjada segons el següent criteri: "*Prim. o menys*", a les primeres dues categories; "*Secund.*", a les dues següents i "*Univers.*" a les altres dues.
10. Clica el botó [+ Afegir condició de recodificació] per afegir cada nova condició de recodificació.
11. La primera condició pot ser: Si el valor de la variable original és 'Sense estudis', aleshores el valor a la variable transformada serà "*Prim. o menys*".
    En el llenguatge de programació de *jamovi*:
    si $source == 'Sense estudis' utilitzar 'Prim. o menys'
12. Cal introduir **cada** condició en una **nova línia**. Clica el botó [+ Afegir condició de recodificació] per a crear la següent condició.
13. Escriu la segona condició:
    si $source == 'Primaria' utilitzar 'Prim. o menys'
14. Introdueix la resta de condicions en llenguatge de programació, clicant [+ Afegir condició de recodificació] tantes vegades com calgui:
    si $source == 'Secundaria' utilitzar 'Secund.'
    si $source == 'F.P.' utilitzar 'Secund.'
    si $source == 'Univers mitjos' utilitzar 'Univers.'
    si $source == 'Univers sup' utilitzar 'Univers.'



Imatge 67. Recodificació de variable qualitativa en qualitativa

15. La condició final, que afecta tots els valors de la variable original que no han estat encara canviats, es pot expressar programàticament com:
    else use
    Aquesta instrucció l'haguéssim pogut fer servir, per exemple, per definir els universitaris, un cop teníem les altres categories definides. Si ho haguéssim fet així, les condicions en llenguatge jamovi haguessin quedat:
    si $source == 'Sense estudis' utilitzar 'Prim. o menys'
    si $source == 'Primaria' utilitzar 'Prim. o menys'
    si $source == 'Secundaria' utilitzar 'Secund.'
    si $source == 'F.P.' utilitzar 'Secund.'
    a la resta utilitzar 'Univers.'
16. Aquesta darrera instrucció diu que tots els valors que no han estat prèviament definits, els passi a *'Univers'*. I com que són els únics que queden per definir, la lògica és correcta.
17. A la part inferior de la finestra de condicions deixa el tipus de mesura com a AUTO per a "copiar" el mateix tipus que la variable original (ordinal).

18. Per a tancar la finestra de regles de transformació, clica la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra.
19. Tal i com pots veure al panell de transformació, la nova variable està basada en els valors originals de la variable ESTUDIS, i utilitza la transformació anomenada "Estudis-Ordinal3" que acabes de crear.
20. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

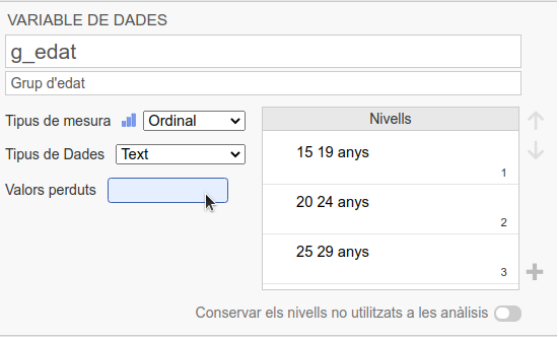
Demana un resum descriptiu simple de les dues variables ESTUDIS i estudis3, per comparar-les i comprovar si la recodificació s'ha generat amb èxit.

Recodificació (III). De variables qualitatives a quantitatives

Si volguéssim transformar una variable qualitativa ORDINAL a quantitativa, el procediment és similar als casos anteriors: crear una variable nova, derivada de la original mitjançant l'aplicació d'una regla de TRANSFORMACIÓ.
Cal tenir en compte, en aquest cas, que les categories recodificades han de ser numèriques, per tant caldrà modificar el tipus de la variable transformada perquè s'ajusti al format adient: ENTER o DECIMAL.

Per a transformar la variable g_edat (que és ordinal/text) a quantitativa cal assignar a cada categoria original, o "interval d'edat" el corresponent valor per la seva "marca de classe". Així, la transformació es resumeix a recodificar *"15 19 anys" com* a 17,5; recodificar *"20 24 anys" com* a 22,5; i *"25 29 anys" com* a 27,5. A cadascuna de les tres etiquetes o nivells, doncs, els assignem el valor (que serà continua/decimal) que correspon al punt mig de l'interval que representen.
Per fer-ho, segueix les passes següents:

1. Doble clic sobre el nom de la variable, que encapçala la columna de dades corresponent. Comprova que, efectivament, la variable "g_edat" està definida com a ORDINAL amb dades de tipus TEXT. Si no és el cas, **modifica'n la definició** fins que s'ajusti a aquests paràmetres.[15]
2. En la mateixa finestra comprova, també, que no té valors perduts.



Imatge 68. Comprovació de definició de variable

3. Selecciona la variable g_edat clicant el nom que encapçala la columna. Clica l'etiqueta "Dades" del menú superior i clica, també, la icona "Trasformar" del submenú de procediments amb dades. Automàticament es crea una nova columna a la dreta de la g_edat, que per defecte anomena "g_edat (2)" i que es basa en la variable d'origen.

15 Pots consultar apartats anteriors d'aquesta mateixa pràctica, on s'ha vist com fer-ho.

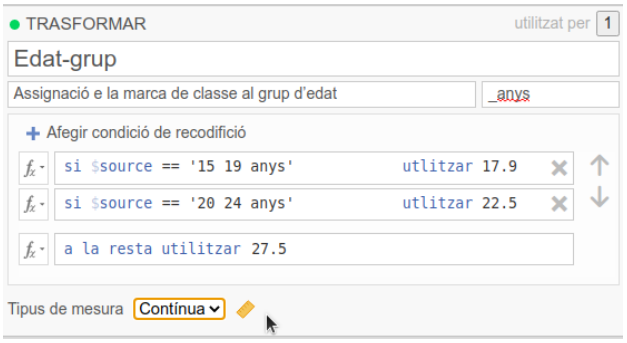
(cc) BY-NC

4. Pots (o no) canviar el nom i incloure una descripció més expressiva, com ara: "Edat estimada a partir de grups d'edat".
5. A continuació cal generar una nova transformació per tal de fer la recodificació de variable qualitativa ordinal (g_edat, 3 intervals) a variable quantitativa decimal (gedat_num, amb 1 decimal). Clica el desplegable que actualment indica [Cap] per a "utilitzant transformació", i llavors selecciona l'opció [Crear Nova Trasformació].
6. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada "Transform X" (el nom depèn, òbviament, del nombre de transformacions fetes anteriorment). La regla que ens dona per defecte és *"els mateixos valors que la variable origen"* (= $source), i un tipus de variable que depèn de la variable original (Auto).
7. Primer **modifica el nom** perquè sigui més explicatiu, canvia "Transform X" per "Edat-grup". Si vols pots introduir a sota una descripció més detallada, com ara "Assignació e la marca de classe al grup d'edat".[16]
8. A continuació, a la dreta de la descripció pots **introduir un sufix** amb el que s'identificaran les noves variables que siguin transformades seguint aquesta mateixa regla: es pot utilitzar "_anys" (sense les cometes) per destacar que la variable transformada serà la versió numèrica.
9. A l'espai corresponent a la condició de recodificació cal **introduir la regla** de transformació que canvia els valors segons el criteri corresponent, especificades en un llenguatge de programació comprensible per *jamovi*. Clica el botó [+ Afegir condició de recodificació] per afegir cada nova condició de recodificació.
10. La primera condició pot ser: "Si el valor de la variable original és '15 19 anys' llavors el valor a la variable transformada serà 17.5". En el llenguatge de programació:
    ```
    si $source == '15 19 anys' utilitza 17.5
    ```

    Compte! Per escriure decimals en *jamovi* has de fer sevir el punt i no la coma (perquè s'està utilitzant el llenguatge R, en anglès).
    Fixa't en la diferència entre l'etiqueta de la variable original, entre cometes perquè és text, i el valor de la nova variable, sense cometes perquè és una xifra.

11. Cal introduir cada condició en una nova línia. Per fer-ho clica el botó [+ Afegir condició de recodificació]. Les següents condicions en llenguatge de programació segueixen la mateixa lògica:
    ```
    Si $source == '20 24 anys' utilitza 22.5
    ```
12. La condició final, que afecta tots els valors de la variable original que no han estat encara canviats (els '25 29 anys'), es pot expressar programàticament com:
    ```
    a la resta utilitza 27.5
    ```
13. Canvia el tipus de mesura a CONTINUA per a definir la nova variable com a quantitativa.

16 Documentar o descriure en detall les transformacions que construeixis ajuda molt quan es volen reutilitzar posteriorment.



Imatge 69. Recodificació de variable qualitativa en quantitativa

14. Per a tancar la finestra de regles de transformació, cal clicar la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra.
15. Tal i com pots veure al panell de transformació, la nova variable està basada en els valors originals de g_edat, i utilitza la transformació anomenada "Edat_grup" que acabes de crear.
16. Si vols comprovar que la transformació és correcta, hauràs de demanar freqüències de la variable original (g_edat) i de la nova variable (gedat_num). Ara bé, si ho intentes veuràs que això no és possible, ja que *jamovi* no permet obtenir freqüències de variables definides com a *Continua*. Tot i que podries fer el canvi a Ordinal, en aquest cas es recomana fer la comprovació utilitzant els menús d'R.

Per fer-ho:

a. A l'apartat "Anàlisis", clica la icona "R" i al desplegable, selecciona l'opció "Rj Editor". Apareix la "consola" d'ordres R i, per defecte, una ordre inactiva d'exemple: `# summary(data[1:3])`
b. Escriu les següents ordres a la consola, en dues línies diferents:
```
table(data$g_edat)
table(data$gedat_num)
```
c. Executa el codi, clicant el petit triangle verd de la part superior dreta de la consola, o clicant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho i comprova que els resultats són coherents.
d. Per tancar el panell central clica la fletxa encerclada de la part superior dreta.

17. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Recodificació (IV). De variables quantitatives a quantitatives

El pas d'una variable quantitativa a una altra de quantitativa no difereix dels anteriors.

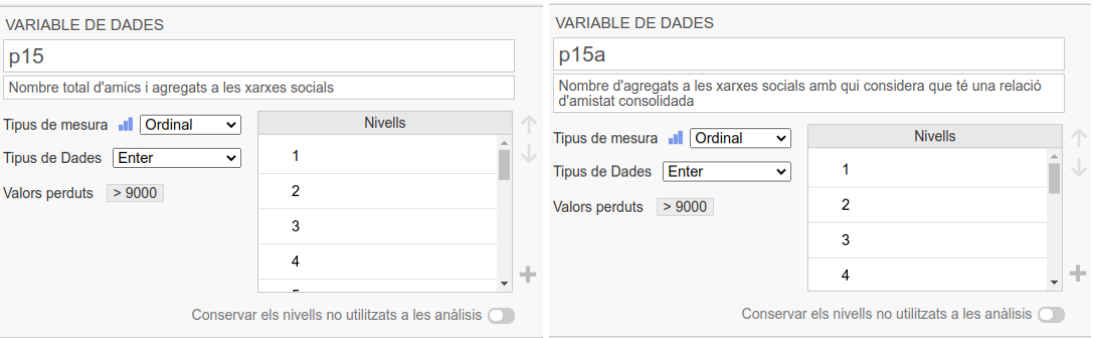
Càlculs amb variables quantitatives

Si volem fer càlculs amb variables quantitatives, creant-ne d'altres que siguin, per exemple, elevar-ne una al quadrat , o aconseguir la seva versió "normalitzada", ho podem fer a partir de menús, utilitzant el procediment de càlcul amb variables ("Calcular"). Aquest procediment permet aplicar un conjunt de funcions matemàtiques i estadístiques als valors de variabls existents.

També permet operar amb més d'una variable per generar noves columnes (variables) amb el resultat d'operacions matemàtiques a partir de les dades d'altres columnes.

Per exemple, disposem de dues variables que ens informen sobre el nombre de persones, contactes, amics i agregats que els entrevistats tenen a les xarxes socials: la p15 indica el nombre

total declarat, mentre que la p15a indica quants d'aquests contactes considera que són una "relació d'amistat consolidada". Ambdues variables han estat definides com a ORDINAL i ENTER, i no tenen valors impossibles, ja que els valors superiors a 9000 (9997, 9998 i 9999) han estat definits com a valors perduts:



Imatge 70. Exemple de variables quantitatives sobre contactes a xarxes socials

Si demanem descriptius d'ambdues obtenim:

Descriptives

	p15	p15a
N	1081	1054
Perduts	338	365
Mitjana	249	42.7
Mediana	175	20.0
Desviació estàndard	333	70.0
Mínim	1	1
Màxim	4600	1000

Imatge 71. Descriptius sobre contactes a xarxes socials

Observa i recorda que.... aquestes dues variables són quantitatives i, per tant, podríem haver tingut la temptació de definir-les com a *Continua/Enter*. El motiu de no fer-ho és poder demanar, si cal, una taula de freqüències d'aquestes variables. Afortunadament, el què ens dona la capacitat de poder fer operacions aritmètiques entre variables és la seva condició d'*Enter* (no la de *Contínua*).

Operacions i càlculs utilitzant una única variable

Anem a estandarditzar (normalitzar) la variable *Nombre d'amics i agregats a les xarxes socials* (p15). Recorda que el procés d'estandardització és el pas d'una variable amb unes mesures qualsevols a una variable que té una mitjana de 0 i una variància d'1. Per fer-ho, cal que a cada valor de la variable en restem la mitjana i dividim el resultat entre la desviació típica.

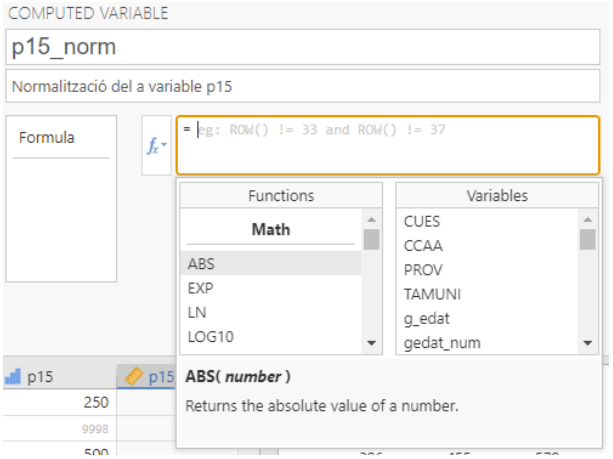
$$Z_i = \frac{x_i - \bar{x}}{S}$$

Per crear aquesta nova variable, calculada a partir de p15, segueix els següents pasos:

1. Comprova, fent doble-clic sobre el nom de la variable, que està definida com *Ordinal/Enter* i que els valors superiors a 9.000 no seran tinguts en compte (estan definits com a *Valors perduts*). Si no és el cas, **modifica'n la definició** fins que s'ajusti a aquests paràmetres.[17]

17 Pots consultar apartats anteriors d'aquesta mateixa pràctica, on s'ha vist com fer-ho.

2. Clicant sobre la variable p15, selecciona-la. Clica l'etiqueta "Dades" del menú superior i, a continuació, selecciona la icona "Calcular" del submenú. Es crea una nova columna a la dreta de la p15, per defecte anomenada amb una combinació de dues lletres i basada en la variable d'origen.
3. Canvia-li el nom. Anomena-la, per exemple, *p15_norm* per denotar és la versió normalitzada de la variable p15. Pots també escriure una descripció més acurada de la variable, on posa *Description*, per exemple, "*Normalització de la variable p15"*.
4. Clicant sobre el lloc on hi ha la funció (f_x) es desplega un menú amb totes les transformacions possibles:

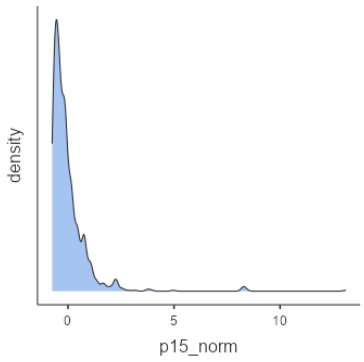


Imatge 72. Menú de càlcul amb variables

5. L'estandardització ve notada per la **funció Z**. Busca-la al menú de Funcions, selecciona-la amb doble-clic. S'escriu codi a la finestra, i el cursos es coloca a dins del parèntesi.
6. A continuació, busca la variable que vols estandarditzar (p15). Selecciona-la amb doble-clic, i s'ecriurà a dins del parèntesi.
7. Amb l'ajut d'aquests menús, en llenguatge de programació comprensible per jamovi li hauràs escrit:
   `Z(p15)`
8. Per executar aquest codi de càlcul, pica la tecla [Intro]. Has de veure com la nova columna calculada s'omple de xifres.
9. Demana un resultat descriptiu de les dues variables, p15 i p15_norm, i comprova que mentre la mitjana de la primera és 249, de la segona és pràcticament 0.[18]
10. Comprova també que la desviació típica de la segona és 1 (i, en conseqüència, la variància serà 1). La nova variable, efectivament, està normalitzada.
11. Tot i que ara la mitjana és 0 i la desviació és 1, podem comprovar a partir de la funció de densitat que no s'ajusta gens a la típica "campana de Gauss", ja que té una forta asimetria cap a la dreta. Pots comprovar com NO és així, hi ha un biaix clar cap a la dreta.

[18] 5.62e-18 és la notació anomenada "científica" de 0,000000000000000000562, amb 18 zeros decimals a la esquerra de 5,62.

(cc) BY-NC



Imatge 73. Gràfica de densitat de la variable P15 normalitzada

12. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

El llenguatge de programació ofereix múltiples funcions per a realitzar els càlculs. Totes tenen la mateixa "sintaxi": paraula-clau (variable-al-qual-s'aplica), p .ex. **`z(p15)`**.
Tal i com hem comentat més amunt, les diferents opcions de *jamovi* es poden consultar (i seleccionar per a utilitzar) clicant el botó [$f_x =$] situat a l'esquerra de la finestra de fórmules.

Operacions i càlculs amb diferents variables

El paquet estadístic *jamovi* permet operar aritmèticament amb dades de diferents variables (columnes) per a cada cas (fila). Utilitzant el procediment general de CÀLCUL (per exemple, una suma, multiplicació, divisió, mitjana...) es pot generar una *nova* variable amb valors funcionalment dependents dels valors d'altres variables existents.

L'exemple que farem servir en aquesta pràctica, és una simple operació entre dues variables, la p15 i la p15a. La variable p15 indica el nombre total de contactes que té la ersona entrevistada a les xarxes socials, mentre que p15a indica quants d'aquests contactes poden ser considerats com a relacions consolidades d'amistat. Ambdues variables han estat definides com a ORDINAL/ENTER, i estan depurades, en el sentit que no tenen valors impossibles, ja que els valors superiors a 9000 (9997, 9998 i 9999) han estat definits com a *Valors perduts*.
Si volem establir la relació entre la quantitat de contactes a xarxes socials i els "veritables amics" per a cada persona que ha respost l'enquesta es pot fer una simple divisió entre les variables p15 i p15a. El quocient o raó p15a/p15 és un indicador del nivell de "contactes fiables" de cada persona enquestada. Quan aquest indicador sigui 1, o proper a 1, voldrà dir que té la màxima confiabilitat amb els contactes de les xarxes, mentre que si és proper a 0 significa una mínima fiabilitat en els contactes de les xarxes.

Per calcular la nova variable CONFIA_xarxes, farem el següent:

1. Tal i com hem fet abans, comprova que la definició d'ambdues variables és correcta. Per fer-ho, en cada variable per separat, fes doble-clic sobre el nom, comprova que està definida com a *Ordinal/Enter* i que els valors superiors a 9000 no seran tinguts en compte (estan definits com a *Valors perduts*). Si no és el cas, **modifica'n la definició** fins que s'ajusti a aquests paràmetres.
2. Clica sobre del nom de la variable p15a, que encapçala la columna de dades, per seleccionar-la.

3. Clica l'etiqueta "Dades" del menú superior i, a continuació, selecciona la icona "Calcular" del submenú. Es crea una nova columna a la dreta de la p15a, per defecte anomenada amb una combinació de dues lletres i basada en la variable d'origen.
4. Canvia el nom a "CONFIA_xarxes". Pots incloure una descripció per recordar que la nova variable està calculada a partir de p15 i la p15a, p. ex. *"Índex de fiabilitat de xarxes (amics/contactes)"*.
5. Clica la finestra de fórmules (a continuació de la igualtat), on caldrà que escriguis la funció que realitza el càlcul de la nova variable.
6. El criteri per a la construcció és dir-li *"Fes un càlcul dividint p15a entre p15"*. En llenguatge comprensible per *jamovi* a la finestra de fórmules:
   `p15a/p15`
   Executa el càlcul prement la tecla [Intro]. Apareixen a la nova columna els valors calculats.
7. Comprova que ara la nova columna conté dades entre 0 i 1. Aquesta comprovació la pots fer mirant directament alguns dels valors de la variable, o demanant descriptius on, tot i no obtenir la taula de freqüències d'aquesta variable perquè és *contínua*, sí que veiem que el mínim és 0,00109 i el màxim és 1:

Descriptives

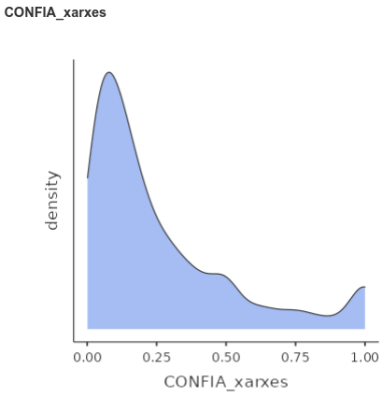
	p15	p15a	CONFIA_xarxes
N	1081	1054	1051
Perduts	338	365	368
Mitjana	249	42.7	0.267
Mediana	175	20.0	0.167
Desviació estàndard	333	70.0	0.271
Mínim	1	1	0.00109
Màxim	4600	1000	1.00

Imatge 74. Descriptius sobre variables originals i variable calculada

Observació. Si és més fàcil per a la interpretació, pots construir aquesta variable en tant per cent en comptes de fer-ho en tant per u. Si t'ho estimes més així, hauràs d'escriure la fórmula: **`P15a/p15*100`**

8. Si vols obtenir la freqüència de la variable construïda, hauràs d'utilitzar codi R:
   a. A l'apartat "Anàlisis", clica la icona "R" i al desplegable, selecciona l'opció "Rj Editor". Apareix la "consola" d'ordres R i, per defecte, una ordre inactiva d'exemple: **`#summary(data[1:3])`**
   b. Escriu la següent ordre a la consola:
      `table(data$CONFIA_xarxes)`
   c. Executa el codi, clicant el petit triangle verd de la part superior dreta de la consola, o clicant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho i comprova que el rang és de 0 a 1.
   d. Per tancar el panell central, clica la fletxa encerclada de la part superior dreta.
9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Pots comprovar la construcció d'aquest índex demanant una representació gràfica de la funció de densitat de la distribució. Tornaràs a veure que es troba entre 0 i 1 i que té una moda absoluta i dues modes relatives:



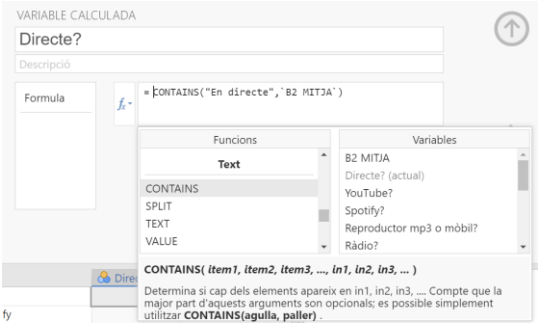
Imatge 75. Diagrama de densitat de l'indicador calculat

EXERCICI

- Genera una variable utilitzant les variables p13a01 i p13a02. Tal i com pots veure en el qüestionari es tracta de dues variables que intenten recollir el temps (diari) que passen a les xarxes socials. Així, p13a01 són les hores a les xarxes socials i p13a02 són els minuts. La nova variable que has de crear és el temps (en minuts) que l'individu es passa connectat a les xarxes socials.
- Per passar d'hores a minuts, òbviament, cal multiplicar les hores per 60, de manera que la "fórmula" o "funció" que cal utilitzar per a la transformació seria:
  `MINUTS_xarxes = (p13a01 * 60) + p13a02`
- A tall d'exemple, el quart individu del qüestionari, que ha respost *p13a01==1* i *p13a02==30*, el resultat serà que, diàriament està 90 minuts connectat a les xarxes socials, (1*60)+30=90 minuts.
- D'aquesta nova variable, calcula totes les mesures de tendència central possibles i representa-la, gràficament, a partir d'un diagrama de caixa.
- Fent servir sintaxi d'R, intenta millorar aquest diagrama (per exemple, que no grafiqui els valors per sobre de 250 minuts).

Càlculs amb variables qualitatives

Els procediment de CÀLCUL a *jamovi* no està restringit a les variables quantitatives, també hi ha funcions disponibles per a variables amb dades textuals. Al menú de funcions disponibles de la finestra emergent de càlcul hi ha una secció anomenada TEXT.



Imatge 76. Finestra CALCULAR, amb funcions de text

Per exemple, amb la funció CONTAINS() es pot consultar si una porció concreta de text es troba o no dins d'una resposta qualitativa, i amb SPLIT() es pot dividir una resposta qualitativa segons un separador concret i extraure informació d'una posició concreta de la sèrie obtinguda.

Operacions (càlculs) amb variables qualitatives multiresposta

El paquet estadístic *jamovi* permet operar amb variables "qualitatives", per extraure informació textual concreta. Un cas típic és el de dades originals d'enquesta que inclouen preguntes amb resposta múltiple que està recopilada amb tota la informació en una única variable nominal textual i amb cada resposta separada per una coma. Alguns dels serveis de realització de formularis i enquestes habitualment utilitzats a la recerca social (Google Forms, Microsoft Forms, Survey Monkey, Lime Survey,...) proporcionen les dades recopilades de preguntes de resposta múltiple en aquest format textual separat per comes.

Per exemple, al qüestionari del CIS corresponent a les dades de "Dades3.omv" s'ha demanat *"p12. ¿Qué red social utilizas principalment en Internet?"* i s'ha obtingut informació en el següent format:

p12m
Facebook, Tuenti
Tuenti
Tuenti
Facebook, Tuenti, Twitter
Bebo, Tuenti
Tuenti, Facebook
Tuenti, Facebook
Facebook, Tuenti
...

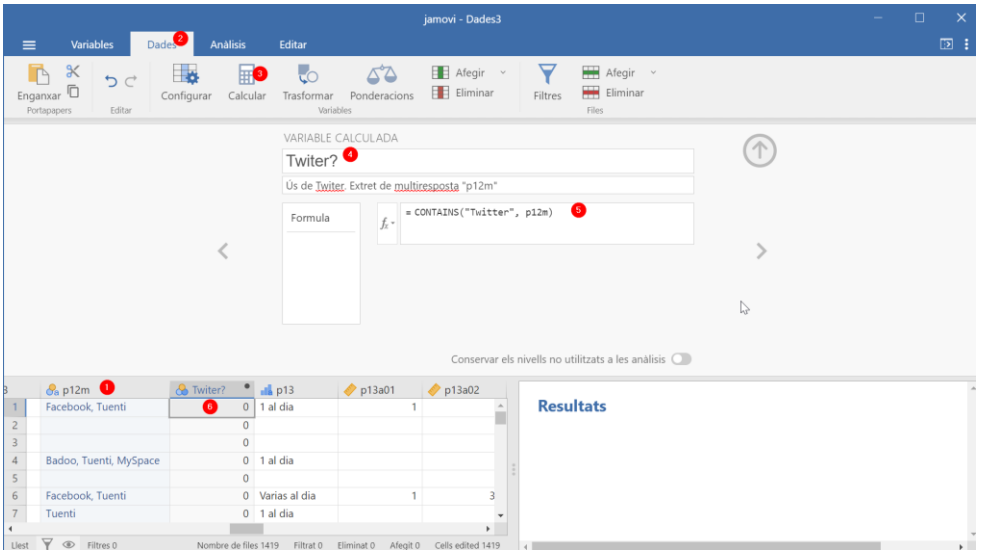
Per analitzar aquest tipus de respostes, prèviament, cal extreure informació corresponent de cadascuna de les unitats rellevants: val a dir, de cadascuna de les possibles respostes múltiples, fins aconseguir "aïllar-les" en variables/columnes pròpies. Aquestes noves variables amb 2 valors, de veritat o de falsedat (codificats respectivament amb les xifres 1 i 0), s'anomenen "dicotòmiques" o "lògiques".

Per a generar noves variables lògiques amb informació sobre si la enquestada utilitza o no cadascuna de les xarxes que interessi, per exemple "Facebook?", "Twiter?", "Tuenti?", etc... segueix el procediment següent:

1. Genera una taula de freqüències de la variable original, per tenir a la vista les diferents respostes. Aquesta mateixa informació es pot aconseguir si es disposa del qüestionari original que indica les multi-respostes. A partir d'aquesta taula es més fàcil determinar

quines son les noves variables lógiques que s'han de crear, i per tant, quantes columnes noves cal afegir a la matriu de dades.

2. Caldrà un procediments "Càlcul" per a cada nova variable o columna, en aquest cas per a cada xarxa social.
3. Clica a sobre del nom que encapçala la columna de les dades, per seleccionar la variable original textual amb la multi-resposta (p12m).
4. Al menú superior, clica l'etiqueta "Dades" i, a continuació, selecciona la icona "Calcular" del submenú. Es crea una nova columna just a la dreta de la variable original, per defecte anomenada amb una combinació de dues lletres.
5. Canvia el nom per a identificar la primera multiresposta, p.e. "Twitter?". Pots incloure una descripció per recordar que la nova variable està calculada a partir de la original.
6. Clica la finestra de fórmules (a continuació de la igualtat), on caldrà que escriguis la funció que realitza el "càlcul" de la nova variable.
7. El criteri per a la construcció és dir-li *"Fes un càlcul determinant si el text "X" es troba a dins del contingut de la resposta a la variable B2 MITJÀ"*. En llenguatge propi de *jamovi* a la finestra de fórmules, escriu:
   `CONTAINS("Twitter", p12m)`
   Executa el càlcul prement la tecla [Intro].



Imatge 77. CALCULAR noves variables des de multiresposta

8. Apareixen a la nova columna "Twitter?" els valors calculats. Comprova que ara la nova columna conté dades 0 i 1 que corresponen a si la resposta múltiple incorpora (1) o no (0) la resposta "Twitter" a la p12m.
9. Si s'obté un missatge d'error (fons vermell), o si es comprova per a casos concrets que aquesta operació no ha tingut l'efecte desitjat, cal revisar la redacció de la funció:
   a) els texts cercats han d'estar "entre cometes dobles" i exactament tal com apareixen a la taula de freqüències, i
   b) els noms de les variables es millor obtenir-los fent clic al menú de formules [f_x] i seleccionant-los (doble-clic) del menú desplegable dret de Variables.
10. Repetir el mateix procediment per a totes les possibles respostes individuals, que generaran tantes noves variables lògiques 0-1.

Les variables així obtingudes, mitjançant càlcul amb la funció CONTAINS(), son de tipus "nominal" amb dades "enter", i per tant se'n pot fer qualsevol tipus de càlcul, d'estadística o de gràfic.

Transformacions complexes

(1) Combinació de recodificacions i càlculs

Sovint ens pot interessar construir índexs a partir de transformacions de variables nominals o ordinals , que no permeten cap càlcul aritmètic. Un exemple el trobem si volem crear la variable QUANTITATIVA "*Per a quants aspectes diferents utilitza Internet"* a partir de la informació disponible a la pregunta 9 del qüestionari, que demana precisament per aquests diferents usos. A la matriu de dades disposem de 14 variables (de la p901 a la p914) que ens informen per separat de cadascun dels usos d'Internet, codificades com a "Si" o "No" en funció de si la persona entrevistada en fa o no aquell ús concret. Per combinar múltiples variables qualitatives i construir una nova variable quantitativa cal combinar dues funcions diferents de *jamovi*: Transformar i Calcular.

Si en comptes de "Si" o "No", tinguéssim valors numèrics (1 i 0, respectivament), seria molt senzill, en tant que sumaríem directament les 14 variables. Seguint aquesta lògica el què fem és:

1. Transformem (TRANSFORMAR) les variables inicials (de la p901 a la p914) en 14 noves variables amb uns o zeros, en funció de si en fa ús o no. Es tracta de transformar variables qualitatives (Si-No) a variables numèriques (1-0), tal i com ja s'ha vist en un apartat anterior de la pràctica.
2. Un cop tinguem les 14 noves variables quantitatives amb 1 o 0, ja podrem fer el càlcul (CALCULAR) sumant les 14 variables, de manera que obtindrem una variable numèrica compresa entre el 0 i el 14. Així la nova variable volem que sigui quantitativa (Contínua/Enter), amb un rang teòric compres entre 0 (no fa cap ús d'Internet) i 14 (utilitza Internet per a 14 coses diferents). Si, a més la dividim entre 14 tindrem un índex amb variació entre 0 i 1.

Així, comencem transformant la "primera" variable nominal a quantitativa, tot creant una nova regla de transformació: *"Si el valor de la variable original és 'Sí' llavors el valor a la variable transformada serà 1, si el valor és 'No' llavors serà 0, i en qualsevol altre cas quedarà 'en blanc' (valor perdut)".* Posteriorment es podrà aplicar aquesta regla a la resta de variables.

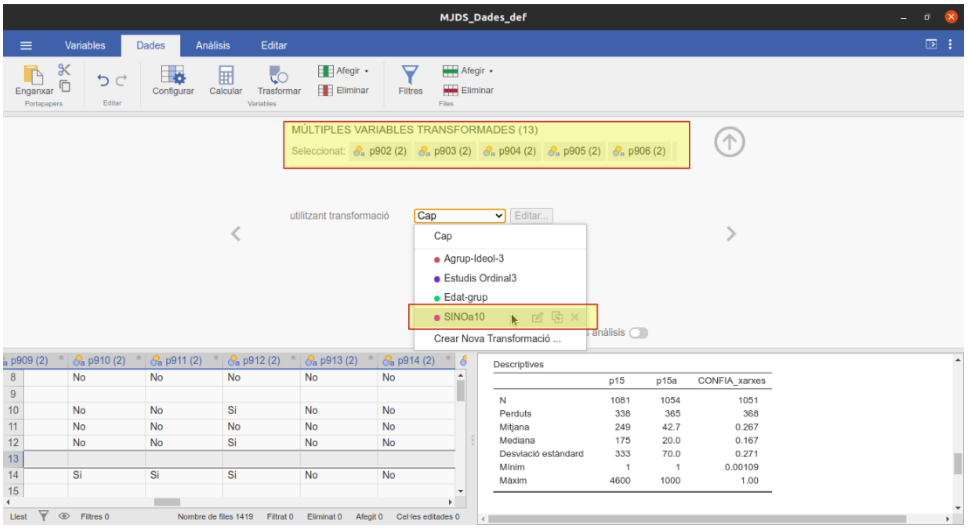
1. Selecciona la variable p901. Clica l'etiqueta "Dades" del menú superior i la icona "Trasformar" del submenú de procediments amb dades. Es crea una nova columna a la dreta de la p901, per defecte anomenada "p901 (2)" i basada en la variable d'origen.
2. Com que es tracta d'una variable utilitària (no en volem fer res, només utilitzar-la per un altre càlculs), no cal que canviïs el nom ni en facis una descripció.
3. A continuació cal crear una nova transformació, que consisteix en recodificar la variable qualitativa nominal (p901, de dues categories, Si i No) en una variable quantitativa entera (p901_c, de dues categories, 0 o 1). Clica el desplegable que actualment indica [Cap] per a "utilitzant transformació", i llavors selecciona l'opció [Crear Nova Trasformació].
4. S'obrirà la finestra de regles de transformació, amb una nova transformació amb una regla molt simple *"els mateixos valors que la variable origen"* (= $source), i un tipus de variable automàticament dependent de la variable original.
5. Primer **modifica el nom** perquè sigui més significatiu, canvia "Transform X" per "SINOa10". Si vols pots introduir a sota una descripció més detallada, com ara "Nominal Sí/No a quantitativa 1/0".
6. A continuació, a la dreta de la descripció cal **introduir el sufix** amb el que s'identificaran les noves variables transformades seguint aquesta mateixa regla: es pot utilitzar "_bin" (sense les cometes) per destacar que la variable transformada serà la versió numèrica Binaria (0,1).

Observa que... aquest sufix sovint és important que l'utilitzem quan vulguem repetir la mateixa transformació vàries vegades, com és el cas.

7. A l'espai corresponent a la condició de recodificació cal **introduir la regla** de transformació que canvia els valors segons el criteri corresponent, especificat en un llenguatge de programació comprensible per *jamovi*. Clica el botó [+ Afegir condició de recodificació] per afegir cada nova condició de recodificació.
8. La primera condició pot ser: "Si el valor de la variable original és 'Sí' llavors el valor a la variable transformada serà 1". En el llenguatge de programació:
   `si $source == 'Sí' utilitzar 1`
9. Cal introduir cada condició en una nova línia, clicant el botó [+ Afegir condició de recodificació] i seguir la mateixa lògica:
   `si $source == 'No' utilitzar 0`
10. En la icona que apareix al costat del nom de la variable podràs comprovar que la nova variable és Nominal/Enter. Als nostres efectes això és positiu, ja que en ser numèrica (Enter) ens permetrà els càlculs que posteriorment haurem de fer.
11. Per a tancar la finestra de regles de transformació, cal clicar a la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra.
12. Tal i com pots veure al panell de transformació, la nova variable està basada en els valors originals de la p901, i utilitza la transformació anomenada "SINOa10" que acabes de crear.
13. Fes freqüències de les dues variables (p901 i p901_bin) i comprova que la transformació ha estat correcta.
14. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Aquesta mateixa transformació cal aplicar-la a les 14 variables de manera que, per cadascuna, generarem una nova columna de dades amb 0 i 1. La forma d'aplicar una regla de transformació <u>a un conjunt</u> de variables és la següent:

1. Cal seleccionar les variables p902 fins a la p914. Per fer-ho, a la matriu de dades, clica sobre el nom de la primera variable a transformar (p902). Queda marcada tota la columna.
2. A la part inferior del panell de matriu de dades hi ha una barra de desplaçament lateral, utilitza-la per a "desplaçar" la matriu cap a la dreta fins que visualitzis la columna corresponent a la "darrera" variable que es vol transformar (p914).
3. Prem la tecla de [Majúscules] al teclat i conjuntament fes clic sobre el nom de la "darrera" variable (p914).



Imatge 78. Selecció de conjunts de variables (per aplicar la mateixa transformació)

4. Ja les tens totes seleccionades. Per a aplicar la mateixa transformació a totes, ves al menú principal superior fes clic a l'apartat "Dades" i clica l'opció "Transformar". Al panell de transformacions múltiples, clica el desplegable de "utilitzant transformació" que indica [Cap] i selecciona SINOa10.
5. S'han generat 13 noves variables quantitatives (amb 1 i 0) basades en les 13 variables nominals originals. Per tancar el panell de transformació clica la FLETXA-AMUNT encerclada de la part superior dreta.
6. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

La selecció múltiple de variables també es pot utilitzar per a definir [Configurar] el tipus de variable si totes són del mateix tipus.
Per fer una selecció múltiple també es pot combinar la tecla [Ctrl] amb clic, d'aquesta forma es pot seleccionar un conjunt de variables no adjacents.

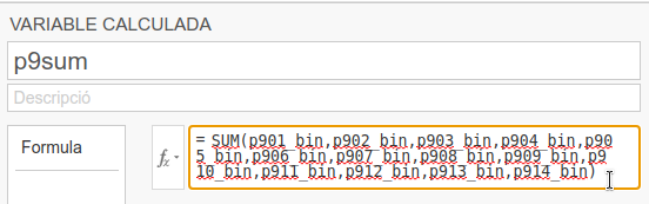
Finalment, tal i com s'ha expressat més amunt, caldrà sumar les noves 14 variables. Totes elles són fàcilment identificables perquè:

(1) el seu nom és el de la seva variable origen amb el sufix _bin, i
(2) l'encapçalament de la columna te un petit punt de color que identifica la regla de transformació utilitzada.

Per a "sumar" les variables (que ara tenen valors 0 i 1) i generar-ne una de nova (nova columna) cal calcular una nova variable (que es pot anomenar, *p9sum*). Cal seguir el següent:

1. Si vols que la ubiqui entre p914_bin i la p1001, clica el nom que encapçala la columna de la variable p914_bin.
2. Clica el menú Data / Afegir / Calcular. Apareix la nova columna de dades.
3. Posa-li un nom curt, p. ex., p9sum.
4. Posa-li una descripció més extensa, p .ex., *"Indicador de diversitat d'usos d'Internet"*.
5. En la finestra de fórmula (al costat del signe =) cal incloure l'operació algebraica que ha de ser realitzada amb els valors de les 14 variables quantitatives binàries. Es pot utilitzar la funció SUM() que està predeterminada en *jamovi* i aprofitar el sistema de construcció assistida de fórmules.

6. Per a seleccionar la funció sumatori: clica el botó central [f_x=]. A la part esquerra del desplegable, on posa "Functions", desplaça cap a baix fins a l'apartat "Statistical" on trobaràs la funció SUM. Clica-la per seleccionar-la (una finestra inferior mostra la seva funció).
7. Fes doble-clic sobre la funció SUM. El sistema *jamovi* l'escriu a la finestra de fórmula i resta a la espera que indiquis quines variables cal sumar.
8. Torna a clicar al botó central [f_x=] i desplaça't pel llistat de "Variables" fins que trobis la primera: p901_bin. Fes-hi doble-clic per escriure-la a la finestra de fórmula.
9. Escriu una coma [,]. Torna a utilitzar el botó [f_x=] i el llistat de variables per a trobar la segona variable a sumar (p902_bin). Amb doble-clic s'escriu a la finestra de fórmula.
10. Continua afegint variables. Cal fer-ho una a una, fins tenir les 14 "variables_bin" que es vol sumar.



Imatge 79. Exemple de la funció SUMATORI amb un conjunt de variables

11. Quan la formula estigui completament "escrita", amb les 14 variables_bin entre comes i dins del parèntesi, prem la tecla [Intro] al teclat per executar el càlcul.
12. Si tot està be, la nova variable p9sum expressa, per a cada persona entrevistada, el nombre d'usos diferents d'Internet.
13. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

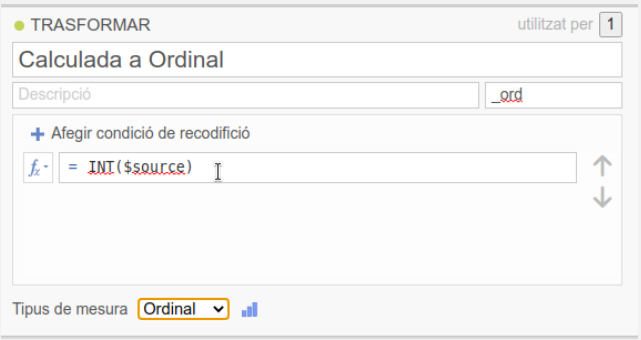
> Tingues en compte que els valors perduts són "contagiosos". Així, quan es fa un càlcul per generar noves variables, una nova fila tindrà valor perdut (en blanc) si qualsevol dels valors de les columnes originals que intervenen en l'operació te el valor perdut.

El resultat d'aquesta operació és una nova variable que mostra que hi ha 39 persones que donen un únic ús a Internet, 93 persones que usen Internet per a 2 funcionalitats, etc. fins arribar a 24 persones (un 1,8%) que utilitzen Internet per a gairebé tot (12 usos diferents). Sembla que la major part de persones enquestades utilitzen Internet per, aproximadament, 5 funcionalitats diferents.

- Pots provar d'explorar descriptivament aquesta nova variable que acabes de construir?
- Pots veure la seva distribució de freqüències? La resposta és que només utilitzant el codi d'R. Però...

PERÒ... com s'aconsegueix fer una taula de freqüències de variables quantitatives sense fer servir codi d'R?
La suma de variables quantitatives (ENTER) per obtenir una nova variable, ens dona una nova variable que també és quantitativa i no hi ha forma de canviar el seu tipus. Per tant, d'entrada, no podrem demanar, en *jamovi*, una taula de freqüències d'aquesta variable calculada.

Es pot TRANSFORMAR qualsevol variable, fins i tot una variable resultat d'un CÀLCUL. Per tant, si es vol "canviar el nivell de mesura" d'una variable calculada, cal seleccionar-la i generar una **nova** variable transformada que sigui una copia d'ella mateixa, en format enter = INT($source), tot **modificant el tipus** de mesura a la part inferior de la pantalla de transformació.



Imatge 80. Exemple de transformació que "canvia" el tipus d'una variable

S'aconsegueix disposar d'una versió "Ordinal" de la variable calculada, i per tant es pot demanar taula de freqüències amb els menús de *jamovi*.

(2) Construcció de tipologies amb la funció IF

Una nova variable tipològica, o classificatòria, se sol construir a partir d'informació continguda en més d'una variable original, de forma que cada categoria de la tipologia correspon a una determinada combinació de categories de les variables originals.
Dins del procés de creació de noves variables mitjançant la transformació de variables (TRANSFORMAR) se sol utilitzar la funció condicional (IF): el valor de la nova variable transformada depèn del valor existent a la variable (o variables) d'origen, en cas que es compleixi un criteri el valor serà un, i si no s'acompleix serà un altre. Amb el seu ús no es fan operacions algebraiques entre variables, com en l'apartat anterior, sinó que implica **operacions lògiques o condicionals**. Aquesta funció IF és fonamental en la construcció de "tipologies". Vegem-ne diferents exemples.

Exemple 1. Anem a veure com combinem dues variables de Text per crear una nova variable. Si observes les preguntes p41 i p42 del qüestionari veuràs que, respectivament, interroguen sobre l'estat civil i la convivència en parella. El nostre objectiu és generar una nova variable que tingui en compte només les persones que afirmen viure en parella (p42=Si) i que, utilitzant l'estat civil, distingeixi entre les unions estables (parelles de fet) i els matrimonis. Si creuem ambdues variables s'obté la següent taula de freqüències creuades:[19]

19 Per generar aquesta taula ves al menú i clica sobre Anàlisis / Freqüècies / Mostres Independents. Un cop ho hagis fet, en la finestra de diàleg situa la p41 a les files i la p42 a les columnes. Per obtenir més informació sobre aquest procediment, vegeu la generació d'aquest tipus de taules a l'apartat "**Error! No s'ha trobat l'origen de la referència.**"d'aquest tutorial.

Taules de Contingència

	p42		
p41	**Si**	**No**	**Total**
Solter/a	219	1029	1248
Casat/da	149	3	152
Separat/da	1	7	8
Divorciat/da	0	4	4
Vidu/a	0	1	1
Total	369	1044	1413

Tal i com s'esmenta més amunt, a partir de la informació disponible a p41 i p42, es vol crear una nova variable que distingeixi als que viuen junts i estan casats (p42=Si i p41=Casat/da, en groc a la taula anterior) dels que viuen junts en cohabitació (en verd a la taula anterior).

El què volem explicar a *jamovi* és alguna cosa com: *Si la p42 és 'Si' i la p41 es 'Casat/da', la nova variable ha de tenir el valor 'Matrimoni', mentre que si la p42 és 'Sí' i la p41 és 'Solter/a' o 'Separat/da' o 'Divorciat/da' o 'Vidu/a', la nova variable ha de tenir el valor de 'Parella de fet'*. De fet aquesta darrera instrucció la resumirem dient-li simplement, que *p41 no sigui 'Casat/da'*.

Com que es tracta de dues instruccions diferents, farem servir dues passes diferents i consecutives.

(1) En la primera crearem una variable que defineixi només els "Matrimonis", mentre que
(2) en la segona crearem una ALTRA variable a la que hi afegirem les "Parelles de fet" a l'anterior.

En el primer pas, per definir els matrimonis (*si la p42 és 'Si' i alhora la p41 és 'Casat/da'*), en llenguatge de programació popi de *jamovi* li haurem de dir el següent:

```
IF (p42=='Si' and p41=='Casat/da' , 'Matrimoni'  ,  NA)
   [----------condició----------]   [-veritat-] [-fals-]
```

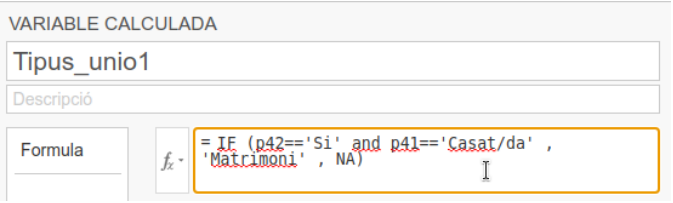
Observa que la funció IF té tres arguments entre cometes, separats per comes. En el primer hi hem d'escriure la condició, en el segon quin resultat volem a la nova variable en cas que es compleixi la condició, i en el tercer el resultat en cas que la condició no es compleixi.

a. Condició: Observa que per expressar una assignació fem servir la doble igualtat (==). Ara bé, es poden fer servir símbols com >, >=, <, >=, != (aquest darrer per indicar no igual a). També és important l'ús de la partícula **and** per indicar que es compleixin dues condicions, així com de la partícula **or**, en cas que vulguem indicar que se'n compleixi una de les dues.
   Tant una partícula com l'altra cal escriure-les en minúscules.
b. Veritat: Resultat en cas que es compleixi. Com que la nova variable serà de Text, li hem escrit 'Matrimoni' entre cometes. Per altra banda, si fos un nombre l'hauríem d'escriure sense cometes.
c. Fals: Resultat en cas que no es compleixi. Aquí volem que deixi en blanc els casos que no compleixen la condició. Part d'aquests casos els omplirem en una segona instrucció. Per indicar-ho fem servir la partícula NA, sense cometes, perquè és una funció pròpia de *jamovi*.

Per crear la **tipologia**, pas a pas, segueix les següents indicacions:

1. Per crear la nova variable a continuació de la p42, situa't a sobre i clica al menú Dades / Calcular. Apareix la nova columna de dades.

2. Posa-li un nom, com per exemple, Tipus_unio1, i si vols una descripció més extensa.
3. A la finestra de fórmula (al costat del signe =) cal incloure la condició que hem escrit més amunt, [20] de manera que la finestra quedarà:



Imatge 81. Càlcul utilitzant el condicional IF i els valors d'altres variables (I)

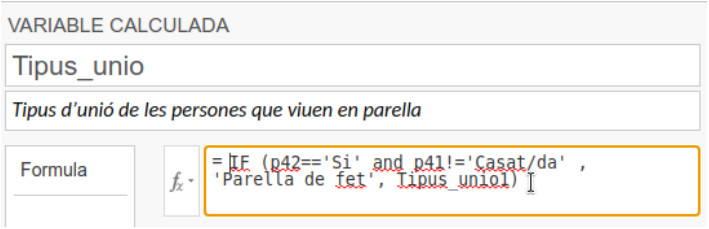
4. En clicar [Intro] ja veuràs la nova variable construïda. En alguns casos puntuals ha escrit *Matrimoni*, mentre que la majoria de casos els ha deixat en blanc. Comprova que els *Matrimoni* es corresponen amb la definició que hem fet, combinant p42 i p43.

Per afegir-hi les parelles de fet, ho farem generant una **nova variable**. El què li volem dir és que si la p42 és 'Si' i la p41 no és 'Casat/da', aleshores escrigui 'Parella de fet', mentre que en cas contrari, deixi la variable Tipus_unio1 tal i com està. En llenguatge de programació propi de jamovi, li haurem de dir:

```
IF (p42=='Si' and p41!='Casat/da' , 'Parella de fet', Tipus_unio1)
```

La diferència amb l'anterior, és que ara, en cas que no es compleixi la condició, li diem que escrigui **el què hi ha a Tipus_unio1**. Això fa que no perdem els *Matrimoni*, que ja havíem definit a la columna anterior.

5. Per crear, doncs, la nova variable, situa't a sobre de Tipus_unio1 i clica al menú Dades / Calcular. Apareix la nova columna de dades.
6. Posa-li un nom, com per exemple, Tipus_unio, i si vols una descripció més extensa com, per exemple, *Tipus d'unió de les persones que viuen en parella* (aquesta variable ja serà la versió definitiva).
7. A la finestra de fórmula (al costat del signe =) cal incloure la condició que hem escrit més amunt, de manera que la finestra quedarà:



Imatge 82. Càlcul utilitzant el condicional IF i els valors d'altres variables (II)

8. En clicar [Intro] ja veuràs la variable construïda. En aquest cas manté la categoria *Matrimoni*, quan ja estava definida, afegeix alguns *Parella de fet*, mentre que encara en deixa molts en blanc. Comprova que els *Parella de fet* es corresponen amb la definició que hem fet, i que els blancs són persones que no viuen en parella (p42=No).

[20] Recorda que has de TECLEJAR la ordre a jamovi. Si proves de retallar a aquest PDF i enganxar a jamovi es generarà un error perquè R no accepta els caracters " i " .

9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

El resultat és la taula següent. Pots comprovar-ho fent Anàlisis / Exploració / Descriptives demanant taules de freqüències:

Descriptives

	Tipus_unio
N	369
Perduts	1050

Freqüències

>

Freqüències de Tipus_unio

Nivells	Freqüències	% del Total	% Acumulat
Parella de fet	220	59.6 %	59.6 %
Matrimoni	149	40.4 %	100.0 %

Imatge 83. Descriptiva de la variable recodificada

Observació important! Tingues en compte que és possible concatenar dues o més funcions IF, de manera que en algun dels dos possibles resultats (en cas que es compleixi la condició o en cas que no es compleixi la condició) hi podem escriure una nova condició (un nou IF). En aquest cas concret això es materialitzaria en no haver de crear una variable intermèdia *(Tipus_unio1)* com hem fet, sinó que podem crear directament la variable final. Per fer-ho, la condició que hauríem d'escriure substituiria, en part, la sortida NA de la primera condició, per la segona condició, deixant els NA pel final. Així:

```
IF (p42=='Si' and p41=='Casat/da', 'Matrimoni',
IF (p42=='Si' and p41 != 'Casat/da', 'Parella de fet', NA))
```

Exemple 2. En alguns casos ens pot interessar pensar les variables com si fossin numèriques, encara que no ho siguin. Això també ho permet la funció IF. El què fa és considerar que a cada categoria o nivell li correspon un número correlatiu, des de l'1 en endavant. Així, per exemple, si fem freqüències de la variable p53, que es correspon amb els *Estudis més alts assolits en l'actualitat*, obtindrem la següent taula, d'on podem deduir que si la tractem com a numèrica el número 1 correspondrà a *Menys 5a esc*, el número 2 a *Primaria*, el 3 a *ESO*... i així correlativament, de manera que l'11 correspondrà a *Postgrau*:

Freqüències de p53

Nivells	Freqüències	% del Total	% Acumulat
Menys 5a esc	1	0.1 %	0.1 %
Primaria	221	15.7 %	15.8 %
ESO	430	30.6 %	46.4 %
FP mig	131	9.3 %	55.7 %
Batxill	279	19.8 %	75.5 %
FP sup	129	9.2 %	84.7 %
Eng tecnic	27	1.9 %	86.6 %
Diplomat	100	7.1 %	93.7 %
Eng sup	7	0.5 %	94.2 %
Llicenciat	70	5.0 %	99.2 %
Postgrau	11	0.8 %	100.0 %

Imatge 84. Descriptiva de la variable ordinal nivell educatiu

(cc) BY-NC

En aquest exemple ho farem servir per construir una TIPOLOGIA d'homogàmia *(aparellament entre persones de característiques similars)* i una de les seves dimensions és l'educativa: hi ha parelles homogàmiques en les que tots dos membres de la parella tenen el MATEIX nivell educatiu, i parelles on l'home té una major nivell d'estudis i d'altres on és la dona (aquí estem pensant només en parelles heterosexuals, que és el que tenim a l'enquesta).

En aquest cas ho mirarem a través del nivell d'estudis dels progenitors: *Estudis assolits pel pare* (p54) i *Estudis assolits per la mare* (p55). Aquestes variables tenen, com a valors textuals, *Menys prim*, *Primaria*, *Secundaria* i *Universit*, de manera que numèricament els correspondrà els nombres 1-4.

Si creuem ambdues variables (recorda... Menú / Anàlisis / Freqüències / Mostres Independents) obtindrem la taula de freqüències "de doble entrada" que hem copiat aquí sota. En aquesta taula, si tenim en compte que p54 són els estudis del pare i p55 els de la mare, pots comprovar que hem marcat en groc els casos on el nivell d'estudis d'ambdós són iguals (P=M), en verd on el pare té més estudis (P>M) i en blau aquells casos on la mare té un nivell d'estudis superior (M>P):

Taules de Contingència

	P55 (mare)				
P54 (pare)	**Menys prim**	**Primaria**	**Secundaria**	**Universit**	**Total**
Menys prim	79	22	4	1	106
Primaria	20	492	87	26	625
Secundaria	4	63	232	50	349
Universit	3	28	76	131	238
Total	106	605	399	208	1318

Si podem pensar que la variable és numèrica, la lògica de construcció de la nova variables és molt senzilla, ja que: aleshores

- Si p54=p55, aleshores P=M (pare i mare tenen el mateix nivell d'estudis). Si comptes els grocs veuràs que t'han de sortir 934 casos.
- Si p54>p55, aleshores P>M (el pare té un nivell d'estudis superior a la mare). Si comptes els verds, veuràs que en sortiran 194.
- Si p55>p54, aleshores M>P (la mare té un nivell d'estudis superior al pare). Si comptes els blaus veuràs que n'hi haurà 190.

Per a crear la **nova variable tipològica** a partir de les dues originals (p54 i p55), ho farem en tres passes, creant dues variables intermèdies:

1. Clica sobre p55 i fes Dades / Calcular. Canvia el nom a, per exemple, *Homog1*.
2. A la finestra de fórmula escriu la funció:
   ```
   IF (p54==p55, 'P=M', NA)
   ```
   Pica la tecla [Intro] i comprova que ha generat una bona part dels casos, i que la definició és correcta.
3. Per a crear una nova categoria, sense perdre l'anterior informació, clica sobre *Homog1* i fes Dades / Calcular. Canvia el nom a, per exemple, *Homog2*.
4. A la finestra de fórmula escriu la funció:
   ```
   IF (p54<p55, 'M>P', Homog1)
   ```
   Pica la tecla [Intro] i comprova que ha generat una bona part dels casos, i que la definició és correcta.
5. Per a crear una nova categoria, sense perdre l'anterior informació, clica sobre *Homog2* i fes Dades / Calcular. La versió que generarem ara ja és la definitiva, de manera que canvia

el nom a, per exemple, *Homogàmia*, i escriu una definició més completa de la variable, per exemple, *Homogàmia educativa dels progenitors*.

6. A la finestra de fórmula escriu la funció:
   IF (p54>p55, 'P>M', Homog2)
   Pica la tecla [Intro] i comprova que ha generat la resta de casos, i que la definició és correcta.
7. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Pots comprovar la distribució d'aquesta nova variable i comprovar que el resultat final s'adequa a allò que ja havíem previst:

Freqüències de Homogàmia

Nivells	Freqüències	% del Total	% Acumulat
P>M	194	14.7 %	14.7 %
M>P	190	14.4 %	29.1 %
P=M	934	70.9 %	100.0 %

Imatge 85. Descriptiva de nova variable "combinada" de p54 i p55

Un cop més... És possible generar la nova variable en una única instrucció que concateni, en aquest cas, tres instruccions IF diferents. Fes-ho creant la nova variable a partir de la funció:

```
IF(p54==p55, 'P=M',IF(p54<p55, 'M>P',IF(p54>p55, 'P>M', NA)))
```

Exemple 3. En aquest exemple volem construir una nova variable a partir de tres de diferents. Observa que la p12 pregunta per les tres xarxes socials que la persona enquestada utilitza més sovint (p1201, p1202 i p1203). La informació es troba "distribuida" en les tres variables. Volem crear una variable que digui, per exemple, *"Si utilitza Facebook o no"*. Tal i com esta construïda la matriu de dades, s'utilitzarà *Facebook* si aquesta xarxa social és esmentada o en la p1201, o en la p1202, o en la p1203.

Aquí el problema no és tant en la categoria dels que fan servir *Facebook*, sinó que rau en la d'aquells que no l'utilitzen, ja que hi ha valors perduts que són més difícils de tractar a partir de la funció IF. Vegem-ho pas a pas:

1. Clica sobre p1203 i fes Dades / Calcular. Canvia el nom a, per exemple, *Facebook1*.
2. A la finestra de fórmula escriu la funció:

```
IF(p1201== 'Facebook' or p1202=='Facebook' or p1203=='Facebook', 'Si', 'No')
```

   Pica la tecla [Intro] i comprova que ha generat una part dels casos. De fet, tots els que fan servir Facebook els ha generat bé, però no així els que no, que n'ha deixat una bona part en blanc (per exemple, si la p1201 no era Facebook i p1202 i p1203 estaven en blanc, els ha deixat en blanc).
   Anem a corregir-ho, canviant tots els blancs per "No" i deixant igual la resta.
3. Clica sobre *Facebook1* i fes Dades / Calcular. Canvia el nom a, per exemple, *Facebook2*.
4. A la finestra de fórmula escriu la funció:
   IF (Facebook1==NA, 'No', Facebook1)

Pica la tecla [Intro] i comprova que ha passat els que estaven en blanc a "No" (si et pares a pensar-hi una mica, això ho haguessis pogut fer a partir d'una recodificació).
Ara, en principi, ja ho tindries. Ara bé, dintre dels que no fan servir Facebook tens dos grups de persones: els que fan servir alguna xarxa social, però no Facebook i els que no fan servir cap xarxa social. Si volguéssim distingir-los hauríem de fer servir la p11, que ens diu, precisament si fa servir o no alguna xarxa social.
Ho podríem matisar continuant amb les següents instruccions.

5. Clica sobre *Facebook2* i fes Dades / Calcular. Canvia el nom a, per exemple, *Facebook3*.
6. A la finestra de fórmula escriu la funció:
   `IF (p11==NA or p11=='Mai', 'No xarxes', Facebook2)`
   Pica la tecla [Intro] i comprova que alguns del que eren *No* els ha passat a *No xarxes*.
   Pots fer freqüències de les tres variables construïdes per poder veure més clarament quin ha estat el procés:

Freqüències de Facebook1

Nivells	Freqüències	% del Total	% Acumulat
Si	928	77.9 %	77.9 %
No	263	22.1 %	100.0 %

Freqüències de Facebook2

Nivells	Freqüències	% del Total	% Acumulat
No	491	34.6 %	34.6 %
Si	928	65.4 %	100.0 %

Freqüències de Facebook3

Nivells	Freqüències	% del Total	% Acumulat
No xarxes	248	17.5 %	17.5 %
No	243	17.1 %	34.6 %
Si	928	65.4 %	100.0 %

Imatge 86. Descriptiva de tres variables amb modificacions acumulatives consecutives

7. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Transformacions amb codi R

Utilitzar codi R pot facilitar la realització de les transformacions i càlculs que s'han vist als apartats anteriors, i a més proporciona accés a tota la potència de R per a la gestió i manipulació de variables.

Com afegir una nova columna, generada amb R,a la matriu de dades de jamovi...
Cal utilitzar el mòdul Rj Editor**+** *(concretament la versió **+**)*

1. Seleccionar la/es variables originals amb les que es vol crear la nova variable.
2. Amb el botó [->] transferir-les a l'espai de treball de la dreta.
3. A la finestra inferior (consola) escriure el codi R corresponent.
4. Fer referencia a variables de la matriu original amb "data$nom-variable".
5. Executar el codi R.

Les noves variables apareixeran a l'extrem dret de la matriu de dades.
En alguns casos, caldrà alguna manipulació posterior a la creació de la noves variables, per a adaptar-les al tipus desitjat (especialmente en el cas de les ordinals).

(cc) BY-NC

Cal tenir en compte... que les noves variables creades per execució de codi R depenen, precisament, d'aquesta execució del mòdul Rj. Com a conseqüència, en cas que s'esborri el codi R de l'arxiu OMV aquestes variables desapareixeran definitivament.
En cas que es desi l'arxiu OMV, amb el codi R, i posteriorment s'intenti obri-lo des d'una versió de *jamovi* que no tingui instal·lat el mòdul Rj les variables no es visualitzaran, fins que s'instali el mòdul requerit.

Seguint l'exemple anterior, per a recodificar una variable quantitativa a una qualitativa, a partir de la variable p33 (ideologia de qui respon, en una escala de esquerra-dreta, de 1 a 10) es genera IDEOLOGIA amb valors perduts (fora del rang 1-10), esquerra(1-4), centre (5) i dreta (6-10).

- A la consola del mòdul Rj Editor+, selecciona la variable p33.
- Per a generar la nova variable IDEOLOGIA amb tots els valor perduts, escriu:
```
data$IDEOLOGIA <- NA
```
- Per assignar a la nova variable uns codis numèrics ordenats, segons les condicions que ha de complir la variable original p33 en aquest cas, escriu:
```
data$IDEOLOGIA[data$p33 < 5] = 1
data$IDEOLOGIA[data$p33 == 5] = 2
data$IDEOLOGIA[data$p33 > 5] = 3
```
- Per assignar a la nova variable IDEOLOGIA les etiquetes textuals corresponents a cada codi numèric, escriu:
```
data$IDEOLOGIA <- ordered(data$IDEOLOGIA,
    levels = c(1,2,3),
    labels = c("Izquierda", "Centro", "Derecha"))
```
- Executa el codi R.
- Per comprovar, revisa la matriu de dades i pots realitzar estadística descriptiva de la nova variable (freqüències, o estadístics, o gràfics).

A la matriu de dades, s'ha creat una nova variable, amb els valors recodificats a partir de les condicions de la variable original (p33) i que es pot analitzar com la resta de variables amb els procediments dels menús de *jamovi* (i també amb R).

Atenció!... No oblidis desar la matriu de dades quan generis noves variables, o es perdran quan surtis de *jamovi*. L'aplicació et demanarà si vols Desar o no quan provis de sortir-ne...excepte en cas de fallada de l'ordinador (bateria, o "sistema penjat").

En el següent exemple, per a recodificar una variable qualitativa a una qualitativa, a partir de la variable original ESTUDIS (ordinal de 6 categories) se'n generarà una de nova que s'anomenarà "ESTUDIOS_UNIV"; amb només dues categories: *Pre-Universitarios*, i *Universitaris*.

- A la consola del mòdul Rj Editor+, selecciona la variable ESTUDIOS.
- Per a generar la nova variable "ESTUDIOS_UNIV" amb tots els valor perduts, escriu:
```
data$ESTUDIOS_UNIV <- NA
```
- Per assignar a la nova variable uns codis numèrics ordenats, segons les condicions que ha de complir la variable original ESTUDIOS en aquest cas, escriu:
```
data$ESTUDIOS_UNIV[data$ESTUDIOS %in% c("Prim","F.P.","Secund")] = 0
data$ESTUDIOS_UNIV[data$ESTUDIOS == "Univers"] = 1
```

- Asegura't que utilitzes, exactament, les etiquetes textuals pròpies de la variable origen, així com les comilles pròpies d'R, i vigila amb els espais en blanc. RECORDA, NO copiis el codi des d'aquest document cap a la consola de Rj, millor que teclejis tot el codi a Rj.
- Per assignar a la nova variable ESTUDIOS_UNIV les etiquetes textuals corresponents a cada codi numèric, escriu:
```
data$ESTUDIOS_UNIV <- ordered(data$ESTUDIOS_UNIV,
    levels = c(0,1),
    labels = c("Pre Universitaris", "Universitaris"))
```
- Executa el codi R.
- Per comprovar, revisa la matriu de dades i pots realitzar estadística descriptiva de la nova variable (freqüències, o estadístics, o gràfics).

Com a exemple de com recodificar amb R una variable qualitativa a una quantitativa, a continuació es transformarà la variable original "g_edat" (que és ordinal/text) tot assignant a cada categoria original, o "interval d'edat" el corresponent valor per la seva *marca de classe*: a "*15 19 anys*" s'assignarà 17; a "*20 24 anys*" s'assignarà 22; i a "*25 29 anys*" s'assignarà 27 (els punts centrals de cada interval d'edat.

- A la consola del mòdul Rj Editor**+**, selecciona la variable "g_edat".
- Per a generar la nova variable "EDAT (estimada)" amb tots els valor perduts, escriu:
```
data$EDAD_estimada <- NA
```
- Per assignar a la nova variable uns codis numèrics ordenats, segons les condicions que ha de complir la variable original ESTUDIOS en aquest cas, escriu:
```
data$EDAD_estimada[data$g_edat == "15 19 anys"] = 17
data$EDAD_estimada[data$g_edat == "20 24 anys"] = 22
data$EDAD_estimada[data$g_edat == "25 29 anys"] = 27
```
- Com que la nova variable "EDAD (estimada) és cuantitativa no requereix cap etiqueta
- Executa el codi R.
- Per comprovar, revisa la matriu de dades i pots realitzar estadística descriptiva de la nova variable (freqüències, o estadístics, o gràfics).

Com a exemple de càlcul amb codi R a continuació s'estandarditzarà (normalitzar) la variable quantitativa "*Nombre d'amics i agregats a les xarxes socials*" (p15). Per fer-ho, cal CALCULAR per a cada valor de la variable original: en restem la mitjana de la variable i dividim el resultat entre la desviació típica. R incorpora directament les funcions per a calcular la mitjana mean() i la desviació típica sd() d'una variable, que s'utilitzaran per fer el càlcul corresponent i generar la nova variable "AMIGOS_norm".

- A la consola del mòdul Rj Editor**+**, selecciona la variable p15 i clica a la fletxa central per afegir-la a l'espai de Variables de treball.
- Per a generar la nova variable, anomenada AMIGOS_norm, amb tots els valor perduts, escriu a la consola inferior:
```
data$AMIGOS_norm <- NA
```
- Primer cal calcular dos estadístics de la variable requerits pel càlcul de la puntuació Z, escriu:
```
media <- mean(data$p15, na.rm=TRUE)
desviacion <- sd(data$p15, na.rm=TRUE)
```
- Ara, per assignar a la nova variable els valors normalitzats per a cada valor original, en aquest cas, escriu:
```
data$AMIGOS_norm <- (data$p15 - media) / desviacion
```
- Executa el codi R.
- Per comprovar, revisa la matriu de dades i pots realitzar estadística descriptiva de la nova variable (freqüències, o estadístics, o gràfics).
```
summary(data$AMIGOS_norm)
```

Si volem establir la relació entre la quantitat de contactes a xarxes socials i els "veritables amics" per a cada persona que ha respost l'enquesta es pot fer una simple divisió entre les variables p15 i p15a. El quocient o raó p15a/p15 és un indicador del nivell de "contactes fiables" de cada persona enquestada.

- A la consola del mòdul Rj Editor+,
- Selecciona les variables p15 i p15a (et pots ajudar amb la icona "lupa" per cercar-les), clica a la fletxa central per afegir-les a l'espai de Variables de treball.
- Per a generar la nova variable, anomenada AMIGOS_norm, amb tots els valor perduts, escriu a la consola inferior:
  `data$contactes_fiables <- NA`
- Ara, per assignar a la nova variable els valors del quocient inclou la formula de càlcul, escriu:
  `data$contactes_fiables <- data$p15a / data$p15`
- Executa el codi R.
- Per comprovar, revisa la matriu de dades, o pots realitzar estadística descriptiva de la nova variable (freqüències, o estadístics, o gràfics).
  `summary(data)`

Es pot utilitzar codi R amb finalitats d'anàlisi de text simple, com en el següent exemple. Al qüestionari del CIS corresponent a les dades de "Dades3.omv" s'ha demanat *"p12. ¿Qué red social utilizas principalment en Internet?"* i s'ha obtingut informació en el següent format *(cada línia inclou les respostes d'una persona a la pregunta multiresposta)*:

"Facebook, Tuenti"
"Badoo, Tuenti, MySpace
"Facebook, Tuenti
"Tuenti"
"Tuenti"
"Facebook, Tuenti, Twitter"
etc...

Per crear noves variables amb informació sobre si una resposta de text conté o no una paraula en concret:

- A la consola del mòdul Rj Editor+, selecciona la variable que conté els text (p12m) i clica a la fletxa central per afegir-la a l'espai de Variables de treball.
- Per a visualitzar les respostes, escriu i executa:
- **`Data$p12m`**
- Genera una nova variable, anomenada FACEBOOK, amb valor lògic de veritat o falsetat en funció de si p12m conté o no el text buscat:
  `data$FACEBOOK <- grepl("Facebook", data$p12m)`
- Executa el codi R.
- Per comprovar, revisa la matriu de dades, o pots realitzar estadística descriptiva de la nova variable (freqüències, o estadístics, o gràfics).
  `summary(data)`
- Repeteix i adpta el codi R per a cada nova columna o variable que vulguis "analitzar", per exemple:
  `data$TUENTI <- grepl("Tuenti", data$p12m)`

A la matriu de dades disposem de 14 variables (de la p901 a la p914) que ens informen per separat de cadascun dels usos d'Internet, codificades com a "Si" o "No" en funció de si la persona entrevistada en fa o no aquell ús concret. Podem utilitzar codi R per a generar una nova variable que indiqui la QUANTITAT d'aparicions de "Si" pels 14 registres de cada individu (fila).

- A la consola del mòdul Rj Editor**+**, selecciona totes les variables que contenen les respostes a la pregunta p9, de la p901 fins a la p914, i clica a la fletxa central per afegir-les a l'espai de Variables de treball.
- Genera la nova variable, anomenada NIVELL_WWW, amb el recompte de quantes vegades ha contestat "si" cada persona entrevistada:
  `data$NIVELL_WWW <- apply(data, 1, function(x) length(which(x=="Si")))`
- Executa el codi R.

Es possible utilitzar una adaptació d'aquesta ordre R per a fer recompte de qualsevol text o valor inclòs en un conjunt de variables.

També és possible utilitzar codi R per a construir tipologies, a partir de combinacions dels valors d'altres variables. Per exemple, per a generar una nova variable que distingeixi entre les unions estables (parelles de fet) i els matrimonis, cal combinar dues variables: p41 i P42.

- A la consola del mòdul Rj Editor**+**, selecciona les variables que contenen informació sobre l'estat civil (p41) i sobre la convivència amb altres persones (p42), i clica a la fletxa central per afegir-les a l'espai de Variables de treball.
- Visualitza la combinació d'aquestes dues variables, escriu i executa:
  `table(data$p41,data$p42)`
- Genera la nova variable, anomenada UNION_MATRIMONIO, amb tots els valors perduts:
  `data$UNION_MATRIMONIO <- NA`
- Assigna els valors de la nova variable ("Matrimonio" o "Pareja de hecho") en funció de la combinació de respostes a les variables originals:
  ```
  data$UNION_MATRIMONIO[data$p41 == "Casado/a" & data$p42 == "Si"] =
  "Matrimonio"
  data$UNION_MATRIMONIO[data$p41 != "Casado/a" & data$p42 == "Si"] = "Pareja
  de Hecho"
  ```
- Executa el codi R.
- Comprova que la distribució de la nova variable indica 149 matrimonis i 220 parelles de fet.

Un altre exemple de construcció de tipologia, basada en informació de la combinació de dues variables és la construcció d'un indicador de "igualtat d'estudis familiars", en aquest cas, informació combinada de *Estudis assolits pel pare* (p54) i *Estudis assolits per la mare* (p55). Si l'objectiu és contruir una nova variable indicadora d'aquestes tres situacions: "P=M" (nivell d'estudis d'ambdós són iguals), "P>M" el pare té més estudis i "M>P" la mare té un nivell d'estudis superior, llavors el codi R adient és el següent:

- A la consola del mòdul Rj Editor**+**, selecciona les variables que contenen informació sobre el nivell d'estudis del pare (p54) i el de la mare (p55), i clica a la fletxa central per afegir-les a l'espai de Variables de treball.
- Visualitza la combinació d'aquestes dues variables, escriu i executa:
  `table(data$p54,data$p55)`

- Genera la nova variable, anomenada IGUALDAD_ESTUDIOS_FAMILIA, amb tots els valors perduts:
  `data$IGUALDAD_ESTUDIOS_FAMILIA <- NA`
- Assigna els valors de la nova variable en funció de la combinació de respostes a les variables originals:
  ```
  data$IGUALDAD_ESTUDIOS_FAMILIA[data$p54 == data$p55] = "P=M"
  data$IGUALDAD_ESTUDIOS_FAMILIA[data$p54 > data$p55] = "P>M"
  data$IGUALDAD_ESTUDIOS_FAMILIA[data$p54 < data$p55] = "M>P"
  ```
- Executa el codi R.
- Comprova que la distribució de la nova variable indica 934 families "igualitaries.

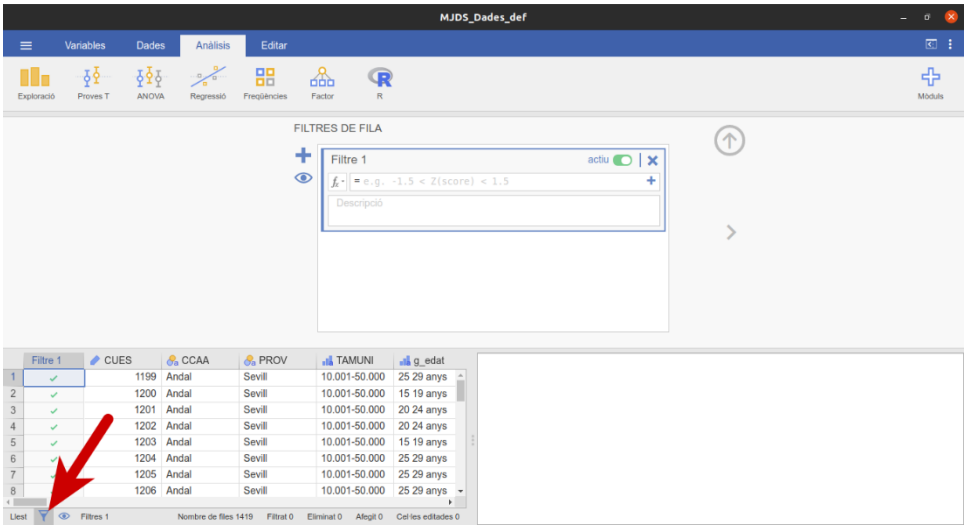
Fixa't! L'ús del codi R permet aconseguir els mateixos resultats que la utilització de les eines de transformacions i càlculs de *jamovi*, però sovint amb més simplicitat i més ràpidament. A més el codi R utilitza recursos generals adaptables a diferents situacions.

Filtrar casos

En determinades circumstàncies pot interessar estudiar un subconjunt de casos de la matriu de dades, per exemple:

1. només les dones (p40), o
2. només els homes majors de 20 anys, o
3. el conjunt "central" de la mostra segons l'edat (és a dir, només els casos que no es "desvien" de l'edat mitjana més d'1,5 desviacions típiques).

Per fer-ho cal utilitzar el sistema de FILTRES de files que *jamovi* implementa a la matriu de dades. Pots trobar els controls de filtrat a la part inferior esquerra de la matriu.

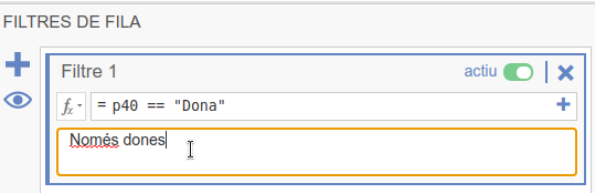


Imatge 87. Ubicació del control de filtratge de casos (selecció)

Per exemple, per a filtrar la matriu de dades de forma que totes les operacions (càlculs, transformacions i anàlisis estadístics) tinguin en compte només les **dones**:

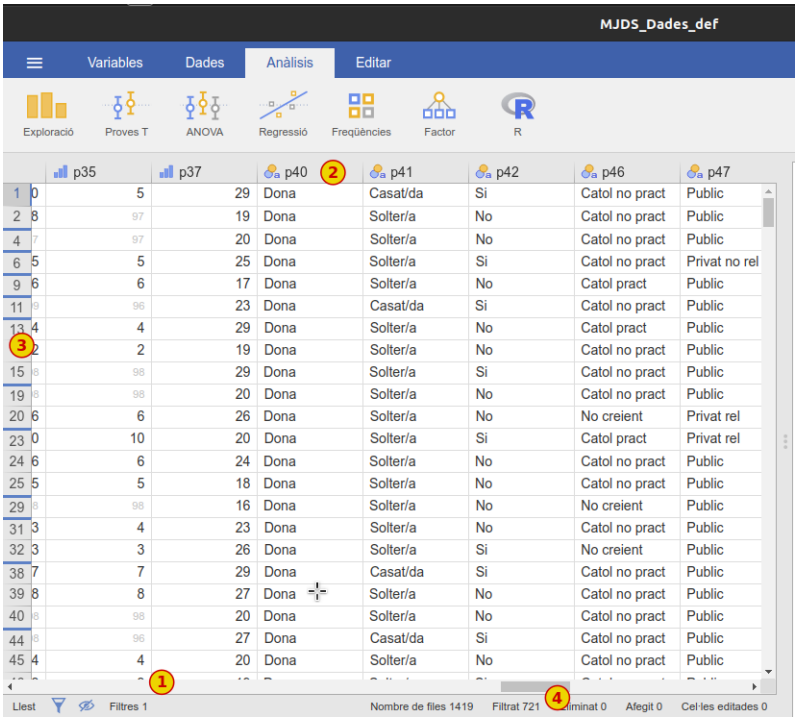
1. Clica la icona de filtratge (el petit embut situat a la part inferior esquerra). S'obre el panell de filtres.

2. Escriu la condició que han de satisfer els casos (files) que quedaran seleccionats:
   `p40 == "Dona"`
   Es útil afegir una descripció especificant la utilitat del filtre.
3. Qualsevol filtre es por activar o desactivar independentment. Tots queden desats en l'arxiu de *jamovi* *.omv



Imatge 88. Pantalla de control de filtres de casos

4. Es poden afegir nous filtres, que combinen amb els anteriors "actius" fent clic a la creu de la part superior esquerra de la finestra.
5. Qualsevol filtre pot estar en estat "actiu" o "inactiu", segons la posició del interruptor corresponent superior dret a la finestra de definició del filtre. Vigila quan activis filtres, si les condicions combinades son incompatibles *jamovi* pot bloquejar-se, i es perd tota la feina des que es va desar per darrera vegada.
6. A la pantalla principal de Dades, a baix al costat de la icona [embut] hi ha una icona [ull] que serveix per a visualitzar o no les files que s'han filtrat. Comprova-ho fent-hi clic repetidament.
7. Quan un o més filtres estan activats, s'indica a la part inferior esquerra (1) i es pot veure en la variable filtrada (2) i en com falten files (3), en concret a la part inferior (4) s'indica quants casos o files (rows) han estat eliminades de la matriu de dades.



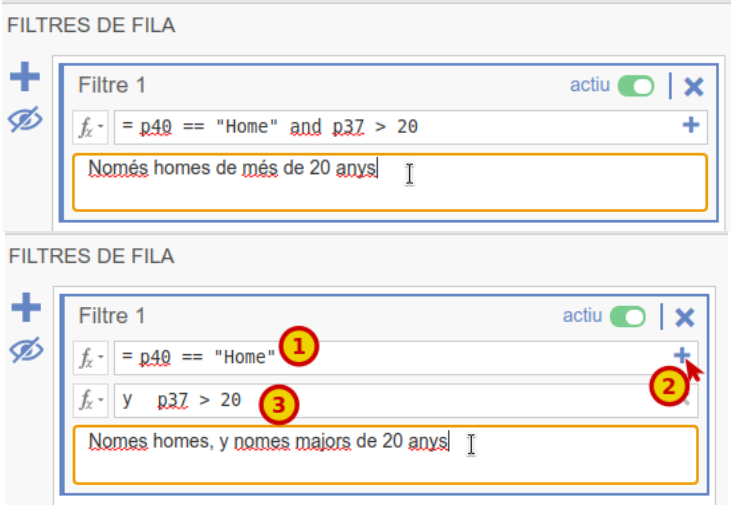
Imatge 89. Exemple de filtre activat

Molt important! Desactiva tots els altres filtres abans de crear-ne un de nou. Hi pot haver conflictes entre filtres de manera que *jamovi* podria entrar en un "bucle" que no es pot resoldre i bloqueja tot el sistema.[21] Per desactivar un filtre (no per esborrar-lo) has de clicar sobre l'interruptor "actiu" situat a dalt a la dreta, deixant aquell filtre inactiu. Per a esborrar un filtre, fes clic a la creu X a la dreta de l'interruptor d'activació.

Es pot aplicar una condició de filtratge complexa, o combinada, utilitzant els connectors que hem vist en l'apartat on hem explicat el condicional IF. Els més habituals són els connectors **and** i **or**.

Per exemple, per a filtrar la matriu de dades de forma que totes les operacions (càlculs, transformacions i anàlisis estadístiques) tinguin en compte només els **homes majors de 20 anys**:

1. Clica la icona de filtratge. S'obre el panell de filtres.
2. Escriu la condició múltiple. Hi ha dues formes de fer-ho:
3. Si volem seleccionar els homes majors de 20 anys, volem que compleixi dues condicions, de manera que utilitzarem el connector **and**, per construir una única condició composta, de la següent manera:
   `p40 == "Home" and p37 > 20`
4. Una alternativa es escriure (1) la primera condició i fer clic (2) al signe + a la dreta d'aquesta condició. Apareixerà una segona finestra on (3) escriure la segona condició composta. (D'aquesta forma només es poden "dissenyar" condicions combinades amb el connector **and**.)
5. Sempre és útil afegir una descripció especificant la utilitat del filtre.



Imatge 90. Exemples de condicions combinades de filtratge de casos

Qualsevol modificació sobre la fórmula d'un filtre triga un temps (que pot ser llarg, **fins i tot de minuts)** en fer-se efectiva, perquè *jamovi* torna a calcular totes les transformacions i anàlisis de l'arxiu per ajustar-los a la submostra seleccionada.
Tots els filtres, actius o no, queden desats en l'arxiu *.omv juntament amb totes les dades, transformacions i resultats de manera que es poden recuperar més endavant i en diferents sistemes.

21 Similarment al "robloqueo" que pateixen els robots als relats d'Isaac Asimov (1950).

(cc) BY-NC

COMPTE! Sempre que es crea o s'activa un filtre (també quan es desactiva) es recalculen totes les anàlisis que hi ha fetes. Això comporta que qualsevol taula o gràfic fet anteriorment, es modifica tenint en compte els filtratges actius.

EXERCICI PROPOSAT: FILTRAR LA POBLACIÓ

- Crea un filtre amb el següent criteri:
  `-1.5 < (p37 - VMEAN(p37)) / VSTDEV(p37) < 1.5`
- Activa'l i comprova com afecta als valors extrems (mínim i màxim) de les distribucions.
- A quin percentatge de la població correspon la mostra que has seleccionat amb aquest filtre?
- Entens que el que has fet és seleccionar la part "central" de la distribució d'edat?

Detecció i depuració d'errors

Quan es comença a treballar amb una matriu de dades, tant és si l'hem construïda nosaltres com si ens ve donada, cal en primer lloc estimar-ne la seva qualitat de manera que es detecti, localitzi i corregeixin els possibles errors o incongruències que hi hagi a les dades. Habitualment, en primer lloc, depurarem les variables per separat, detectant valors impossibles o altament improbables, i, a continuació, comprovarem que hi hagi una absoluta congruència entre cadascuna de les variables i la resta.

Errors en variables individuals

Per avaluar possibles errors univariables cal estudiar les freqüències de totes les variables, tant de les quantitatives (ORDINAL - ENTER) com les qualitatives (NOMINAL o ORDINAL - TEXT). La simple observació d'aquestes freqüències (o d'un *summary* d'R) permet detectar valors impossibles o altament improbables.

1. Clica Anàlisis / R / Rj Editor... Escriu l'ordre corresponent d'R:
   `summary(data)`
2. Clica el triangle verd a la part superior dreta del panell consola R, per executar-la.

Quan la variable és **qualitativa** l'error es detecta fàcilment. Observa, per exemple, la p1001:

1. Clica Anàlisis / R / Rj Editor... Escriu l'ordre corresponent d'R:
   `table(data$p1001)`
2. Clica el triangle verd a la part superior dreta del panell consola R, per executar-la.

Per error, en dues ocasions s'han introduït els codis "SI" i "si", en comptes del correcte "Si".

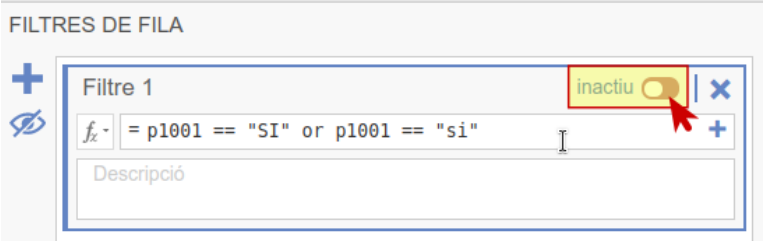
```
  Si    No   SI   si
 592   707    1    1
```

Si volem localitzar els qüestionaris on succeeix l'error, el més efectiu és crear un FILTRE per seleccionar els individus que només continguin aquests dos casos.

1. Clica la icona de filtratge. S'obre el panell de filtres. Si tens algun filtre activat, desactiva'l.
2. Escriu la condició que han de satisfer els casos que vols seleccionar:
   `p1001 == "SI" or p1001 == "si"`
3. Com que han quedat exclosos tots els casos que no compleixen la condició, escriu, a la Consola d'R:
   `table (data$CUES)`

I veuràs que l'error es troba en els qüestionaris 592 i 790.

4. Desactiva el filtre, clicant-hi a sobre i passant-lo a inactiu:



Imatge 91. Desactivació d'un filtre (inactive)

L'acció de localitzar en quin qüestionari es troba l'error és possible perquè disposem de la variable CUES, un identificador dels qüestionaris. Ara, a més, i si disposéssim dels qüestionaris en paper podríem consultar quin és el valor real i corregir la matriu de dades (en aquest cas això seria prescindible, ja que és molt obvi que la informació correcta d'aquests dos qüestionaris era "Si".

Anem a fer el mateix per a la **variable quantitativa p6** (edat que tenia en començar a utilitzar Internet), tenint en compte que la enquesta s'ha fet a joves de fins a 29 anys):

1. Clica Anàlisis / R / Rj Editor... Escriu l'ordre corresponent d'R:
   `table(data$p6)`
2. Clica el triangle verd a la part superior dreta del panell consola R, per executar-la.
3. Observa que hi ha un seguit de valors que són impossibles. En concret, els valors 39 i 49, ja que no hi pot haver cap valor superior a l'edat màxima de l'enquesta, 29 anys.
4. Clica la icona del filtratge i escriu la condició que han de satisfer els casos que vols seleccionar:
   `P6 > 29`
5. Com que han quedat exclosos tots els casos que no compleixen la condició, escriu, a la Consola d'R:
   `table (data$CUES)`

I veuràs que l'error es troba en els qüestionaris 728 i 784.

6. Desactiva el filtre, clicant-hi a sobre i passant-lo a inactiu.

Tal i com et pots imaginar, en aquest cas sí que tindria sentit anar als qüestionaris en paper, sí els tinguéssim, i comprovar quin és el veritable valor, i canviar-lo; ja que no podem suposar quin ha estat l'error.

(cc) BY-NC

Errors en la congruència entre variables

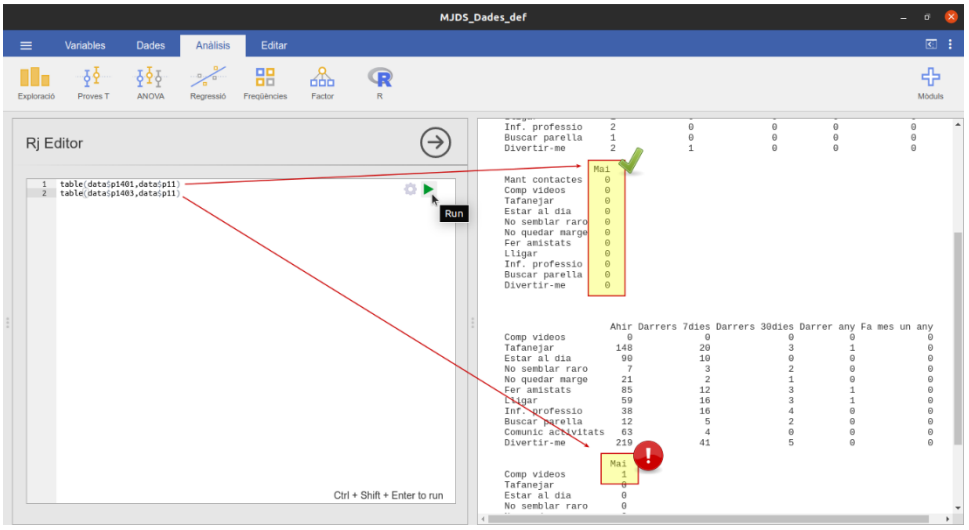
Per avaluar possibles errors que afecten "creuaments" de parelles de variables haurem de determinar, en primer lloc, quins creuaments poden dur-nos a incongruències. En la nostra base de dades, per exemple, podríem trobar incongruències entre les següents variables:

- La Comunitat Autònoma i la província han de ser congruents.
- Haver dit que no has utilitzat mai Internet (p3=No) hauria de comportar que en totes les preguntes que depenen d'aquesta (fins a la p28) hi hagués com a mínim 118 valors perduts.
- Respondre a p11 que no has fet servir mai cap xarxa social (o no contestar la pregunta) ha de comportar un seguit de valors perduts (118+125+5=248) en totes les variables que fan referència a l'ús de xarxes socials (de la p12 a la p15).
- A la pregunta p15, el nombre de contactes a les xarxes socials (p15) ha de ser igual o superior al nombre d'aquests contactes que es consideren amistats consolidades (p15a).
- I molts altres...

Per detectar possibles incongruències ho solem fer mitjançant "taules de doble entrada" (taules de contingència). Per exemple, si creues la p11 amb la p1401 veuràs que no hi ha cap incongruència, és a dir, ningú dels que han contestat que no fan servir mai les xarxes socials ha respost aquesta pregunta, mentre que si creues la p11 amb la p1403 sí que hi ha una incongruència, atès que s'observa que hi ha un individu que ha respost que no fa servir xarxes socials i ens diu que les utilitza per compartir vídeos.

Per fer aquestes comprovacions ràpides és millor utilitzar directament les ordres d'R executades des de la consola.

1. Clica Anàlisis / R / Rj Editor... Escriu les ordres corresponent d'R:
   `table(data$p1401,data$p11)`
   `table(data$p1403,data$p11)`
2. Clica el triangle verd a la part superior dreta del panell consola R, per executar-les.



Imatge 92. Exemple de control d'incongruència a les dades, mitjançant taules creuades

3. Observa que a la columna corresponent a *Mai*, hi ha un cas que indica *Comp videos*.

4. Clica la icona del filtratge i escriu la condició que han de satisfer els casos que vols seleccionar:
   `p11 == "Mai" and p1403 == "Comp videos"`
5. Com que han quedat exclosos tots els casos que no compleixen la condició, escriu, a la Consola d'R:
   `table(data$CUES)`

I veuràs que l'error es troba en el qüestionari 447.

6. Desactiva el filtre, clicant-hi a sobre i passant-lo a inactiu.

Els connectors and i or:

Te n'hauràs adonat que, als filtres que combinen condicions, a vegades fem servir el connector **or** i altres el connector **and**. Per saber quin utilitzar has de tenir molt present el què estàs buscant. Es tracta de tenir clar si les dues condicions han de ser excloents o incloents.

Exemple 1. En el cas anterior de la variable p1001, volem seleccionar els individus que han respost SI, així com els que han respost si. Ara bé, ningú pot haver respost al mateix temps aquests dos valors o nivells, de manera que cal escriure **or**.
Si haguéssim escrit **and** li estaríem dient, precisament que filtri les persones que al mateix temps han dit SI i han dit si.

Exemple 2. En el cas de la taula de contingència, volíem trobar un individu que complís alhora totes dues condicions , de manera que cal utilitzar **and**.
Si haguéssim escrit **or** ens hagués seleccionat els individus que complissin alguna de les dues condicions).

EXERCICIS PROPOSATS: Transformació de dades

Després de seguir completament l'anterior apartat del tutorial, pots utilitzar l'arxiu resultant (o bé, obrir a *jamovi* l'arxiu *MJDS_Dades_Exercicis.omv*) i respondre les següents qüestions.

(A) Observa que en la variable que ens parla de la probabilitat de quedar amb algú que hagi conegut a través d'Internet (p16) hi ha quatre respostes vàlides: *Molt prob*, *Bastant prob*, *Poc prob* i *Gens prob*. Reconverteix aquestes quatre categories en dues: *Probable* (inclou el molt i el bastant) i *Improbable* (inclou el poc i el gens). Enganxa en un processador de textos la freqüència d'aquesta variable abans i després de la conversió.

(B) El total de contactes que tenen a les xarxes socials (p15) és una variable numèrica que va des d'un contacte fins a 4.600 contactes. A partir d'aquesta genera una variable que tingui tres categories: *Pocs amics* (<100), *Bastants amics* (100-500) i *Molts amics* (>500). En aquest cas enganxa en el processador informació sobre la mitjana d'amics, la mediana i la desviació típica; així com la freqüència de la nova variable.

(C) L'anterior variable compta amb 338 perduts que es corresponen a persones que no fan servir les xarxes Socials. Crea una nova variable que tingui en compte aquests valors perduts, de manera que tingui quatre categories: Cap amic (Missing de l'anterior variable), *Pocs amics* (<100), *Bastants amics* (100-500) i *Molts amics* (>500). Enganxa en un processador freqüències de la nova variable, així com la sintaxi que has fet servir. Per generar aquesta nova variable, explora el funcionament de la funció **IFMISS**.

(D) A partir de les variables p13a01 i p13a02 (respectivament, hores diàries i minuts diaris que l'individu està connectat a les xarxes socials) genera una variable que ens digui **les hores (amb decimals)** que l'individu està connectat a les xarxes socials. Per fer-ho tingues en compte que per passar de minuts a hores decimals cal dividir els minuts entre 60 (o multiplicar-los per 1/60). Fent aquesta operació, per exemple, els 30 minuts es converteixen en 0,5 hores.

Fes un resum de la variable (*summary*) i enganxa'l en un processador de textos.

Respon també les següents qüestions:

- Quina és la moda d'hores de connexió a les xarxes socials?
- El 50% dels que es connecten menys estona, quantes hores es connecta?
- El 10% dels que es connecten més estona, quantes hores es connecta?
- Quant val la mediana?
- I el coeficient de variació de la variable?

En el processador de textos, respon les anteriors preguntes i enganxa els resultats de jamovi que et permeten contestar les preguntes.

(E) A la pregunta 10 del qüestionari (variables de la p1001 a la p1011) es demana els efectes que té Internet en el temps que dediquen a altres activitats (estar sense fer res, dormir, passejar...). Volem crear un índex que ens digui la incidència que té Internet en el temps dedicat a altres activitats. Per fer-ho compta **a quantes activitats** ha afectat Internet. Enganxa en un processador de textos les tres mesures de tendència central de la variable. A partir del mòdul d'R, genera una taula de freqüències que contingui els valors absoluts i els percentatges. Interpreta-la.

(F) A la pregunta 14 es demana quin ús fan de les xarxes socials, i es recull la informació en tres variables p1401, p1402 i p1403. Observa que hi ha resultats que textualment diuen "*Lligar*" i "*Buscar parella*". Crea una variable que ens informi de quanta gent fa servir les xarxes socials amb aquesta finalitat i quanta gent no. Enganxa en un processador de textos la sintaxi que has utilitzat per a la construcció, així com una freqüència de la nova variable.

(G) La variable p3 ens informa sobre el fet d'utilitzar o no Internet (Sí, No), mentre que la p5 ens diu, per als que fan servir Internet, amb quina freqüència es connecten. Observa que en la p3 hi ha 118 persones que no fan servir Internet, mentre que en la p5 hi ha 119 perduts, que es corresponen a les 118 persones que no fan servir Internet més una altra que no ha contestat la pregunta.

Combinant ambdues, construeix una variable que tingui <u>tres categories o nivells</u>: "Fa servir Internet a diari", "Fa servir Internet, però no a diari", i "No fa servir Internet". Aquí sota es mostra el resultat final, per tal que puguis comprovar si la construcció que has fet és o no correcta. Enganxa en un processador de textos la sintaxi que has utilitzat i una taula de freqüències amb la variable final.

Descriptives

Descriptives

	Internet
N	1419
Perduts	0

Freqüències

Freqüències de Internet

Internet	Freqüències	% del Total	% Acumulat
No fa servir Internet	119	8.4 %	8.4 %
Internet a diari	1066	75.1 %	83.5 %
Internet, no a diari	234	16.5 %	100.0 %

Imatge 93. Resultat de la "combinació" de p3 amb p5

(H) Recodifica la variable p15 (quants contactes o amics o persones agregades tens en total en les teves xarxes socials?) en quatre grups que tinguin el mateix nombre de casos. Per fer-ho, demana primer els quartils de la variable i, posteriorment, fes l'agrupació. Comprova fent freqüències de la nova variable que el resultat és correcte i enganxa aquestes freqüències en un processador de textos.

(I) A partir de les variables TAMUNI i PROV, filtra la base de dades perquè contingui només els residents a les ciutats de Madrid i Barcelona. Un cop fet el filtratge, demana freqüències de la comunitat autònoma i comprova que només hi ha residents a Catalunya i Madrid (42 i 90, respectivament). Enganxa en un processador de textos la sintaxi que has fet servir per a crear el filtre, així com la freqüència de les dues variables (TAMUINI i PROV). Un cop hagis acabat l'exercici, recorda de desactivar el filtre.

(J) Identifica tres errors que hem comès en la variable p63 (nacionalitat). Localitza en quins qüestionaris es troben. Canvia, en la base de dades original, aquests dos valors pel seu codi corresponent. Copia la freqüència de la variable corregida en un processador de textos.

(K) Un cop corregida la variable p63 (nacionalitat), comprova que en creuar-la amb la p63a (des de quan té la nacionalitat espanyola?), es donen incongruències en cinc casos (segons p63 tenen la nacionalitat espanyola i segons p63a només tenen una altra nacionalitat. Localitza en quins qüestionaris es troben aquestes cinc incongruències. Genera una variable nova que es digui p63a_correg, que sigui igual a l'anterior p63a, excepte en aquestes cinc incongruències, que els assignaràs el valor perdut (NA). En el processador de textos copia, en primer lloc, el creuament de les dues variables on es vegi la incongruència. Enganxa-hi també la freqüència de la variable p63a i de la p63a_correg, així com la sintaxi utilitzada per a la correcció.

(L) Crea una variable comparant la ideologia del pare i de la mare (p34 i p35, respectivament). La nova variable ha de tenir tres categories: (1) Pare i mare tenen exactament la mateixa ideologia (2) El pare és més d'esquerres que la mare i (3) La mare és més d'esquerres que el pare. Enganxa en un processador de textos, una taula de freqüències de la nova variable, així com el procediment utilitzat.

(M) Observa les variables p15 (contactes que té a les xarxes socials) i p15a (contactes a les xarxes que són amics consolidats). Crea una variable que ens digui quants amics té a les xarxes que no poden considerar-se amics consolidats.

Un cop construïda la variable calcula'n la mitjana. Demana també la mitjana de la variable *amics consolidats a les xarxes* (p15a). Digues si, majoritàriament, els contactes a les xarxes estan formats per amics consolidats o per amics no consolidats.

(cc) BY-NC

8. Inferència amb una variable utilitzant *jamovi*

L'objectiu d'aquest apartat del Tutorial és guiar en la realització d'operacions amb *jamovi* relatives a la inferència matemàtica o estimació de paràmetres poblacionals a partir de dades mostrals. Primer es tractarà la construcció d'intervals de confiança per a una variable, i a continuació test d'hipòtesis d'una variable en relació a un paràmetre "teòric". En tots dos casos es distingirà entre tipus de variables diferents: estimació de proporcions per a variables nominals/ordinals, estimació de mitjanes per a variables contínues.

Utilitzareu, en tot moment, l'arxiu de dades *MJDS_Dades_def.omv*. Es tracta d'una matriu creada a partir de l'estudi del CIS número 2889, titulat "*Actitudes hacia las tecnologías de la información y la comunicación*" de 12 de maig de 2011.[22] De totes les variables de l'estudi només se n'han conservat algunes, les que podeu consultar en el document *MJDS_QüestionariCIS.pdf*, que necessitareu per anar seguint el tutorial pràctic.

A l'espai docent Campus Virtual (CV) de l'assignatura hi pots trobar els arxius requerits: *MJDS_Dades_def.omv*, i *MJDS_QüestionariCIS.pdf* corresponents a l'enquesta del CIS.

1. Descarrega aquests dos arxius a una carpeta local del teu PC.
2. Fes doble-clic sobre el fitxer *MJDS_Dades_def.omv*
3. Una vegada obert disposes d'una matriu de dades, amb 1.419 individus entrevistats i 70 variables ben definides (tipus de mesura i de data correctes i valors perduts definits).

Construcció d'un interval de confiança per a una proporció (π).

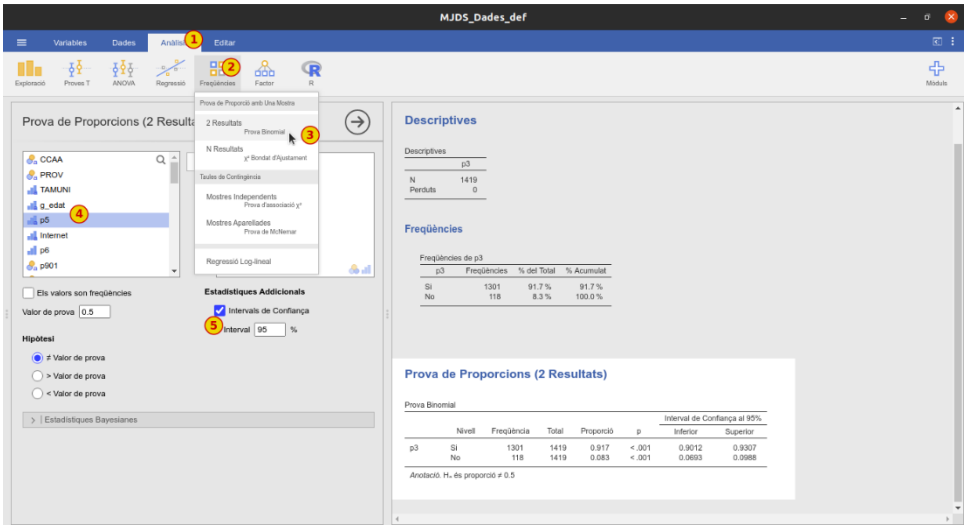
Un interval de confiança per a una proporció ens serveix per estimar, a partir d'una proporció coneguda $P_n(x_i)$ –corresponent a dades d'una mostra–, entre quins dos valors es pot situar la proporció corresponent per a tota la població (π) –de la que s'ha estret la mostra. Així, per exemple, si volem saber quin percentatge de població ha fet servir Internet en els darrers tres mesos, el procediment habitual és preguntar a una mostra representativa i, a partir de l'*estimador puntual* P_n*("S'ha connectat...")* o percentatge de persones que a la mostra s'ha connectat a Internet en els darrers tres mesos, construir un interval de confiança. Direm que aquest interval molt probablement (amb un 95% de confiança, per exemple) contindrà el percentatge real (o poblacional) de gent que ha fet servir Internet en els darrers tres mesos.

El procediment estadístic per a obtenir intervals de confiança per a proporcions amb el sistema de menús de *jamovi* és un test de comparació d'una proporció observada, anomenat "Prova de Proporció amb Una Mostra - 2 Resultats - Prova binomial" (*2 Outcomes Binomial Test*, en anglès).

1. Comprova que la variable p3 està ben definida en *jamovi*. Fes doble-clic a sobre del nom que encapçala la seva columna. S'obre el panell de definició, que hauria d'especificar que és de tipus NOMINAL / TEXT. Si no és així, modifica la definició de la variable p3.
2. Clica l'etiqueta "Anàlisis" del menú superior de *jamovi*.
3. Selecciona l'opció "Freqüències", cap a la dreta del submenú de procediments estadístics.

22 Es pot accedir a les dades originals a <http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=12624>. Tot i que originàriament les dades es van descarregar del servidor del CIS en un format compatible amb l'aplicació comercial d'anàlisi SPSS, van ser convertides en una base de dades de R (amb extensió *.rdata) i modificades convenientment per facilitar la aquesta pràctica.

(cc) BY-NC



Imatge 94. Càlcul d'interval de confiança per a una proporció.

4. Al menú que es desplega, selecciona l'opció "2 Resultats – Prova Binomial". S'obre un nou panell central de definició del procediment. A l'esquerra del panell hi ha un llistat amb totes les variables disponibles (identificades amb la corresponent icona de nivell de mesura). Selecciona la p3. A la part central superior hi ha un botó FLETXA-DRETA, fent clic "passaràs" la variable a l'espai dret del panell on apareixen totes les variables amb les que es generarà l'anàlisi.
5. Per defecte aquest procediment només genera 4 resultats per a cadascuna de les dues categories: recompte de casos (Freqüència), casos totals a la mostra (Total), proporció de casos respecte al total (Proporció), i p-valor d'aquesta proporció (p).
6. Per a generar intervals de confiança cal activar una opció situada a sota del llistat dret de variables disponibles, a l'apartat **Estadístiques Addicionals**. Clica el requadre per a activar "Intervals de Confiança".

Els intervals de confiança (les dues columnes de la dreta de la taula) estan expressats en "tant per 1", i si es multipliquen per 100 passen a ser els percentatges que habitualment s'utilitzen per a expressar estimació de proporcions.

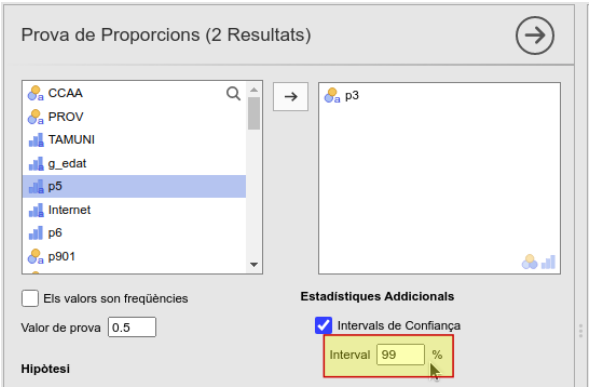
Prova Binomial

						Interval de Confiança al 95%	
	Nivell	**Freqüència**	**Total**	**Proporció**	**p**	**Inferior**	**Superior**
p3	Si	1301	1419	0.917	< .001	0.9012	0.9307
	No	118	1419	0.083	< .001	0.0693	0.0988

Anotació. H_a és proporció ≠ 0.5

El resultat que obtenim ens permet assegurar a partir de les dades de mostra disponibles, i amb un 95% de confiança (o probabilitat d'encert), que la proporció de persones de tota la població que en els darrers tres mesos s'ha connectat a Internet es troba entre el 90,12% i el 93,07%.

Fixa't que al panell de definició del procediment "2 Resultats – Prova Binomial" es pot canviar el nivell de confiança, i construir un interval amb un 99,9% de confiança, per exemple. Si fos el cas hauríem escrit 99 a l'apartat "Interval [ ] %".



Imatge 95. Control del nivell de significativitat d'intervals de confiança

EXERCICI PROPOSAT: INTERVALS DE CONFIANÇA PER PROPORCIONS

- Demana a *jamovi* que calculi un interval de confiança del 90%, del 95% i del 99% per a la resposta "No" de la variable p3. Escriu què vol dir (com cal llegir-lo) cada resultat. Quin interval és més "ample"? Per què?
- Construeix amb jamovi un interval de confiança poblacional del 95% per a la proporció de persones "menors de 25 anys". Escriu què vol dir el resultat.

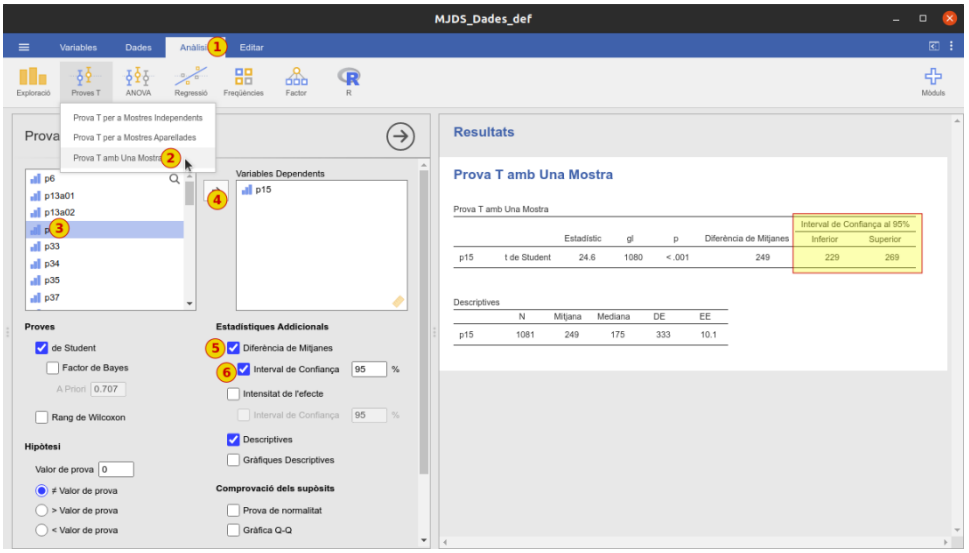
Recorda! Desa freqüentment l'arxiu actualitzat en format propi de *jamovi* (.omv).

Construcció d'un interval de confiança per a la mitjana (μ).

A continuació, veuràs com construir un interval de confiança per a una mitjana, per tant es tracta d'un cas que implica inferència d'una variable del tipus anomenat CONTINU (o d'escala, o numèric). Quan construïm un interval de confiança per a una mitjana ho fem amb l'objectiu d'obtenir informació sobre el valor que, molt probablement, pren la mitjana poblacional. És l'estratègia típicament quantitativista per arribar a conclusions sobre una població general a partir d'una mostra particular.

Per exemple, per demanar en *jamovi* un interval de confiança de la variable p15 (*Nombre d'amics i agregats a les xarxes socials*) cal fer:

1. Comprova que la variable està definida (Configurar) com a *Tipus de Dades=ENTER*. Si no és així, modifica la definició de la variable p15.
2. Comprova que no hagi valors perduts d'usuari. Si cal, elimina'ls.
3. Selecciona a menú Anàlisis / Proves T / Proves T amb Una Mostra… .
4. S'obre el panell de definició del procediment. Situa la variable (p15) de la que vols calcular l'interval a l'espai *Variables Dependents.*
5. Per defecte no es calcula l'interval de confiança. Cal que activis l'opció "Diferència de Mitjanes" i l'opció "Interval de Confiança" a l'apartat **Estadístiques Addicionals**.
6. Per visualitzar els estadístics descriptius cal activar l'opció "Descriptives" a l'apartat **Estadístiques Addicionals**.
7. Per defecte, el nivell de confiança és 95%, però, si cal, es pot canviar a la "finestra" corresponent.



Imatge 96. Càlcul d'interval de confiança per a una mitjana

El resultat és:

Prova T amb Una Mostra

						Interval de Confiança al 95%	
		Estadístic	gl	p	Diferència de Mitjanes	Inferior	Superior
p15	t de Student	24.6	1080	< .001	249	229	269

Per tant, es pot afirmar, amb un nivell de significació del 5% (que és el mateix que dir, amb un 95% de confiança), que la mitjana de contactes a les xarxes socials pel conjunt de la població estarà entre 229 amics i 269 amics.

Compte! Abans de construir l'interval de confiança és convenient comprovar que la variable estigui "depurada", és a dir que no hi hagi valors impossibles. Per fer-ho demanarem freqüències de la variable i comprovarem que tots els valors són lògics i, en cas contrari, definirem com a perduts els valors impossibles. Recorda que per fer-ho, malgrat que la variable sigui numèrica, cal que estigui definida a *jamovi* com a *Tipus de mesura=ORDINAL*.

Codi R per a la construcció d'intervals de confiança (mitjana).

El mateix procediment, càlcul d'un interval de confiança per a una mitjana, es pot realitzar utilitzant directament la sintaxi R:

1. Obre la consola R a *jamovi*: menú Anàlisis / R / Rj Editor...
2. Tot i que la instrucció que cal teclejar, és la següent *(amb p15, i confiança del 95%)*:
   `t.test(data$p15, conf.level=0.95)`
   ...aquesta instrucció ens donarà un ERROR , ja que per R la variable p15 no és numèrica.
3. Per tant, en primer lloc l'haurem de reconvertir. Una manera de fer-ho utilitzant R és escrivint les dues instruccions següents (la primera genera, a partir de p15, la variable

anomenada *p15_n* que serà numèrica, i la segona demana directament l'interval de confiança d'aquesta nova variable):

```
data$p15_n<-as.numeric(as.character(data$p15))
t.test(data$p15_n, conf.level=0.95)
```

4. Executa la instrucció R clicant el triangle verd de la part superior dreta del panell, o bé teclejant la combinació [Ctrl] + [Majus.] + [Intro] al teclat.

Tal i com pots deduir, per canviar el nivell de confiança demanat, per exemple al 99%, caldrà escriure i executar la següent variant de la instrucció anterior (en vermell el canvi introduït):

```
t.test(data$p15_r, conf.level=0.99)
```

De manera que ara, el nou resultat serà:

```
        One Sample Proves T

data:  data$p15_n
t = 25, df = 1080, p-value <2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
        228.9 268.6
sample estimates:
mean of x
          248.7
```

- Quin dels dos intervals és més ample, amb 95% o amb 99% ?
- Com es relaciona això amb la significativitat o error acceptable ?
- Com és que aquest resultat no es dona en català, a diferencia del demanat amb els menus de *jamovi* ?

EXERCICIS PROPOSATS: Intervals de confiança

Després de seguir completament el present apartat del tutorial sobre intervals de confiança, pots posar en pràctica el teu aprenentatge realitzant els següents exercicis.

Seguint amb el mateix arxiu de dades, *MJDS_Dades_def.omv*, respon les següents qüestions.

(A) A partir de la pregunta 13 del qüestionari (variable p13) es demana que facis la següent estimació puntual: quin percentatge dels que utilitzen les xarxes socials ho fa com a mínim una vegada al dia? Un cop la tinguis, construeix-ne (i interpreta) un interval de confiança utilitzant un nivell de confiança del 95%.

(B) A partir de la pregunta 6 del qüestionari, digues quin percentatge dels entrevistats que fan servir Internet, va començar a fer-ne ús abans dels 12 anys? Tingues en compte, abans de construir l'interval, que cal depurar la variable, atès que l'enquesta s'ha fet a joves de 15 a 29 anys, i que no hi pot haver ningú que hagi començat a fer servir Internet després dels 29 anys. A partir d aquest percentatge construeix un interval de confiança, amb tres nivells de confiança diferents: un 50%, un 95% i un 99%. Comenta quines conseqüències té aquesta diferència tan gran entre els intervals.

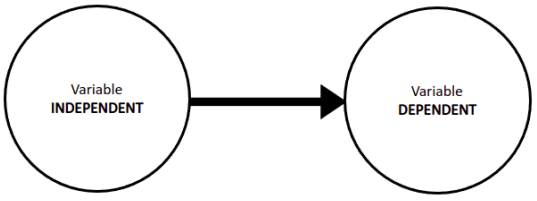
(C) A partir de p13a01 i de p13a02 (respectivament, hores i minuts d'utilització de les xarxes socials) genera una variable que sigui el temps (en minuts) dedicat a les xarxes socials. Quina és l'estimació

puntual del temps dedicat a les xarxes socials? Construeix i interpreta un interval de confiança per aquesta mitjana, amb un 95% de confiança.

9. Inferència de relació entre variables

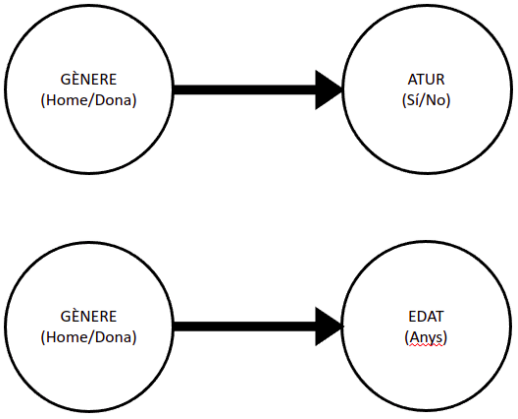
Les seccions anteriors del present Tutorial pràctic estan dedicades a la descripció i la inferència en relació a UNA variable. Aquesta és una part fonamental dels estudis socials: el coneixement de la distribució quantitativa dels valors d'un indicador concret o característica entre la població. Correspon a hipòtesis DESCRIPTIVES respecte a la població, del tipus "més d'un 20% es troba en situació d'atur", o "l'edat mitjana se situa a prop dels 30 anys".

Quan es consideren hipòtesis EXPLICATIVES cal treballar amb més d'una variable. Una d'elles (anomenada DEPENDENT) presenta una distribució concreta de valors que no és aleatòria, sinó que està determinada –fins a cert punt– per la distribució d'una altra (o altres) variables. Segons un model d'anàlisi determinat, fonamentat en un marc de referència teòric, aquestes altres variables "explicatives" (anomenades INDEPENDENTS) tenen una relació amb la variable dependent que pot ser de simple covariació o d'associació i influència causal.



Imatge 97. Model d'anàlisi per a contrastar la relació entre 2 variables.

Aquesta relació explicativa també pot ser contrastada empíricament amb dades quantitatives, fins al punt d'afirmar o desmentir hipòtesis. Aquesta estratègia correspon a hipòtesis EXPLICATIVES respecte a la població, del tipus "més dones que homes es troben en situació d'atur", o "l'edat mitjana de les dones és superior a la dels homes".



Imatge 98. Dos exemples d'anàlisi de la relació entre 2 variables.

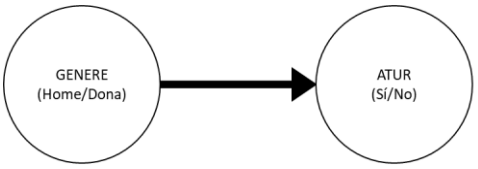
En aquest cas, cal utilitzar les eines estadístiques corresponents a l'anàlisi bivariable (objecte de la resta del Tutorial) o multivariable.

En analitzar la relació entre dues variables, i en funció de la seva mètrica, tenim quatre opcions diferents, que es poden expressar de la següent forma:

COMBINACIÓ DE VARIABLES	OPERACIONS ESTADÍSTIQUES
Variable qualitativa → Variable qualitativa	Taula de contingència. Test de la khi quadrat
Variable qualitativa → Variable quantitativa	Comparació de mitjanes. Anàlisi de la variància
Variable quantitativa → Variable quantitativa	Anàlisi de regressió. La regressió lineal
Variable quantitativa → Variable qualitativa	Regressió logística

10. Anàlisi de taules de contingència mitjançant jamovi

En aquesta secció, seguint el Tutorial pas a pas, pots adquirir competència per a utilitzar el programari *jamovi* per analitzar la relació entre dues variables qualitatives mitjançant el procediment Taules de Contingència, així com fer-ne la representació gràfica pertinent. També podràs interpretar la informació que s'obté de *jamovi* per tal de donar resposta a les hipòtesis de relació entre les variables d'acord amb un model d'anàlisi bàsic (aspecte descriptiu i de contingut de l'anàlisi), i determinar la significació d'aquesta informació per tal d'extrapolar-la al conjunt de la població (aspecte inferencial de l'anàlisi).



Imatge 99. Exemple de model d'anàlisi propi de Taula de Contingència.

La idea general de la inferència estadística és avaluar fins a quin punt unes diferències que s'han observat en una mostra són significativament extrapolables al conjunt de la població i no es poden atribuir a l'atzar, tot i que les dades provenen d'una mostra o subconjunt de població.

En bona part de les anàlisis bivariables, quan parlem de diferències ens referim a diferències entre els grups que genera una variable respecte d'una altra. Així, per exemple, si disposem de la variable sexe i ús d'Internet, quan parlem de diferències ens referim a les possibles diferències en l'ús d'Internet entre homes i dones. En aquests casos dir que hi ha diferències SIGNIFICATIVES entre homes i dones en l'ús d'Internet permet parlar de la relació entre el sexe i l'ús d'Internet per a la població en general.

Observació! Parlar de relació entre dues variables comporta assignar a una de les variables el paper de variable independent (o explicativa) i a l'altra el de variable dependent (o resposta). La variable independent és la que ens generarà els grups de població comparables.

En aquesta secció es continuarà treballant amb dades d'una matriu creada a partir de l'estudi del CIS número 2889, titulat "*Actitudes hacia las tecnologías de la información y la comunicación*" <http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=12624> de 12 de maig de 2011. De totes les variables de l'estudi només se n'han conservat algunes, que han estat depurades, de manera que no hi hagi errors ni valors que puguin ser considerats com a perduts, i se n'han escurçat algunes etiquetes per facilitar-ne la lectura i la manipulació de dades.

Cal consultar el document associat *MJDS_Qüestionari.pdf* per anar seguint el Tutorial pas a pas.

Per a disposar de les matrius de dades corresponents:

1. Localitza a l'espai Moodle de l'assignatura l'apartat corresponent a aquest apartat dedicat a TAULES DE CONTINGÈNCIA (o localitza els següents arxius als adjunts de MUJADES).
2. Descarrega al teu PC els arxius necessaris: *MJDS_Qüestionari.pdf*, *MJDS_Dades_def.omv*, i *MJDS_reincid.omv*.
3. Executa l'aplicació *jamovi* al teu ordinador.

4. Clica el menú general (dalt, esquerra).
5. Selecciona l'opció Obrir / Aquest PC. Al panell central i clica [Navega] (dalt, dreta).
6. S'obre una finestra per seleccionar l'arxiu de dades a obrir. Tria l'arxiu *MJDS_Dades_def.omv*.

Recorda que, en funció del tipus de variable, *jamovi* permet la realització o no de determinats processos estadístics. Cal tenir definit de forma adient el tipus de variable per poder realitzar els procediments d'anàlisi pertinents.

Sempre cal revisar la definició de cadascuna de les variables quan s'obre (o s'importa) amb *jamovi* un arxiu de dades, especialment en el cas que sigui del seu propi format (.omv). Com ja saps, l'aplicació *jamovi* prova d'"endevinar" el tipus de la variable a partir de les dades contingudes a l'arxiu. Aquesta identificació automàtica no sempre és l'òptima.

Recorda desar el resultat de les teves modificacions en format propi de *jamovi*.

- Utilitza el menú general / Desa com a ... i desa a una ubicació que puguis recuperar en un altre moment l'arxiu en format propi de *jamovi* (*MJDS_Dades_treb.omv*, per exemple).

Estadística descriptiva en les taules de contingència.

Per analitzar la relació entre dues o més variables de mètrica qualitativa (variables mesurades a nivell nominal o ordinal) *jamovi* disposa del procediment Taules de Contingència al menú *Anàlisis / Freqüències*. Aquest menú proporciona, com a resultats, taules de distribució conjunta de freqüències. Les taules contenen càlculs de diverses mesures que avaluen l'existència d'associació entre les variables i la intensitat d'aquesta associació estadística.

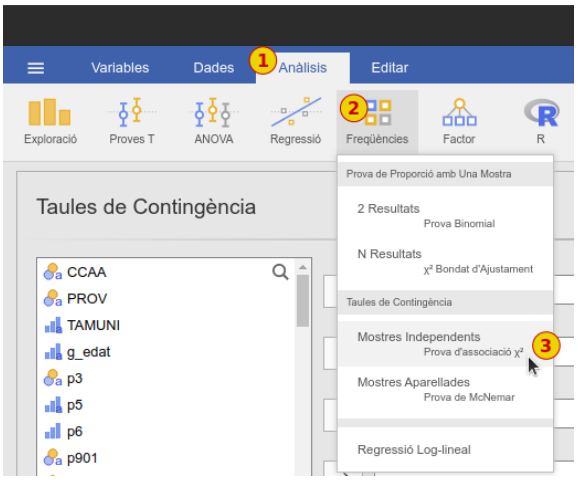
En aquest exemple explorem la possible relació entre la utilització d'Internet (p3) i l'estatus socioeconòmic (ESTATUS). Una taula de contingència (TdC) o taula de freqüències creuades, té dues dimensions, una variable se situa en horitzontal amb les seves categories a les columnes de la taula, i l'altra en vertical amb les seves categories com a files. Cal decidir doncs quina variable apareixerà a les files i quina a les columnes.

La possible relació entre aquestes dues variables implica una comparació de mostres "independents", anomenada així perquè és impossible que un mateix cas (persona) estigui **alhora** en dues categories diferents de la variable. Hi ha una altra situació en què les mostres són "dependents" o "aparellades": quan les dues variables fan referència a observacions en moments de temps diferents pel **mateix** cas o persona (p.e., "Estatus socioeconòmic actual" i "Estatus socioeconòmic fa 10 anys").

Compte!. Fixa't que els termes "independent" i "dependent" (o aparellades), referit a les mostres o subgrups que es volen comparar, té un significat TOTALMENT DIFERENT que quan la mateixa terminologia fa referència al paper d'una variable en la relació bivariable.

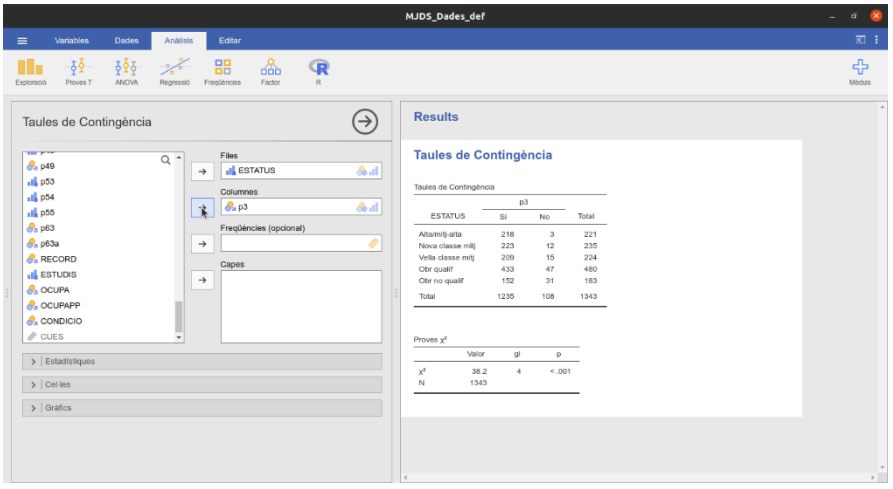
Per a demanar com a resultat una taula de contingència (TdC) amb *jamovi*:

1. Comprova els tipus de variables i de dades: la p3 ha de ser NOMINAL/TEXT, i l'ESTATUS ha de ser ORDINAL/TEXT. Si no és així, modifica la seva definició ("Setup"). Utilitza el que has après anteriorment al Tutorial: pots ajudar-te de l'arxiu *"MJDS_Qüestionari.pdf"* per a aquesta tasca, i per a introduir les descripcions per a cada variable.
2. Selecciona el procediment de Taules per a mostres Independents, a menú Anàlisis / Freqüencies / Taules de Contingència / Mostres Independents – Prova d'associació χ^2 .



Imatge 100. Menú per generar taules de contingència amb mostres independents.

3. Selecciona la variable p3 i mou-la a l'espai dret per a "Columns" (columnes).
4. Al final del llistat esquerre, selecciona la variable ESTATUS i mou-la a l'espai dret per a "Rows" (files).



Imatge 101. Taula de contingència simple, mostres independents

Tal i com pots veure, no cal res més per a generar un resultat de taula de contingència o de freqüències creuades. Les xifres corresponen al recompte d'efectius, l'anomenada "freqüència observada", per a cada combinació de categories de les dues variables (o "cel·la" o "casella").

Taules de Contingència

	p3		
ESTATUS	**Si**	**No**	**Total**
Alta/mitj-alta	218	3	221
Nova classe mitj	223	12	235
Vella classe mitj	209	15	224
Obr qualif	433	47	480
Obr no qualif	152	31	183
Total	1235	108	1343

Les taules de contingència no tenen en compte els valors "perduts". Pots comprovar que la suma total d'aquesta taula és 1.343, tot i que sabem que la mostra conté informació per a 1.419 persones.

A aquesta versió simple de taula de freqüències OBSERVADES (el recompte real de casos a la mostra) es poden afegir d'altres elements que enriqueixen les possibilitats d'anàlisi. Altres tipus de freqüències opcionals, son:

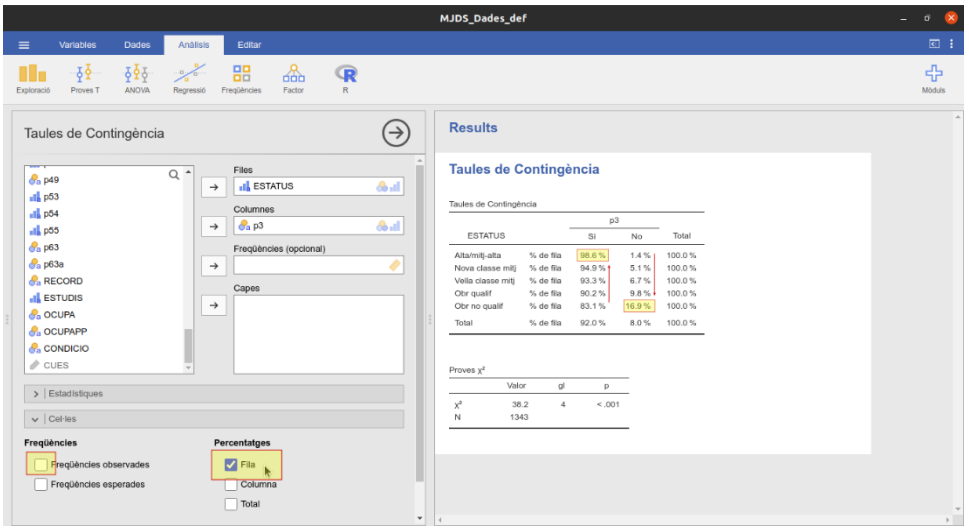
- *Observades*. Freqüència absoluta i realment observada per a cada casella.
- *Esperades*. Freqüència "prevista" o "calculada si es donen condicions d'independència". Ens fa una estimació dels efectius que seria esperable trobar a cada cel·la si NO hi hagués CAP relació entre les dues variables.
- *% de fila*. Freqüència relativa, o Percentatge respecte de la fila, que en el nostre cas és la variable ESTATUS. Ens diu com es distribueix l'ús d'Internet (p3) en cadascun dels nivells d' ESTATUS. En la primera fila llegim que d'entre el total dels que tenen estatus alt o mitjà-alt, un 98,6% han fet servir Internet en els darrers tres mesos, i un 1,4% no. Per construcció, els percentatges de cada fila han de sumar el 100%.
- *% de columna*. Freqüència relativa, o Percentatge respecte de la columna, en el nostre cas la utilització d'Internet. Ens diu quin és l'ESTATUS dels que fan servir Internet i quin és l'estatus dels que no el fan servir. En la primera columna llegirem que un 17,7% dels que fan servir Internet tenen un ESTATUS alt o mitjà-alt, un 18,1% pertanyen a les noves classes mitjanes, un 16,9% a les antigues classes mitjanes, i així successivament. El total de la columna suma el 100%.
- *% del total*. Freqüència relativa, o Percentatge respecte del total. Quin percentatge del total d'entrevistats compleixen ambdues característiques, per exemple ser obrer qualificat i fer servir Internet. D'entre el total d'entrevistats, un 16,2% pertany a la categoria C.A/C.M-A i fa servir Internet, un 0,2% pertany a la categoria C.A/C.M-A i no fa servir Internet, i així successivament. La suma de tots aquests percentatges suma el 100% de la taula.

(cc) BY-NC

Observació! Els percentatges que serveixen per avaluar la relació entre dues variables són sempre els percentatges respecte de la variable independent. Així, per avaluar la relació entre dues variables qualitatives el més habitual és demanar només un percentatge, el que es construeix respecte de la variable independent. En aquest cas és l'ESTATUS (variable ubicada a files), per tant, en aquest cas cal calcular els % per fila.

A grans trets, en l'exemple anterior:

1. Clica a sobre de la taula. S'obre el panell de definició de la taula de contingència.
2. Clica el desplegable de les opcions > Cel·les.
3. Activa l'opció de percentatges [ ] Files.
4. Desactiva l'opció de recomptes [ ] Freqüències observades.



Imatge 102. Percentatges per fila (V.I.) a una Taula de Contingència.

Si llegim els percentatges respecte de l'estatus (*% de fila*) veiem que l'ús d'Internet "descendeix" (passa de Sí a No) a mesura que també ho fa l'estatus (variable ordinal). En el cas dels extrems, veiem que l'ús és d'un 98,6% entre els de classe alta/mitjana alta, mentre que el menor ús correspon als obrers no qualificats, entre els quals només un 83,1% fa servir Internet.

Taules de contingència descriptives amb R

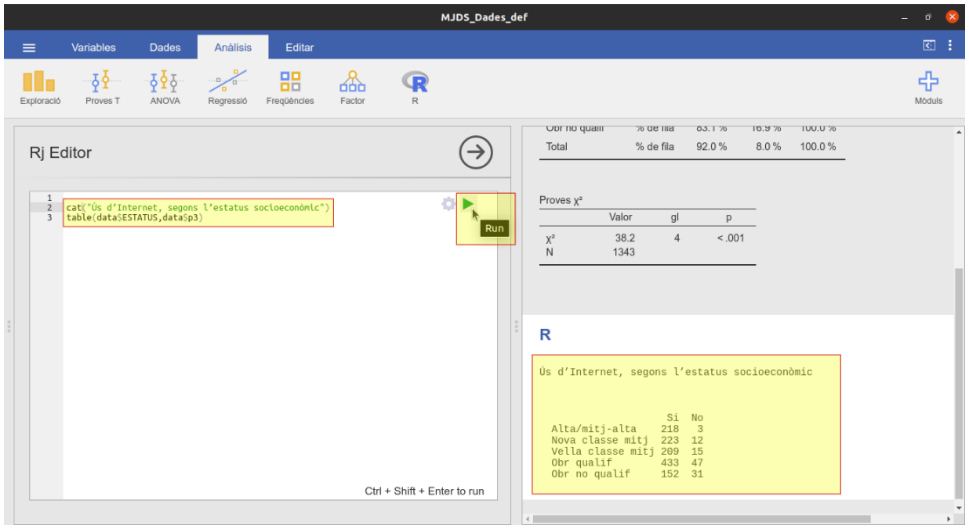
Per a obtenir el resultat d'una **taula de contingència directament executant codi R** amb l'extensió (Rj) de *jamovi* cal:

1. Tenir instal·lat el mòdul o extensió de jamovi anomenat Rj Editor.[23]
2. A l'apartat *Anàlisis*, clicar la icona *R*. Al desplegable, seleccionar l'opció *Rj Editor*,
3. Escriu, poc a poc, les dues ordres per a generar una taula amb les variables p3 i ESTATUS a la consola:

23 Al tutorial Pràctica 1 es pot trobar la forma d'instal·lar aquest i d'altres mòduls o extensions per augmentar les prestacions de jamovi.

```
cat("Ús d'Internet, segons l'estatus socioeconòmic")
table(data$ESTATUS,data$p3)
```

4. Com has pogut veure, *jamovi* t'ajuda proposant ordres i variables disponibles.
5. Aquestes simples ordres que has escrit en llenguatge R signifiquen: "mostra el text: 'Ús d'Internet segons l'estatus socioeconòimic' ", i "mostra una taula de freqüències de les variables ESTATUS a files i p3 a columnes de la matriu de dades".
6. Per a executar el codi que hi ha escrit a la consola cal clicar al petit triangle verd de la part superior dreta de la consola. Alternativament, també es pot executar picant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho.
7. Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.



Imatge 103. Taula de contingència generada amb codi R

8. Per tancar el panell central, igual que en el cas de qualsevol altre procediment d'anàlisis de *jamovi*, cal clicar a la fletxa encerclada de la part superior dreta.
9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Per a generar una **taula de percentatges amb codi R**, en aquest cas cal tenint en compte que la variable independent (ESTATUS) està ubicada a les files de la taula.

Molta atenció a com la taula definitiva es va construint, pas a pas, a partir d'"objectes" intermedis, o taules parcials. Recorda que en codi R s'utilitza <- per "anomenar" o "desar en memòria" objectes com ara valors, o taules, o textos.

1. A la consola R de *jamovi* afegeix les ordres següents per a generar diverses versions de taules de % per files amb les variables p3 i ESTATUS:

```
TITOL <- cat("Ús d'Internet, segons l'estatus socioeconòmic (% files)")
TCont <- table(data$ESTATUS,data$p3)
TPerc <- prop.table(TCont,1)
TP2dec <- round(100*TPerc,2)
```

2. Aquestes ordres que has escrit en llenguatge R signifiquen:
   - "genera l'objecte anomenat TITOL que serà 'Ús d'Internet, segons l'estatus socioeconòmic (% files)'",

- "genera l'objecte anomenat TCont que serà una taula de freqüències de les variables ESTATUS a files i p3 a columnes de la matriu de dades",
- "genera l'objecte anomenat TPerc que serà una taula de proporcions calculada a partir de l'objecte anomenat TCont, les proporcions són per a cada fila (,1) " [per a demanar el càlcul per columna cal indicar (,2)]
- "genera l'objecte anomenat TP2dec que serà 100 multiplicat per cada element d'una taula de proporcions calculada a partir de l'objecte anomenat TPerc, amb valors arrodonits a 2 decimals (,2)"

3. Per a executar el codi que hi ha escrit a la consola cal clicar al petit triangle verd de la part superior dreta de la consola. Alternativament, també es pot executar picant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho.
4. NO APAREIX RES AL PANELL DE LA DRETA. Per què, en aquest cas, les ordres només han generat resultats (objectes) emmagatzemant-los a la memòria. Calen altres ordres R per a VISUALITZAR aquest resultats. Escriu i executa, una a una, les següents ordres:
```
cat(TITOL)
TPerc
TP2dec
```
5. Apareixeran els resultats al panell de la dreta, integrat amb la resta de resultats de l'arxiu. Comprova com la taula definitiva (TP2dec) està generada a partir de les taules anteriors.
6. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Atenció! Tal i com has pogut veure, el codi R permet més control sobre el contingut i el disseny de cada taula. També que una taula definitiva es pot construir a poc a poc reutilitzant resultats anteriors (identificats pel seu nom com a "objecte R").

La representació gràfica de les taules de contingència

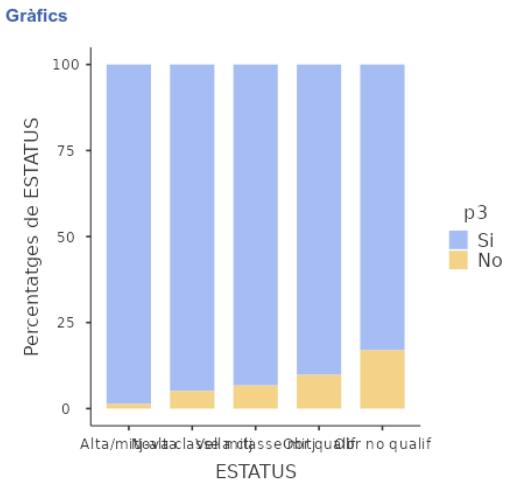
Passant a l'aspecte de la representació gràfica de les dades, el més habitual és dibuixar la relació entre ambdues variables (és a dir, els percentatges construïts respecte de la variable independent) mitjançant l'anomenat **gràfic de columnes apilades al 100%**. Cada columna representa el 100% de cada categoria o nivell de la variable independent situada a les columnes, mentre que cadascuna d'aquestes columnes està dividida proporcionalment a la distribució de la variable dependent situada a les files.

Per poder generar amb *jamovi* aquest tipus de gràfics:

1. Si encara no existeix la Taula de Contingència, segueix les instruccions anteriors per a generar-la de nou.
2. Si ja existeix un resultat Taula de Contingència, clic a sobre de la taula. S'obrirà a la part esquerra la finestra de definició del procediment, amb les opcions que s'han utilitzat per generar el resultat existent.
3. Fes clic a la pestanya inferior > Gràfics. Es desplegaran les opcions de generació de representació gràfica associada a taules de contingència.
4. Activa l'opció [ ] Gràfic de barres.

Nota: Sí, tot i que la gràfica representa COLUMNES, s'anomena arreu "gràfica de BARRES"...

5. Al Tipus de barres, selecciona l'opció ( ) Apilats.
6. A l'eix Y, selecciona l'opció ( ) Percentatges, clica al desplegable i selecciona "de fila".
7. El resultat mostra cinc columnes, una per a cada nivell o categoria de la variable independent ESTATUS. Cada columna representa el total (100%) dels casos d'ESTATUS, i internament està dividida en dos seccions proporcionals als dos nivells de la variable US D'INTERNET (p3).



Imatge 104. Exemple de gràfica de columnes apilades al 100% amb jamovi

Com veus, hi ha un problema amb la superposició de les etiquetes dels nivells o categories de la variable representada a l'eix X (horitzontal). És una de les limitacions de fer servir els sistema simple de menús de *jamovi*.

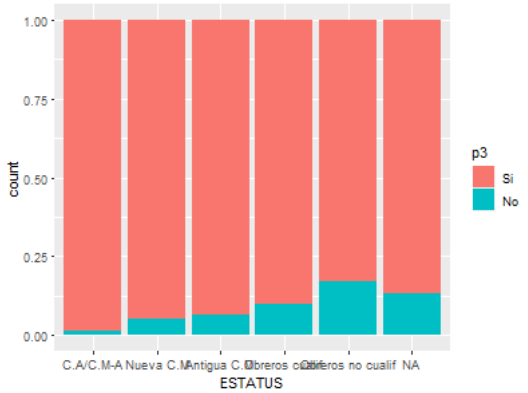
Gràfiques de taules de contingència, amb R

La programació R permet un control més detallat dels elements del gràfic. Per a generar el gràfic anterior, a partir d'una matriu de dades que contingui una variable "ESTATUS" a les files, una variable "p3" a les columnes i titulat "Ús d'Internet segons Estatus" és la següent:[24]

VERSIÓ SIMPLE

```
library(ggplot2)
ggplot() +
   geom_bar(aes(y=..count.., x=ESTATUS, fill=p3),
   data=data, position = position_fill())
```

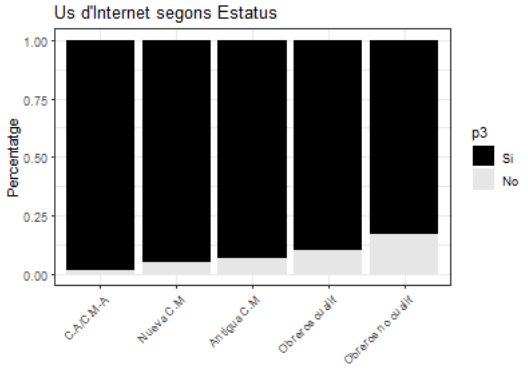
24 Cal executar la instrucció que "activa" la part de R que realitza els gràfics: **`library(ggplot2)`** , només una vegada abans d'executar el primer gràfic. Per a la resta de gràfics ja no cal executar-la, tot i que aquí l'hem inclosa a cada gràfic.



Imatge 105. Representació gràfica de taula de contingència (SIMPLE)

VERSIÓ MILLORADA VISUALMENT **(i eliminant els valors perduts NA)**

```
library(ggplot2)
ggplot() +
    geom_bar(aes(y = ..count..,x = ESTATUS, fill = p3),
    data = subset(data, !is.na(ESTATUS)),
    position = position_fill()) +
    ggtitle("Ús d'Internet segons Estatus") +
    xlab("") + ylab("Percentatge") +
    scale_fill_grey(start = 0.4, end = .8) +
    theme_bw() +
    theme(axis.text.x=element_text(angle=45, hjust=1, size=rel(1.3))) +
    theme(axis.text.y=element_text(size=rel(1.2)))
```



Imatge 106. Representació gràfica de taula de contingència (MILLORADA)

Fixa't en com les opcions milloren la representació gràfica:

- S'utilitzen només els valors de la variable ESTATUS que no (!) són perduts (na) amb
  `!is.na(ESTATUS)`
- S'afegeix un títol a l'eix vertical, amb
  `ylab("Percentatge")`
- No s'utilitza més que blanc, negre i un gradient de gris, amb
  `scale_fill_grey(start = 0.4, end = .8) + theme_bw()`

- S'ajusta l'angle de les etiquetes de l'eix horitzontal, perquè no se solapin i siguin totes visibles, i s'incrementa la mida de la lletra amb
  `theme(axis.text.x=element_text(angle=45, hjust=1, size=rel(1.3)))`
- S'incrementa la mida de la lletra de l'eix vertical amb
  `theme(axis.text.y=element_text(size=rel(1.3)))`

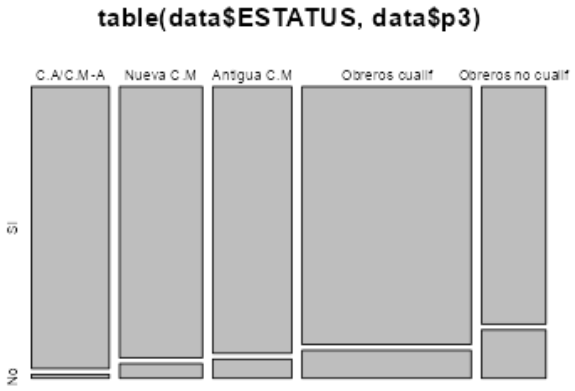
Atenció! En aquest cas només la variable ESTATUS contenia valors perduts (na), de manera que només els hem eliminat en aquesta variable. Si hi hagués hagut perduts en ambdues, a la tercera fila hauríem d'haver escrit:

```
data = subset(data, !is.na(ESTATUS) & !is.na(p3)),
```

Per a generar d'altres gràfiques similars només cal copiar la instrucció, canviar els elements propis pels del nou gràfic i executar des de la consola.

Recorda: Segons la versió de *jamovi* que tinguis instal·lada, aquest tipus de gràfic es pot aconseguir utilitzant el menú. Tot i així, no permet manipular cap dels elements que sí permet la consola d'R, com és posar un títol, canviar els colors, posar etiquetes més llegibles...

Aquestes gràfiques de barres apilades escalades al 100% son adequades per a visualitzar la distribució en quant a un parell de variables de tipus qualitatiu. Amb codi R es pot generar una altra modalitat de representació gràfica, anomenada **gràfica de mosaic**, que indica el mateix d'una forma molt més simple.



Imatge 107. Representació gràfica de taula de contingència (MOSAIC)

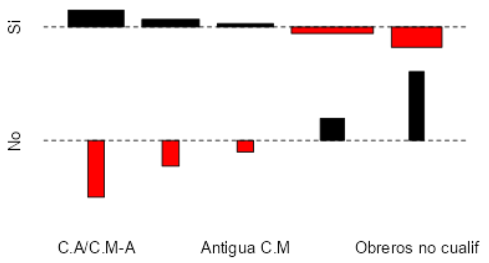
Per a generar aquesta modalitat de gràfica cal carregar la libreria anomenada "vcd", escriure i executar el següent codi R:

```
library("vcd")
mosaicplot(table(data$ESTATUS,data$p3))
```

Aquestes gràfiques de mosaic permeten afegir fàcilment un element interpretatiu a les representacions de taules de contingència. Amb la opció ASSOCPLOT (gràfica d'associació) es pot

identificar visualment a quines categories hi ha freqüències observades superiors (negre) o inferiors (vermell) al que es podries esperar sota la hipòtesi de NO DEPENDENCIA entre les dues variables considerades, val a dir, mostren la estructura detallada d'associació de la taula:

```
assocplot(table(data$ESTATUS,data$p3))
```



Imatge 108. Representació gràfica de taula de contingència (GRAFICA D'ASSOCIACIÓ)

En aquest cas, es pot veure com "destaca" especialment el no ús de WWW dels obrers qualificats i no qualificats, així com l'ús en classes Alta i Mitjà-Alta.

La inferència estadística amb dues variables no quantitatives: El test de la khi quadrat de Pearson.

L'objectiu de l'estadística inferencial, expressada aquí a partir del test de la khi quadrat (χ^2), és avaluar fins a quin punt aquestes diferències –que s'han trobat, recordem-ho, en una mostra–, són extrapolables al conjunt de la població. La conclusió pot ser:

- Les diferències **són significatives**. Arribarem a aquesta conclusió quan les diferències difícilment siguin degudes a l'atzar. Si unes diferències observades a la mostra no són degudes a l'atzar, és que es donen a la població general i, per tant, és molt probable que existeixi relació entre les dues variables considerades.
- Les diferències **no són significatives**. Les diferències poden ser degudes a l'atzar, atès que les dades han estat obtingudes d'una mostra. En aquest cas, les conclusions derivades de l'anàlisi no es poden extrapolar al conjunt de la població.

L'anàlisi complert de la relació entre dues variables qualitatives (nominal o ordinal) segueix els clàssics QUATRE passos successius i necessaris de qualsevol test d'hipòtesis:

A. Es pot fer l'anàlisi de la khi quadrat? Cal comprovar si es compleixen les **condicions d'aplicació** d'aquest test.

   >> Si no es compleixen, caldrà recodificar les variables i/o aplicar un altre test.

B. Hi ha **associació estadística significativa** entre ambdues variables ? Les diferències observades, són significatives? Cal comprovar que el p-valor sigui ≤ 0,05 *(en cas que estiguem treballant amb un nivell de confiança del 95%)*.

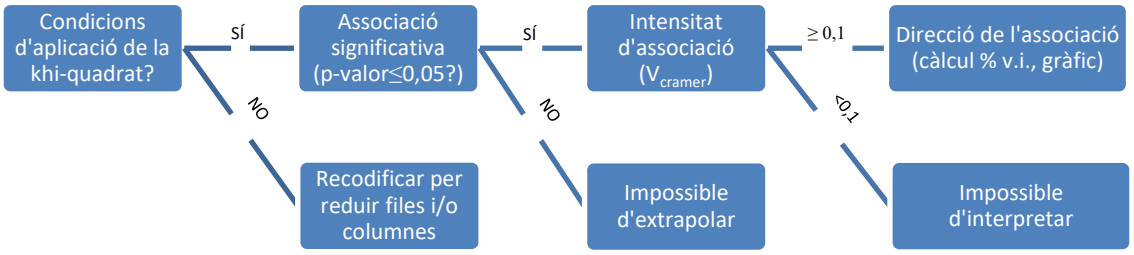
   >> Si no ho és, els resultats de l'anàlisi són veritat per a la mostra, però no es pot dir si també ho són per a la població general d'on s'ha extret la mostra.

C. Si hi ha associació estadística, quina és la **seva intensitat**? Cal avaluar la "força" de l'associació a partir d'un "indicador d'associació" (en aquest cas la *V de Cramer*).

>> Si la intensitat és molt baixa, es pot donar la relació com a pràcticament inexistent.

D. Si la intensitat d'associació es suficient, **"en quina direcció" va la relació**? Cal avaluar el comportament de la variable DEPENDENT quan la variable (ordinal) INDEPENDENT creix, o cal determinar el "perfil típic" corresponent a la relació entre les dues variables.

Si en qualsevol d'aquests passos s'arriba a la conclusió que no es pot continuar (per no associació, o intensitat molt baixa) s'acaba l'anàlisi. En el cas d'incompliment de les condicions d'aplicació, es pot transformar una o les dues variables fins aconseguir que es compleixin les condicions, i llavors realitzar el test corresponent.



Imatge 109. Realització d'un test d'hipòtesi amb taula de contingència

A) Les condicions d'aplicació en el test de la khi quadrat de Pearson.

El test de la khi quadrat té una restricció deguda a la relació entre la grandària de la mostra (*n*) i la mida de la taula a analitzar (*IxJ*, total de files mutiplicat pel total de columnes). Aquesta relació es concreta en el què s'anomenen les "freqüències esperades", és a dir, el nombre de casos que esperaríem en cada cel·la si es complís la hipòtesi nul·la (H_0: *No hi ha diferències. No hi ha relació*). En termes generals aquesta condició ens diu que **no hi pot haver freqüències esperades inferiors a 5** casos.

Alguns autors no són tant restrictius, i consideren que el nombre de cel·les esperades inferior a 5 no comporta cap problema, mentre no superi el 20% de les cel·les de la taula.

Per a demanar taules de contingència amb indicació de les freqüències esperades ("Esperades"):

1. Clica la taula de freqüències anterior, que analitza la possible relació entre ESTATUS i Ús d'internet (p3). Al panell de definició de la taula de contingència, clica el desplegable de les opcions > Cel·les.
2. Activa l'opció de percentatges [ ] Freqüències esperades.
3. Desactiva la resta d'opcions de recomptes.
4. El resultat es mostra al panell de la dreta:

Taules de Contingència

ESTATUS		p3 Si	p3 No	Total
Alta/mitj-alta	Esperades	203	17.8	221
Nova classe mitj	Esperades	216	18.9	235
Vella classe mitj	Esperades	206	18.0	224
Obr qualif	Esperades	441	38.6	480
Obr no qualif	Esperades	168	14.7	183
Total	Esperades	1235	108.0	1343

Tal i com pots veure, cap dels valors esperats és inferior 5 (el mínim és 14,7). Per tant, es compleixen les condicions d'aplicació del test de la khi quadrat i **es pot continuar amb l'anàlisi.**

Més endavant s'explica com solucionar el problema de NO compliment de les condicions.

B) Test de significació estadística khi quadrat de Pearson

El valor que ens informa d'aquesta influència de l'atzar és la **significació del test de la khi quadrat**, també anomenat *p-valor*. Aquesta significació ens informa sobre la probabilitat que les diferències observades siguin degudes a l'atzar:

- quan la significació de χ^2 és **gran**, no podem descartar que les diferències siguin degudes a l'atzar i, per tant, no les podrem extrapolar a la població,
- mentre que una significació de χ^2 **petita** significa que les diferències difícilment seran degudes a l'atzar i, per tant, les diferències són extrapolables a la població.

Habitualment el llindar entre el què és petit o gran se sol situar en el 5%. Si el p-valor és igual o menor (p-valor ≤ 0,05) direm que podem acceptar que hi ha diferències significatives entre les categories (i per tant, que hi ha relació entre les variables) amb un 95% de confiança. De fet, com més petita sigui la significació més improbable és que l'atzar ens hagi jugat una mala passada i més probable és que les diferències es donin veritablement al conjunt de la població.

Recorda! Si analitzes la relació entre dues variables amb un nivell de confiança del 95% (vol dir, assumint un risc tant petit com el 5% d'equivocar les teves conclusions), cal que al test d'inferència el p-valor sigui ≤ 0,05 per a considerar les conclusions "extrapolables".

En general, el nivell de confiança del 95% és acceptable a Ciències Socials. En altres disciplines (més restrictives en quant als requeriments d'evidència empírica) només és acceptable un nivell de confiança del 99%; cosa que significa que s'exigeix un p-valor ≤ 0,01 per poder extrapolar les conclusions.

Per defecte, *jamovi* ja calcula la significació de la relació entre les dues variables quan es demana una Taula de Contingència. En el cas anterior, anàlisi de la relació entre ús d'Internet (p3) i l'estatus

socioeconòmic (ESTATUS), ha generat la taula de freqüències creuades i a continuació la taula amb el test de la khi quadrat.

Proves χ^2

	Valor	gl	p
χ^2	38.2	4	1.01e-7
N	1343		

Si no vols que ens doni aquesta informació, al panell de definició del procediment "Taula de contingència" cal desplegar l'apartat "Estadístiques" i desmarcar l'opció [ ] χ^2.

Recorda! Als resultats d'R (i *jamovi*) el **marcador decimal és el punt** i no la coma. A més, per estalviar espai, sovint s'elimina el 0 sencer d'un nombre decimal petit. Així 0,001 apareix com a .001

En aquest cas, pots observar un p-valor molt petit, menor de 0,001, de manera que les diferències observades entre els diferents estatus socioeconòmics no poden ser degudes a l'atzar i, per tant, es pot assegurar amb un risc d'error molt petit, que l'estatus socioeconòmic està relacionat amb la utilització d'Internet per a la població en general.

El resultat d'aquesta etapa és que hi ha significació, i per tant es pot continuar amb l'anàlisi.

En cas que no hi hagi significació estadística es poden treure conclusions de la taula, però nomes seran aplicables a les persones entrevistades en la mostra (*n*). En aquest cas no es podrà inferir cap conclusió a la població (*N*) a partir de les dades disponibles.

C) Intensitat de l'associació entre variables nominals (V de Cramer).

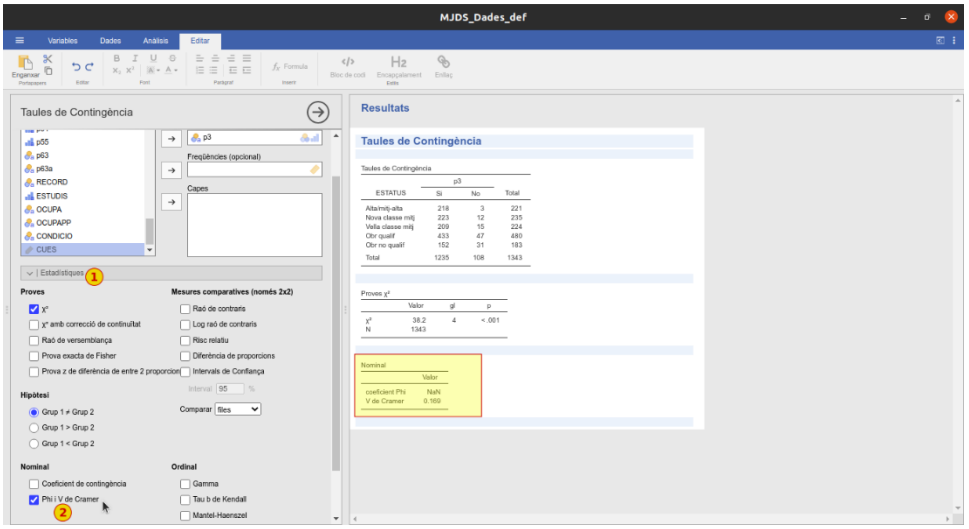
Un cop establert que hi ha relació significativa entre ambdues variables se sol calcular una mesura per avaluar la intensitat d'aquesta relació. Es habitual fer-ho a partir de l'indicador d'associació bilateral **V de Cramer**, que varia entre 0 i 1: Un valor de 0 equivaldria a dir que no hi ha cap relació (o amb valors propers a 0, que és molt baixa) mentre que un valor proper a 1 hauria de ser llegit en termes d'associació forta entre les variables. En ciències socials no és habitual trobar valors molt elevats d'aquest indicador, de manera que superar el valor de 0,3 ja es considera una intensitat elevada.

Interpretació de la V de Cramer

Valor	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Associació	Nul·la	Baixa	Moderada	Elevada							

Per a demanar a *jamovi* taules de contingència amb indicació de l'indicador V de Cramer:

1. Clica la taula de freqüències anterior, que analitza la possible relació entre ESTATUS i Ús d'internet (p3). Al panell de definició de la taula de contingència, clica el desplegable de les opcions > Estadístiques.
2. A l'apartat "Nominal", activa l'opció "[ ] Phi i V de Cramer".



Imatge 110. Càlcul de V de Cramer amb jamovi.

El resultat és 0,169. Per tant hi ha una intensitat "Baixa-Moderada" de l'associació entre les variables. Mentre aquest valor no sigui inferior a 0,1 es pot continuar amb l'anàlisi.

En cas que la intensitat sigui pràcticament nul·la ($V_{cramer} < 0{,}1$) no cal continuar amb l'anàlisi.

D) Avaluació del sentit de l'associació

Si es comprova que es compleixen les condicions d'aplicació del test de la khi quadrat (A), i si el test indica significació estadística (B), i si l'indicador V mostra associació moderada o elevada entre les dues variables analitzades (C), llavors es pot passar a la fase final de l'anàlisi.

Abans, en fer la representació gràfica de la taula de contingència sobre *Ús d'Internet segons ESTATUS socioeconòmic*, s'ha pogut veure com a major estatus es dona un major ús d'Internet, i contràriament a menor estatus menor ús d'Internet.

En funció del coneixement teòric, del model d'anàlisi previ, i de les hipòtesis explicatives que han inspirat l'atenció a aquestes dues variables, es pot considerar que, més enllà de la simple associació estadística entre les variables hi ha una certa relació de determinació entre elles. La diferència és que l'associació no pressuposa que una de les variables incideixi o condicioni (o sigui la causa de) l'altra, contràriament la incidència pressuposa un "sentit" o una "direcció" concreta de l'explicació.

En sentit estricte, que una variable determini l'altra, vol dir que en algunes categories socioeconòmiques tothom fa servir Internet mentre que en altres no hi ha ningú que es connecti. En aquest cas, podem suposar que la posició d'ESTATUS incideix en l'ús d'Internet (i no a l'inrevés, el fet de començar o deixar d'usar Internet no condiciona un canvi de classe social).

Aquesta suposició és el fonament de considerar la variable "Ús d'Internet" com a DEPENDENT de l'altra, i la variable ESTATUS com a INDEPENDENT (o causal) en aquest model explicatiu bivariable.

(cc) BY-NC

Si existeixi justificació teòrica, com la que s'acaba d'apuntar, es pot procedir a analitzar amb *jamovi* el sentit de la incidència. Per això es pot utilitzar qualsevol de dues alternatives:

A. una anàlisi del percentatges (disponible als menús de *jamovi*), o
B. una anàlisi de residuals (disponible executant R a la consola de *jamovi*).

(A) Fer una **anàlisi dels percentatges** d'una taula de contingència a la qual s'ha identificat una variable com a dependent i una altra como a independent implica:

- Demanar una taula creuada de percentatges, calculats en el sentit que la variable independent (si aquesta variable surt a la taula en sentit vertical, en columnes, calcular els % per a cada columna), i a continuació,
- Aquests percentatges s'han de comparar en el sentit de la variable dependent (si aquesta variable surt a la taula en sentit horitzontal, en files, comparar els % per a cada fila).

Per demanar una taula creuada de percentatges amb *jamovi*, cal:

1. Selecciona el procediment de Taules per a mostres Independents, a menú Anàlisis / Freqüències / Taules de Contingència / Mostres Independents – Prova d'associació χ^2 .
2. Selecciona la variable DEPENDENT (p3) i mou-la a l'espai dret per a Files.
3. Selecciona la variable INDEPENDENT (ESTATUS) i mou-la a l'espai dret per a Columnes.
4. Clica per a desplegar l'apartat > Cel·les.
5. Com que la variable independent (ESTATUS) ara està en el sentit de les COLUMNES, cal activar l'opció de Percentatges per [ ] Columna.
6. Per tenir una visió més clara, desactiva tota la resta d'opcions.
7. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Taules de Contingència

		ESTATUS					
p3		**Alta/mitj-alta**	**Nova classe mitj**	**Vella classe mitj**	**Obr qua-lif**	**Obr no qualif**	**Total**
Si	% de co-lumna	98.6 %	94.9 %	93.3 %	90.2 %	83.1 %	92.0 %
No	% de co-lumna	1.4 %	5.1 %	6.7 %	9.8 %	16.9 %	8.0 %
To-tal	% de co-lumna	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %

Regla del "creuament": Si la variable independent (possible causa) està a les columnes els percentatges es calculen per columnes i es comparen per a cada fila de la variable dependent (possible efecte). Contràriament, si la variable independent (possible causa) està a les files es procedeix a l'inrevés: càlcul per files i comparació per columnes.

En aquest cas pots comprovar a la taula de percentatges que, per exemple, qui USA INTERNET (fila "Si") té una distribució que va minvant des del màxim (Classe Alta / Mitjana-Alta: 98,6%) fins al

mínim (Obrers no qualificats: 83,1%) i és un decreixement constant a tot el rang de la variable ESTATUS (98,6 > 94.9 > 93,3 > 90,2 > 83,1). Això mostra clarament el patró d'intensitat d'ús d'Internet dependent de la pertinença a una classe socioeconòmica.

(B.1) Fer una **anàlisi NUMÉRICA dels residus ajustats** d'una taula de contingència a la qual s'ha identificat una variable com a dependent i una altra com a independent implica:

- Demanar a R una taula de residus ajustats de la variable dependent en funció de la variable independent, i a continuació,
- Detectar visualment les desviacions de les freqüències esperades (*residus*) més elevades en nombres absoluts, per a identificar **QUINES categories** contribueixen més a la determinació d'una variable per l'altra, així com analitzar el patró de creixement/decreixement dels residuals.

Per demanar a R una taula de residus ajustats amb *jamovi*, cal:

1. Poder executar codi R amb l'extensió (Rj) de *jamovi* que implica tenir instal·lat el mòdul o extensió de jamovi "Rj Editor".[25]
2. A l'apartat Anàlisis, clica a la icona "R". Al desplegable, seleccionar l'opció "Rj Editor",
3. Escriu, a poc a poc, les ordres per a generar la taula de residus ajustats amb les variables p3 i ESTATUS a la consola. Recorda, si fas retalla/enganxa possiblement hi haurà errors; cal escriure tota l'ordre:

```
taula1<-table(data$p3,data$ESTATUS)
taula2<-chisq.test(taula1)
taula3<-taula2$stdres
taula3
```

4. Aquestes simples ordres que has escrit en llenguatge R signifiquen:
   - *"Genera un objecte anomenat taula1 que conté la taula de freqüències observades del creuament de la variable p3 a les files i l'ESTATUS a les columnes",*
   - *"Genera un objecte anomenat taula2 què conté la taula de RESIDUS a partir de les dades de l'objecte taula2",*
   - *"Genera un objecte anomenat taula3 què conté la taula de residus AJUSTATS a partir de les dades de l'objecte taula2",*
   - *"Mostra l'objecte anomenat taula3".*
5. Per a executar el codi que hi ha escrit a la consola cal clicar al petit triangle verd de la part superior dreta de la consola. Alternativament, també es pot executar picant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho.
6. Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.
7. Per tancar el panell central, igual que en el cas de qualsevol altre procediment d'anàlisis de jamovi, cal clicar a la fletxa encerclada de la part superior dreta.
8. Recorda desar l'arxiu jamovi freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

25 Al tutorial Pràctica 1 es pot trobar la forma d'instal·lar aquest i d'altres mòduls o extensions per augmentar les prestacions de jamovi.

(cc) BY-NC

El resultat és la següent taula de residus ajustats (o "corregits"):

	Alta/mitj-alta	Nova classe mitj	Vella classe mitj	Obr qualif	Obr no qualif
Si	3.9978	1.8217	0.8111	-1.7588	-4.7629
No	-3.9978	-1.8217	-0.8111	1.7588	4.7629

A la taula de residus ajustats es pot veure, per a cada cel·la o creuament de la taula, la diferència entre les freqüències observades i les esperades en condició d'independència. Aquests valors estan sotmesos a una transformació matemàtica que fa que segueixen una distribució normal, de mitjana 0 i variància 1. Els residuals ajustats permeten saber, per a cadascuna de les cel·les, si hi ha un nombre de casos observats significativament[26] més gran (>1,96) o més baix (<1,96) que d'esperats en condicions d'independència de les variables. Això permet identificar l'existència de categories que *"contribueixen"* en una mesura elevada a la diferència i, per tant, a la dependència entre les variables.

En aquest cas, pots veure com a la cel·la estatus Alt/Mitjà-Alt i NO ús d'Internet, hi ha un residual negatiu i molt significatiu (3,9978). Això vol dir que en aquesta cel·la s'hi acumulen menys casos dels esperats, de manera que els casos realment observats d'estatus Alt/Mitjà-Alt tenen un **major** ús d'Internet que el conjunt. Correlativament, a la part superior del gràfic es pot veure que hi ha més casos dels esperats (hi ha més gent d'estatus Alt/Mitjà-Alt que fa ús d'Internet). En l'extrem contrari es dona una situació simètrica, és a dir una major acumulació d'individus d'estatus Obrer no qualificat que NO utilitzen Internet (4,7629). Per altra banda, es pot observar com existeix un patró de "creixement" lineal que permet concloure que el nivell d'estatus socioeconòmic incideix de manera directa en el nivell d'ús d'Internet, de manera que a major estatus, major ús d'Internet.

(B.2) Fer una **anàlisi GRÀFICA dels residuals** d'una taula de contingència a la qual s'ha identificat una variable com a dependent i una altra com a independent implica:

- Demanar a R un gràfic d'associació de la variable dependent en funció de la variable independent, i a continuació,
- Detectar visualment (per colors extrems: vermell i blau) les desviacions de les freqüències esperades (*residuals*) més elevades, per a identificar **QUINES categories** contribueixen més a la determinació d'una variable per l'altra, així com analitzar el patró de creixement/decreixement dels residuals.

Per demanar a R un gràfic d'associació i residuals amb *jamovi*, cal:

1. Poder executar codi R amb l'extensió (Rj) de jamovi que implica tenir instal·lat el mòdul o extensió de jamovi "Rj Editor".[27]
2. A l'apartat Anàlisis, clica la icona "R". Al desplegable, seleccionar l'opció "Rj Editor",
3. **OPCIÓ A:** Escriu, a poc a poc, l'ordre per a generar el gràfic d'associació amb les variables p3 i ESTATUS a la consola. Recorda, si fas retalla/enganxa possiblement hi haurà errors; cal escriure tota la ordre:

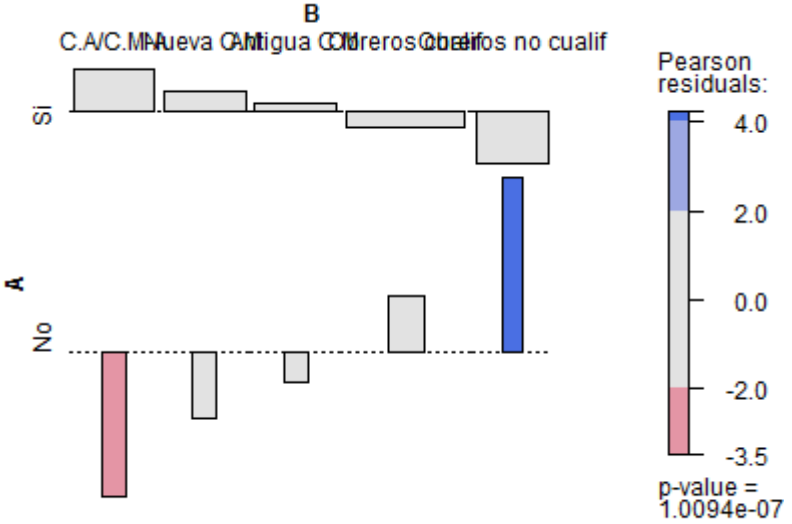
```
library(vcd)
assoc(table(data$p3,data$ESTATUS), shade=TRUE)
```

[26] Si es el cas que s'està treballant amb un nivell de confiança del 95%, que es correspon a un valor Z de 1,96.
[27] Al tutorial Pràctica 1 es pot trobar la forma d'instal·lar aquest i d'altres mòduls o extensions per augmentar les prestacions de *jamovi*.

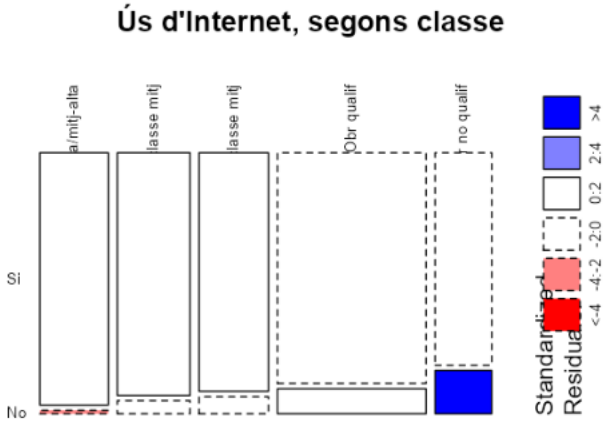
4. **OPCIÓ B:** Escriu, a poc a poc, l'ordre per a generar el gràfic de mosaic amb les variables p3 i ESTATUS a la consola. Recorda, si fas retalla/enganxa possiblement hi haurà errors; cal escriure tota la ordre:

```
library("graphics")
mosaicplot(table(data$p3,data$ESTATUS), shade = TRUE, las=2,
           main="Ús Internet, segons classe")
```

5. Aquestes simples ordres que has escrit en llenguatge R signifiquen:
   - *"Activa a R la possibilitat de fer gràfics i càlculs per analitzar variables categòriques", i*
   - *"Mostra un gràfic amb la variables p3 com a dependent i la variable ESTATUS com a independent o causa, acoloreix els valors residuals segons la seva desviació a la freqüència esperada".*
6. Per a executar el codi que hi ha escrit a la consola cal clicar al petit triangle verd de la part superior dreta de la consola. Alternativament, també es pot executar picant alhora tres tecles: [Ctrl] + [Majúscules] + [Intro]. Fes-ho.
7. Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.
8. Per tancar el panell central, igual que en el cas de qualsevol altre procediment d'anàlisi de *jamovi*, cal clicar a la fletxa encerclada de la part superior dreta.
9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.



Imatge 111. Representació gràfica (associació) dels residuals d'una taula de contingència.



Imatge 112. Representació gràfica (mosaic) dels residuals d'una taula de contingència

En aquestes does representacions gràfiques alternatives es pot veure, en horitzontal, les dues categories de la variable dependent (Sí i No) amb una línia discontínua on coincidirien les freqüències observades i les esperades en condicions d'independència de les variables. A partir d'aquesta línia base, en vertical es mostren columnes per a cada categoria de la variable independent (ESTATUS), que representen **els residuals**.[28]

A més les columnes estan acolorides (seguint una escala que es presenta a la dreta) segons el grau o intensitat de desviació de casos observats representat en unitats de desviació estàndard. Destaquen especialment les categories que es desvien "significativament" més de 2 desviacions típiques per sobre (blau) o per sota (vermell). Això permet identificar visualment l'existència de categories que "contribueixen" en una mesura elevada a la diferència i, per tant, a la dependència entre les variables.

En aquest cas, pots veure com en la cel·la estatus Alt/Mitjà-Alt i NO ús d'Internet, hi ha un residual negatiu i molt significatiu (vermell). Això vol dir que en aquesta cel·la s'hi acumulen menys casos dels esperats, de manera que els casos realment observats d'estatus Alt/Mitjà-Alt tenen un **major** ús d'Internet que el conjunt. Correlativament, a la part superior del gràfic es pot veure que hi ha més casos dels esperats (hi ha més gent d'estatus Alt/Mitjà-Alt que fa ús d'Internet)

En l'extrem contrari es dona una situació simètrica, és a dir, una major acumulació d'individus d'estatus Obrer no qualificat que NO utilitzen Internet (blau). Per altra banda, es pot observar com existeix un patró de "creixement" lineal que permet concloure que el nivell d'estatus socioeconòmic incideix en el nivell d'ús d'Internet de manera directa: a major estatus, major ús d'Internet.

Solució als problemes d'incompliment de les condicions d'aplicació.

Les freqüències esperades no apareixen per defecte a les taules de contingència generades per *jamovi*. De manera que és imprescindible, abans de fer el test, demanar que es mostrin les freqüències esperades, per tal d'avaluar si hi pot haver algun problema de violació de les condicions d'aplicació. Per exemple, si volem analitzar la incidència del nivell d'estudis (ESTUDIS) en el fet d'utilitzar Internet com a font d'informació (p20), cal demanar prèviament una taula de freqüències esperades.

Fes el què s'ha indicat anteriorment a l'apartat "Les condicions d'aplicació en el test de la khi quadrat". El resultat ha de ser aquest:

Taules de Contingència

p20		ESTUDIS						Total
		Sense estudis	Prima-ria	Secun-daria	F.P.	Univers mitjos	Uni-vers sup	
Fonamental	Esperades	0.4056	222.7	110.7	102.2	50.29	35.69	522.0
Secund, import	Esperades	0.4406	241.9	120.3	111.0	54.63	38.77	567.0
Secund, poc import	Esperades	0.0963	52.9	26.3	24.3	11.95	8.48	124.0
No, font inform	Esperades	0.0575	31.6	15.7	14.5	7.13	5.06	74.0
Total	Esperades	1.0000	549.0	273.0	252.0	124.00	88.00	1287.0

[28] L'ordre R "assoc" només suporta el càlcul dels residus de Pearson (no estandarditzats, ni tipificats, ni corregits).

(cc) BY-NC

Hi ha una categoria d'estudis, els que no tenen estudis, que només té un individu. Sembla coherent pensar que no es puguin extrapolar diferències entre grups quan un d'aquests conté només un individu. Precisament el problema rau exclusivament en aquesta categoria, tal i com ens mostren les freqüències esperades: totes les d'aquesta categoria són clarament inferiors a cinc (0,4056, 0,4406, 0,0963 i 0,0575), mentre que no hi ha cap problema amb la resta de cel·les, en tant que en totes hi ha més de cinc casos esperats.

Quan ens trobem amb aquest problema la solució és **reduir la taula agrupant** la categoria que ens genera problemes amb la que té una "major proximitat" amb ella (des d'un punt de vista conceptual); en aquest cas amb els que tenen estudis de Primària. Un cop fet, totes les freqüències esperades han de ser superiors a 5 i ja podrem analitzar la taula tal i com hem fet en l'anterior apartat.

EXERCICI

1. Crea una nova variable, basada en la variable original ESTUDIS, mitjançant una **transformació** que recodifiqui els valors de forma que "Sense estudis" i "Primaria" s'unifiquin en una nova categoria "Prim. o menys", mentre que la resta de categories d'ESTUDIS es mantinguin igual a la nova variable.[29]
2. Demana la taula de contingència de freqüències esperades de p20 amb la variable ESTUDIS original i amb la variable recodificada.
3. En cas que es compleixin les condicions, continua amb l'anàlisi de la relació entre aquestes dues variables fins on puguis arribar (condicions? significació? intensitat? sentit?).

*La taula de contingència 2*2 (la correcció de Yates).*

En casos extrems, quan treballem amb mostres molt petites, ens podem trobar que, fins i tot agrupant al màxim, segueixin sense complir-se les condicions d'aplicació. Així, en taules 2*2 (*2 files x 2 columnes*), quan alguna cel·la esperada és inferior a 5, **però totes són superiors a 3**, s'aplica una correcció al test de la khi quadrat de Pearson anomenada *correcció de Yates*, i que *jamovi* anomena com a l'opció [ ] χ^2 *amb correcció de continuïtat*.

Anem a veure-ho amb un exemple extret d'un fitxer provinent del món de la criminologia, que només conté dues variables, el fitxer *MJDS_reincid.omv*, que trobaràs al campus. En aquest fitxer hi ha gravades dues variables: *faltes* i *delicte*. La primera fa referència ha haver tingut faltes de comportament durant la condemna (Si/No), mentre que la segona ens diu si ha reincidit o no, dos anys després de sortir de la presó.

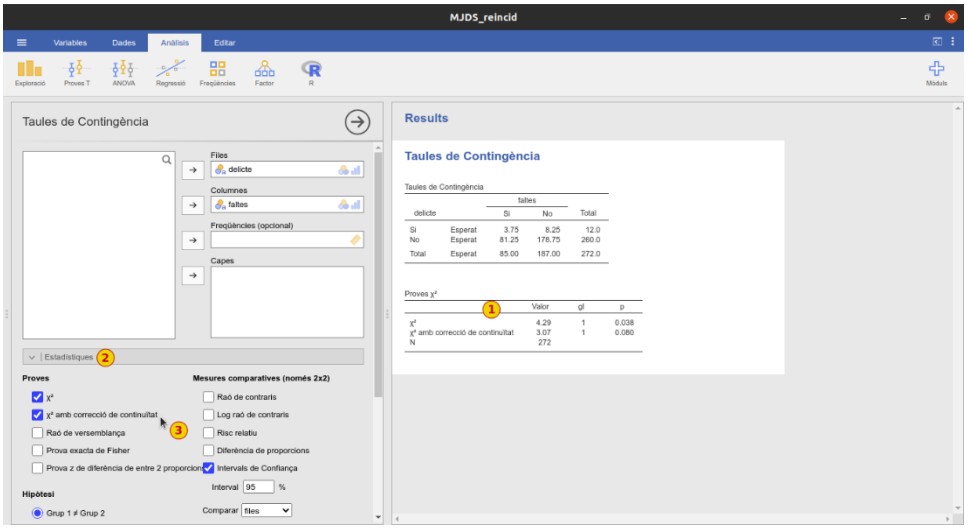
1. Comprova els tipus de variables i de dades: han de ser NOMINAL/TEXT. Si no és així, modifica la seva definició ("Setup").
2. Selecciona el procediment de Taules per a mostres Independents, a menú Anàlisis / Freqüencies / Taules de Contingència / Mostres Independents – Proves d'associació χ^2 .
3. Selecciona la variable "faltes" (independent) i mou-la a l'espai dret per a Columnes.
4. Selecciona la variable "delicte" (dependent) i mou-la a l'espai dret per a Files.
5. Clica per a desplegar l'apartat > Cel·les.
6. Activa l'opció de recompte [ ] Freqüències esperades.
7. Per tenir una visió més clara, desactiva tota la resta d'opcions de > Cel·les.
8. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

29 Les transformacions de variables i recodificació es van veure a a la pàgina núm 60 d'aquest tutorial.

Taules de Contingència

		faltes	
delicte		**Si**	**No**
Si	Esperat	3.75	8.25
No	Esperat	81.25	178.75
Total	Esperat	85.00	187.00

Comprova que es tracta d'una taula de 2*2 i que hi ha una cel·la amb una freqüència esperada inferior a 5, però superior a 3 (3,75), de manera que al moment de fer el test de khi quadrat cal demanar la correcció de Yates.



Imatge 113. Chi-quadrat amb correcció de Yates per a taules 2x2.

Per a demanar la correcció de Yates o correcció per continuïtat:

1. Clica la taula generada, al panell dret de resultats.
2. Al panell de definició del procediment (esquerra) desplega l'apartat > Estadístiques.
3. A l'opció de tests activa [ ] *χ^2 amb correcció de continuïtat*.
4. No desactivis la resta d'opcions de > Estadístiques.
5. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general ≡ / Desar.

Proves χ^2

	Valor	gl	p
χ^2	4.29	1	0.038
χ^2 amb correcció de continuïtat	3.07	1	0.080
N	272		

D'on concloem que **no** podem assegurar amb un 95% de confiança (p-valor = 0,080) que el fet d'haver comés faltes o no durant la condemna incideix en el risc de reincidir en sortir de presó.

(cc) BY-NC

Atenció! Si s'hagués utilitzat en aquest cas, incorrectament, el test "normal" de khi quadrat de Pearson que, com pots veure a la taula dona un p-valor de 0,038, s'hagués arribat a la conclusió errònia que SÍ que hi ha relació significativa, amb un 95% de confiança, entre les dues variables. Compte amb casos com aquest, on el criteri general del test khi quadrat (p-valor ≤ 0,05) es compleix o no en funció de si s'utilitza la fórmula de càlcul de la khi quadrat "normal" o la corregida per a taules 2*2.

Utilitzar l'indicador o procediment incorrecte ens pot portar a conclusions falses (i eventualment, com en aquest cas, pot tenir conseqüències sobre persones si s'apliquen les conclusions falses)

Molt important! Aquesta correcció de Yates només és aplicable en una taula 2*2. En altres casos, si no es compleixen les condicions d'aplicació, cal aconseguir que es compleixin agrupant alguna de les variables.

EXERCICIS PROPOSATS: Taules de contingència

Després de seguir aquest apartat del tutorial, sobre taules de contingència, obre en jamovi el fitxer *MJDS_Dades_def.omv* i respon les següents qüestions:

(A) A partir del creuament de la variable p16 i el sexe (p40), contesta aquestes tres preguntes.

a) Quin percentatge del total d'entrevistats és dona i considera molt probable que quedés amb algú que ha conegut a través d'Internet?
b) Quin percentatge dels que consideren gens probable que quedessin amb algú que han conegut a Internet, és dona?
c) Quin percentatge dels homes considera que no és gens probable que quedessin amb algú que han conegut a Internet?

Enganxa, en un processador de textos, el creuament així com la resposta a les tres qüestions.

(B) A partir dels estudis de l'entrevistat (ESTUDIS), genera una variable que tingui només tres categories: primària o menys, secundària (FP+Secundària) i Universitaris. A partir del creuament gràfic d'aquesta variable i la freqüència d'ús d Internet (p5) digues quin és l'ordre de major a menor ús d'Internet. És a dir digues qui són els que tenen un major ús d'Internet quins en segon lloc, i així successivament. Enganxa el gràfic corresponent en un processador de textos. Tingues en compte que per fer-ho hauràs d'eliminar els NA de les dues variables.

(C) Comprova a la mostra les diferències entre homes i dones en l'ús d'Internet (p3). Avalua amb un 95% de confiança, si aquestes diferencies es poden extrapolar a la població. En cas afirmatiu explica com són, representa-les gràficament, calcula la V de Cramer i analitza els residuals. Enganxa els principals resultats en un processador de textos i comenta'ls

(D) Volem fer el mateix amb la variable edat. Observa que en el fitxer en tenim dues, g_edat i p37. Quina de les dues faràs servir?

Repeteix el mateix procediment que en l'exercici anterior: digues si podem estar segurs que l'edat incideix en l'ús d'Internet i, en cas afirmatiu, explica les diferències representant-les gràficament, calcula la V de Cramer i analitza els residuals ajustats. Enganxa els principals resultats en un processador de textos i comenta'ls.

(E) Fes el mateix utilitzant com a variable independent la mida del municipi (TAMUNI) i estudia la seva relació amb la el fet d'utilitzar Internet com a principal font d'informació (p20). Observa que en aquest cas l'agrupació no és imprescindible. Calcula la V de Cramer i analitza els residuals. Enganxa els principals resultats en un processador de textos i comenta'ls.

(F) A partir de les preguntes 3 i 5 del qüestionari crea una variable sobre l'ús general d'Internet que tingui 4 categories: (1) no el fa servir mai o gairebé mai, (2) el fa servir alguna vegada, però no diàriament, (3) el fa servir un cop al dia, i (4) el fa servir més d'un cop al dia. Analitza, en primer lloc, la relació d'aquesta nova variable amb la nacionalitat (p63) arribant a totes les conclusions possibles.
Analitza-la també en funció de la relació amb l'activitat (p49) arribant a totes les conclusions que et permeti l'anàlisi.

(G) Volem saber si hi ha diferències entre homes i dones en l'ús que fan d'Internet per buscar parella o lligar. Com que no disposes de la variable l'hauràs de construir.[30] Per fer-ho, basa't en la pregunta 14 del qüestionari. Considera que els que no han fet servir les xarxes socials formaran part del grup que no les fa servir per lligar ni buscar parella.
Arriba al màxim de conclusions possibles en la relació entre aquesta variable i el sexe (p40).

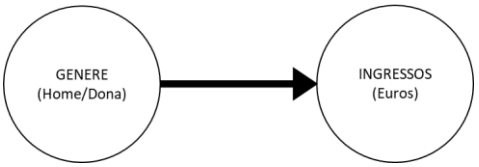
(H) Tal i com mostra la pregunta 21 del qüestionari un dels usos d'Internet és el d'obtenir informació sobre l'actualitat política. Digues si hi ha alguna relació entre aquesta variable i la ideologia política (p33), arribant a totes les conclusions possibles. Fes el test corresponent amb un 99% de confiança.

(I) Des de la *Consellería de Política Social* de la *Xunta de Galicia* ens transmeten la seva preocupació perquè creuen que entre els joves hi ha moltes diferències en l'ús d Internet (p3) en funció del sexe. A partir de la variable CCAA selecciona els residents a Galícia i comprova si tenen motius per preocupar-se o no.

[30] A la pregunta 14 es demana quin ús fan de les xarxes socials, i es recull la informació en tres variables p1401, p1402 i p1403. Observa que hi ha resultats que textualment diuen "Lligar" i "Buscar parella".

11. Comparació de mitjanes amb jamovi

En aquesta secció, seguint el Tutorial pas a pas, pots adquirir competència per a utilitzar el programari *jamovi* analitzant la relació entre una variable independent qualitativa i una variable dependent quantitativa (de la qual se'n pot calcular la mitjana) mitjançant els procediments o menús anomenats "Proves T" i "ANOVA unidireccional", així com per fer-ne la representació gràfica pertinent. També podràs interpretar la informació que s'obté de *jamovi* per tal de donar resposta a les hipòtesis de relació o associació entre aquestes dues variables d'acord amb un model d'anàlisi bàsic (aspecte descriptiu i de contingut de l'anàlisi), i determinar la significació d'aquesta informació per tal d'extrapolar-la al conjunt de la població (aspecte inferencial de l'anàlisi).



Imatge 114. Exemple de model d'anàlisi de comparació de mitjanes.

Per poder seguir el Tutorial pas a pas, en aquesta secció es treballarà amb el fitxer de dades *MJDS_Dades_def.omv* que és el mateix que en la pràctica anterior, i que es basa en el qüestionari *MJDS_Qüestionari.pdf*, que necessitareu per anar seguint la practica.

Per a disposar de la matriu de dades corresponent:

1. Localitza a l'espai Moodle de l'assignatura l'apartat corresponent a aquest apartat dedicat a COMPARACIÓ DE MITJANES.
2. Descarrega al teu PC els arxius necessaris: *MJDS_Qüestionari.pdf* i *MJDS_Dades_def.omv*.
3. Executa l'aplicació jamovi al teu ordinador.
4. Clica el menú general (dalt, esquerra).
5. Selecciona l'opció Obrir / Aquest PC. Al panell central, clicla [Navega] (dalt, dreta).
6. S'obre la finestra on has de seleccionar l'arxiu de dades a obrir. Tria l'arxiu *MJDS_Dades_def.omv*.

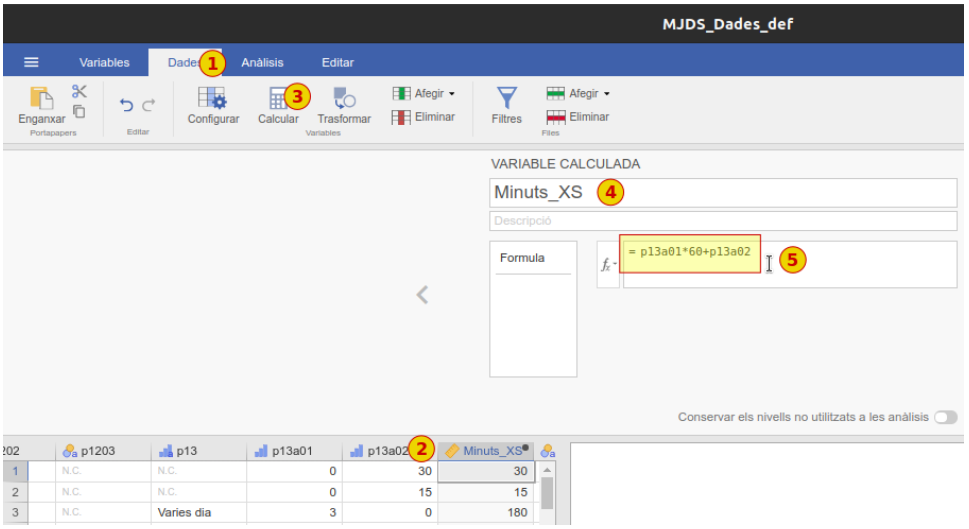
Recorda que, en funció del tipus de variable, *jamovi* permet la realització o no de determinats processos estadístics. Cal tenir definit de forma adient el tipus de variable per a poder realitzar els procediments d'anàlisi pertinents.

Com ja saps, en el moment de processar una ordre d'obertura d'un arxiu de dades que no sigui del seu propi format, l'aplicació *jamovi* prova "d'endevinar" el tipus de la variable a partir de les dades contingudes a l'arxiu. Aquesta identificació automàtica no sempre és l'òptima. Malgrat això, el nostre fitxer ha estat depurat correctament i les variables tenen el format adequat i els valors perduts s'han identificat correctament.

En aquesta pràctica, a banda de les variables que hi ha gravades a la matriu de dades, es treballarà amb la variable *Total de minuts a les xarxes socials*, que hauràs de construir prèviament (repassa els apartats anteriors, on s'especifica com definir una variable nova). Aquesta variable la construirem a partir de la p13a01 i de la p13a02 (hores:minuts de connexió a les xarxes socials).

Fes-ho de la següent manera:

1. Situa't a sobre de p13a02 i clica el menú Dades / Calcular.
2. Escriu un nom a la variable (per exemple, Minuts_XS) i defineix-la seguint l'operació aritmètica corresponent:



Imatge 115. Càlcul de nova variable a partir de valors de dades originals.

Introducció als tests de comparació de mitjanes. Quin test cal fer?

Seguint amb la idea general de la inferència estadística (avaluar fins a quin punt unes diferències que s'han observat en una mostra es poden atribuir a l'atzar o es poden inferir a la població), en el context de les anàlisis bivariables, quan parlem de diferències ens referim a diferències entre els grups que genera una variable respecte d'una altra. Per exemple, si disposem d'una mostra de dades de les variables *gènere* i *ingressos*, quan parlem de diferències ens referim a les possibles diferències entre la mitjana d'ingressos dels homes i la mitjana d'ingressos de les dones. En aquests casos, dir que hi ha diferències SIGNIFICATIVES entre homes i dones en els ingressos a la mostra permet parlar de la relació entre el sexe i la retribució econòmica per a la població en general.

Recorda! Parlar de relació entre dues variables comporta assignar a una de les variables el paper de variable independent (o explicativa) i a l'altra el de variable dependent (o resposta). La variable dependent és la que ens generarà els grups de població comparables.

L'objectiu principal d'aquest capítol és mostrar què hem de fer si volem comparar les **mitjanes d'una variable quantitativa** entre els diferents grups o **categories d'una variable qualitativa**. Ens trobem davant la possible relació entre dues variables, una de mètrica qualitativa, que farà el paper de variable independent (en *jamovi*, de tipus NOMINAL o ORDINAL, amb dades de TEXT), i una altra de quantitativa, la variable dependent (en *jamovi*, de tipus ORDINAL, amb dades ENTER o DECIMAL). De fet, la diferencia entre el test de la khi-quadrat i la comparació de mitjanes rau, precisament, en la mètrica de la variable dependent –qualitativa en el primer cas i quantitativa en el segon.

Tècnicament, en fer inferència sobre la comparació de mitjanes es distingeix en funció dels grups que genera la variable independent, dos o més de dos. En el primer cas ens trobem davant d'un Proves T de comparació de dues mitjanes independents, mentre que en el segon ens trobem davant d'una anàlisi de la variància.[1]

En conseqüència, la primera pregunta que ens haurem de fer si volem fer inferència sobre mitjanes és, quants grups genera la variable independent? Un cop resposta aquesta pregunta, en funció del

nombre de categories de la variable qualitativa considerada com a independent, tenim dues famílies de tests que comparen mitjanes:

1. Tests paramètrics
2. Tests no paramètrics

Tests paramètrics. Són els més recomanats, en tant que treballen amb la informació "complerta", directament amb els valors de la variable quantitativa. Ara bé, per a la seva aplicació tenen uns requisits que prèviament haurem de comprovar si es compleixen. El requisit es fonamenta en la mida de la mostra dels diferents grups i/o en la normalitat de les distribucions. Quan es parla de grups es refereix als que genera la variable qualitativa independent i que poden ser dos o més de dos. Per tal de poder fer un test paramètric la condició és que els grups han de ser prou grans o ajustar-se a una distribució normal.[2]

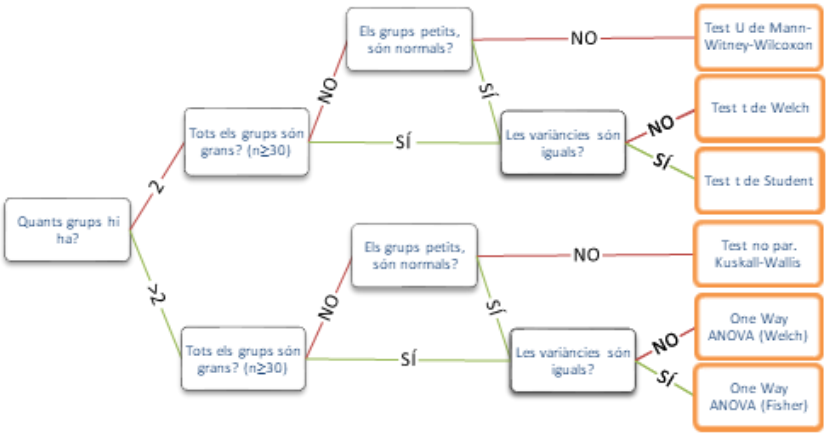
A banda d'aquesta qüestió, si ens centrem en la família de tests paramètrics, disposem de dues possibilitats més de test, en funció de com són les variàncies dels grups considerats:

1. Variàncies iguals ("homoscedasticitat"): Si en tots els grups hi ha la mateixa variància.
2. Variàncies diferents ("heteroscedasticitat"): Si algun dels grups presenta una variància diferent dels altres.

Tests no paramètrics. Tenen l'inconvenient que en la seva aplicació es perd molta informació, en tant que no treballen amb els valors quantitatius sinó que ho solen fer amb el rang, amb la posició que ocupen les observacions. En aquest sentit només els farem servir quan, per qüestions de mostra o normalitat, no sigui possible fer un test paramètric.

Observació! La majoria d'autors considera que si un grup té *30 casos o més* ja es pot considerar gran, de manera que si tots els grups tenen 30 observacions o més farem un test paramètric, sense que sigui necessari interrogar-nos sobre la possible normalitat. Ara bé, si algun/s grup/s no arriba/en a aquesta mida, per poder fer el test paramètric cal exigir la normalitat d'aquest/s grups/s; i, si no es compleix en algun d'ells, fer un test no paramètric.

Esquemàticament, les diferents situacions es poden expressar seguint aquest arbre de decisions:



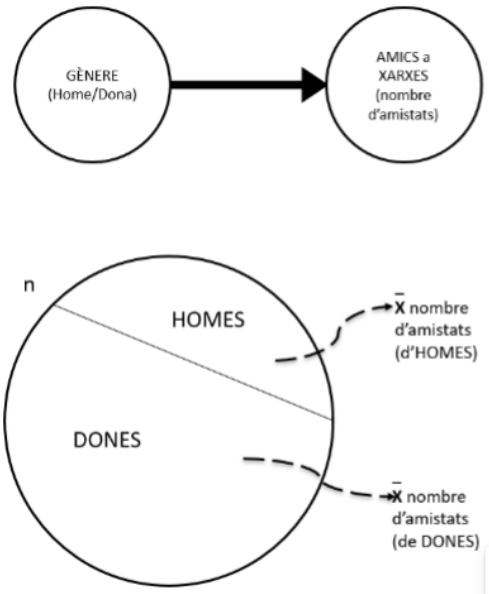
Imatge 116. Anàlisi de comparació de mitjanes: Selecció del test correcte

Així doncs, hi ha una seqüència de preguntes que són importants i que ens farem abans d'establir quin és el test més adequat:

1. Quants grups tenim, dos o més de dos?
2. Algun dels grups és petit?
3. Els grups petits, són normals?
4. Si tots els grups són grans o normals, hi ha igualtat de variàncies?

L'estadística descriptiva en la comparació de mitjanes i els intervals de confiança

En analitzar la relació entre una variable qualitativa i una de quantitativa pensem sempre que la primera, la qualitativa, és la independent i la quantitativa la dependent. Per exemple, es vol analitzar el nombre d'amics a les xarxes socials (p15) en funció del sexe (p40).

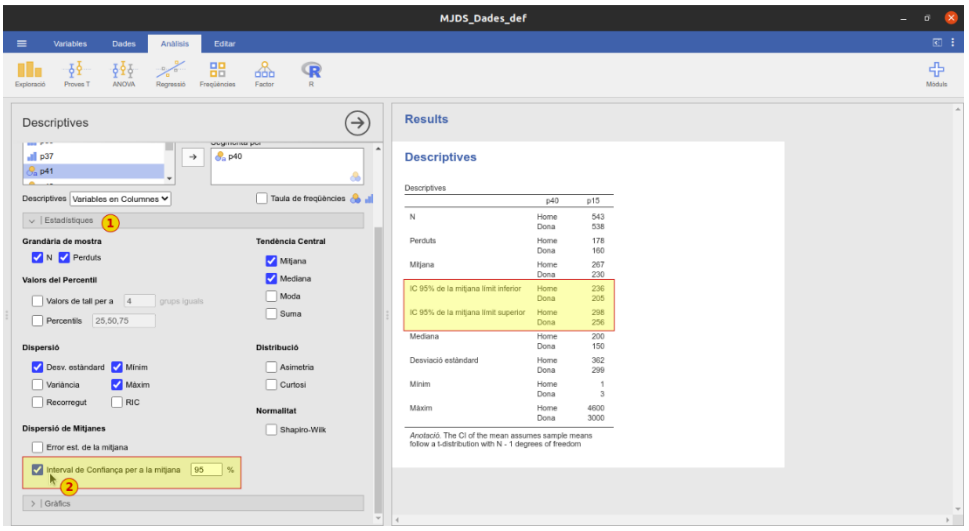


Imatge 117. Exemple d'anàlisi de comparació de 2 mitjanes, mostres independents.

Per descriure aquesta relació, és habitual comparar alguns estadístics descriptius per a cada grup que genera la variable independent. Per exemple, compararíem la mitjana d'amistats a les xarxes socials dels homes amb la mitjana per a les dones. Per fer-ho:

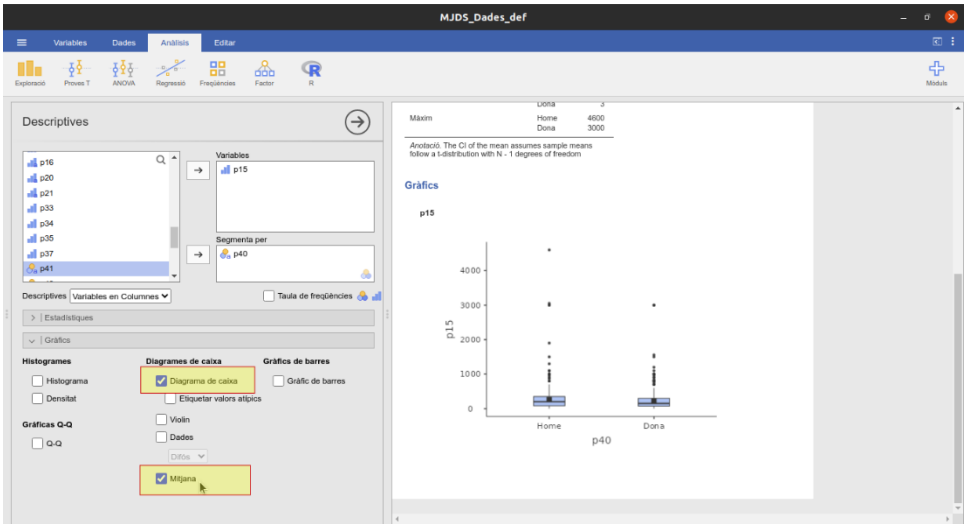
1. Comproveu el tipus de les variables i de dades: Contactes (p15) ha de ser ORDINAL i ENTER, i Sexe (p40) ha de ser NOMINAL i TEXT.
2. Cliqueu el menú Anàlisis / Exploració / Descriptives.
3. Situeu la variable quantitativa dependent (p15) a l'espai superior *Variables*.
4. Col·loqueu la variable qualitativa independent (p40) a l'espai inferior *Segmenta per*.
5. Per defecte ens dona els estadístics descriptius més comuns: la mitjana, la mediana, la desviació típica el mínim i el màxim, per a cada grup que genera la variable independent. Si te'n calen altres, cal que els activis.
6. Per altra banda, si voleu construir intervals de confiança per a la mitjana d'amistats dels homes i de les dones separadament, cal que activeu, a la pestanya > Estadístiques, l'opció

[ ] *Interval de confiança per la mitjana*. En aquesta mateixa pestanya teniu l'opció de canviar el nivell de confiança que, per defecte, és del 95%:



Imatge 118. Comparació descriptiva de dues mitjanes de mostres independents.

7. Si voleu visualitzar la representació gràfica que millor resumeix la relació entre ambdues variables ho haureu de fer mitjançant un gràfic de caixes múltiples, amb una caixa per a cada grup de la variable qualitativa. Per fer-ho, dintre de la pestanya **> Gràfics**, activeu a les opcions de Diagrama de caixa (centre) tant [ ] Diagrama de caixa com [ ] Mitjana:



Imatge 119. Diagrames de caixa per a dos grups independents.

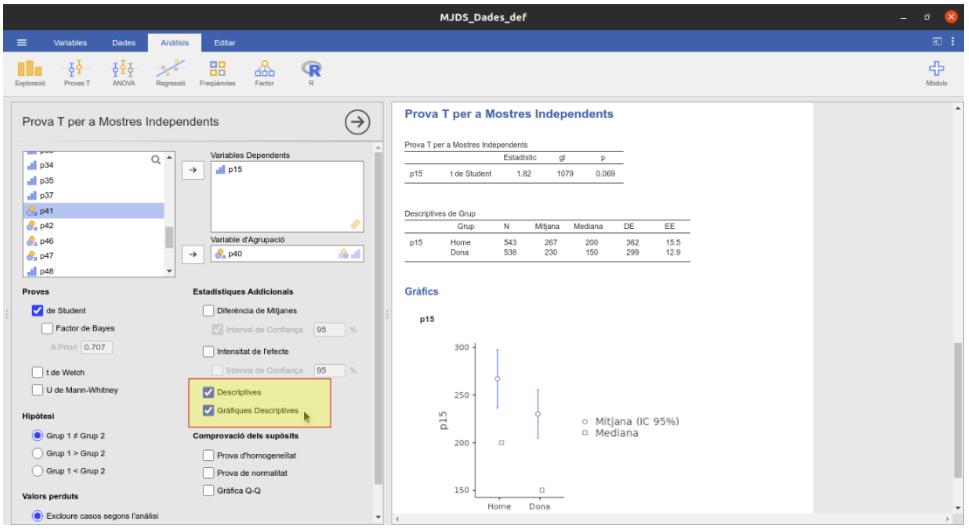
El resum gràfic de la informació ens mostra la posició de les medianes i dels dos quartils, juntament amb el rang interquartíl·lic. Per a interpretar-lo cal recordar (1) que el rectangle de la caixa està delimitat en alçada (eix Y) pel quartils 1r i 3r de la variable quantitativa (percentils 25 i 75) i inclou el 50% central de la distribució de casos; (2) que la línia horitzontal interior indica la mediana (percentil 50); (3) que els punts indiquen valors "desviats" (*outliers*) que queden fora de la centralitat de la distribució; i (4) que les corresponents mitjanes es representen com a petits requadres –habitualment, prop de la mediana-.

Per altra banda, si llegim els resultats numèrics veurem que, a la mostra, la mitjana de contactes a xarxes socials dels homes (267 amistats) és més alta que la de les dones (230 amistats). A més,

podem assegurar amb un 95% de confiança que la mitjana per als homes es troba entre les 237 i les 297 amistats, mentre que la mitjana per a les dones es troba entre les 205 i les 256 amistats.

Els intervals de confiança que acabem de construir, també els podem representar gràficament:

1. Cliqueu el menú Anàlisis / Proves T / Prova T per a Mostres Independents.
2. Situeu la variable quantitativa dependent (p15) a l'espai superior *Variables Dependents*.
3. Col·loqueu la variable qualitativa independent (p40) a l'espai inferior *Variable d'Agrupació*.
4. Podeu visualitzar alguns estadístics descriptius activant l'opció [ ] *Descriptives* a l'apartat **Estadístiques Addicionals**.
5. Per tal de visualitzar la representació gràfica dels intervals de confiança haurem d'activar l'opció [ ] *Gràfiques Descriptives* a l'apartat **Estadístiques Addicionals**.

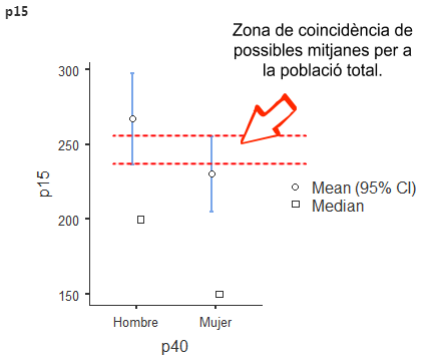


Imatge 120. Gràfic de BARRES D'ERROR per a comparació de dues mitjanes de mostres independents.

La variable independent està representada a l'eix horitzontal, com a dos grups en funció de les categories de p40, i la variable dependent està representada a l'eix vertical. Com indica la llegenda lateral del gràfic, els cercles mostren la posició de les mitjanes (*Mean*) de la variable p15 per a cada grup de p40, mentre que els requadres mostren les posicions de les medianes (*Median*). A més, unes línies de color blau s'estenen amunt i avall dibuixant el rang de l'interval de confiança, per a les mitjanes de cada grup.

La interpretació d'aquesta visualització gràfica és:

1. A la mostra obtinguda, hi ha evidència de diferències entre la mitjana de contactes a xarxes socials que tenen homes i dones, respectivament 267 i 230.
2. Però aquestes diferències no són prou elevades com per poder assegurar que hi ha diferències entre homes i dones al conjunt de la població, perquè els intervals de confiança en part se superposen.



Imatge 121. Solapament d'intervals de confiança.

A partir dels resultats d'aquest gràfic, que visualitzen els intervals de confiança construïts anteriorment, es podria donar el cas que la mitjana "real" d'amistats de la població d'homes fos 248 i la mitjana "real" d'amistats de la població de dones fos 252.

1. Identifica aquests dos punts al gràfic, tots dos situats dins dels respectius intervals de confiança.
2. Comprova que aquesta "realitat poblacional" dona una perspectiva diferent del resultat obtingut a la mostra.

Compte! Els valors extrems (*outliers*) molt elevats, tot i que siguin poc nombrosos, poden tenir un efecte considerable sobre les mitjanes corresponents. Els gràfics de caixa són una bona eina per a detectar els problemes d'aquest tipus (que cal corregir eliminant aquests valors), mentre que la representació gràfica dels intervals de confiança, en tant que concentren l'atenció en la part central de la distribució i no són idonis per a detectar aquests problemes de biaix.

El test de Shapiro-Wilk per a la Normalitat de la variable.

Tal i com s'ha esmentat anteriorment, per poder fer un test de tipus paramètric (que és el més recomanat) cal que les mostres siguin "suficientment grans", o bé que la distribució "no difereixi molt" d'una distribució normal. Quan no es compleix la condició, és a dir quan en algun grup poc nombrós no hi ha normalitat, és recomanable fer servir un test no paramètric.

Recordatori! La majoria d'autors considera que si els grups superen els 30 casos ja és suficient per a considerar-los grans. Així, només avaluarem la normalitat en els grups que no arribin a aquest nombre de casos.

Per avaluar la normalitat farem la prova de normalitat de *Shapiro-Wilk* que ofereix *jamovi*. Aquest test avalua les següents hipòtesis:

H0: la distribució és normal
H1: la distribució no és normal

S'interpreta de manera que si el *p-valor* és menor (o igual) a 0,05 considerem evidència de suport a H_1 (no hi ha normalitat), mentre que si és superior a 0,05 podem assumir la normalitat (amb un 95% de nivell de confiança).

Per exemple, si volem avaluar la possible relació entre la Comunitat Autònoma de residència (CCAA) i el temps (en minuts) dedicat a les xarxes socials (variable construïda anteriorment, Minuts_XS), el primer que faríem seria estimar la **mida de la mostra per a cada grup** que genera la variable Comunitat Autònoma. L'opció més raonable és fer-ho "dividint" la variable Minuts_XS amb les categories de la variable CCAA a partir del procediment *Descriptives*, tal i com hem vist en l'apartat anterior:

1. Cliqueu el menú Anàlisis / Exploració / Descriptives
2. Situeu la variable quantitativa dependent (Minuts_XS) a l'espai superior *Variables*.
3. Col·loqueu la variable qualitativa independent (CCAA) a l'espai inferior *Segmenta per*.
4. Com que només ens interessen els resultats sobre la mida dels grups (N), podem obtenir una taula més "neta" que la que apareix per defecte desplegant el menú d'opcions estadístiques [> Estadístiques] i desmarcant tots els resultats que no siguin N a l'apartat **Grandària de la mostra**. Així obtenim:

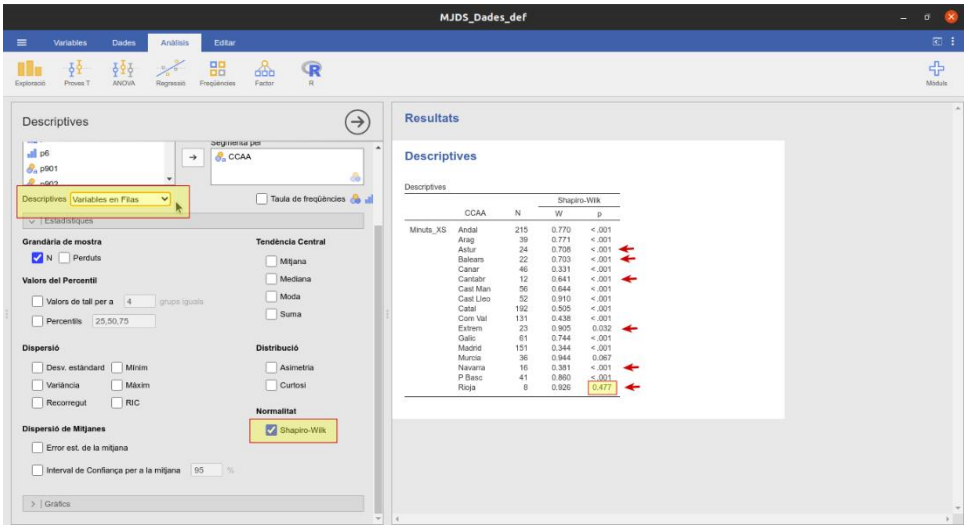
Descriptives

	CCAA	Minuts_XS
N	Andal	215
	Arag	39
	Astur	24
	Balears	22
	Canar	46
	Cantabr	12
	Cast Man	56
	Cast Lleo	52
	Catal	192
	Com Val	131
	Extrem	23
	Galic	61
	Madrid	151
	Murcia	36
	Navarra	16
	P Basc	41
	Rioja	8

Observa que hi ha diversos grups que són poc nombrosos, en tant que tenen menys de 30 observacions vàlides en la variable *Minuts_XS*. Són els casos d'Astúries, Balears, Cantàbria, Extremadura, Navarra i la Rioja.

Si es vol fer un test que avaluï si les diferències observades a la mostra són significatives, primer hem de saber si podem assumir que **aquests sis grups** segueixen una distribució normal. Recorda que si tant sols un d'aquests grups no segueix una distribució normal caldrà fer un test no paramètric, en tant que les condicions d'aplicació del test paramètric ens diuen que tots els grups han de ser grans o normals. Per avaluar la normalitat, farem:

1. Cliquem la taula resultat que acabem de generar.
2. Despleguem el menú d'opcions, clicant a sobre de > *Estadístiques*
3. A la part inferior dreta, a l'apartat **Normalidad**, activem l'opció *[ ] Shapiro-Wilk.*
4. Per a obtenir una taula més compacta, pots activar "Variables en files" a l'opció "Descriptives" que trobaràs just a sota de llistat de variables disponibles.



Imatge 122. Test de Normalitat Shapiro-Wilk

Observa que s'amplia la taula de resultats amb dues informacions més, l'estadístic de contrast utilitzat en fer el test de Shapiro-Wilk (a la taula, *Shapiro-Wilk W*) i el p-valor associat a aquest test (a la taula, *Shapiro-Wilk p*). Si observem els sis grups que ens generaven problemes, podem resumir la informació de la següent manera:

	N	p-valor$_{S\text{-}W}$
Astúries	24	<0,001
Balears	22	<0,001
Cantàbria	12	<0,001
Extremadura	23	=0,032
Navarra	16	<0,001
La Rioja	8	=0,477

Així doncs, dels sis grups petits, només en un podem assumir normalitat, el de la Rioja. En conseqüència, hauríem de fer el test no paramètric corresponent.

ATENCIÓ! El què succeeix en la prova de normalitat de Shapiro-Wilk amb els grups grans (n≥30) no és rellevant, atès que ja compleixen un dels dos supòsits per a poder fer el test paramètric.

En resum! En els test habituals (amb un nivell de confiança del 95% de confiança), si en la prova de Shapiro-Wilk observem que algun grup petit (n<30) té un p-valor$_{S\text{-}W}$≤0,05, conclourem que no hi ha normalitat i, en conseqüència, serà suficient per haver de fer un test no paramètric, independentment del què passi en la resta de grups.

La comparació de dues mitjanes independents.

Ens trobem en aquest cas quan tenim una variable qualitativa independent que genera **dos grups** i ens preguntem si hi ha diferències entre ambdós respecte d'una variable quantitativa. Per exemple,

conviure o no en parella, comporta diferències en el nombre d'amics que es tenen a les xarxes socials?, o bé, i respecte del sexe, el nombre d'amics que tenen homes i dones és diferent?

Recorda! Tots els test d'hipòtesis són test d'inferència: proven d'avaluar si, donades determinades diferències entre mitjanes de grups a una mostra, aquestes diferències són significatives, o prou grans per ser molt improbable que siguin degudes a l'atzar. Si es fa un test d'hipòtesis es pot afirmar (amb el nivell de confiança corresponent) que hi haurà diferències en el conjunt de la població i, per tant, es podrà afirmar que hi ha prou evidència estadística de la relació, d'associació o de determinació, entre les dues variables considerades.

Per a comparar la mitjana d'una variable resposta (el nombre d'amics a xarxes, p15) entre dos grups independents (conviuen o no en parella, p42) se sol fer servir la prova anomenada "T d'Student per a mostres independents". En aquest cas, el que volem contrastar es pot expressar en un test d'hipòtesis de diferents maneres, que venen a dir el mateix:

H_0: En la població, la mitjana d'amics a la xarxa social dels que conviuen en parella és la mateixa que la mitjana dels que no hi conviuen: $\mu_{si} = \mu_{no}$

H_1: En la població, la mitjana d'amics a la xarxa social dels que conviuen en parella és diferent que la mitjana dels que no hi conviuen: $\mu_{si} = \mu_{no}$

H_0: En la població, el fet de conviure o no en parella no incideix en el nombre d'amics a les xarxes socials.

H_1: En la població, el fet de conviure o no en parella incideix en el nombre d'amics a les xarxes socials.

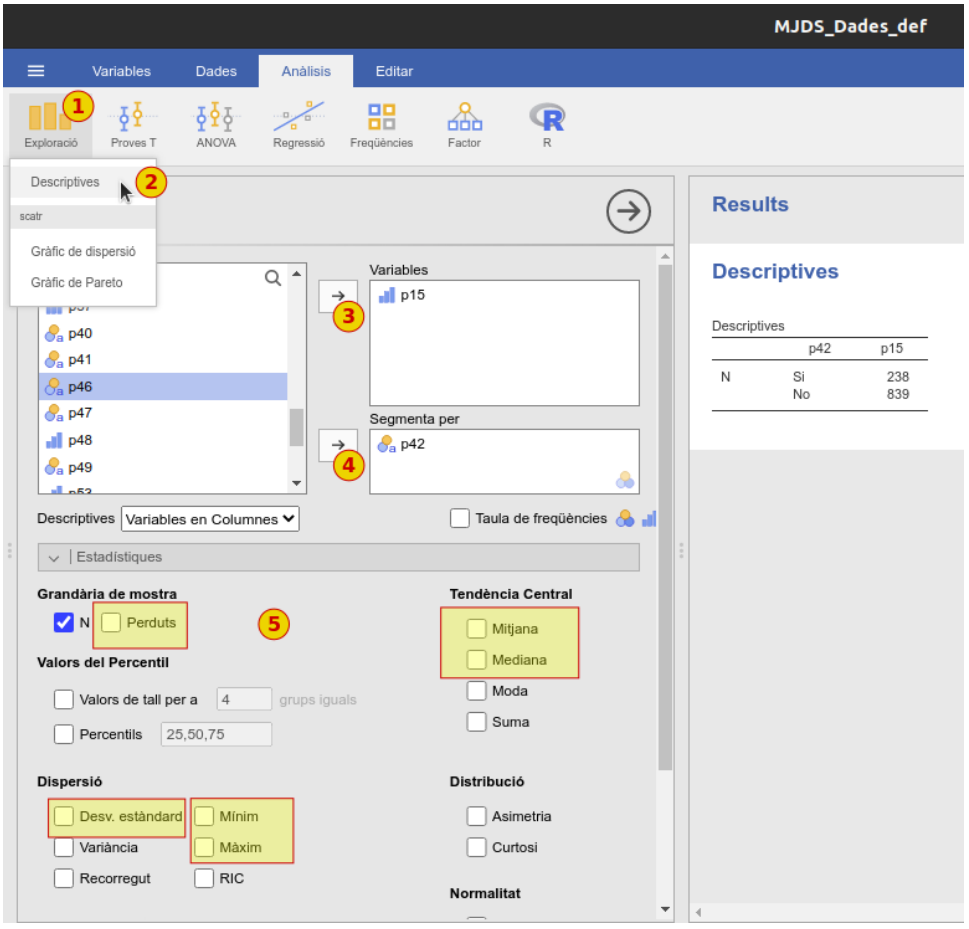
H_0: En la població, no hi ha relació entre el fet de conviure o no en parella i el nombre d'amics a les xarxes socials.

H_1: En la població, hi ha relació entre el fet de conviure o no en parella i el nombre d'amics a les xarxes socials.

Si hi ha evidència estadística per a la primera hipòtesis (H_0) és que la situació de convivència **no** comporta diferències en el nombre d'amics a les xarxes; per tant, **no hi ha relació** entre la forma de convivència i el nombre d'amics. Per altra banda, si hi ha evidència estadística per a la segona hipòtesis (H_1) és que la situació de convivència **sí** que comporta diferències en el nombre d'amics a les xarxes; per tant, **hi ha relació** entre la forma de convivència i el nombre d'amics.

El que cal fer, en primer lloc, és avaluar si la mostra és suficientment gran per poder fer un test paramètric o si cal comprovar la normalitat d'algun dels dos grups. Per avaluar-ho utilitzarem el procediment que hem fet servir abans:

1. Clica el menú Anàlisis / Exploració / Descriptives…
2. Situa la variable quantitativa dependent (p15) a l'espai superior *Variables*.
3. Col·loca la variable qualitativa independent (p42) a l'espai inferior *Segmenta per*.
4. Com que només ens interessen els resultats sobre la mida dels grups (N), podem obtenir una taula més "neta" que la que apareix per defecte desplegant el menú d'opcions estadístiques [> Estadístiques] i desmarcant tots els resultats que no siguin N a l'apartat **Grandària de mostra**.



Imatge 123. Estimació de la mida dels grups o mostres per comparar 2 mitjanes.

Veiem que els dos grups són grans (238 i 839), de manera que podem prescindir de l'anàlisi de la normalitat, i passar directament a fer un test paramètric.

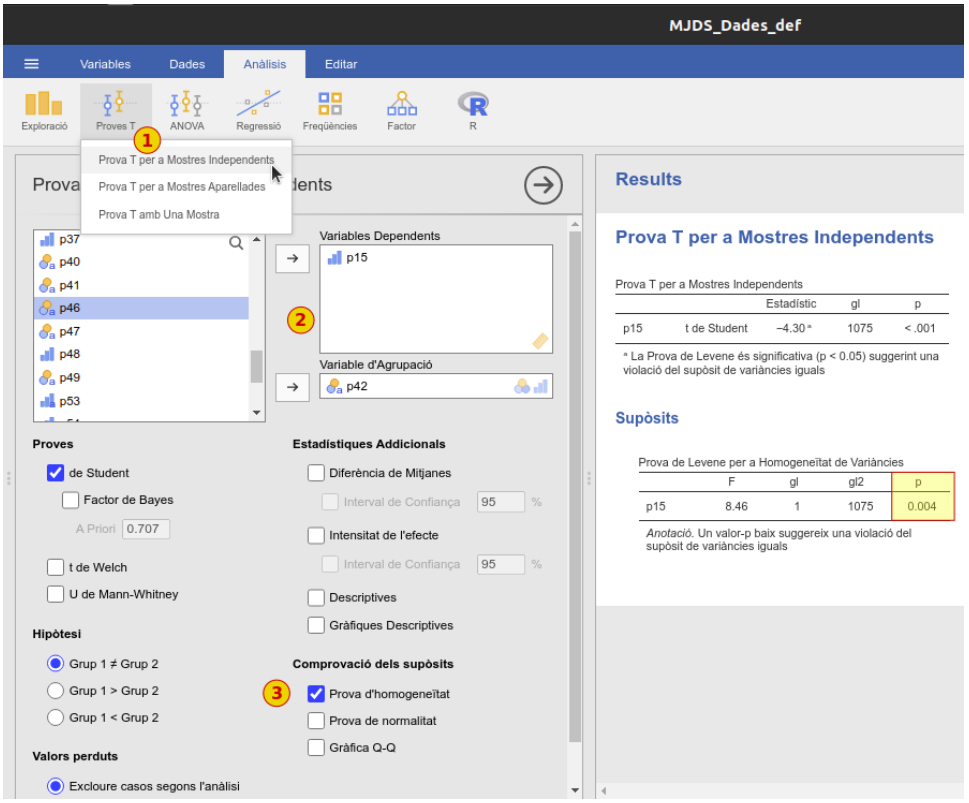
Tal i com s'esmenta en la introducció, hi ha dues maneres diferents de fer aquesta comparació paramètrica: (1) quan les variàncies són iguals i (2) quan no ho són; de manera que en primer lloc haurem de comprovar si es dona la condició d'aquesta igualtat entre variàncies (homoscedasticitat). Per fer-ho durem a terme un primer test que ens resolgui aquesta qüestió, una prova que s'anomena ***test de Levene*** per a la igualtat de variàncies.

En aquest test de Levene les hipòtesis que plantejarem seran:

H_0: $\sigma^2_{sí} = \sigma^2_{no}$ (les dues variàncies són iguals)
H_1: $\sigma^2_{sí} \neq \sigma^2_{no}$ (les dues variàncies són diferents)

Per fer el test de comparació de dues variàncies (Levene) amb *jamovi*:

1. Cliqueu el menú Anàlisis / Proves T / Prova T per a Mostres Independents.
2. Situeu la variable quantitativa dependent (p15) a l'espai superior *Variables Dependents*.
3. Col·loqueu la variable qualitativa independent (p42) a l'espai inferior *Variable d'Agrupació*.
4. Per defecte es mostren resultats del test t de Student. A l'apartat inferior dreta **Comprovació del supósits** cal activar l'opció [ ] *Prova d'homogeneïtat*.



Imatge 124. Test de Levene, per a comprovar igualtat de variàncies.

De la taula resultant només n'interpretarem el p-valor. En aquest cas pren el valor de 0,004, de manera que hi ha prou evidencia estadística per a concloure que **les variàncies siguin diferents**, i així li ho haurem d'indicar quan fem el test de comparació de mitjanes.

Compte! No perdem de vista que aquests tests (normalitat i homoscedasticitat) no ens informen de la relació entre les dues variables, sinó que són tests previs per a poder saber quin test és el més adequat.

Un cop assumit que (1) les mostres són grans, i (2) que les variàncies són diferents *(heteroscedasticitat)* ja podem fer el test corresponent de comparació de dues mitjanes independents : el Test de Welch:[3]

1. Cliqueu el menú Anàlisis / Proves T / Prova T per a Mostres Independents.
2. Situeu la variable quantitativa dependent (p15) a l'espai superior *Variables Dependents*.
3. Col·loqueu la variable qualitativa independent (p42) a l'espai inferior *Variable d'Agrupació*.
4. Per defecte es mostren resultats del test t de Student. A l'apartat superior esquerra **Proves** cal activar l'opció [ ] *t de Welch*. Per a obtenir una taula més "neta" es pot desactivar l'opció [ ] *de Student*, o deixar-la activada donat que genera informació rellevant sobre el test de Levene.
5. Per tal d'obtenir informació més complerta és convenient activar també l'opció *Descriptives* a l'apartat **Estadístiques Addicionals**.

Hem obtingut el següent:

Prova T per a Mostres Independents

		Estadístic	gl	p
p15	T de Welch	-4.94	481	< .001

Descriptives de Grup

	Grup	N	Mitjana	Mediana	DE	EE
p15	Si	238	167	100	268	17.4
	No	839	271	200	345	11.9

Pots comprovar a la taula de DESCRIPTIVES que els que conviuen en parella tenen menys amics (en mitjana 167) que els que no hi conviuen (271). De fet, la pregunta que formula el test és si aquesta diferencia és suficient per a poder assegurar, amb un 95% de confiança, que en la població el nombre d'amics d'uns i altres difereix. El resultat del *test t de Welch*, indica que el p-valor (< 0.001) és clarament inferior al llindar de significació establert, en tant que fer el test amb un nivell de confiança del 95% és equivalent a acceptar una significació màxima del 5%.

La decisió estadística haurà de ser que hi ha suficient evidencia estadística (p-valor<0,001) per afirmar que el nombre d'amics **és diferent** en funció de si es viu en parella o no, i, com a conclusió, que no viure en parella comporta un nombre superior d'amics a les xarxes socials.

Recordatori! En tots els tests fets amb un nivell de confiança del 95%, si observem un p-valor≤0,05 ens quedem amb la hipòtesi alternativa, mentre si el p-valor>0,05 ens quedem amb la nul·la. A més, la hipòtesi nul·la sempre ha d'estar formulada en termes d'igualtat (normalitat, homoscedasticitat, igualtat de mitjanes) mentre que l'alternativa en termes de diferència (no normal, heteroscedasticitat, diferència de mitjanes).

Comprova a l'apartat **Proves** del panell de definició del procediment *Test T per a Mostres Independents* que hi ha tres diferents opcions de càlcul de test, cadascuna d'elles adient per a un tipus de situació:

- **U de Mann-Whitney**: Test no paramètric. L'utilitzarem si algun dels dos grups és petit i no normal.
- **t de Student**: Test paramètric en cas d'homoscedasticitat (per defecte a *jamovi*). L'utilitzarem si els dos grups són grans o bé normals i, a més, es compleix la igualtat de variàncies (avaluada a partir del test de Levene).
- **t de Welch**: Test paramètric en cas d'heteroscedasticitat. L'utilitzarem si els dos grups són grans o bé normals i, en el test de Levene, hem obtingut que les variàncies són diferents.

Ves molt en compte amb... Possiblement te n'hauràs adonat que a la finestra corresponent al *Test T per a Mostres Independents* hi ha una opció que diu [ ] *Prova de normalitat*. Tingues molt present que la normalitat per a cada grup no la pots generar a partir d'aquesta pestanya, sinó tal i com hem fet abans, utilitzant el menú Anàlisis / Exploració / Descriptives.

La pestanya *Prova de normalitat* que veiem aquí, es tracta d'una normalitat per al conjunt de la variable, i no per a cadascun dels grups que genera la variable independent (que és el que nosaltres necessitem).

La comparació de més de dues mitjanes independents: l'anàlisi de la variància (ANOVA).

La pregunta que ens formularem en aquest cas és molt similar a l'anterior. L'únic que canvia és que la variable explicativa qualitativa te més de dues categories i, per tant, genera més de dos grups a comparar. En aquest cas ens preguntem si el temps diari dedicat a les xarxes socials (Minuts_XS) difereix segons l'edat, que ha estat agrupada en **tres** categories (g_edat).

En tant que volem comparar una variable resposta (el temps a les xarxes socials) dependent entre més de dos grups independents (15-19 anys, 20-24 anys i 25-29 anys) hem de fer servir la prova anomenada com a anàlisi de la variància[4] (que utilitza les sumes de quadrats, que en definitiva són els variàncies, per a comparar mitjanes). En aquest cas, el que volem contrastar ho podem expressar en forma de test d'hipòtesis plantejant les hipòtesis de tres maneres diferents:

H_0: En la població, la mitjana de temps dedicat a les xarxes socials és la mateixa en els tres grups d'edat: $\mu_{15\text{-}19} = \mu_{20\text{-}24} = \mu_{25\text{-}29}$

H_1: En la població, la mitjana de temps dedicat a les xarxes socials no és la mateixa en els tres grups d'edat, sinó que alguna mitjana difereix de les altres: $\mu_i \neq \mu_j$, per algun i, j

H_0: En la població, l'edat no incideix en el temps dedicat a les xarxes socials.

H_1: En la població, l'edat incideix en el temps dedicat a les xarxes socials.

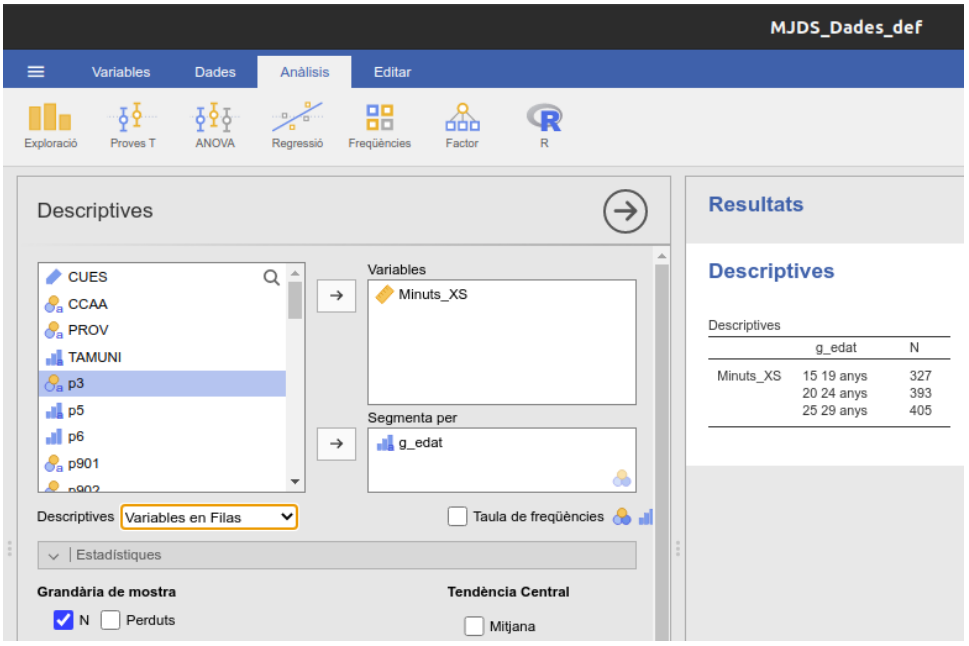
H_0: En la població, no hi ha relació entre l'edat i el temps dedicat a les xarxes socials.

H_1: En la població, hi ha relació entre l'edat i el temps dedicat a les xarxes socials.

Si hi ha evidència estadística per a la primera hipòtesis (H_0) és que l'edat no comporta diferències en el temps dedicat a les xarxes socials; per tant **no hi ha relació** entre l'edat i el temps dedicat a les xarxes socials. Si hi ha evidència estadística per a la segona hipòtesis (H_1) és que l'edat comporta diferències en el temps dedicat a les xarxes socials; per tant **hi ha relació** entre l'edat i el temps dedicat a les xarxes socials.

Tal i com passa amb la comparació de dues mitjanes, aquí també tenim dues famílies de tests, el paramètric i el no paramètric. Anàlogament, farem el test paramètric si tots els grups són grans o normals, mentre que, en cas contrari (algun grup és petit i no normal), farem un test no paramètric. Així, el primer que hem de comprovar és la mida dels grups de la mostra mitjançant el procediment d'anàlisi exploratori descriptiu:

1. Comprova que la variable "g_edat" estigui ben definida: de tipus ORDINAL, amb dades TEXT, i amb les categories ben ordenades.
2. Demana a *jamovi* estadística Descriptiva de la variable "Minuts_XS" agrupada per "g_edat",
3. Per millorar la visualització, activa l'opció *Variables en Files* al desplegable just a sota el llistat de variables disponibles (esquerra).

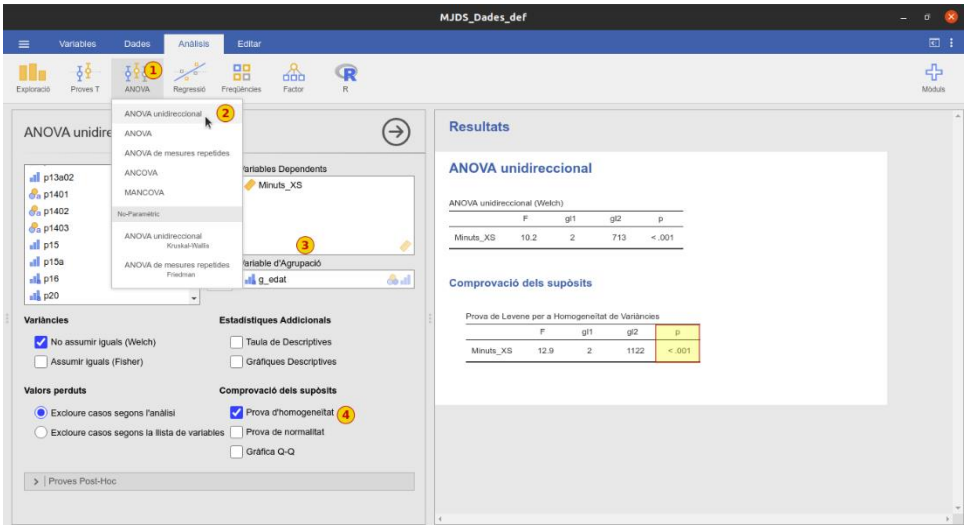


Imatge 125. Estimació de la mida dels grups o mostres per comparar DIVERSES mitjanes.

Pots veure que els tres grups tornen a ser grans ($n_i \geq 30$), de manera que podem fer directament un test paramètric. En aquest cas també hi ha dos tests paramètrics possibles, en funció de si les variàncies de tots els grups són o no iguals. Per tant, en primer lloc, haurem de tornar a realitzar el *test de Levene* per a la igualtat de variàncies que ens donarà informació sobre quin és el test adient per a aquest cas.

Per a fer test de Levene en el cas d'ANOVA:

1. Clica el menú Anàlisis / ANOVA / ANOVA Unidireccional. S'obre el panell de definició del procediment d'anàlisi de variàncies simple.
2. Situa la variable dependent (Minuts_XS) a l'espai superior *Variable Dependent*, i la variable independent (g_edat) a l'espai inferior *Variable d'Agrupació*.
3. Per defecte es mostren resultats del test ANOVA per a variàncies diferents (Welch). Però cal comprovar si aquesta *heteroscedasticitat* és certa o no.
4. Per a realitzar el test de Levene, a l'apartat inferior dreta **Comprovació del supòsits** activa l'opció [ ] *Prova d'homogeneïtat*.



Imatge 126. Anàlisi de Variàncies simple (*One-way*) amb control d'homocedasticitat.

El resultat d'aquesta comprovació de supósits ens el torna a donar en forma de p-valor. En aquest cas pren un valor menor de 0,001 de manera que hi ha evidència suficient per afirmar que les variàncies no són iguals: cal fer el test ANOVA en cas de variàncies diferents.

En el panell de definició del procediment anterior, apartat **Variables**, es pot comprovar que hi ha DOS tests possibles:

- el test paramètric en cas de variàncies diferents: *"No assumir iguals (Welch's)"*, i
- el test paramètric en cas de variàncies iguals: *"Assumir iguals (Fisher's)"*.

Per defecte, *jamovi* marca el primer que, en aquest cas, és l'adient. Hem obtingut:

ANOVA unidireccional (Welch)

	F	gl1	gl2	p
Minuts_XS	10.2	2	713	< .001

El p-valor en aquest cas és molt petit (<0,001) per tant es pot afirmar que podem **acceptar que hi ha relació (p- valor<0,001)** entre l'edat i el temps que es dedica a les xarxes socials.

Tal i com indiquen les hipòtesis que hem plantejat, dir que hi ha relació entre edat i temps a les xarxes socials és equivalent a dir que el temps dedicat a les xarxes socials és diferent en funció de l'edat. Ara bé, l'anàlisi que hem dut a terme fins ara, no ens diu entre quins grups es donen les diferències observades. Per veure-ho hem de comparar cada possible parella de grups, el que s'anomena **comparacions múltiples a posteriori**, o comparacions 2 a 2 (*Proves post-hoc*). Aquestes comparacions són equivalents a plantejar les tres hipòtesis següents:

H_0: $\mu_{15\text{-}19} = \mu_{20\text{-}24}$ (el grup d'edat 15-19 i el 20-24 es comporten igual).
H_1: $\mu_{15\text{-}19} \neq \mu_{20\text{-}24}$ (el grup d'edat 15-19 i el 20-24 es comporten diferent).

H_0: $\mu_{20\text{-}24} = \mu_{25\text{-}29}$ (el grup d'edat 20-24 i el 25-29 es comporten igual).
H_1: $\mu_{20\text{-}24} \neq \mu_{25\text{-}29}$ (el grup d'edat 20-24 i el 25-29 es comporten diferent).

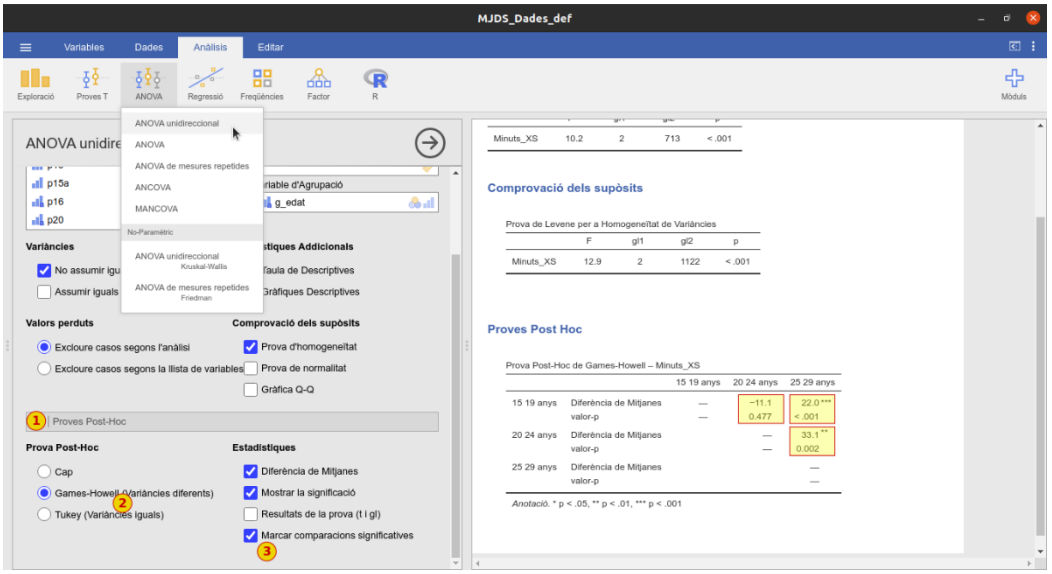
H_0: $\mu_{15\text{-}19} = \mu_{25\text{-}29}$ (el grup d'edat 15-19 i el 25-29 es comporten igual).
H_1: $\mu_{15\text{-}19} \neq \mu_{25\text{-}29}$ (el grup d'edat 15-19 i el 25-29 es comporten diferent).

Com més grups tinguem més comparacions 2 a 2 haurem de fer (k^2).

Per avaluar-les amb jamovi cal:

1. Al mateix panell de definició del procediment *ANOVA Unidireccional*, clica a sobre de la pestanya desplegable **> Proves Post-Hoc** per activar les comparacions múltiples a posteriori, o comparacions 2 a 2.
2. Per defecte no fa cap comparació. Cal activar el tipus adient, segons les conclusions de la prova d'homogeneïtat prèvia: en aquest cas, *Games-Howel (Variàncies diferents)* perquè hi ha variàncies diferents segons la prova Levene.
3. Per avaluar més fàcilment la significatividad estadística, a l'apartat **Estadístiques** activa l'opció [ ] *Marcar comparacions significatives*.

(cc) BY-NC



Imatge 127. Comparacions entre grups en un ANOVA simple.

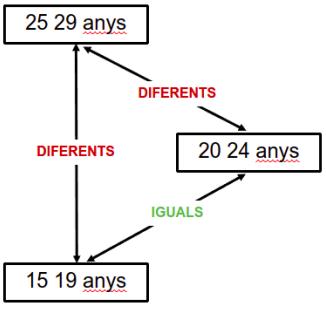
Els resultats es mostren a la següent taula on situa les diferents categories de la variable en files i columnes, de manera que el punt de creuament ens dona la significació de la diferència entre els grups corresponents:

Prova Post-Hoc de Games-Howell – Minuts_XS

		15 19 anys	20 24 anys	25 29 anys	
15 19 anys	Diferència de Mitjanes	—	-11.1	22.0	***
	valor-p	—	0.477	< .001	
20 24 anys	Diferència de Mitjanes		—	33.1	**
	valor-p		—	0.002	
25 29 anys	Diferència de Mitjanes			—	
	valor-p			—	

Anotació. * p < .05, ** p < .01, *** p < .001

Així, el valor 0,477 és la significació entre els grups 15-19 i 20-24, de manera que concloem que **no hi ha** diferències significatives entre aquests dos grups (p-valor>0,05). Per la seva banda, el valor <0,001 correspon a la significació entre els grups 15-19 i 25-29 i, per tant, podem concloure que **sí que hi ha** diferències significatives entre aquests dos grups. Per acabar, el valor 0,002 ens diu que entre les edats 20-24 i 25-29 **també hi ha** diferències significatives.



Imatge 128. Interpretació de proves post-hoc entre grups (ANOVA).

A manera de resum, direm que hi ha dos grups d'edat pràcticament iguals (15-19 i 20-24) en quant al nombre d'amics a les "xarxes socials" i que el grup que és significativament diferent dels altres és el de 25 a 29 anys, el d'edat més elevada.

Si n'avaluem les mitjanes per a cada grup (que podem obtenir a la taula generada per *Analysis / Exploració / Descriptives*), veurem que aquest grup destina a les xarxes socials un temps menor que la resta: el grup 25-29 anys destinen a les xarxes socials uns 60,3 minuts diaris, mentre que els de 15-19 i els de 20-24 hi destinen més temps (82,3 i 93,4 minuts, respectivament) i no hi ha diferències significatives entre ambdós.

El fet que el p-valor sigui gran (superior a 0,05) no comporta que assegurem que hi ha igualtat. Simplement, el que diem és que volíem demostrar que dos grups eren diferents, però que no hi ha prou evidencia estadística per a afirmar-ho. És en aquest sentit que sovint diem que hem observat diferències no significatives.

Important. El procediment per demanar les comparacions grup a grup **<u>només té sentit</u>** en el cas que puguem assegurar que, en general, hi ha diferències.
Aquest procediment el farem exactament igual tant si es tracta d'un test paramètric com no paramètric, com si es tracta d'un test suposant homoscedasticitat o heteroscedasticitat.

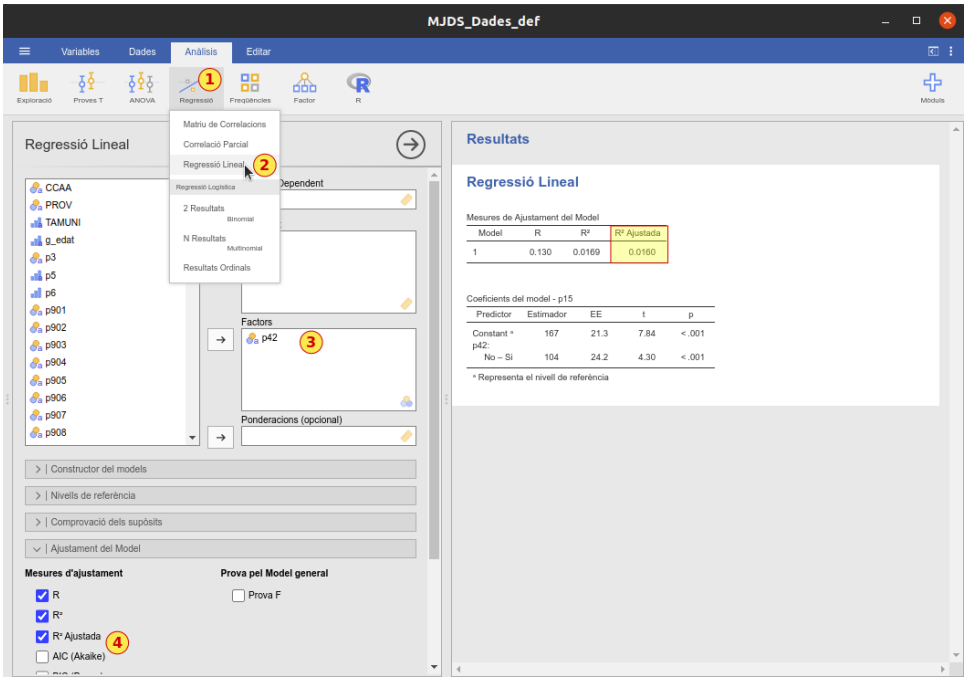
La intensitat de la relació: El coeficient de determinació.

Mentre que a les taules de contingència s'avalua la intensitat de la relació a partir de la *V de Cramer*, quan comparem mitjanes ho fem mitjançant l'estadístic R^2, anomenat com a *coeficient de determinació*. La diferència més important respecte de la *V de Cramer* és que R^2 es pot expressar en forma de percentatge, i ens diu **quina part de la variable dependent ve explicada** per la variable independent.

Així, un cop hem comprovat (mitjançant test de comparació de mitjanes) que dues variables estan relacionades, cal demanar la intensitat de la relació a partir de R^2. Per fer-ho utilitzarem una opció diferent al procediment ANOVA: el procediment *Regressió Lineal*.

Continuant amb el cas d'analitzar si conviure o no amb algú (p42) està relacionat amb el nombre d'amics que es tenen a les xarxes socials (p15), cal fer:

1. Clica el menú Anàlisis / Regressió / Regressió Lineal. S'obre el panell de definició del procediment *Regressió lineal*.
2. Situa la variable dependent (p15) a l'espai superior *Variable Dependent*, i la variable independent (p42) a l'espai inferior *Factors* (que admet variables no numèriques).
3. Per defecte es mostren resultats de la R^2 sense ajustar (una mesura de variància explicada criticada per ser sensible a la mida de la mostra i al nombre de variables explicatives).
   Per a demanar el càlcul del coeficient de determinació ajustat clica a la pestanya desplegable **> Ajustament del Model** i activa l'opció [ ] *R^2 Ajustada*:



Imatge 129. Intensitat de la relació (R^2) per a comparació de mitjanes.

El resultat és:

Mesures de Ajustament del Model

Model	R	R^2	R^2 Ajustada
1	0.130	0.0169	0.0160

Coeficients del model - p15

Predictor	Estimador	EE	t	p
Constant [a]	167	21.3	7.84	< .001
p42:				
No – Si	104	24.2	4.30	< .001

[a] Representa el nivell de referència

Observa que s'ha afegit a la taula inicial el valor del coeficient de determinació ajustat, 0,0160; que és el mateix que dir una variància explicada de l'1,6%. Aquesta intensitat és molt baixa: la "convivència" explica només un 1,6% de la variabilitat del "total d'amics a les xarxes socials".

EXERCICI PROPOSAT

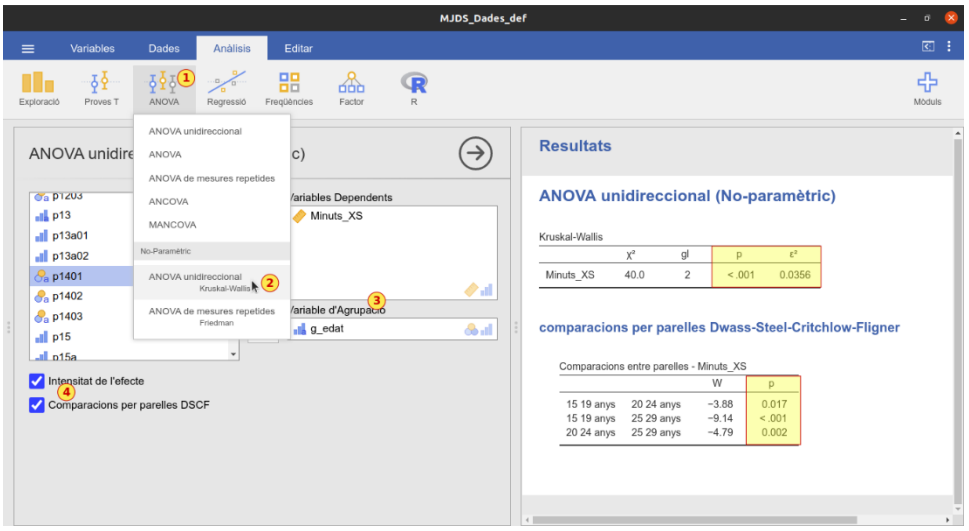
- Tingues en compte que aquest procediment també el pots fer servir en la comparació de més d'una mitjana. Comprova, per exemple, quin % de la variabilitat associada al temps dedicat a les Xarxes Socials (Minuts_XS) queda explicat per l'edat (edat_g).
- Llavors, és més explicativa la edat, o el model de convivència ?
- I en el cas del gènere?, explica aquest més o menys que la edat i que la convivència ?

La prova no paramètrica de comparació de més de dos grups: La prova de Kruskal-Wallis

Tot i que la lògica és exactament la mateixa que en l'anterior test, el test no paramètric de comparació de més de dues mitjanes, conegut com a *test de Kruskal-Wallis*, es troba a un lloc diferent que el test paramètric.

Així, si tenim una comparació de més de dues mitjanes, i hem d'aplicar un test no paramètric:

1. Clicarem el menú Anàlisis / ANOVA / No Paramètric / ANOVA Unidireccional - Kruskal-Wallis. S'obre el panell de definició del procediment *ANOVA unidireccional (No-paramètric)*.
2. Situarem la variable dependent (p.e., *Minuts_XS*) a l'espai superior *Variable Dependent*, i la variable independent (p.e., *g_edat*) a l'espai inferior *Variable d'Agrupació*. Obtindrem el p-valor.
3. Si el p-valor és significatiu (p-valor ≤0 ,05) avaluarem la diferència entre els grups clicant la pestanya [ ] *Comparació per parelles DSCF,* i avaluarem la intensitat de la relació clicant sobre la pestanya [ ] *Intensitat de l'efecte*:



Imatge 130. Prova no paramètrica de comparació de més de dos mitjanes (Kruskal-Wallis).

Observació! Veuràs que la intensitat de la relació l'avalua amb l'indicador eta-quadrat ε^2. La seva interpretació és exactament la mateixa que la que hem fet servir amb l'indicador R^2.

EXERCICIS PROPOSATS: Comparació de mitjanes

Després de seguir aquest apartat del tutorial, sobre comparació de mitjanes, obre en jamovi el fitxer *MJDS_Dades_def.omv* i respon les següents qüestions:

(A) Tal i com hem fet a la sessió de teoria, construeix la variable minuts diaris dedicat a les Xarxes Socials (Minuts_XS). A més, agrupa la variable nivell d'estudis (ESTUDIS) en tres categories: Primària o menys, Secundària (F.P.+Secundària) i Universitaris (mitjos+superiors).

Descriu la relació existent entre les dues variables generades a partir de les mesures de tendència central més adequades. Representa gràficament la relació entre ambdues variables.

(B) Comprova si hi ha diferències significatives entre homes i dones (p40) respecte dels amics que tenen a les xarxes socials (p15). Fes servir el procediment explicat per a avaluar-ho, comprovant prèviament si podem fer un test paramètric, així com la igualtat de variàncies. Planteja les hipòtesis i arriba a conclusions amb un 95% de confiança. En cas de diferències significatives, explica'n el sentit i la intensitat.

(C) Volem analitzar si hi ha diferències en el nombre d'amics (p15) que es tenen a les xarxes socials respecte de l'estat civil (p41). Si observes l'estat civil veuràs que només té sentit tractar amb tres categories: solters, casats i la resta. Agrupa l'estat civil en aquestes tres categories, i analitza si podem assegurar, amb un 95% de confiança que l'estat civil incideix en el nombre d'amics. Planteja les hipòtesis i arriba a totes les conclusions possibles. Enganxa en un processador de textos tots els resultats que necessitis per arribar a una conclusió i comenta'ls.

(D) A partir dels 14 ítems de la pregunta 9, genera un indicador d'ús d'Internet. Es tractarà d'un indicador quantitatiu, que pot anar des del 0 (que voldrà dir que no fa servir Internet per cap de les qüestions que li hem proposat) a un màxim de 14 (que significarà que el fa servir per a les 14 coses proposades). Fes-ho de manera que als valors en blanc (no fan servir les xarxes socials) els correspongui un 0.[5]

Per tal de comprovar si l'has fet bé, aquí tens el resultat dels descriptius d'aquest indicador:

Descriptives

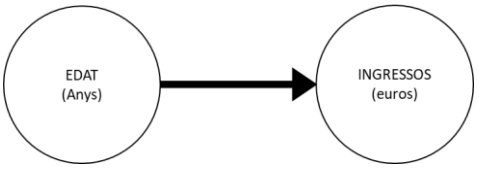
	g_edat
N	1419
Perduts	0
Mitjana	5.21
Mediana	5
Mínim	0
Màxim	12

Un cop construït l'indicador, comprova la relació amb les següents variables: sexe (p40), estatus socioeconòmic (ESTATUS), nacionalitat (p63) i viure en parella (p42). Planteja les hipòtesis corresponents i fes totes les anàlisis amb un 90% de confiança.

(cc) BY-NC

12. Regressió lineal simple mitjançant jamovi

En aquesta secció, seguint el Tutorial pas a pas, pots adquirir competència per a utilitzar el programari *jamovi* analitzant la relació entre dues variables quantitatives, una independent i l'altra dependent mitjançant el procediment *Regressió* que calcula la correlació i genera models de regressió lineal, així com per fer-ne la representació gràfica pertinent. També podràs interpretar la informació que s'obté de *jamovi* per tal de donar resposta a les hipòtesis de relació entre aquestes dues variables d'acord amb un model d'anàlisi bàsic (aspecte descriptiu i de contingut de l'anàlisi), i determinar la significació d'aquesta informació per tal d'extrapolar-la al conjunt de la població (aspecte inferencial de l'anàlisi).



Imatge 131. Exemple de model d'anàlisi de regressió lineal simple.

L'objectiu principal de la pràctica és comparar les distribucions de **DUES variables quantitatives**. En els capítols anteriors d'aquest manual, havíem recomanat que les variables quantitatives estiguessin definides, en *jamovi*, com a tipus de mesura ORDINAL (i data ENTER o DECIMAL). El motiu era que això ens permetia, per exemple, visualitzar les variables utilitzant taules de freqüències, unes taules que no ens mostra quan la variable està definida com a tipus CONTINOUOS que, d'entrada, sembla el tipus més lògic per a variables quantitatives. Doncs bé, algun dels procediments que utilitzarem en aquest capítol presenten algun problema si la variable no està definida, precisament, de tipus CONTINOUOS, de manera que haurem d'anar modificant el tipus de mesura en funció de les nostres necessitats.

En el capítol es tracta, per tant, la possible relació entre variables de mètrica quantitativa i s'utilitzarà el procediment d'estimació de la recta de regressió lineal simple com a introducció de la lògica de modelització.

En aquest apartat es treballarà amb el fitxer de dades originals *MJDS_Dades_def.omv* que és el mateix que en els dos capítols anteriors, i que es basa en el mateix qüestionari, que necessitareu per anar seguint la pràctica. Descarrega al teu PC els arxius necessaris: *MJDS_Qüestionari.pdf* i *MJDS_Dades_def.omv*.

Descripció de la relació entre dues variables quantitatives

L'objectiu principal dels models de regressió lineal simple és avaluar la relació entre dues variables quantitatives, la variable *resposta* o dependent i la variable *explicativa* o independent. El coneixement d'aquesta relació ens permetrà fer **prediccions** sobre la variable resposta, una predicció d'un valor numéric que té el seu fonament en la recta que millor *ajusta* les dades, anomenada *recta de regressió* i generada o estimada a partir del *diagrama de dispersió*. En aquest exemple analitzarem la relació entre l'edat en anys (p37) i el nombre d'amics que es tenen a les xarxes socials (p15).

En primer lloc, clicant sobre cadascuna de les dues variables, canvia la seva tipologia a variables de tipus CONTINUA i dades tipus ENTER. Anteriorment, aquestes dues variables s'havien definit com

a ORDINALS (per poder demanar les seves taules de freqüències) peró ara cal definir-les com a CONTINUES per poder realitzar els procediments d'anàlisi corresponents al model de regressió lineal.

Compte! Tot i que aquí s'explicarà primerament l'anàlisi descriptiva i després la inferència, la lògica analítica és la contrària. És a dir, en una anàlisis real:

(1r) Assegurem si hi ha o no relació significativa **per a tota la població** (inferència, p-valors).

(2n) Només **en el cas que hi hagi** relació, interpretem tots els paràmetres que siguin interpretables; mentre que si no podem assegurar la relació, no cal interpretar descriptivament els paràmetres.

Per tal de descriure la relació entre dues variables quantitatives, se solen utilitzar tres estratègies:

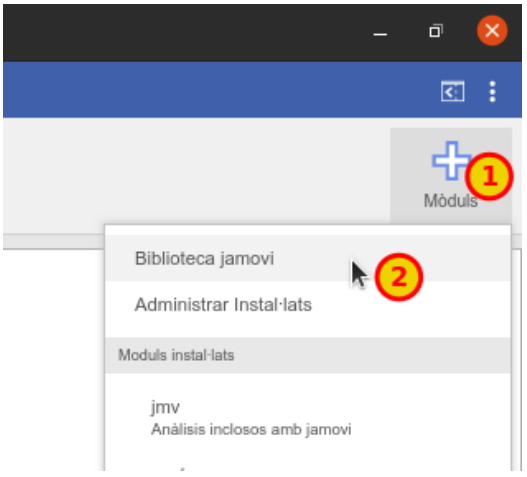
1. Dibuixar gràficament el *núvol de punts* que representa la relació, utilitzant l'anomenat diagrama de dispersió o diagrama XY.
2. Avaluar el nivell o intensitat de la relació entre les dues variables a partir de l'anomenat coeficient de correlació de Pearson (*r*).
3. Estimar els coeficients que determinen la *recta de regressió* (*a* i *b*), així com l'indicador d'ajustament del núvol de punts a la recta, el coeficient de determinació (r^2).

El diagrama de dispersió

Tal i com succeeix en altres representacions gràfiques, en els diagrames XY sempre representarem la variable independent a l'eix de les X (horitzontal) i la variable dependent a l'eix de les Y (vertical). En el nostre exemple es corresponen, respectivament, a l'edat (p37) i al nombre d'amics que es tenen a les xarxes (p15). A més cal especificar el tipus de corba per fer l'ajustament, de les tres possibilitats cal triar la *lineal*.

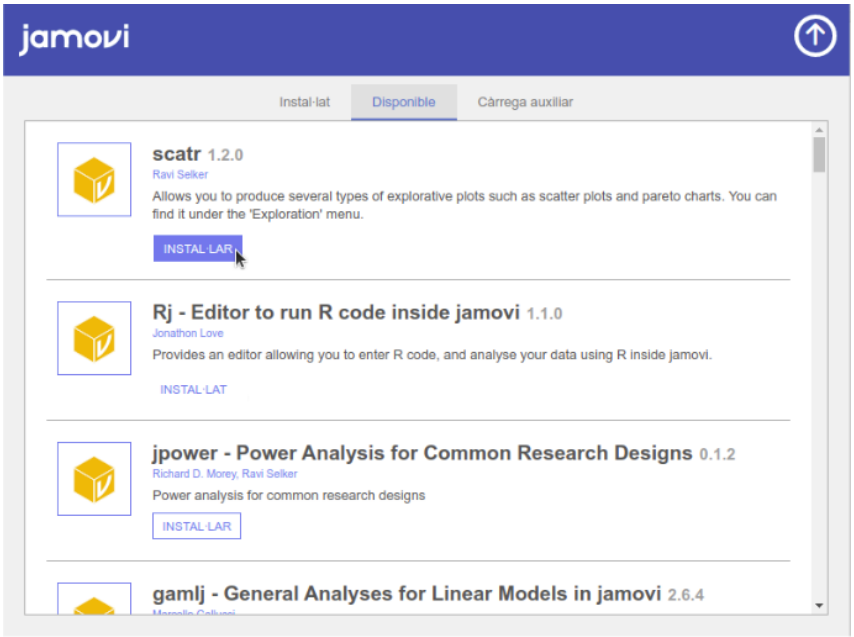
Per a generar un diagrama de dispersió o diagrama XY (*scatter plot*), en *jamovi*), cal instal·lar un mòdul addicional anomenat *scatr*, si és que no s'ha fet prèviament:

1. Comprova si el módul ja està instal·lat, a Menú / Anàlisi / Exploració.
   En cas que no estigui instal·lat, segueix els següents pasos.
2. Assegureu-vos que l'ordinador disposa de connexió a Internet.
3. Per a carregar el mòdul *scatr*, a l'extrem dret del menú cliqueu la icona + *Móduls*. Es desplega el llistat d'opcions Seleccioneu *Biblioteca jamovi*.



Imatge 132. Accés al sistema de gestió d'extensions

4. S'obre la finestra d'instal·lació de mòduls o extensions disponibles. Baixeu pel llistat fins que trobeu el mòdul *scatr*. Cliqueu el botó blau [INSTALAR]



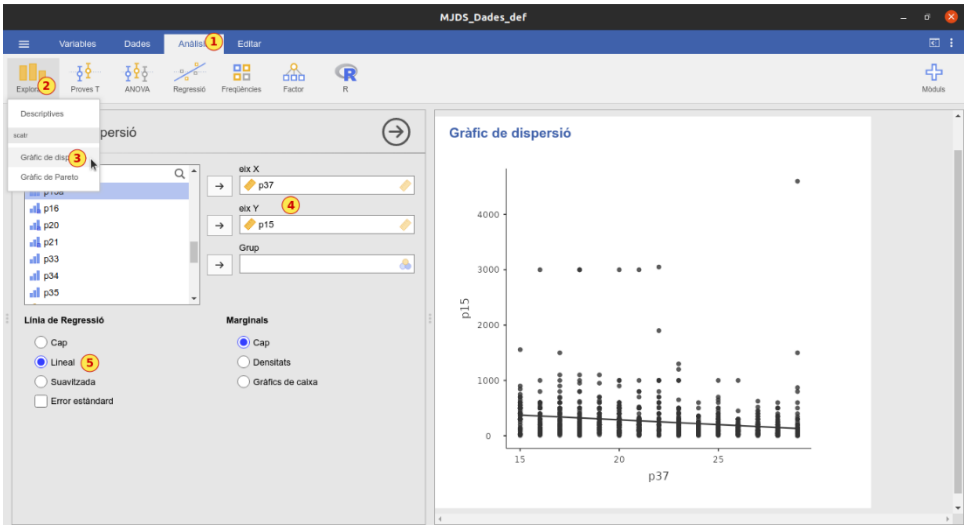
Imatge 133. Instal·lació de mòdul per a generar diagrames XY

Els mòduls o extensions de *jamovi* s'instal·len descarregant-los des d'Internet, de manera que cal tenir connexió per poder-los instal·lar. Una vegada instal·lat un mòdul, apareixeran noves opcions als menús de *jamovi*.

Compte! Si l'arxiu *.omv s'obre des d'un altre ordinador cal que aquest PC tingui instal·lat el mateix mòdul per poder visualitzar els resultats.

Per a generar un diagrama de dispersió, o diagrama XY (*scatter plot*) amb menús de *jamovi:*

1. Una vegada carregades les dades de *MJDS_Dades_def.omv* cal comprovar que les dues variables estiguin ben definides. Feu doble-clic al nom de p15, que encapçala la columna de dades: és convenient que sigui de tipus CONTINUA i les dades ENTER. Feu el mateix per a la p37.
2. Cliqueu la FLETXA-AMUNT ENCERCLADA de la part superior dreta per a tancar la finestra de definició de variables.
3. Ara cliqueu el menú Anàlisis / Exploració / Gràfic de Dispersió. S'obre la finestra de definició del diagrama de dispersió.
4. Situeu la variable p37 a l'espai superior dret *Eix X* i la variable p15 a l'espai central *Eix Y*.
5. Per defecte, només es dibuixa el núvol de punts. Per a afegir una línia d'ajustament, seleccioneu l'opció *( ) Lineal* a l'apartat inferior esquerre **Línia de Regressió**. Es genera el gràfic següent:



Imatge 134. Diagrama de dispersió o XY amb jamovi.

Tal i com hem vist en capítols anteriors, les representacions gràfiques de *jamovi* són molt rígides, no es poden modificar, de manera que qualsevol canvi comporta haver d'utilitzar codi R. Així, per exemple, si voguéssim incloure títols al gràfic de dispersió o als eixos, o controlar l'escala dels eixos, és imprescindible utilitzar codi R des de la consola de *jamovi*. Abans de fer-ho tingues en compte que has de tenir instal·lat el mòdul *Rj Editor*.

Per a generar gràfics de dispersió amb R (no cal tenir instal·lat el mòdul d'extensió *scatr*)

1. Cliqueu el Menú / Anàlisis / R / Rj Editor
2. Generarem, en primer lloc, la versió més senzilla del gràfic XY amb la variable p37 (edat) a l'eix horitzontal i p15 (nombre d'amics a Xarxes) a l'eix vertical:

```
library(ggplot2)
ggplot(data)+
aes(x=p37, y=p15)+
geom_point()
```

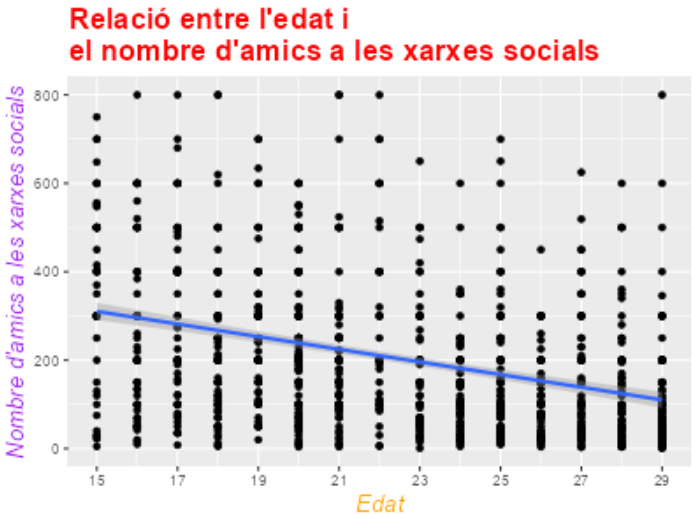
3. Anem a completar l'anterior instrucció (1) afegint-hi títols, (2) canviant l'escala de l'eix de les Y per tal de veure més clarament la relació, (3) dibuixant la recta que passa més a prop de tots els punts i (4) canviant el format dels eixos, entre d'altres:

```
library(ggplot2)
ggplot(data)+
aes(x=p37, y=p15)+
geom_point()+
geom_smooth(method="lm")+
ggtitle("Relació entre edat i \namics a les xarxes socials")+
labs (x="Edat", y="Nombre d'amics a les xarxes socials")+
scale_y_continuous(limit = c(0,800)) +
scale_x_continuous(breaks=seq(15, 29, 2)) +
theme (plot.title = element_text(size=rel(1.5), face="bold", color="red")) +
theme(axis.title.x = element_text(face="italic", colour="orange", size=rel(1.2)))+
theme(axis.title.y = element_text(face="italic", colour="purple", size=rel(1.2)))
```

4. Comprova que per enllaçar les diferents instruccions, al final de cadascuna hem afegit el símbol de suma (+). Cadascuna d'aquestes instruccions que hem afegit, fan el següent:

- Dibuixa la recta de regressió:
  `geom_smooth(method="lm")`
- Afegeix un títol al gràfic (el símbol \n serveix perquè ens faci un salt de línia):
  `ggtitle("")`
- Afegim un títol a cadascun dels eixos:
  `labs (x="", y="")`
- Li diem que l'eix vertical comenci al 0 i acabi al 800:
  `scale_y_continuous(limit = c(0,800))`
- En l'eix de les X li demanem que escrigui un nombre de cada dos:
  `scale_x_continuous(breaks=seq(15, 29, 2))`
- Definim les característiques del títol. En concret li demanem que ampliï la mida de la lletra (un 50% més que la que té per defecte), l'escrigui en negreta i de color vermell:
  `theme (plot.title = element_text(size=rel(1.5), face="bold", color="red"))`
- En quant a les característiques dels eixos demanem que la mida sigui un 20% superior a l'habitual, els escrigui en cursiva, i de colors taronja i morat:
  `theme(axis.title.x = element_text(face="italic", colour=" ", size=rel(1.2)))`
  `theme(axis.title.y = element_text(face="italic", colour=" ", size=rel(1.2)))`

5. El gràfic resultant hauria de quedar així:



Imatge 135. Diagrama XY i recta amb codi R (ggplot)

Aquest gràfic de dispersió ens mostra que hi ha una certa correlació inversa (o negativa) entre l'edat i el nombre d'amics a les xarxes socials, de manera que les persones més joves són les que tenen més amics, i aquests van descendint a mesura que augmenta l'edat. A més, observem una forta dispersió de les dades, de manera que la correlació no serà massa intensa.

Compte! Si el gràfic no et surt com el que aquí es mostra és probable que no hagis canviat les variables p15 i p37 a tipus de mesura CONTINOUOS. I és que aquest és un dels procediments sensibles al fet que la variable sigui de tipus ORDINAL, atès que en aquest cas no serà capaç de dibuixar els eixos correctament, com a nombres enters.

A R habitualment hi ha moltes instruccions alternatives que poden generar resultats similars. Per exemple, també pots fer servir la instrucció `plot` d'aquesta forma:

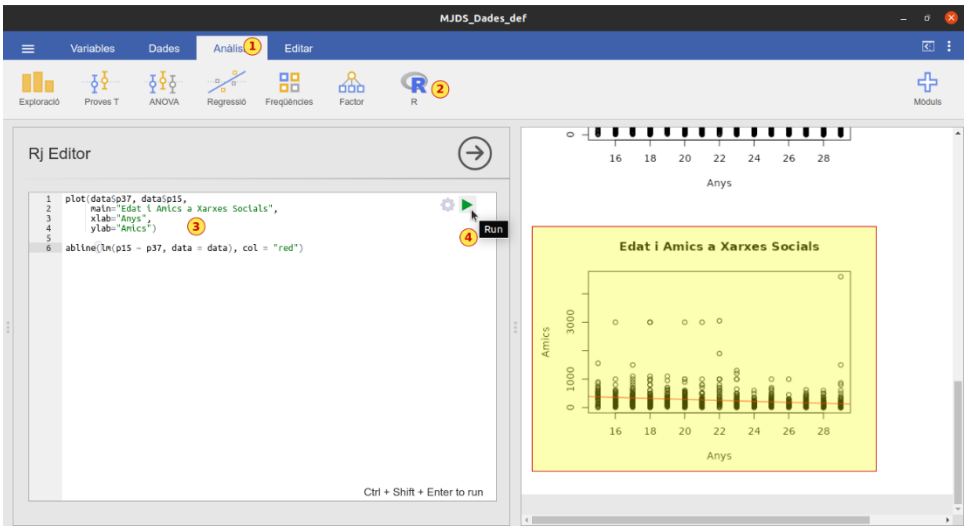
1. Fes clic a Menú / Anàlisis / R / Rj Editor...

2. Escriu en codi R per generar un gràfic XY amb Edat (p37) a l'eix horitzontal i Nombre d'amics a Xarxes (p15) a l'eix vertical:

```
plot(data$p37, data$p15,
     main="Edat i Amics a Xarxes Socials",
     xlab="Anys",
     ylab="Amics")
```

Les ordres en llenguatge R es poden escriure tot seguit, o separar-les en línies consecutives, com a l'exemple anterior. Fins i tot es pot aprofitar l'espai o marge esquerre per a "endressar" d'una forma comprensible la ordre: separant en línies diferents les diverses opcions addicionals.
En el cas anterior, també es podria haver escrit tot seguit:
`plot(data$p37, data$p15, main="Edat i Amics a Xarxes Socials", xlab="Anys", ylab="Amics")`

3. Executa el codi R: clica al tringle verd (o combinació de teclat [Ctrl]+[Mayús]+[Intro])
4. Per afegir una línia de regressió cal afegir una segona instrucció a la consola i executar:[31]
   `abline(lm(p15 ~ p37, data = data), col = "red")`



Imatge 136. Diagrama XY i recta amb codi R (plot)

5. Fes clic a la FLETXA-DRETA ENCERCLADA per a tancar la finestra de consola.

La correlació entre dues variables

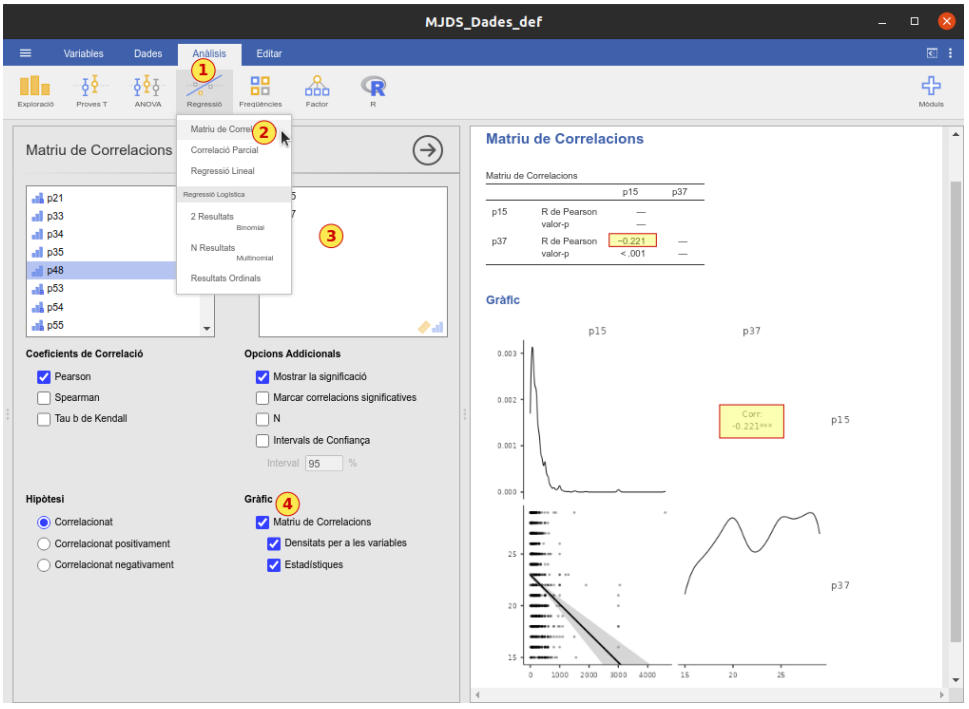
Per avaluar de manera més precisa la magnitud del grau d'associació lineal entre dues variables quantitatives ho fem mitjançant el coeficient de correlació de Pearson (*r*), un coeficient que no té unitats i que pren valors compresos entre –1 i +1.

Per obtenir aquest coeficient amb els menús de *jamovi*:

1. Cliqueu el Menú / Anàlisis / Regressió / Matriu de Correlacions. S'obre el panell de definició del procediment càlcul de correlacions entre parelles de variables,
2. Situeu totes dues variables, p37 i p15, a l'espai superior dret,

[31] El símbol ~ es pot aconseguir amb la combinació de teclat [Alt Gr] + [4] i a continuació [Espai].

3. Addicionalment, activeu les tres opcions sota l'apartat dret inferior **Gràfic**, per dipossar d'informació rellevant sobre la distribució de cada variable.



Imatge 137. Coeficient de correlació (r) amb jamovi.

El resultat bàsic ens dona un coeficient de –0,221 que ens mostra una correlació **inversa** entre l'edat i el nombre d'amics, de manera que a major edat el nombre d'amics a les xarxes socials és menor. El valor no és molt elevat, el que mostra una **correlació no molt intensa** entre ambdues variables. Cal tenir en compte, però, que en ciències socials quan avaluem variables individuals no solem obtenir coeficients massa elevats.

El resultat gràfic inferior mostra el diagrama de dispersió o XY d'aquestes dues variables, més la corba de densitat de cadascuna, on podem veure com la p15 està molt esbiaxada cap a valors baixos (o el que es el mateix, existeixen alguns pocs valors elevats que podrien ser "extrems" pendents de depurar) mentre que la p37 es distribueix de forma bi-modal però prou uniforme a tot el rang de valors.

Observació! No confonguis els conceptes i tingues en compte que l'existència de correlació entre dues variables no implica causalitat entre elles. Pensa, per exemple, que hi ha correlació estadística molt elevada entre la longitud dels dos braços de cada individu, però que una mesura o variabe no és en cap cas la causa de l'altra. Tingues en compte que el mateix argument es pot aplicar, per exemple, a la correlació entre indicadors de "inmigració" i de "delinqüencia".

El model de regressió lineal simple

La correlació entre dues variables –el coeficient r de Pearson– es pot avaluar sempre que disposem de dues variables quantitatives, atès que no comporta que hi hagi una variable dependent i una d'independent, simplement ens mostra si les dues variables canvien al mateix temps. En canvi, el model de regressió lineal –tal i com passa també amb el diagrama de dispersió– comporta ja pensar en una variable dependent (Y = *efecte o resposta*) i una d'independent (X = *causa o tractament*).

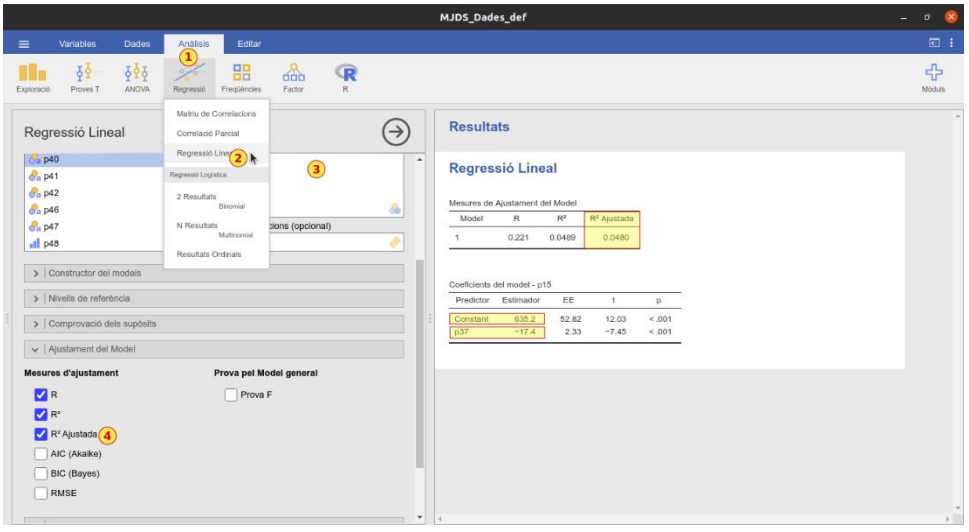
En aquests casos, la intensitat de la relació ens la dona el *coeficient de determinació* (r^2), un estadístic que ens diu quina part de la variabilitat total de la variable dependent queda explicada per la variable independent. Es tracta d'un coeficient que varia entre 0 i 1, i que podem passar a percentatge per indicar precisament això, quina proporció de la variabilitat total de Y és explicada per X.

A més, sol ser habitual que ens interessi quina és la funció (o fórmula) de la recta de regressió, la recta que *jamovi* ha dibuixat en el gràfic de dispersió. Aquesta funció ens permetrà predir, per a cada valor de la variable independent, el valor que pot prendre, en mitjana, la variable dependent.

El model funcional que correspon a la recta de regressió lineal simple depèn només de dos paràmetres, la **constant** i el **pendent** (expressats, respectivament amb les lletres *a* i *b*). Així, en termes generals, la recta de regressió s'expressa amb la funció: Y = *a* + b*X* que, en el nostre context, ho traduïm per VI = *a* + *b*(VD).

Per demanar a *jamovi* la estimació dels paràmetres *a* i *b* (que ens determinaran la funció o línia de regressió) i el coeficient de determinació cal fer servir el procediment *Regressió Lineal*.

1. Cliqueu el Menú / Anàlisis / Regressió / Regressió Lineal. S'obre el panell de definició.
2. Seleccioneu la variable dependent p15 –amics (Y, serà considerat l'efecte)– i situeu-la a l'espai superior dret *Variable Dependent*.
3. Seleccioneu la variable independent p37 –edat (X, serà considerat la causa)– i situeu-la a l'espai central dret *Covariant* (que admet únicament variables numèriques, com pots veure per la petita icona groga).
4. Cliqueu sobre el desplegable **> Ajustament del Model** i seleccioneu l'opció [ ] R^2 *Ajustada.*
5. Per defecte, es generen dues taules de resultats:



Imatge 138. Model de regressió lineal simple i R^2 ajustada, amb jamovi.

La taula superior conté les informacions sobre l'ajustament del model de regressió estimat (recta) en relació al conjunt de dades disponibles (núvol de punts): es tracta de la correlació de Pearson (notat com a *R*) entre ambdues variables, el coeficient de determinació (notat com a R^2), i el

coeficient ajustat (que notem com a R^2 *Ajustada*). El valor 0,0480 del coeficient de determinació ajustat el podem llegir com que *"la variable edat explica fins a un 4,8% de la variabilitat donada en el nombre d'amics a les xarxes socials"*.

Per altra banda, els paràmetres per construir la funció o recta de regressió es poden obtenir a la columna *Estimador* de la taula inferior: la *b* o pendent (*p37*) i la *a* o *Constant* (*Intercept*, en anglés). Així podem escriure funció de la recta que millor s'ajusta (o descriu) les dades disponibles d'aquesta forma:

Nombre d'amics = 635,2 - 17,4 x Edat

Aquesta funció permet estimar el nombre d'amics a les xarxes socials de qualsevol persona en funció de la seva edat i, per tant, mostra la relació entre ambdues variables en forma de *pendent de la recta*.

Compte! La *constant* es llegeix com el valor de la variable Y quan X=0. És molt corrent que aquesta constant no sigui interpretable. De fet, només ho serà si en la mostra tenim valors mesurats en X=0. En aquest cas vist com a exemple no té sentit; considerant la variable p37 (edat) voldria dir que hem entrevistat persones de 0 anys i que han dit *quants amics tenen a les xarxes socials*. En altres casos sí que pot tenir sentit, peró cal analitzar-ho detenidament.

El pendent és l'augment de la variable Y en incrementar en una unitat la variable X. En aquest cas direm que, per a cada ANY d'edat de més, el nombre d'amics a les xarxes incrementa, en mitjana, en –17,4. És a dir, amb cada any més d'edat descendeix en 17,4 el nombre d'amics a les xarxes. (Es tracta de diferents individus de la mostra, per tant no és que a mesura que un individu envelleixi perdi amics a les xarxes... és que els individus amb menys edat en tenen més amics mentre que els individus amb més edat en tenen menys)

El signe del pendent serà sempre el mateix que el del coeficient de correlació de Pearson, de manera que també indica si es tracta d'una relació directa o inversa. Òbviament, el signe ha de ser també el del pendent de la recta dibuixada en el diagrama de dispersió.

En termes generals, la funció de la recta ens permet, entre altres coses, estimar o predir el valor de la variable Y a partir de diferents valors que pugui prendre la variable X. En el nostre cas seria, per exemple, estimar, en mitjana, el nombre d'amics a l'edat 23:

Nombre d'amics = 635,2 – (17,4 x Edat) = 635,2 – (17,4 x 23) = 635,2 – 400,2 = **235,0 amics**

Compte! Les estimacions fetes a partir de la recta tenen una lògica semblant a la constant. Així, si dèiem que no podem interpretar la constant si no disposàvem de valors propers a edat=0, tampoc no podrem fer estimacions fora del rang del qual tinguem valors. En aquest cas, per exemple, on la variable edat està compresa entre 15 i 29 anys, no té sentit fer estimacions o prediccions per a edats inferiors als 15 anys ni superiors als 29.

Com pots comprovar al panell corresponent de *jamovi*, el procediment *Regressió Lineal* permet moltes més opcions, per exemple, incorporar més d'una variable quantitativa explicativa (*Covariant*), així com generar models amb variables independents qualitatives (*Factors*), o combinar models que continguin tot tipus de variables independents. Totes aquestes possibilitats, i la personalització del model són objecte de l'anàlisi de dades MULTIVARIABLE (fora de l'abast d'aquest tutorial).

Inferència estadística en els models de regressió

En els anteriors apartats pots comprovar que hem estimat quatre paràmetres: el coeficient de correlació (*r*), el coeficient de determinació (r^2), la constant (*a*) i el pendent (*b*).

La seva estimació per a tota la població, a partir de la mostra disponiible, comporta la possibilitat de formular quatre tests d'hipòtesis diferents, que serien:

- Test sobre el coeficient de correlació:
  - H_0: R = 0 → La correlació entre X i Y és 0. No hi ha relació entre X i Y.
  - H_1: R ≠ 0 → La correlació entre X i Y és diferent de 0. Hi ha relació entre X i Y.
- Test sobre el coeficient de determinació:
  - H_0: $R^2 = 0$ → La X explica un 0% la Y. No hi ha relació entre X i Y.
  - H_1: $R^2 \neq 0$ → La X explica més d'un 0% la Y. Hi ha relació entre X i Y.
- Test sobre la constat:
  - H_0: A = 0 → Quan X=0, la Y també és 0.
  - H_1: A ≠ 0 → Quan la X=0, la Y no és 0.
- Test sobre el pendent:
  - H_0: B = 0 → Els canvis en X no comporten cap canvi en Y. No hi ha relació entre X i Y.
  - H_1: B ≠ 0 → Els canvis en X comporten algun canvi en Y. Hi ha relació entre X i Y.

Observa que... per distingir els valors mostrals (obtinguts a la nostra mostra) dels valors poblacionals (els que es donen realment a la població), ho fem utilitzant minúscules en el primer cas i majúscules en el segon.
De fet, la lògica de la inferència és sempre la mateixa: nosaltres volem estimar un paràmetre poblacional (per exemple, R^2), però no podem entrevistar tota la població. Per això, fem una enquesta a una mostra reduïda i obtenim un estimador (r^2) que intenta, d'alguna manera, parlar-nos del paràmetre en qüestió, però que no és el paràmetre.

Tal i com escrit en la mateixa formulació de les hipòtesis, la conseqüència final de tres d'aquests tests és la mateixa. És a dir, (1) afirmar que no hi ha correlació, (2) afirmar que una variable explica el 0% de l'altra o (3) afirmar que si una variable canvia l'altra no ho fa, és, en el fons, exactament el mateix. Els tres casos comporten, simplement, que no hi ha relació entre les variables. Així, **els tests sobre R, R^2 i B són exactament el mateix**. I aquest és el resultat que ens importa a nosaltres, en tant que el nostre objectiu en la inferència és, precisament, avaluar si podem o no afirmar que en la població hi ha relació entre ambdues variables:

Matriu de Correlacions

Matriu de Correlacions

		p15	p37
p15	R de Pearson	—	-0.221
	valor-p	—	< .001
p37	R de Pearson		—
	valor-p		—

Regressió Lineal

Mesures de Ajustament del Model

Model	R	R^2	R^2 Ajustada
1	0.221	0.0489	0.0480

Coeficients del model - p15

Predictor	Estimador	EE	t	p
Constant	635.2	52.82	12.03	< .001
p37	-17.4	2.33	-7.45	< .001

Si observes el resultat dels tres tests veuràs que la significació és molt petita, gairebé 0,000,[32] de manera que podem assegurar, amb un 95% de confiança (p-valor<0,001)[33] **que hi ha relació entre l'edat i el nombre d'amics** que es tenen a les xarxes socials. Així doncs, els resultats que s'han generat en l'anterior apartat es poden interpretar en termes de tota la població i no només en termes de la mostra observada: són estadísticament significatius.

En quant al test sobre la constant (*a* o *Constant*), el seu resultat sol ser poc rellevant per nosaltres. En primer lloc, perquè si la constant no és interpretable, el resultat del seu test té una transcendència molt limitada i, en segon lloc, i fins i tot si és interpretable, el fet que si X=0, Y també sigui 0, no ens diu res de la relació entre les variables.

Observació! El resultat que donarà *jamovi* pels tests sobre R, R^2 i B serà **SEMPRE** exactament el mateix, essent l'únic canvi el nombre de decimals que aproxima. En canvi, el resultat sobre la constant té una altra lògica que, si com aquest cas és també el mateix, és per casualitat.

Els supòsits del model: L'anàlisi de residuals

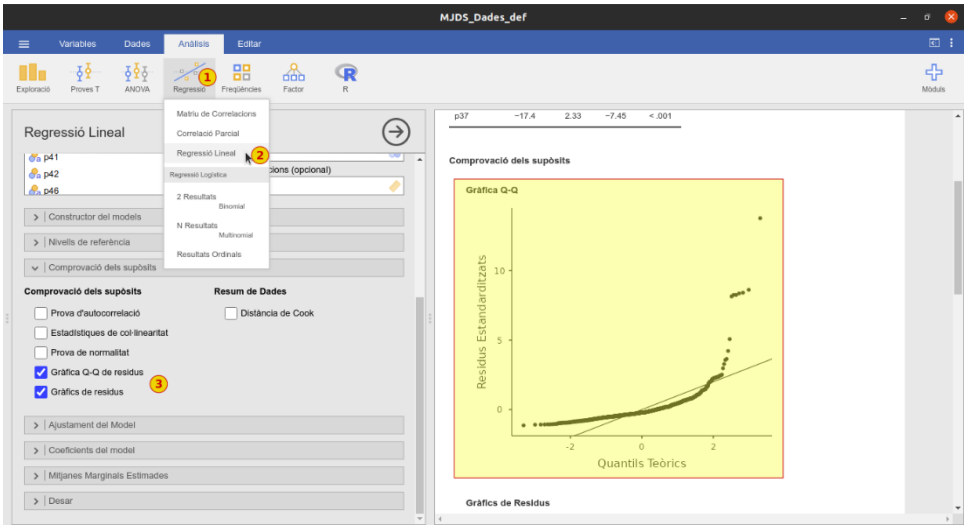
Recorda que en l'anàlisi de taules de contingència parlàvem també dels *residuals* que no eren més que la diferència entre el valor observat i el valor esperat ("esperat" si es complia la hipòtesi nul·la). En el cas de la Regressió Lineal, els residus són la diferència entre el valor observat i el valor predit per la recta de regressió. Pot ser considerat com l'*error aleatori* observat. L'estudi d'aquests residus es fa, bàsicament, a partir de diagrames anomenats "Gràfiques de Residus" (en anglès, *Residual plots*) de dispersió XY entre els valors residuals i els valors predits per la recta de regressió. En termes generals per a validar el model de regressió mitjançat l'estudi dels residuals és recomanable assegurar-se que no hi ha cap tipus de patró en el gràfic de residuals i prediccions, i comprovar cadascun dels supòsits del model, que són:

32 En ser el valor tan petit, en els tres casos jamovi escriu p-valor<0,001.
33 Podríem assegurar-ho fins i tot amb un 99,9% de confiança.

1. El model està correctament especificat.
2. Els residus són **independents** els uns dels altres (*Col·linearitat*, molt rellevant en anàlisi multivariable, però no tant en la regressió simple amb només una variable independent).
3. En mostres petites, hi ha **normalitat** en els residus (*Gràfica Q-Q de residus*)
4. Els residus tenen una mitjana de zero i una variància constant (*Gràfics de residus*).
5. No hi valors anòmals.

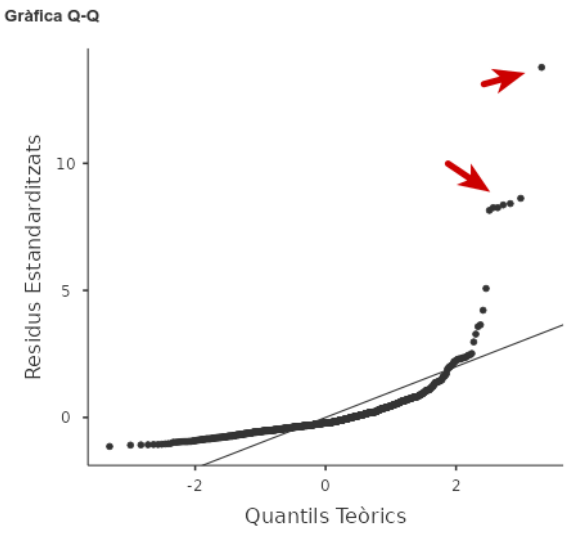
El sistema actual de menús de *jamovi* permet controlar només la Normalitat i la homoscedasticitat dels residus. Per fer-ho:

1. Cliqueu la taula de resultats de la Regressió Lineal corresponent. S'obrirà el panell de definició del procediment.
2. Desplegueu l'apartat **> Comprovació del supòsits** i activeu les opcions [ ] *Gràfica Q-Q de residus*, i [ ] *Gràfics de residus*.



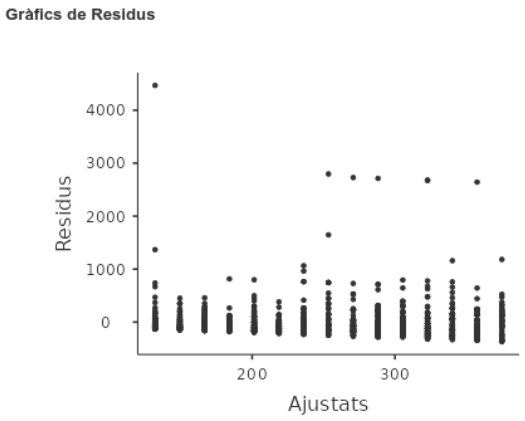
Imatge 139. Anàlisi de residuals al model RLS. Gràfic Q-Q.

El primer dels gràfics que obtenim és el que ens parla de la possible normalitat. Si la distribució fos normal, els punts es distribuirien exactament sobre la recta del gràfic *Q-Q Plot*. En aquest cas, pots veure que els valors més petits, però sobretot els més elevats, es desvien (s'allunyen) molt de la recta, de manera que probablement el problema el tindrem amb aquests valors, on la variable nombre d'amics pren els valors més elevats:



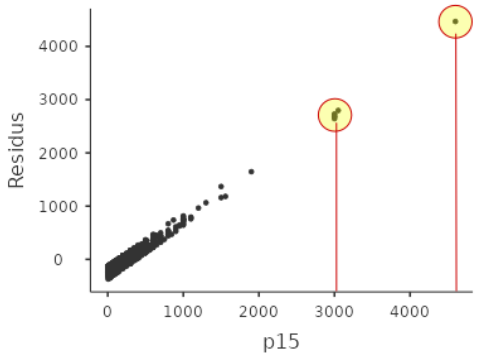
Imatge 140. Anàlisi de residus al model RLS: Gràfic Q-Q

El segon dels gràfics –*Residuals plots*– representa els residus (diferència entre el valor real i el valor ajustat/predit per la recta) respecte dels valors ajustats. Les condicions del model ens diuen que els residus (1) han de tenir una mitjana de zero i (2) una variància constant (homoscedasticitat). La variància serà constant si la distribució dels punts al voltant d'una recta horitzontal (que cal imaginar passant pel valor 0 de la escala vertical de *Residuals*) és més o menys homogènia. Si algun punt es troba molt per sobre o molt per sota de la recta de referència es tractarà d'un valor atípic. Veiem que hi ha uns quants casos molt allunyats de la recta i que ens impedeixen comprovar estrictament si la variància és constant:



Imatge 141. Anàlisi de residus al model RLS: Gràfic de Residus.

Analitzats els gràfics sembla clar que hi ha registres que poden condicionar molt els resultats i que poden ser un problema en tant que es desvien molt de la norma. Quins són aquests casos?



Imatge 142. Identificació de casos extremadament desviats.

Els anteriors gràfics semblen indicar que part del problema es pot resoldre si s'eliminen de l'anàlisi els casos identificats com a *extrems* (*outliers*), situats més enllà de 2 desviacions típiques del punt central. En concret, respecte a la variable p15 es pot identificar un cas proper al valor de 4.500 contactes a les xarxes socials i diversos casos propers a 3.000 contactes. Aquests valors, que són casos possibles, presenten la màxima desviació i ens condicionen fortament els resultats, de manera que podríem provar de treure'l de les anàlisis i comprovar què passa amb la resta.

Cal recordar que la recta de regressió (o funció predictiva) té la forma p15 = 635,2 - 17,4*p37. Per a calcular la regressió lineal sense tenir en compte els casos que es desviïn més de la mitjana (en termes de desviació estàndard), caldrà crear una nova variable a la matriu de dades. L'objectiu és IDENTIFICAR els casos que es troben fora del rang de desviació recomanat, que habitualment són 2 desviacions típiques.

Per fer això en *jamovi*, generem una nova variable on els casos més desviats els haguem definit com a valors perduts (NA). Ho farem en quatre passes:

1. A partir de l'edat, i aplicant la funció que hem obtingut (paràmetres de la recta), hem de calcular la predicció d'amics a les xarxes socials per a cada individu. Per fer-ho:
    a. Clica el nom de la variable p15.
    b. Clica el menú Dades / Calcular
    c. Al panell de càlcul, anomena la nova variable **p15_pred** i a la finestra de fórmula escriu el codi corresponent a "Si el valor de la variable amics (p15) és perdut, llavors aquesta variable també ho serà i, si no ho és, el valor d'aquesta variable serà el resultat de 635,2 menys 17,4 multiplicat pels anys d'edat (p37)". En codi R:
    `IFMISS (p15, NA, 635.2-17.4*p37)`
2. A continuació hem de calcular la desviació o diferència del nombre d'amics reals respecte dels amics predits per la recta:
    a. Clica el nom de la nova variable *p15_pred*,
    b. Clica el menú Dades / Calcular
    c. Al panell de càlcul, anomena la nova variable **p15_desv** i a la finestra de fórmula escriu el codi corresponent a "El valor d'aquesta variable és la diferencia absoluta (sense signe) entre el nombre d'amics observat (p15) i el que prediu la regressió (p15_desv)". En codi R:
    `ABS (p15-p15_pred)`

3. Cal calcular l'estandardització de la desviació anterior.
    a. Clica el nom de la variable *p15_desv*.
    b. Clica el menú Dades / Calcular
    c. Al panell de càlcul, anomena la nova variable **p15_desvEst** i a la finestra de fórmula escriu el codi corresponent a *"El valor d'aquesta variable serà el valor estandarditzat de la desviació del nombre d'amics real i predit*". En codi R:
    `Z (p15_desv)`
4. Anem ara a crear la variable final transformant *p15_desvEst* per tal que incorpori només els valors de p15 dins del rang de desviació acceptable (dues desviacions estàndards):
    a. Clica el nom de la variable *p15_desvEst*
    b. Clica el menú Dades / Calcular
    c. Al panell de càlcul, anomena la nova variable **p15_trim2SD** i a la finestra de fórmula escriu el codi corresponent a "Si el valor de la variable p15_desvEst és més gran de 2 o és més petit de -2, que la nova variable sigui NA i, en cas contrari, que sigui el valor que conté la variable p15". En codi R:
    `IF (p15_desvEst > 2 or p15_desvEst < -2, NA, p15)`

Observació! En realitat es podia haver fet en un sol pas, però s'ha detallat tota la seqüència per fer-ho més comprensible pas a pas. De fet, fer-ho en un únic pas és bastant complicat de seguir. Ara bé, les tres primeres instruccions sí que es poden escriure com una de sola:
`Z (ABS (p15 - IFMISS(p15,NA,635.2-17.4*p37)))`

"Retallar" una variable implica eliminar un conjunt de casos indesitjables per l'anàlisi. Aquesta seqüència de transformacions crea una variable final **p15_trim2SD** que és una versió de la p15 "retallada" en els seus extrems per sobre i per sota de 2 desviacions típiques i que pot ser incorporada a un nou model de regressió lineal simple.

Atenció! Comprova de quin tipus és la variable final **p15_trim2SD**. Si aquesta variable no és de tipus CONTINUA llavors no es podrà utilitzar dins d'una anàlisi de regressió lineal com a variable dependent. Recorda que el tipus de cada variable està representat per una petita icona al costat del seu nom a l'encapçalament de la columna corresponent.

Es pot aconseguir generar una versió CONTINUA de qualsevol variable ORDINAL, fins i tot amb les variables generades mitjançant càlculs sobre altres) aplicant una transformació molt simple a la mateixa.

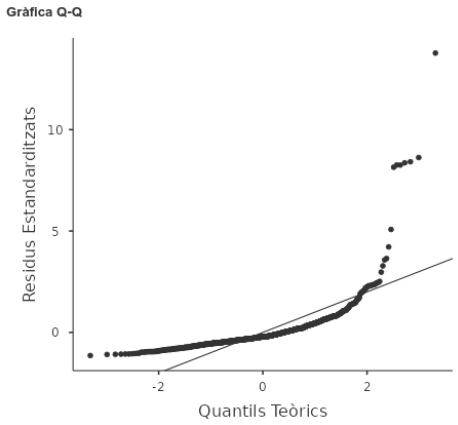
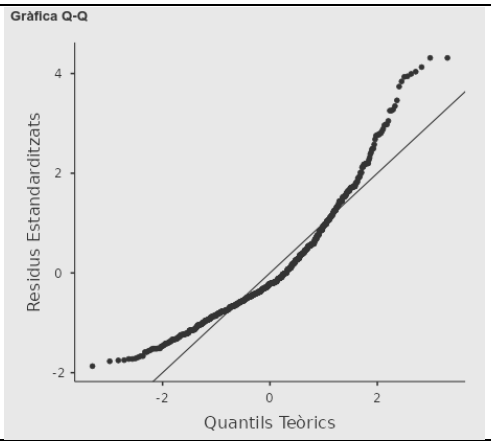
1. Clica el nom de la variable **p15_trim2D**.
2. Clica el menú Dades / Transformar
3. Al panell de transformació, clica a "utilitzant transformació" / Crear nova transformació.
4. A la finestra de transformació, estableix com a "Sufix de la variable" _C
5. No canviïs la fórmula predeterminada = $source.
6. Al desplegable "tipus de mesura", selecciona "Contínua".

Imatge 143. Canvi en el tipus de variable calculada, mitjançant transformació.

Ara la nova variable és contínua, i per tant ja es pot utilitzar. Repeteix l'anterior procediment de regressió lineal simple (MLS), que has fet abans per analitzar ara la influència de la variable edat (p37) sobre el nombre d'amics, però ara amb la versió retallada i contínua (p15_trim2SD_C). Si ho fas i compares els resultats anteriors amb els actuals veuràs que:

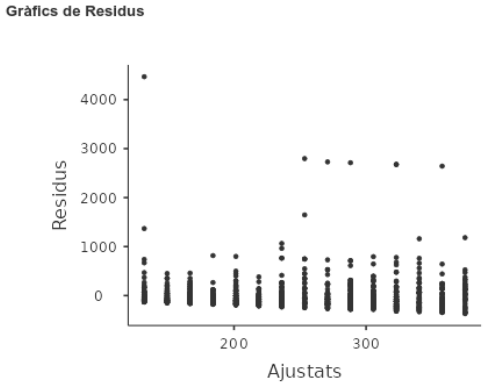
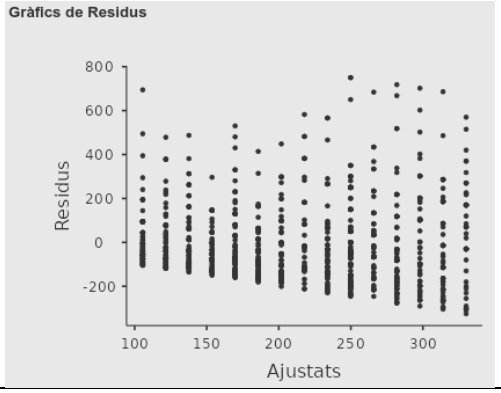
Model	**R**	**R^2 ajustada**	**Variància explicada**	**Recta de regressió**
Original	0,221	0,048	4,8%	Amics = 635,2 – 17,4(Edat)
Retallat	0,364	0,133	13,2%	Amics = 570,5 – 16,0(Edat)

I en quant als canvis en els supòsits del model, en concret per la normalitat:

Model original (p15)	Model retallat (p15_trim2SD_C)
	

Imatge 144. Canvis en la normalitat, després d'eliminar *outliers*.

Mentre que per a l'homoscedasticitat obtenim que:

Model original (p15)	Model retallat (p15_trim2SD_C)
	

Imatge 145. Canvis en la igualtat de variàncies, després d'eliminar *outliers*.

Podem interpretar, per tant, que és preferible i més acceptable el model amb la variable retallada, ja que és significativament més explicatiu, i compleix millor els supòsits del model de regressió lineal simple. La diferencia entre els dos models ha estat l'exclusió de 25 casos (un 2,3% de la mostra), els que mostraven una desviació superior a ±2 desviacions típiques.

EXERCICIS PROPOSATS: Regressió lineal simple

Després de seguir aquest apartat del tutorial, sobre regressió lineal simple, obre en *jamovi* el fitxer *MJDS_Dades_def.omv* i respon les següents qüestions:

(A) Volem saber si l'edat a la que els joves es connecten per primer cop a Internet ha canviat. Per fer-ho analitza la relació que es dona entre l'edat actual (p37) i l'edat que tenia en el moment de connectar-se per primer cop a Internet (p6). Abans de respondre les següents preguntes, comprova que la variable p6 no conté valors impossibles i, en cas que en tingui, passa'ls a NA (tingues en compte que es tracta de la variable *Edat d'inici d'ús d'Internet* i que l'enquesta s'ha fet a joves compresos entre els 15 i els 29 anys). A continuació, respon el següent:

(A.1) Quina és la variable dependent i quina la independent?

(A.2) En termes generals, el test que ens plantegem és el següent:

- H0: Entre la població jove, l'edat al connectar-se per primer cop a Internet, **no** ha canviat en el temps
- H1: Entre la població jove, l'edat al connectar-se per primer cop a Internet, ha canviat en el temps

Aquest test es pot formular en termes dels paràmetres del model. Quins són aquests paràmetres? Formula els tres tests, i arriba a conclusions sobre cadascun d'ells.

(A.3) Escriu la funció que explica una variable a partir de l'altra, i interpreta els elements de la funció.

(A.4) Representa gràficament la relació entre ambdues variables. Fes-ho a partir de l'*Rj Editor*, afegint un títol al gràfic, als eixos de coordenades i dibuixant la recta de regressió.

(A.5) Digues quant val el coeficient de determinació i interpreta'l.

(A.6) Estima quina és l'edat a la que van començar a fer servir Internet els que tenen 23 anys. Estima també l'edat d'inici a Internet pels que tenen 33 anys.

(A.7) Comenta breument els supòsits del model a partir de la seva representació gràfica.

(A.8) Seguint el mateix procediment que hem fet servir a la pràctica, elimina els individus que tinguin un residu estudentitzat fora del rang [–2, 2]. Comprova: (1) que els supòsits han millorat lleugerament i (2) que aconseguim augmentar una mica el percentatge d'explicació de la variable dependent.

(B) Moltes investigacions assenyalen que veure la televisió assíduament afecta al rendiment escolar dels nens i nenes. Per a contrastar aquesta qüestió, hem preguntat a una mostra de 10 estudiants del darrer curs de primària el nombre d'hores a la setmana que dediquen a veure la televisió. A més, també tenim la nota mitjana de l'última avaluació. Les dades obtingudes han estat:

Hores setmanals de TV	9	10	4	10	8	6	4	9	5	3
Nota mitjana	5	6	8	5	4	6	9	8	8	9

(B.1) Entra les dades al *jamovi* i dibuixa el núvol de punts. Digues si et suggereix una relació directa o inversa.

(B.2) Planteja el test d'hipòtesis que creguis més oportú i contrasta'l amb un 95% de confiança.

(B.3) Amb un 99% de confiança, a quina conclusió hauries arribat? Creus que aquest resultat i l'anterior són incongruents? Quin és el màxim nivell de confiança amb el que podries acceptar que hi ha relació entre ambdues variables?

(B.4) Quin percentatge de la variabilitat d'una variable és explicada per la variabilitat de l'altra? Com s'anomena a aquest valor?

(B.5) Calcula la recta de regressió i interpreta'n els resultats.

+ MENUS =

JAMOVI és un sistema de tractament de dades i anàlisi estadístic que et proporciona totes les funcionalitats bàsiques per a la recerca social amb mètodes quantitatius.
És software lliure, gratuït, multiplataforma (Windows, Linux, Mac) i ampliable mitjançant "mòduls".
Està basat en el llenguatge R, i és una bona forma d'introduir-se a l'ús d'aquest potent llenguatge d'anàlisi de dades.

Pots trobar molta més informació a <https://www.jamovi.org/download.html>

(cc) BY-NC