



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

MUJADES v1.3



2017-2021

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

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

GRIERA ARTIGUES, Oriol (Dep. de Sociologia, UAB)

Aquest material docent es publica sota llicència *Creative Commons*, 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.

Contingut

1. <i>jamovi</i> : Tractament i anàlisi de dades quantitatives	7
2. Instal·lació del programari <i>jamovi</i>	9
Mòduls d'extensió de <i>jamovi</i>	10
Com afegir el mòdul per a operar directament amb codi R?	10
3. Interfície gràfica de <i>jamovi</i>	14
Menús de les modalitats de treball.....	15
1. Menú d'arxius.....	15
2. Modalitat de gestió de variables	17
3. Modalitat de gestió de dades	18
4. Modalitat d'anàlisi de dades	20
5. Modalitat de comentaris de text.....	21
6. Activador de pantalla central	22
7. Menú de configuració de <i>jamovi</i>	23
4. Lectura i escriptura d'arxius de dades amb <i>jamovi</i>	25
Introducció directa de dades.....	25
Enganxar dades des d'altres aplicacions	28
Càrrega de les dades d'arxius.....	31
5. Definició de dades amb <i>jamovi</i> . Tipus de variables i valors perduts	35
Correspondència entre nivell de mesura de les variables i tipus de variable/dada en <i>jamovi</i>	35
Introducció a les no respostes. La definició de valors perduts en <i>jamovi</i>	39
La definició de valors perduts en <i>jamovi</i> (I). Variables nominals i ordinals.....	40
La definició de valors perduts en <i>jamovi</i> (II). Variables quantitatives	43
6. Estadística descriptiva i exploratòria bàsica amb <i>jamovi</i>	46
Exploració bàsica descriptiva de les variables	46
Exploració de variables amb codi R.....	48
Resultats de l'exploració de variables: Afegir comentaris i exportar	54
Modalitat comentari de text.....	54
Exportació de resultats cap a altres aplicacions	55
Descripció gràfica de variables	57
7. Transformació i depuració de dades mitjançant <i>jamovi</i>	63
Recodificació de variables.....	63
Recodificació (I). De variables quantitatives a qualitatives.....	64
Recodificació (II). De variables qualitatives a qualitatives	70
Recodificació (III). De variables qualitatives a quantitatives.....	71
Recodificació (I). De variables quantitatives a quantitatives	74
Càlculs amb variables quantitatives	74
Operacions i càlculs utilitzant una única variable	75
Operacions i càlculs entre variables diferents.....	76
Transformacions complexes (1): Combinació de recodificacions i càlculs.....	78
Transformacions complexes (2): Construcció de tipologies amb la funció IF.....	83

Filtrar casos.....	89
Detecció i depuració d'errors	91
Errors en variables individuals	91
Errors en la congruència entre variables	93
Exercicis proposats: Transformació de dades.....	95
8. Inferència amb una variable utilitzant <i>jamovi</i>	98
Construcció d'un interval de confiança per a una proporció (π).	98
Construcció d'un interval de confiança per a la mitjana (μ).	101
Codi R per a la construcció d'interval de confiança (mitjana).	102
9. Inferència de relació entre variables	104
10. Anàlisi de taules de contingència mitjançant <i>jamovi</i>	106
L'estadística descriptiva en les taules de contingència.	107
La representació gràfica de les taules de contingència	112
La inferència estadística amb dues variables no quantitatives: El test de la khi quadrat de Pearson.	114
A) Les condicions d'aplicació en el test de la khi quadrat de Pearson.....	115
B) Test de significació estadística khi quadrat de Pearson	116
C) Intensitat de l'associació entre variables nominals (V de Cramer).	117
D) Avaluació del sentit de l'associació	118
Solució als problemes de compliment de les condicions d'aplicació.....	123
La taula de contingència 2*2 (la correcció de Yates).	124
11. Comparació de mitjanes amb <i>jamovi</i>	128
Quin test cal fer?	129
L'estadística descriptiva en la comparació de mitjanes.	131
El test de Shapiro-Wilk per a la Normalitat de la variable.	134
La comparació de dues mitjanes independents.....	136
La comparació de més de dues mitjanes independents: L'anàlisi de la variància (ANOVA).	141
La intensitat de la relació: El coeficient de determinació.	145
Exercicis proposats: Comparació de mitjanes	147
12. Regressió lineal simple mitjançant <i>jamovi</i>	149
Descripció de la relació entre dues variables quantitatives.....	149
El diagrama de dispersió	150
La correlació entre ambdues variables	154
El model de regressió lineal simple.....	154
Inferència estadística en els models de regressió.....	157
Els supòsits del model: L'anàlisi de residuals.	158
Exercicis proposats: Regressió lineal simple	162
ANNEX. Del paper a la matriu de dades: Flux de treball amb dades de qüestionaris	164
Eines	164
Google FORMULARIS (GF)	164
Prerequisits.....	164

Operació (5 fases).....	164
Flux de Tasques	165
(A) Construcció del Qüestionari	165
(B) Aplicació del Qüestionari.....	167
(C) Gestió de les Respostes	167
(D) Exportació de les Respostes	168
(E) Importació de les Respostes a <i>jamovi</i>	169

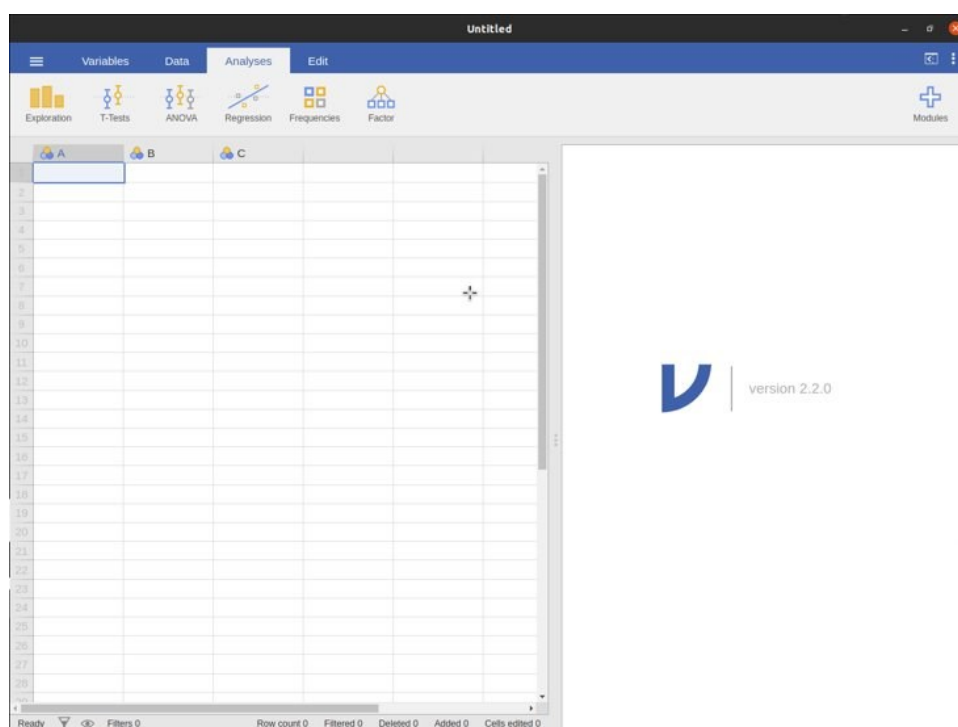
Aquesta versió del tutorial utilitza imatges de pantalla de la versió 2.2.0 de *jamovi*.



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 quantitativs. 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’executa el codi d’R associat, de forma transparent pels usuaris. Així es pot aprofitar tota la potència analítica d’R sense la necessitat de conèixer en detall el seu funcionament, programació, instal·lació i manteniment.

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*).

Imatge 2. Esquema funcional de *jamovi*

Tanmateix, *jamovi* permet també “programar” **directament** utilitzant en llenguatge R, mitjançant un dels mòduls disponibles per ampliar les capacitats bàsiques del sistema.

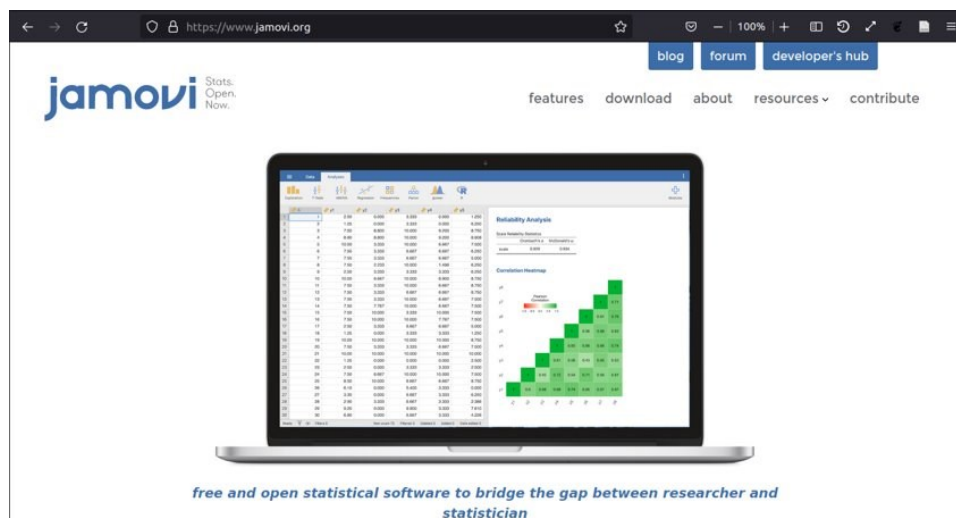
Als diferents tutorials MUJADES s'introduirà, el primer lloc, les diferents operacions mitjançant el sistema de menús i, a continuació, es mostrarà el codi R per aconseguir el mateix resultat (bé sigui al mòdul Rj de *jamovi*, o directament a R). D'aquesta forma, *jamovi* esdevé una eina per a l'anàlisi de dades quantitatives tant pel seu ús directe com per la seva utilitat per 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 recomana complementar aquest tutorial amb els següents recursos (*gratuïts, en anglès*):

- 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://blog.jamovi.org/2018/10/25/learning-statistics-with-jamovi.html>

2. Instal·lació del programari *jamovi*

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



Imatge 3. Website de jamovi (6 oct. 2021)

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.
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.



Imatge 4. Icona de *jamovi*

² A data d'octubre de 2020 la versió de demostració està accessible a <<https://cloud.jamovi.org/>>, amb la versió més actualitzada però amb capacitat de donar serveis a poques connexions alhora. Es recomana la instal·lació de versions locals per a treballar amb seguretat.

Mòduls d'extensió de jamovi

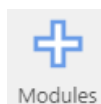
L'aplicació *jamovi* inclou un conjunt de procediments d'anàlisi estadística inicial que es pot ampliar amb els anomenats "modules". Dependent del llenguatge R, *jamovi* segueix la mateixa filosofia:

- 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 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 complets procediments estadístics propis i especialitzats en disciplines concretes (que requereixen del seu propi manual d'instruccions i interpretació).

Els mòduls estan disponibles, i 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 ja funcionen des de la instal·lació local.

La pàgina web <<https://www.jamovi.org/library.html>> manté un llistat dels mòduls. També és possible descarregar-los directament (per a diversos sistemes operatius) i llavors fer la "instal·lació local" (*library sideload*) a *jamovi*. Dins de la mateixa aplicació *jamovi* hi ha una opció de menú "Modules" que permet la gestió d'aquestes expansions, inclosa la seva descàrrega si es disposa de connexió WWW.



Imatge 5. Accés a la gestió d'extensions

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

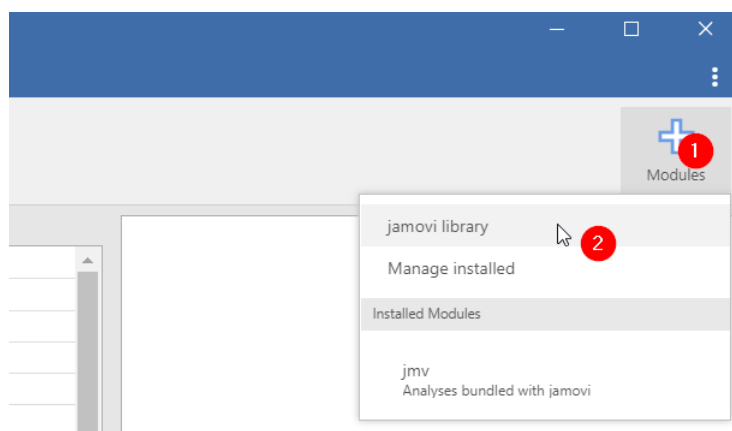
Com afegir el mòdul per a operar directament amb codi R?

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 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 jamovi**".

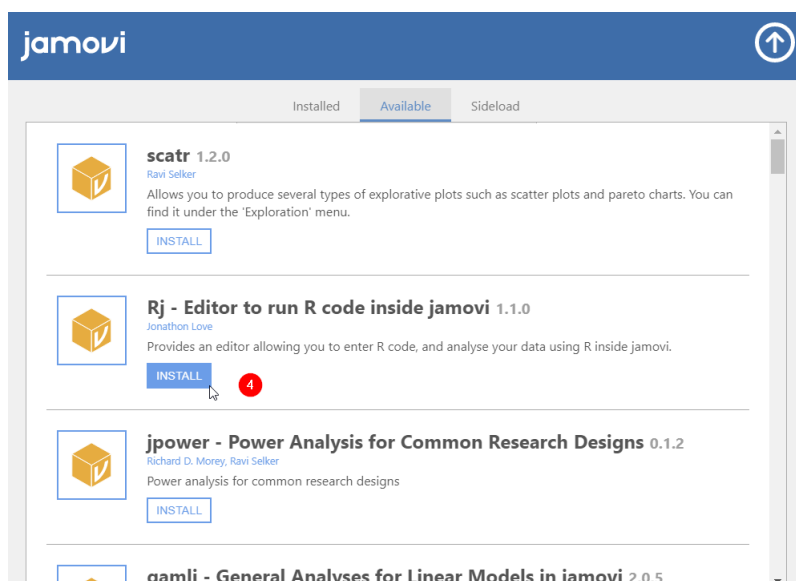
Per fer-ho, pas a pas:

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



Imatge 6. Accés al sistema de gestió d'extensions

- 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 7. Catàleg d'extensions (*jamovi library*)

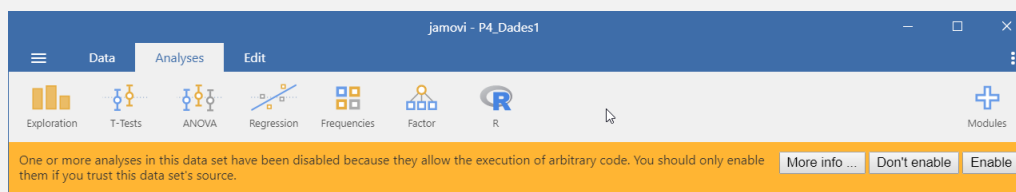
- Clica el botó [INSTALL]. Apareix un missatge mentre s'està fent la descàrrega. Quan culmina el procés, el catàleg indica “INSTALLED”.
- Tanca la finestra del catàleg de mòduls (*jamovi library*) clicant la icona fletxa-amunt-encerclada de la part superior dreta.
- Al menú gràfic de *jamovi* ha aparegut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

Imatge 8. Menú gràfic ampliat amb mòdul *Rj-Edit*

Ara el mòdul corresponent ha quedat instal·lat al 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 cas de missatge d'error o advertiment, cal respondre [Enable].



Imatge 9. 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:

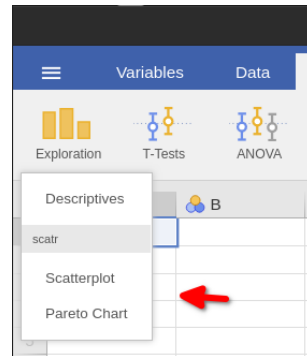
1. A la pantalla principal de *jamovi*, al menú gràfic superior, extrem dret, clica la icona "Modules".
2. Al desplegable selecciona l'opció "Manage Installed". Apareix un catàleg de les extensions instal·lades localment al sistema, amb possibilitat d'amagar-les del menú [HIDE], i de desinstal·lar-les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

Pots comprovar que hi ha una extensió impossible d'eliminar... "**jmv – Analyses 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 operativa).



EXERCICI

1. A la pantalla principal de *jamovi*, al menú gràfic superior, extrem esquerra, clica la icona "Exploration". 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 "Exploration" ?
(Aquestes dues noves opcions s'utilitzaran al llarg del tutorial, no continuïs fins a tenir-les disponibles).



Imatge 10. Menú de Descriptiva, ampliat amb mòdul “sctr”

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

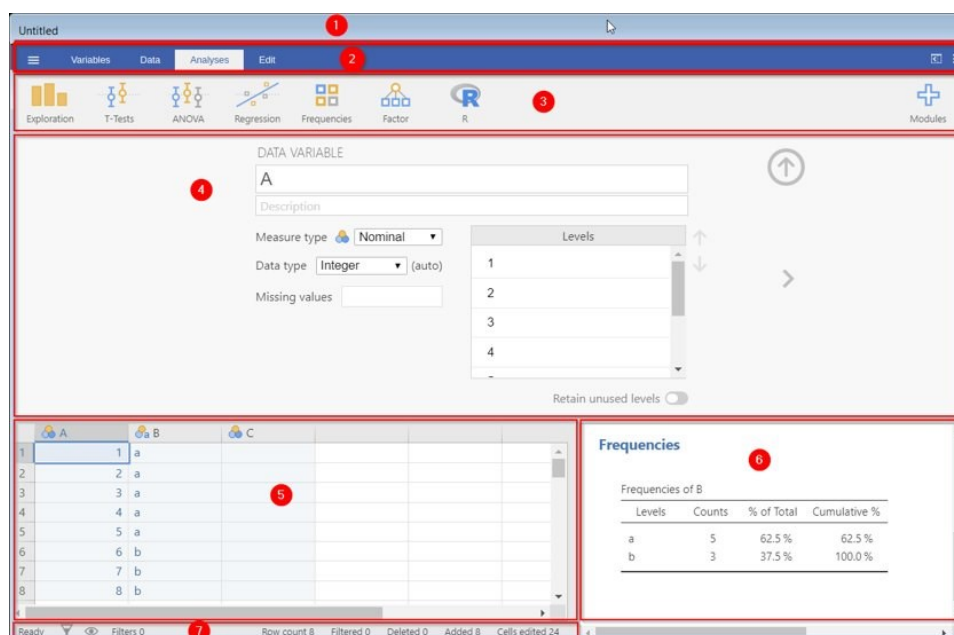
- A) trobis que alguns mòduls d'extensió ja estiguin prèviament disponibles, i
- B) si instal·les nous mòduls al sistema “desapareguin” en reiniciar la màquina (*per les necessàries polítiques de seguretat dels PCs públics*).

3. Interfície gràfica de *jamovi*

L'aplicació *jamovi* no és més que una complexa interfície gràfica d'usuari (GUI) per executar operacions en el llenguatge de programació R. Així, 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 amigable i amb màxima facilitat d'ús sense haver d'aprendre programació.

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 11. 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.

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 12. Menús d'opcions generals de *jamovi*

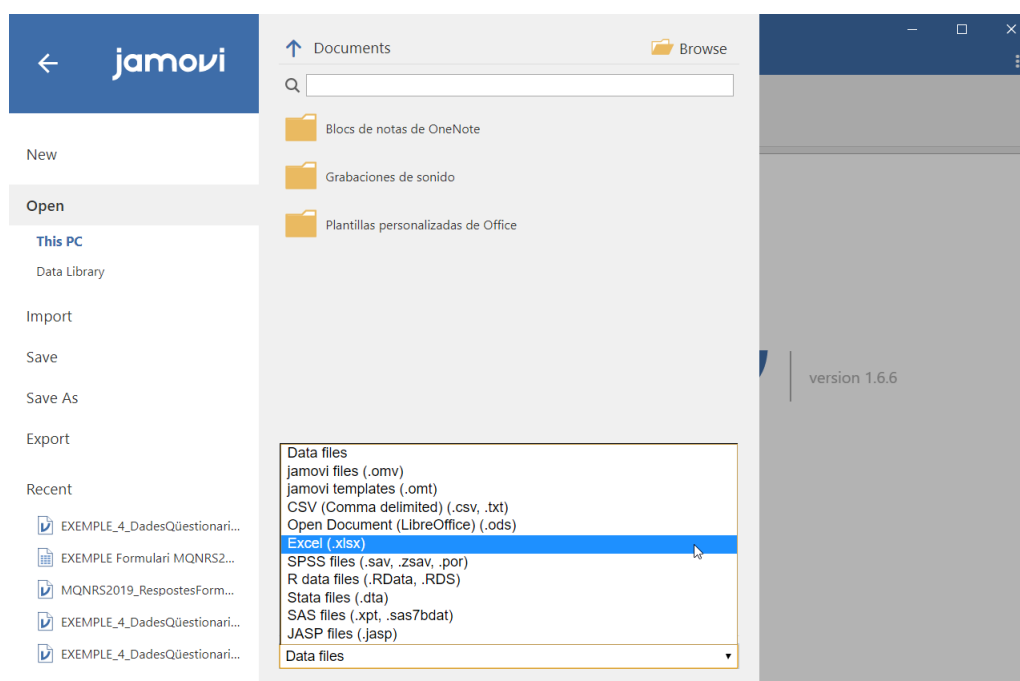
La barra de color blau conté els activadors d'aquest menú de modalitat i altres opcions, respectivament, d'esquerra a dreta:

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. “Data” Activa la modalitat de GESTIÓ de DADES: Per a treballar amb les dades i variables.
4. “Analyses” Activa la modalitat d'ANÀLISI de DADES: Per a demanar resultats d'operacions.
5. “Edit” 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.
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).



Imatge 13. Menú d'arxius

- **New:** Obre una nova instància de *jamovi*, sense tancar l'actual.
- **Open:** 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 reproduïen totes les operacions que es van fer en l'arxiu original fins al moment dels seu "tancament".

- **Import:** 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.
- **Save / Save As:** 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 identifications 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, "Save" el desarà amb aquest nom a la mateixa ubicació actualitzant l'anterior.
- **Export:** Genera diferents tipus d'arxius amb continguts específics, segons el format seleccionat per a l'exportació:

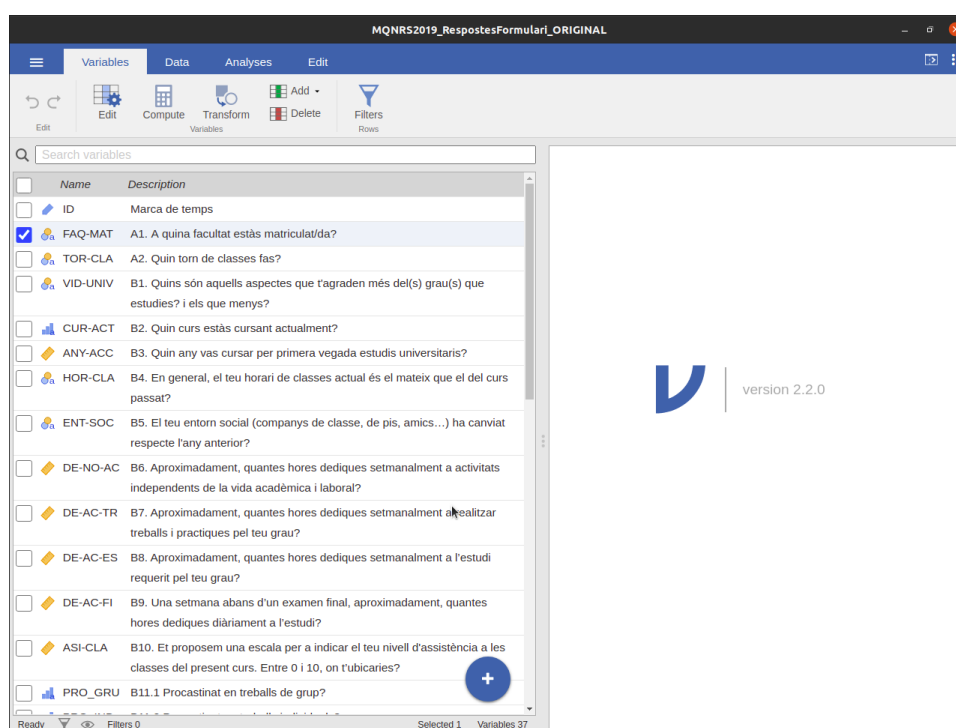
- Resultats: En el cas d'exportacions cap a PDF, a HTML, o a LaTeX (bundle zip), es poden generar pàgines amb totes les taules i gràfics dels resultats i amb els textos adjunts.
- Dades: En el cas d'exportacions cap a formats d'aplicacions d'anàlisi (SPSS, Stata, SAS, CSV o R) el què es genera són arxius 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) reproduïx de nou totes les operacions que es van realitzar anteriorment, per la qual cosa en ocasions pot trigar minuts en completar l'obertura. 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 14. Modalitat gestió de variables

- **Edit:** 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 ("levels").
- **Compute:** 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).

- **Transform:** 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).
- **Add / Delete:** 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.
- **Filters:** 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 l’encapçalament de columna 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 / deseleccionar 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’organitzcen visualment com una matriu de files i columnes.

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’escala combinades).

The screenshot shows the Jamovi software interface with a data matrix on the left and a 'Descriptives' panel on the right. The data matrix has columns labeled C-FI, ASI-CLA, PRO-GRU, PRO-IND, PRO-EXA, CAR-AVA, and ES. The 'Descriptives' panel displays summary statistics for the selected variables.

Descriptives

	PRO-EXA	PRO-GRU	ASI-CLA
N	535	535	536
Missing	1	1	0
Mean			7.71
Median			8.00
Standard deviation			1.95
Minimum			0
Maximum			10

Frequencies

Frequencies of PRO-EXA

Levels	Counts	% of Total	Cumulative %
Mai	134	25.0 %	25.0 %
Algunes vegades	220	41.1 %	66.2 %
Moltes vegades	137	25.6 %	91.8 %
Sempre	44	8.2 %	100.0 %

Frequencies of PRO-GRU

Levels	Counts	% of Total	Cumulative %
Mai	209	39.1 %	39.1 %
Algunes vegades	215	40.2 %	79.3 %
Moltes vegades	87	16.3 %	95.5 %
Sempre	24	4.5 %	100.0 %

Imatge 15. Modalitat gestió de dades

- **Clipboard:** 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 porta-papers).
- **Edit:** Registre històric per a desfer o refer accions d'edició de dades.
- **Variables**
 - **Setup:** 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 (*"Levels"*).
 - **Compute:** 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).
 - **Transform:** 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).
 - **Add / Delete:** 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.
- **Rows (files, casos)**

- **Filters:** 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").
- **Add / Delete:** 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.



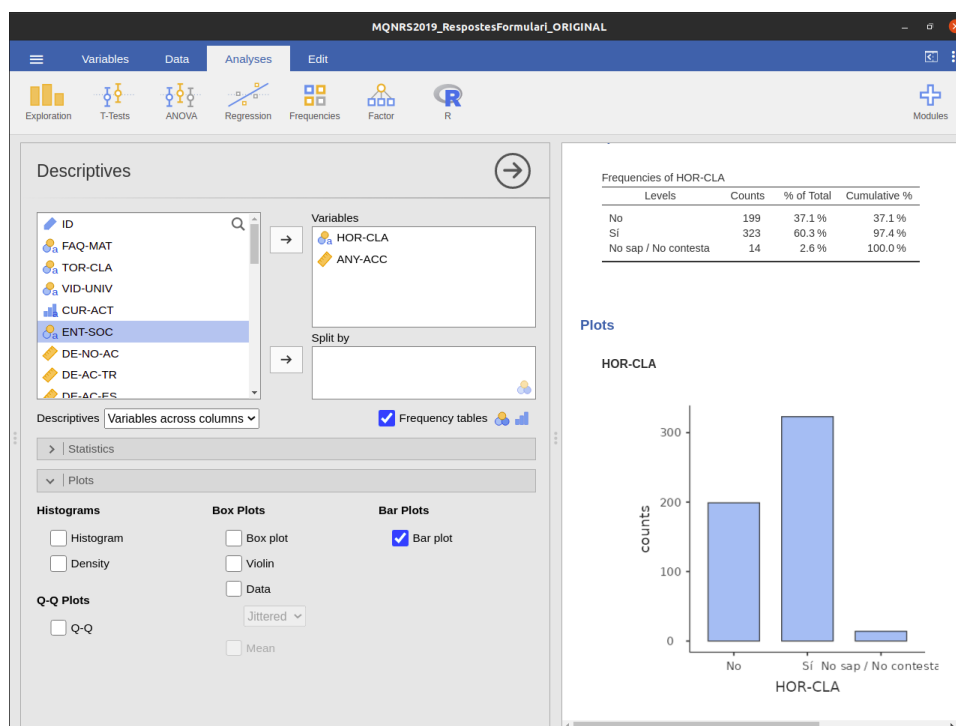
A efectes de llenguatge R, les dades de la matriu de *jamovi* són un objecte R del tipus **data.frame** anomenat **data\$**

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 16. Disposició en modalitat d'operacions estadístiques

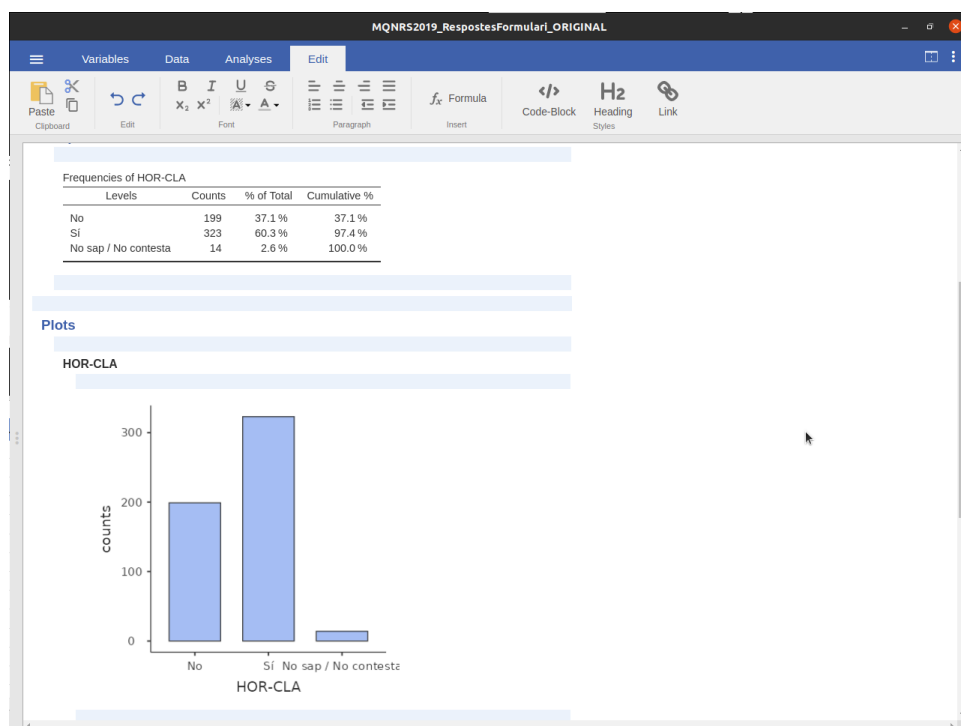
Clicant les icones del submenú gràfic desplega un menú amb els procediments estadístics disponibles, que han de ser utilitzats específicament per a condicions concretes de les dades. Correspon a qui realitza l'anàlisi conèixer les condicions d'aplicabilitat de cadascun dels procediments: *jamovi* no 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.

5. Modalitat de comentaris de 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 17. 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 "Edit", 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 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 complet** 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 món científic: "complet", "transparent" i "compacte".

Per aquest motiu es demanarà prioritàriament com a format als lliuraments que calgui fer.

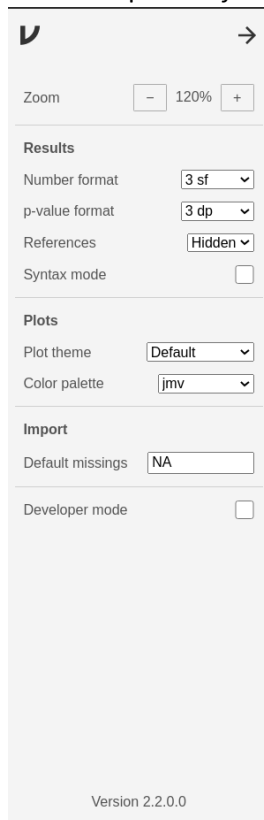
6. Activador de pantalla central

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.

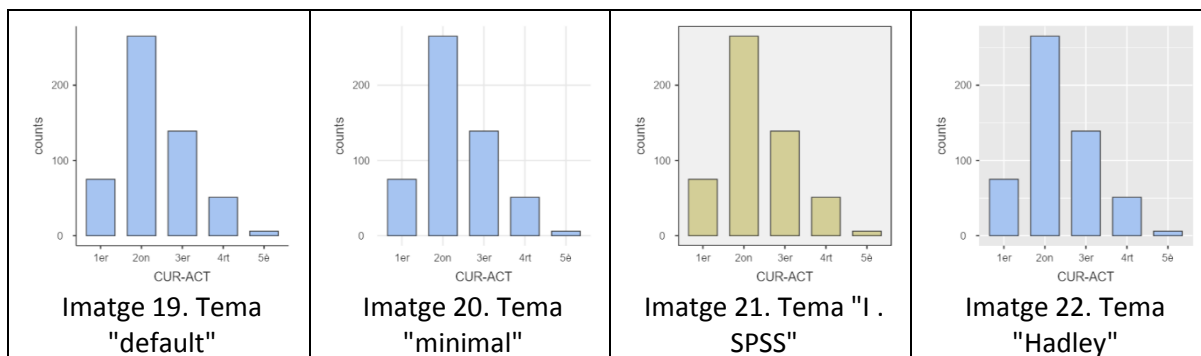
7. Menú de configuració de *jamovi*

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 18. Menú de configuració

- **Zoom:** Permet ajustar la mida de visualització del conjunt d'elements de l'aplicació.
- **Results**
 - **Number format:** Permet triar el format dels decimals a les taules, bé en forma “relativa” de díigits significatius (sense tenir en compte els zeros a l'esquerra, així el format 3sf genera 0,123 i 0,00458) o en forma “absoluta” (arrodonint a un nombre determinat de decimals inclosos els zeros, així el format 3dp genera 0,123 i 0,005).
 - **p-value format:** Permet igualment triar el format dels decimals pels càlculs de valors-p de significativitat, bé en forma “relativa” o en forma “absoluta”.
 - **References:** 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.
 - **Syntax mode:** 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.
- **Plots**
 - **Plot Theme:** Diverses configuracions pels resultats gràfics generats.



- **Color palette:** Diversos conjunts de colors aplicables als resultats gràfics.
- **Import**
 - **Default missings:** 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.
 - **Developer mode:** 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.

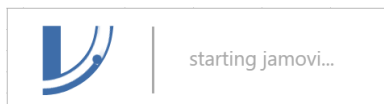
A continuació, el tutorial t'acompanyarà en la realització del següent cas pràctic:

- 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.



Imatge 23. Indicador d'inici

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 la variable**, assignant-li un nom (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...

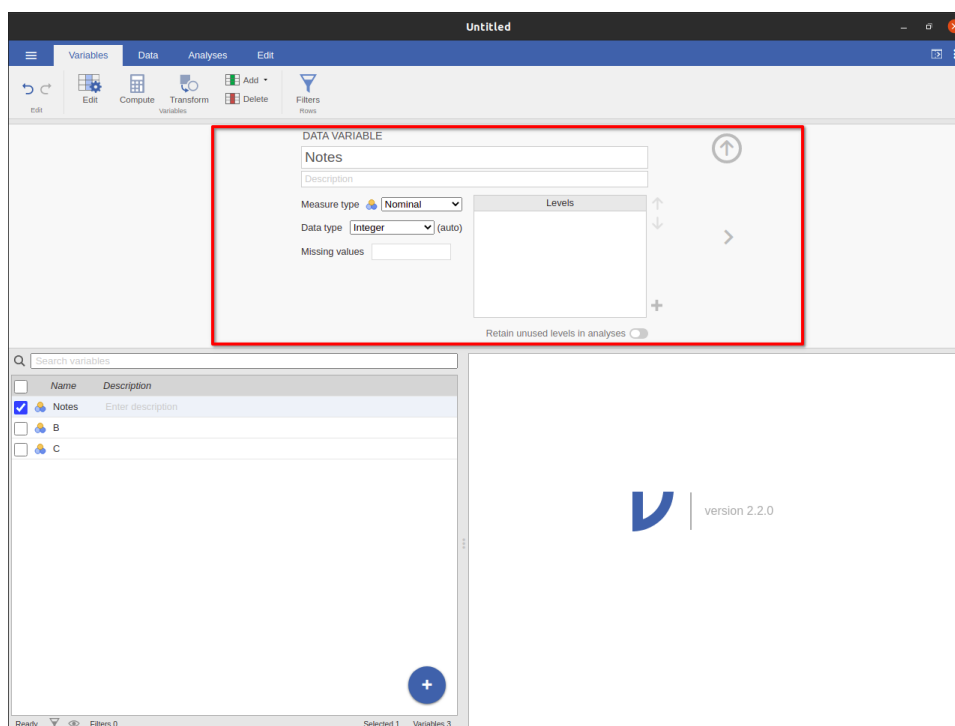
OPCIO A: Clica el menú / Variables

Tecleja el nom desitjat, substituint les lletres assignades automàticament pel sistema, per exemple: *Notes*. A continuació fes doble-clic a qualsevol lloc de la primera línia o variable.

OPCIO B: Clica el menú / Data / Setup

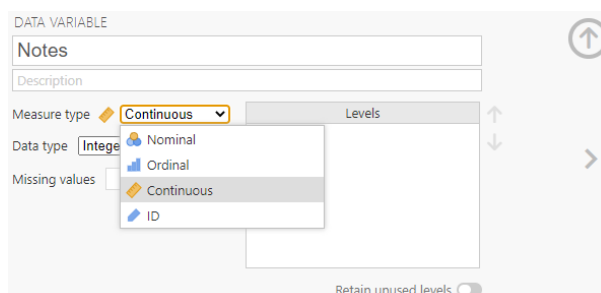
(alternativament, doble-clic a l'encapçalament de la primera columna (lletra A))

4. S'obre l'àrea de definició de variables.



Imatge 24. 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). Es recomana 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 "Description" 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 "Measure type" (tipus de variable) clica el desplegable i selecciona "Continuous".³
9. A "Data type" (nivell de mesura de les dades) selecciona "Integer" (és l'opció per defecte)
10. Pots continuar definint d'altres variables, clicant a les icones laterals > i < (alternativament, fent clic a l'encapçalament de la columna corresponent)
11. Pots tancar l'àrea de definició de variable clicant la fletxa-amunt-encerclada.
12. En el cas de variables de tipus nominal o ordinal, caldrà comprovar les categories de la variable ("Levels"). Veure en detall el següent capítol "5. Definició de dades amb *jamovi*: Tipus de variables i valors perduts".



Imatge 25. Definició d'una variable numèrica sense decimals

³ Per a més informació sobre "Measure Type" i "Data Type", consulta l'apartat posterior "5. Definició de dades amb *jamovi*: Tipus de variables i valors perduts".

13. Comprova que (A) la línia superior de la definició de variables, o (B) l'encapçalament de la primera columna, inclou el nom, 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 ("Cut") 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.

	Notes	B	C
33	4		
34	8		
35	5		
36	8		
37	8		
38	4		
39	2		
40	8		
41	1		
42	8		
43			
44			

Imatge 26.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, on fora millor una introducció "per files" que no està disponible.

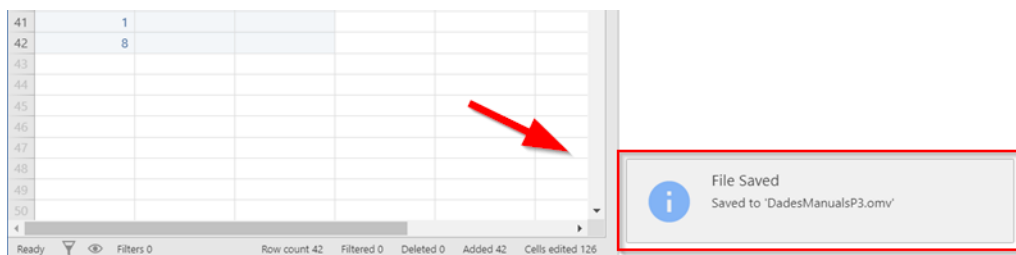
RECOMANACIÓ: Si has d'introduir manualment les dades de diferents qüestionaris en paper, és millor que (1) ho facis prèviament en un full de càlcul (no directament en *jamovi*), ja que permet la introducció contínua per files i, (2) posteriorment, utilitzis qualsevol dels propers dos procediments per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

Tot i que no forma part del procediment d'introducció de dades, avancem en el procediment de DESAR les dades introduïdes en un arxiu nou en el format propi de *jamovi*.

Per a desar un arxiu propi de *jamovi*:

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

3. Al desplegable, selecciona l'opció "Save". S'obrirà una finestra d'emmagatzematge.
4. Tecleja el nom desitjat, p. ex., "DadesManuais" (recomanable sempre sense espais)
5. Clica "Browse", per a seleccionar la ubicació 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 27. Missatge indicació procés de desar arxius

8. Comprova que la barra superior de l'aplicació ara indica "jamovi - DadesManuais".

Enganxar dades des d'altres aplicacions

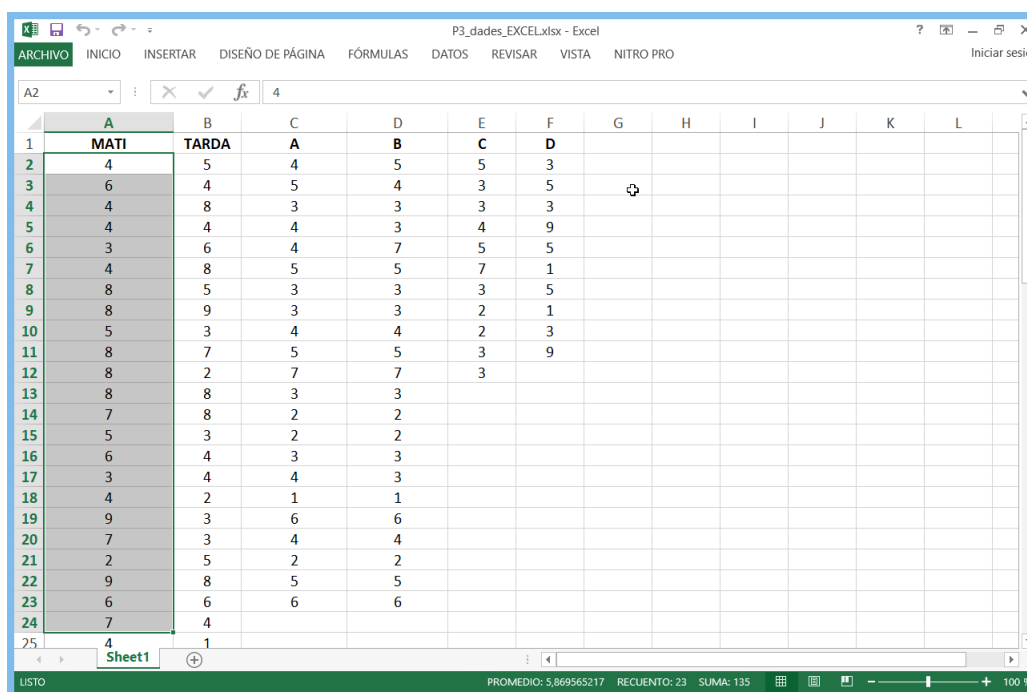
En ocasions interessa analitzar amb *jamovi* dades que es troben disponibles directament en altres aplicacions, o publicades en format electrònic o en línia. Per exemple, les dades poden ser a un arxiu de full de càlcul, o publicades a una pàgina web.

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 integrar dins d'una matriu de *jamovi*.

A l'espai docent Campus Virtual (CV) de l'assignatura hi pots trobar l'arxiu *MJDS_dades_EXCEL.xlsx*, un full de càlcul on hi ha dades corresponents a diverses distribucions.

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



	A	B	C	D	E	F	G	H	I	J	K	L
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 28. Exemple de dades disponibles a un full de càlcul

Observa que la columna B conté les mateixes dades de qualificació del curs pel grup de tarda que anteriorment has introduït manualment. A la columna A hi ha dades corresponents a la mateixa prova pel grup de matí de l'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 matí o a la tarda. Tot plegat s'utilitzarà per a practicar les operacions disponibles amb *jamovi* i el porta-papers del sistema operatiu.



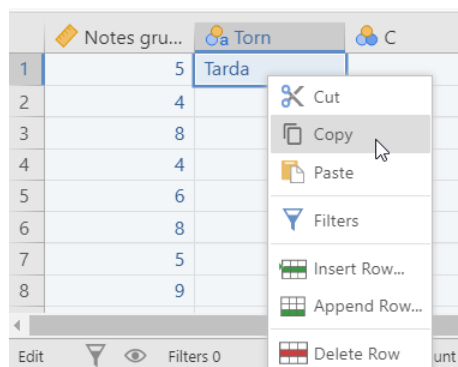
EXERCICI

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

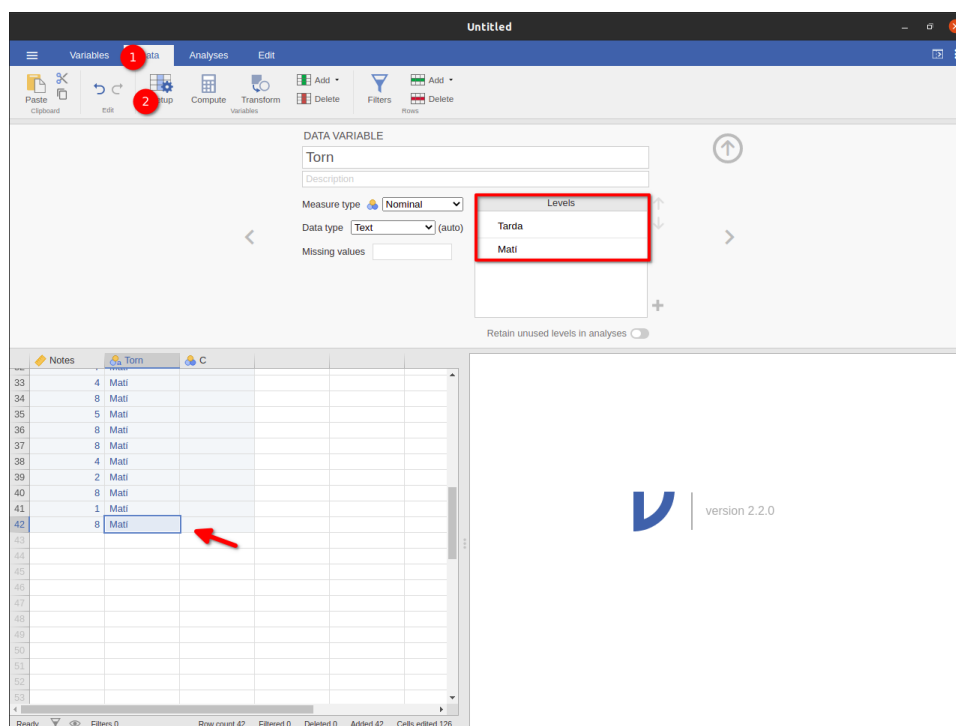
1. A la finestra de dades, clica la casella (buida) del primer individu per a aquesta segona variable "Torn" que has creat.
2. Introdueix manualment el valor de la categoria corresponent: Tarda.

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”.
4. Es desplega el menú contextual. Selecciona l’opció “Copy” (alternativament, combinació de tecles [Ctrl]+[C]).



Imatge 29. 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 “Paste” 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 A24).
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 grup tarda”.
10. Clic-dreta, al menú contextual opció “Paste” (alternativament, combinació de teclat [Ctrl]+[V]). Ara totes les dades 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 + “Copy” + seleccionar múltiples cel·les + “Paste”) per acabar d’introduir ràpidament les dades dels 98 individus.



Imatge 30. Retalla/Enganxa dades matriu

Si ara obres la finestra d'identificació de variables, amb Menú / Variables, o amb Menú / Data / Setup (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 problemes d'errors de teclat.

Compte! Per a *jamovi* “Matí” és un valor (“Level”) totalment diferent de “matí” o de “Mati”.

Càrrega de les dades d'arxius

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

ID	FAQ-MAT	TOR-CLA	VID-UNIV	CUR-ACT	ANY-ACC	HOR-CLA	ENT-SOC	DE-NO-AC	DE-AC-TR	
1	10/24/2019 ...	Facultat de ...	De tarda	Li agrada la ...	2on	2017	No	No	7	4
2	10/24/2019 ...	Facultat de ...	De tarda	Li agrada qu...	2on	2017	No	No	8	3
3	10/24/2019 ...	Màrqueting i ...	De mati	Menys: mate...	2on	2018	Sí	No	14	5
4	10/25/2019 ...	Escola d'En...	De mati	Le gusta el h...	2on	2018	Sí	Sí	27	0
5	10/25/2019 ...	EUSS Engin...	De mati	Més: Les prà...	2on	2015	Sí	No	14	2
6	10/25/2019 ...	Estudis d'Ar...	De mati	Més: les assi...	1er	2019	Sí	Sí	10	27
7	10/25/2019 ...	Facultat de ...	De mati	Més: Tota la ...	2on	2018	Sí	No	25	10
8	10/26/2019 ...	Facultat de ...	De tarda	Menys: histò...	1er	2019	Sí	Sí	10	2
9	10/26/2019 ...	Facultat de ...	De mati	El que més ...	2on	2018	No	No	14	7
10	10/26/2019 ...	Facultat de ...	De mati	El que més ...	2on	2018	Sí	Sí	12	5
11	10/26/2019 ...	EINA (adscri...	De mati	No m'agrada...	2on	2018	No	No	30	5
12	10/27/2019 ...	Facultat de ...	De mati	més: literatur...	2on	2018	Sí	No	3	10
13	10/28/2019 ...	Facultat de ...	De mati	El que més: ...	3er	2017	Sí	No	0	10
14	10/28/2019 ...	Facultat de ...	De mati	El que més: l...	4rt	2014	Sí	Sí	0	1
15	10/28/2019 ...	Facultat de ...	De mati	El que més l...	2on	2018	No	No	7	2
16	10/28/2019 ...	Facultat de ...	De tarda	el que meny...	2on	2018	Sí	Sí	34	3
17	10/28/2019 ...	Facultat de ...	De tarda	més: interpr...	2on	2018	Sí	No	6	1
18	10/28/2019 ...	Facultat de ...	De tarda	Menys: molt ...	2on	2018	Sí	No	10	7
19	10/28/2019 ...	Facultat de ...	De tarda	més: alguna ...	2on	2018	Sí	No	8	1
20	10/28/2019 ...	Facultat de ...	De tarda	el que meny...	2on	2018	Sí	No	3	3
21	10/28/2019 ...	Facultat de ...	De tarda	Mes: es el c...	2on	2018	Sí	No	14	14
22	10/28/2019 ...	Facultat de ...	De mati	M'agrada la ...	2on	2018	Sí	No	0	7
23	10/28/2019 ...	Facultat de ...	De tarda	Lo que mas l...	1er	2019	No	Sí	5	2
24	10/28/2019 ...	Facultat de ...	De tarda	El que més l...	2on	2018	Sí	Sí	2	16
25	10/28/2019 ...	Facultat de ...	De tarda	El que més l...	2on	2018	No	Sí	16	16
26	10/29/2019 ...	Facultat de T...	De mati	El que més: l...	2on	2018	Sí	Sí	3	10
27	10/29/2019 ...	Facultat de T...	De mati	El que més: ...	4rt	2012	Sí	No	14	5
28	10/29/2019 ...	Facultat d'Ec...	De tarda	El que més: ...	4rt	2016	No	No	2	6
29	10/29/2019 ...	Facultat d'Ec...	De tarda	El que més: ...	2on	2018	Sí	Sí	10	5

Imatge 31. Matriu de dades a jamovi

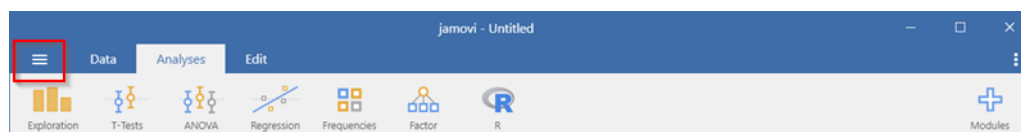
L'opció del menú d'arxius "Import" permet AFEGER dades **noves a les ja existents** a l'espai de treball. D'aquesta forma es poden afegir casos o variables per a poder AJUNTAR diferents arxius "parcials" de dades fent servir l'opció d'importació i seleccionant (amb Ctrl+Clic) els diferents arxius que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

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*.

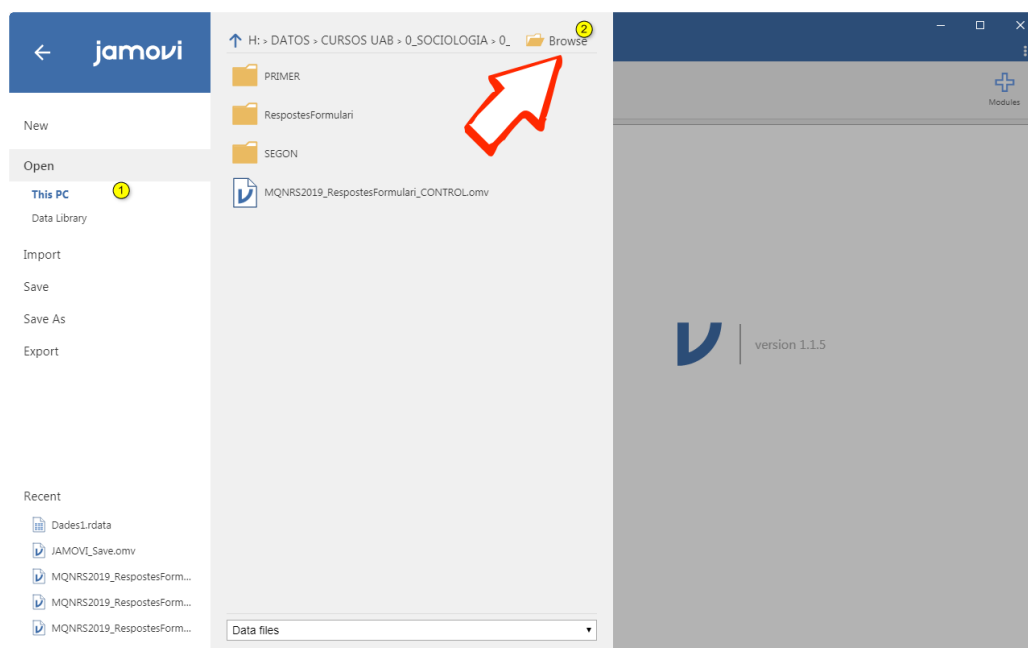
A l'espai docent Campus Virtual (CV) de l'assignatura hi pots trobar l'arxiu *MJDS_Dades.rdata*, un arxiu de dades corresponents a una enquesta del CIS.

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 32. Ubicació menú d'arxius

5. Selecciona l'opció Open / This PC. Al panell central, clica [Browse] (dalt, dreta)



Imatge 33. 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 i 70 variables.

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 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) ⁴



EXERCICI

- A l'espai Moodle corresponent (PRÀCTICA 3) pots trobar dos arxius format *jamovi* que contenen dades de la mateixa enquesta del CIS: *MJDS_Dades0_A.omv* i *MJDS_Dades0_B.omv*.
En el moment d'entrar les dades es va repartir la tasca entre dues persones o dues empreses de servei, per accelerar el procediment (*gravació de dades en paral·lel, per*

⁴ *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>>.

estalviar temps), i ara cada arxiu conté les MATEIXES variables i UNA PART dels casos d'enquestes contestades.

- 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*. Utilitza l'opció "Import" del menú d'arxius, tot seleccionant **ahora** els dos arxius. Comprova que el total sigui de 1.419 enquestes, i desa l'arxiu unificat amb el nom que vulguis.

NOTA: En la versió actual la IMPORTACIÓ 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.

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, des del punt de vista del nivell de mesura, amb quin tipus de variable ens trobem. 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 en *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 hi pots trobar els arxius *MJDS_QüestionariCIS.pdf* (que reproduïx una mostra del 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 de tot el tutorial.
3. Obre amb *jamovi* l'arxiu amb la matriu de dades (respostes),
4. Comprova en el qüestionari del CIS que hi ha exemples de variables dels diferents nivells 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).⁵

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:

		Tipus de variable			
		"Continuous" 	"Ordinal" 	"Nominal" 	"ID" 
Tipus de les dades	"Integer"	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:

⁵ 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.

- **“Continuous”**: 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 textuais corresponents 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 convenient tenir definit de forma adient el tipus de variable per poder realitzar els procediments pertinents.

Segons el tipus **de les dades**, *jamovi* distingeix entre:

- **“Integer”**: 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.⁶
- **“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 la seva forma interna d'emmagatzemar les dades, la qual cosa té conseqüències en la generació de resultats en els processos 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.

En quant a la correspondència entre el nivell de mesura i les classes de variables/dades en *jamovi*, és útil 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=“Continuous” i data=“Integer”, aquesta definició en pot portar algun problema, ja que si la mesura és

⁶ 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 .

Continuous no permet, per exemple, fer una taula de freqüències. Per aquest motiu, en un moment inicial les definirem com a mesura="Ordinal" i data="Integer" i avaluarem, en alguns casos, la possibilitat de canviar la definició al llarg de l'anàlisi.

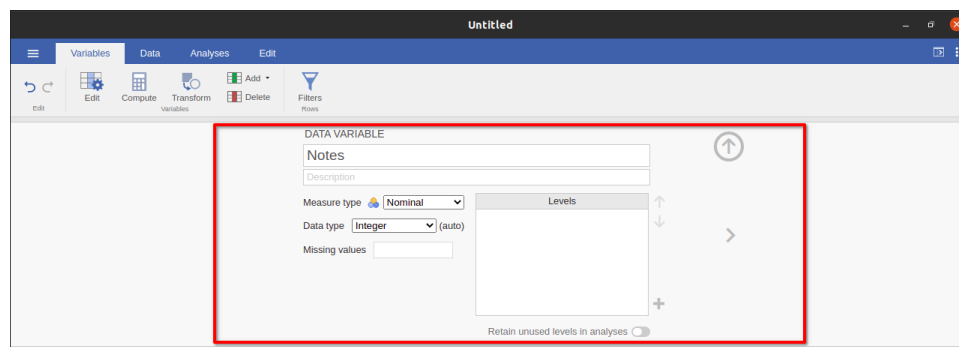
RECORDA: A *jamovi*, les variables nominals son de mesura Nominal amb dades tipus Text, les variables ordinals son de mesura Ordinal amb dades tipus Text, i les variables quantitatives son de mesura Ordinal amb dades Integer (o Decimal si tenen decimals).

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.

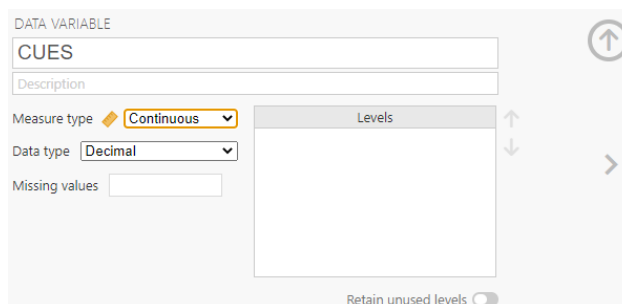
A continuació el tutorial t'acompanyarà en la identificació de variables a partir de l'arxiu R.

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 "Data" del menú superior, i a continuació clica la icona "Setup". Comprova que 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 34. Accés a l'àrea d'identificació de variables

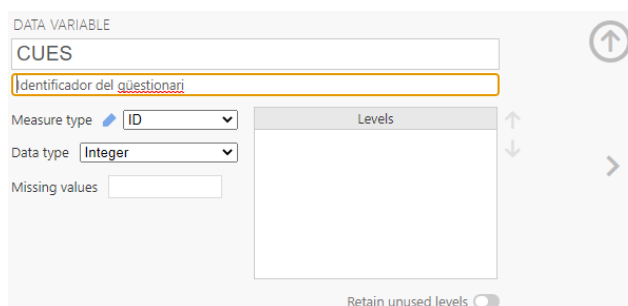
3. El panell d'identificació i definició de variables es pot amagar fent clic a la fletxa encerclada que es mostra a la seva part superior dreta.
4. Es pot passar seqüencialment d'una variable a la següent o anterior clicant les fletxes > < que es mostren a ambdós laterals del panell.
5. En aquest cas,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



Imatge 35. Identificació automàtica (Open) de la variable CUES

Tal i com pots veure, *jamovi* ha identificat automàticament la “CUES” com a variable numèrica (*Continuous+Decimal*) a partir de les dades (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 treure les dades, per tant, per a definir correctament d'aquesta variable fes el següent:

1. Clica sobre ID, per canviar el tipus de variable,
2. Canvia també el tipus de dada, i passa-la a “Integer”,
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”.



Imatge 36. 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 “Integer” en quant al tipus de dada. Cosa que és **incorrecta**, ja que haurien d'estar definides com a “Text”. Per tant, per a definir-les correctament fes el següent:

1. No canviïs el tipus de variable,
2. Clica sobre el selector desplegable de “Data type”, 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 variable, TAMUNI i g_edat, observa que han estat definides automàticament com a “Nominal” i “Integer”. En ambdós casos, ambdues coses són **incorrectes**, ja que es tracta de variables “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 de variables “Text”. Per tant, cal canviar ambdues característiques. En cadascuna de les variables, fes:

1. Clica sobre Ordinal, i canvia el tipus de variable,
2. Clica sobre el selector desplegable de “Data type”, i canvia el tipus a “Text”,⁷

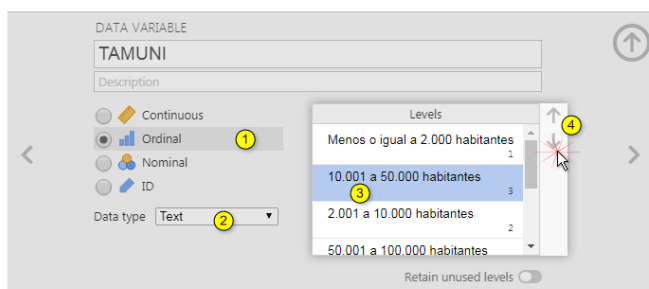
⁷ Tot i que representen intervals numèrics, les “etiquetes” de cada categoria són en realitat textuales.

3. A la descripció (sota del nom de la variable) pot ser útil afegir, respectivament, “Mida del municipi” i “Grup d’edat”, i
4. En el cas concret de les variables ordinals cal comprovar si l’ordre intrínsec de les categories que ha atorgat *jamovi* és el correcte.

En el cas de les variables de tipus “Ordinal” és possible que *jamovi* 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 jerarquia intrínseca de la variable.

Per **modificar l’ordre** de les categories d’una variable “Ordinal”

- Clica a sobre d’una de les categories (“Levels”) i, quan quedi seleccionada, fes servir les fletxes que apareixen a la part dreta de la finestra de categories per completar l’ordenació cap amunt o avall de la jerarquia.



Imatge 37. 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), però **NOMÉS** és possible si la variable ordinal te les dades definides com tipus “Text” (no funciona per a dades definides com a tipus “Integer” o “Decimal”).

Introducció a les no respostes. 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 *Missing values* (valors perduts) i que hi podem accedir.

Sovint algunes persones entrevistades deixen de contestar alguna pregunta 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, 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, no procedeix preguntar quants amics té a les xarxes socials, si abans ens ha dit que no disposa de 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.

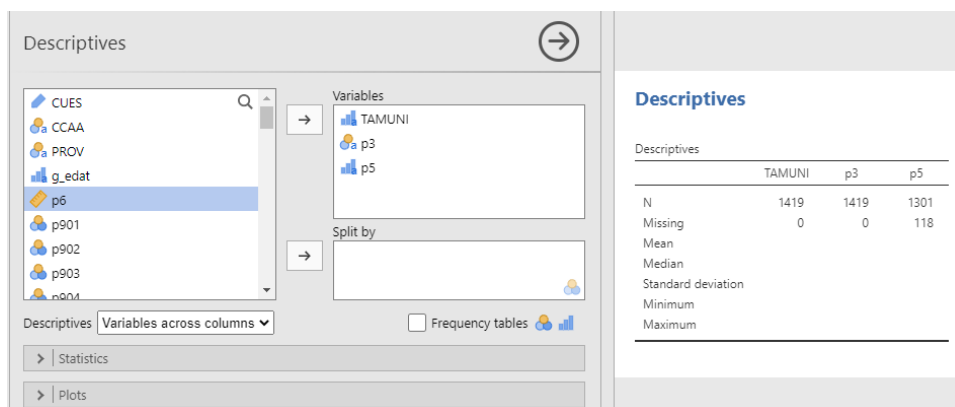
- Si la persona ha decidit no contestar, ja sigui perquè no ho sap o perquè no vol respondre, se sol assignar un **codi especial**: en variables de text sol ser “N.S.” o “N.C.”, mentre que en variables numèriques se solen deixar codis que no tinguin un significat substantiu, per exemple 98 i 99, si aquests 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, directament, el significat de “valor perdut” en la matriu de dades, i el considera com a “*missing*” 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.

La 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 de les variables TAMUNI, que ja tenim ben definida, i p3 i p5, que prèviament definirem:

1. Entra a 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. A la pestanya “Analyses” selecciona l'opció “Exploration”, a l'extrem esquerra.
3. Al menú desplegable, selecciona l'opció “Descriptives”. S'obre un panell de definició del procediment.



Imatge 38. Panell típic de definició de 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 TAMUNI, p3 i p6. 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.
5. Com pots veure, a la finestra de resultats de la dreta van apareixer, immediatament i de forma actualitzada, els resultats (segons el tipus de variable).

6. 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. Tot i així, com que en aquest cas estem analitzant tres variables de text, no ens fa cap càlcul estadístic i només ens mostra el nombre de casos i el recompte dels perduts.
7. Per a generar taules de freqüències cal activar específicament aquesta opció: clica el botó a sota del llistat dret de variables disponibles "[] Frequency tables". 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.
8. Si vols amagar el panell central, pots clicar a la FLETXA-EN-CERCLE de la part superior dreta.
9. Si vols visualitzar el panell central de definició d'un procediment, fes doble-clic a sobre de qualsevol resultat al panell dret de resultats.

Si et fixes en el descriptius, i referint-nos a les variables p3 i p5, veuràs que p3 no conté cap valor que es correspongui a un perdut ("Missing") 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 més avall, si no contestava a p5 li ha assignat el codi "N.C.":

Descriptives

Descriptives	TAMUNI	p3	p5
N	1419	1419	1301
Missing	0	0	118
Mean			
Median			
Standard deviation			
Minimum			
Maximum			

Frequencies of p3

Levels	Counts	% of Total	Cumulative %
Si	1301	91.7 %	91.7 %
No	118	8.3 %	100.0 %

Frequencies of p5

Levels	Counts	% of Total	Cumulative %
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 %

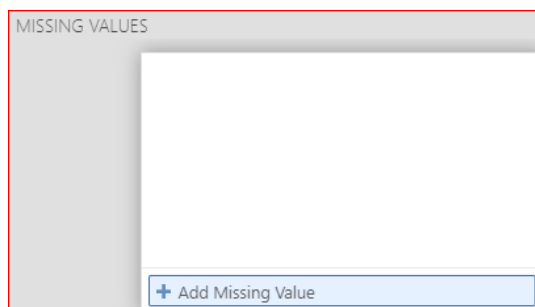
Imatge 39. Resultat típic d'anàlisi descriptiva, amb valors perduts

ATENCIÓ: Recorda que *jamovi* només "executa" ordres del llenguatge 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 el descartarem, el passarem a "valor perdut".

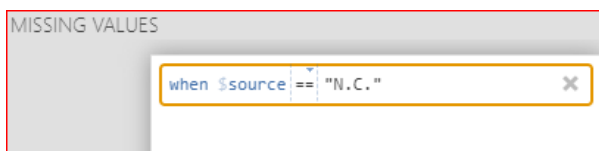
L'actual versió de *jamovi* permet definir valors perduts en la mateixa definició de la variable. I ho fa en la pestanya per hem comentat abans, anomenada "Missing values". Per fer-ho:

1. Fes doble clic a la variable p5, per tornar a obrir el panell de definició de la variable.
2. Clica sobre la finestra que hi ha a la dreta de Missing values, i s'obrirà una nova finestra de diàleg.
3. Clica la pestanya [+ Add Missing Value]



Imatge 40. Panell per definir valors perduts

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



Imatge 41. Exemple d'escriptura al panell per definir valors perduts

El què has fet és definir com a perduts, els valors que en la variable de treball (*\$source*) equivalguin a (==) "N.C.". Observa que hem fet servir una doble igualtat, i no una igualtat, 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 la pestanya [+ Add Missing Value], i escriuries una altra categoria (per exemple, "N.S.", si fos al cas).
6. Per tornar a la pantalla DATA VARIABLE, 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 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 només es fa sobre els 1.300 casos vàlids:

Descriptives			
Descriptives			
	TAMUNI	p3	p5
N	1419	1419	1300
Missing	0	0	119
Mean			
Median			
Standard deviation			
Minimum			
Maximum			

Frequencies of p5			
Levels	Counts	% of Total	Cumulative %
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 42. 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 i clicar sobre la creueta que hi ha a dalt a l'esquerra.

La definició de valors perduts en *jamovi* (II). Variables quantitatives

La lògica dels valors perduts quan la variable és quantitativa és pràcticament idèntic 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. 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="Integer"*.

1. Entra a p33 (doble clic) i veuràs que, per defecte, *jamovi* l'ha definida, incorrectament, com *mesura="Continuous"* i *data="Decimal"*.
2. Canvia-ho i anomena-la *Autoposicionament polític (1-10)*, com a *mesura="Ordinal"* i *data="Integer"*.
3. A la pestanya "Analyses" selecciona l'opció "Exploration", 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 envia-la a la dreta (finestra "Variables"). Activa l'opció "[] Frequency tables".

Els resultats obtinguts són els següents:

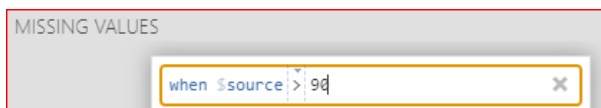
Descriptives		Frequencies of Autoposicionament política (1-10)			
Autoposicionament política (1-10)		Levels	Counts	% of Total	Cumulative %
N	1419	1	55	3.9 %	3.9 %
Missing	0	2	60	4.2 %	8.1 %
Mean	38.2	3	157	11.1 %	19.2 %
Median	6	4	131	9.2 %	28.4 %
Standard deviation	44.7	5	264	18.6 %	47.0 %
Minimum	1	6	98	6.9 %	53.9 %
Maximum	99	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 43. Resultat típic d'anàlisi descriptiva, amb valors perduts

En aquest cas, on no hem definit cap valor perdut, observem que la mitjana d'autoposicionament polític és de 38,2, un valor impossible en una variable que va d'1 a 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, si volem saber, per exemple quina és la mitjana de la variable o altres indicadors estadístics.

Per fer-ho seguirem la mateixa lògica que abans:

1. Fes doble clic a la variable "Autoposicionament polític (1-10)", per tornar a obrir el panell de definició de la variable.
2. Clica sobre la finestra que hi ha a la dreta de Missing values, i s'obrirà una nova finestra de diàleg; i clica la pestanya [+ Add Missing Value]
3. En el text "when \$source...", omple'l de la següent manera:



Imatge 44. 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, a més, com que es tracta d'un valor numèric, no l'has descriure entre cometes.

4. Per tornar a la pantalla DATA VARIABLE, 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		Frequencies of Autoposicionament polític (1-10)			
Autoposicionament polític (1-10)		Levels	Counts	% of Total	Cumulative %
N	907	1	55	6.1 %	6.1 %
Missing	512	2	60	6.6 %	12.7 %
Mean	4.59	3	157	17.3 %	30.0 %
Median	5	4	131	14.4 %	44.4 %
Standard deviation	1.90	5	264	29.1 %	73.5 %
Minimum	1	6	98	10.8 %	84.3 %
Maximum	10	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 45. 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. L'important és que ara consten com a *Missing values*, 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. Ajudant-te de la funció estadística freqüències, defineix, també, en les variables que s'escaigui, valors perduts.
5. Recorda **desar** sovint els resultats de les teves modificacions. Utilitza el menú d'arxiu / Save As ... 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.

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:

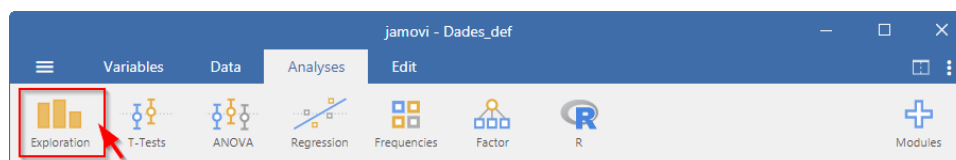
- Un qüestionari realitzat pel CIS (fitxer *MJDS_QüestionariCIS.pdf*).
- La matriu de dades corresponent, on hi ha les variables correctament identificades, amb el tipus de variable i els valors perduts definits, i que està gravada en format de *jamovi* (*MJDS_Dades_def.omv*). Aquesta matriu de dades prové de la transformació de la matriu original, que era en format R (*Dades.rds*) i que hem transformat en el capítol 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 de variables) però vols continuar –immediatament– amb el tutorial... pots obrir l’arxiu *MJDS_Dades_def.omv* disponible a l’espai CV de l’assignatura.

Obre el teu arxiu anterior, o el disponible al CV, per poder seguir amb el tutorial a partir d’aquest punt.

Tal i com ens avançàvem en l’anterior capítol, 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 46. Ubicació del menú d’anàlisi estadística DESCRIPTIVA

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):

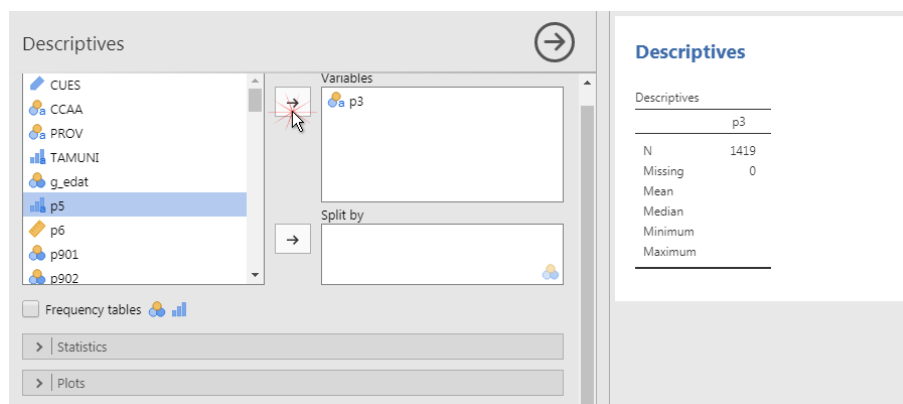
- Variable nominal, definida com a Nominal/Text: *Ús d'internet en els darrers tres mesos*, (p3). No té valors perduts definits.
- Variable ordinal, definida com a Ordinal/Text: *Freqüència d'ús d'Internet*, (p5). El codi "N.C." consta com a valor perdut.
- Variable quantitativa, definida com Ordinal/Integer: *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, així com la definició de valors perduts (*Missing values*).

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.

2. Clica l'etiqueta "Analyses" del menú superior i selecciona l'opció "Exploration", 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 47. Panell típic de definició de procediments estadístics

3. 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 superior 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.
4. Fes el mateix amb les variables p5 i p33.
5. 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).
6. 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.

7. Per a generar taules de freqüències cal activar específicament aquesta opció: Fes clic a el botó a sota del llistat dret de variables disponibles “[] Frequency tables”. 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.
8. Si vols amagar el panell central, pots clicar a la FLETXA-EN-CERCLE de la part superior dreta.
9. Si vols visualitzar el panell central 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. Coincideix amb la idea prèvia que intuïtivament tens sobre les característiques de la població espanyola objecte de l’estudi?
3. Desa l’arxiu actualitzat en format propi de *jamovi* (.omv), posant-li un nom que t’ajudi a identificar-lo (p. ex., *Dades_CIS1.omv*).
4. Tanca totalment *jamovi*.
5. Torna a obrir amb *jamovi* l’arxiu que acabes de desar. Comprova com s’han conservat les dades JUNTAMENT amb les taules que has generat.

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 / Analyses / Exploration / Descriptives, i activar l’opció “Frequency tables”
- Els estadístics descriptius QUANTITATIUS per defecte (mitjana aritmètica, mediana, desviació, mínim i màxim) només ens els dona si la variable està definida com a “Data type=Integer” o com a “Data type=Decimal”; mai si està definida com a “Data type=Text”.
- Pots veure tots els estadístics que ofereix *jamovi* clicant la secció desplegable “Statistics”, a la part central de la finestra de definició de descriptius.

Exploració de variables amb codi R



Com a alternativa a l'ús dels menús de *jamovi*, es pot fer servir directament el llenguatge de programació orientat a l'anàlisi de dades "R" 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(data)*") 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 R / Rj-Editor.
3. Escriure l'ordre en el llenguatge R, i
4. Executar la consola d'R per a generar el resultats.

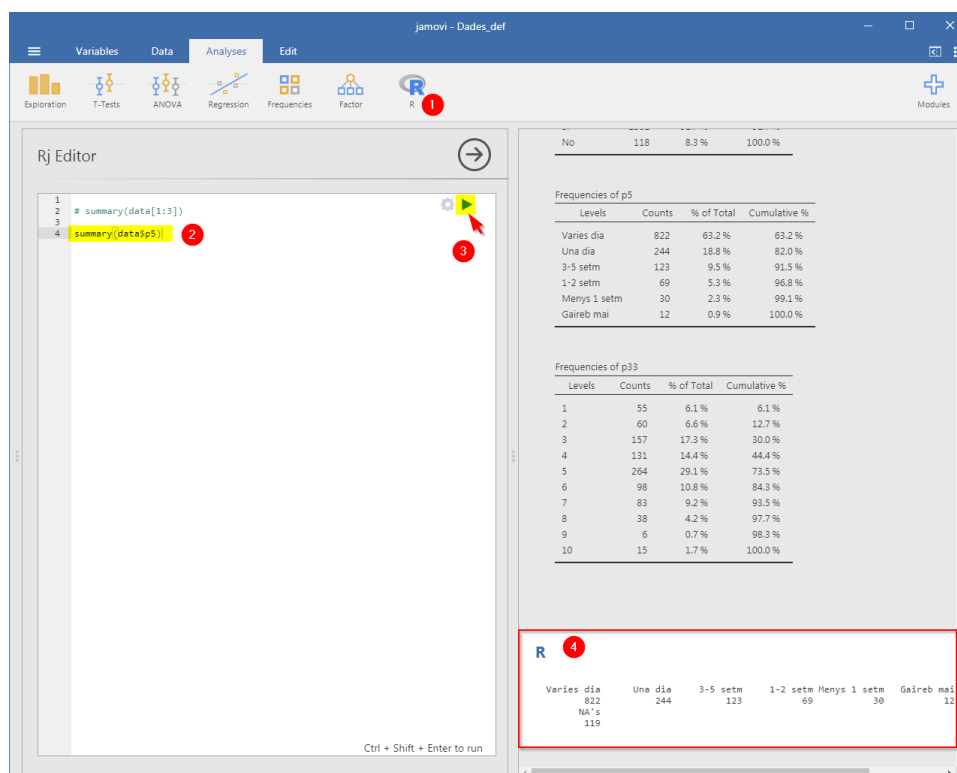
Per a instal·lar el mòdul o extensió (Rj) que permet l'execució, des de *jamovi*, de codi R, cal disposar de connexió a Internet i seguir les passes indicades a la secció anterior "Com afegir el mòdul per a operar directament amb codi R?". Una vegada instal·lat un mòdul aquest quedarà integrat al teu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

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

1. A l'apartat "Analyses", clica la icona "R". Al desplegable, selecciona l'opció "Rj Editor",
2. Apareix la "consola" d'ordres R, i per defecte, una ordre inactiva d'exemple:
`# summary(data[1:3])`
3. Qualsevol línia que comenci per # es considera inactiva (està de color verd),
4. Situa el cursor a qualsevol línia de la consola diferent de la que comença per #, i escriu, 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.

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 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 48. Exemple resultat d'execució codi R

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

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 PROBABLE que doni errors. Això es degut a que has enganxat codis PDF que semblen adients per a R quan 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 de nou.

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: Analyses / R / Rj Editor...
2. Per a generar una taula de freqüències simple de la variable 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 la variable que està a la matriu de dades i es diu p3".
4. Executa el 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 simple no disposa ni de títol ni d'altres càlculs com ara freqüències relatives o acumulades.



EXERCICI

- Genera, amb codi R, una taula de freqüències simple de la variable ORDINAL/Text p5 (Freqüència d'ús d'Internet) i de la variable Ordinal/Integer 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$p33))`
2. Per a generar la taula vertical, que contingui les freqüències relatives de la variable P33, escriu i executa el següent codi a la consola:
`cbind(prop.table(table(data$p33)))`
3. I si volem que ens doni els percentatges, només hem de multiplicar la taula per 100:
`cbind(prop.table(table(data$p33))*100)`

- 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)`
- El què li hem dit, en llenguatge R, és: Fes una taula vertical que tingui dues columnes. La primera ha de contenir una taula de freqüències absolutes de la variable p33 i la segona que contingui les freqüències relatives (multiplicades per 100), de la taula de freqüències absolutes de la variable p33.



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 el nombre **9**. Així, si li diem **A+1** donarà com a resultat **10**. Per anomenar qualsevol resultat a R cal indicar el nom desitjat, seguit de **<-** seguit de l'operació corresponent.

Utilitzant diferents funcions, i la possibilitat d'anomenar i re-utilitzar resultats, es pot combinar una sèrie d'ordres d'R per a generar una taula vertical amb diversos resultats a les columnes:

- 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 al final:

```
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)
```
- Observa que a les primeres línies de codi hem generat i anomenant objectes, que utilitzem en operacions posteriors.
- 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.
- Aquesta seqüència d'ordres significa, en llenguatge R:
 Anomena **dades** a la columna de 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 freqüències relatives d'allò anomenat **n_i** (és a dir, l'anterior taula de freqüències absolutes) multiplicades 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),
 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**
- Si executes 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 llenguatge R. Com que R “escriu” els resultats “en anglès” els punts (.) que apareixen a les xifres dels resultats són separacions decimals (que nosaltres denotem amb comes).



EXERCICI

- Genera el mateix tipus de taula per a les variables p3 i p5.
- Observa que si tens la instrucció complerta, només hauràs de canviar, en cada cas, el nom de la variable.

Estadística descriptiva en 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 dues coses:

- Les variables quantitatives amb les que treballem amb *jamovi* no són, pròpiament, numèriques, ja que les hem definides com a Ordinal/Integer. Per tant, el primer que haurem de fer, utilitzant directament el codi R és canviar-les, tal i com ho fem aquí, on generem una còpia exacta, però numèrica de la variable:

```
data$p33num<-as.numeric(as.character(data$p33))
```

 Li estem dient que creï la variable p33num (al fitxer data) que sigui la versió numèrica de la p33, passada prèviament a character.
- Si demanes qualsevol estadístic descriptiu d'una variable que conté valors perduts, cal dir-li 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.

Anem a veure com ho faríem 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** ($\bar{x}$). 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.

- Els **quartils** (Q_1 , Q_2 , Q_3). Escriu a la consola, i executa:
`quantile(data$p15num, na.rm = T)`
La persona que té menys amics i la que en té més, tenen, respectivament 1 amic i 4.600 amics. A més, el 25% que menys amics té, en té 77 o menys. Hi ha un 50% d'individus que tenen 175 amics o menys i un 50% que en tenen 175 o més (mediana o quartil segon). El 75% que menys amics té, en té 300 o menys (anàlogament, el 25% que més en té, té 300 amics o més).
- El **coeficient de variació** (CV), que es correspon a la desviació típica dividida per la mitjana. Escriu a la consola, i executa:
`sd(data$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 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), 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 a la programació en R.



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 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 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, 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 textual de les taules generades per *jamovi* NO es pot editar, però és possible incorporar text propi, notes i títols a aquest resultats generats. Per fer-ho cal entrar en la modalitat COMENTARI de resultats mitjançant el menú corresponent (menú "Edit").

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 "Add Note".

1. Clica el menú "Edit".

2. A la finestra de resultats, queden destacats en blau 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 àrea editable. 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 "Heading" 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.

The screenshot shows the Jamovi interface with a data table on the left and a results panel on the right. The results panel displays a frequency table for variable p3, which is highlighted with a red box. Below the table, a comment is added in blue text.

Frequency Table for p3:

Levels	Counts	% of Total	Cumulative %
Si	1301	91.7 %	91.7 %
No	118	8.3 %	100.0 %

Comment: Com es pot veure, només una mínima part de la població (8.3%) no fa servir Internet i per tant es pot entendre que la població espanyola té un nivell prou elevat de "alfabetització digital".

Imatge 49. Exemple de comentaris afegits a resultats estadístics

Exportació de resultats cap a altres aplicacions

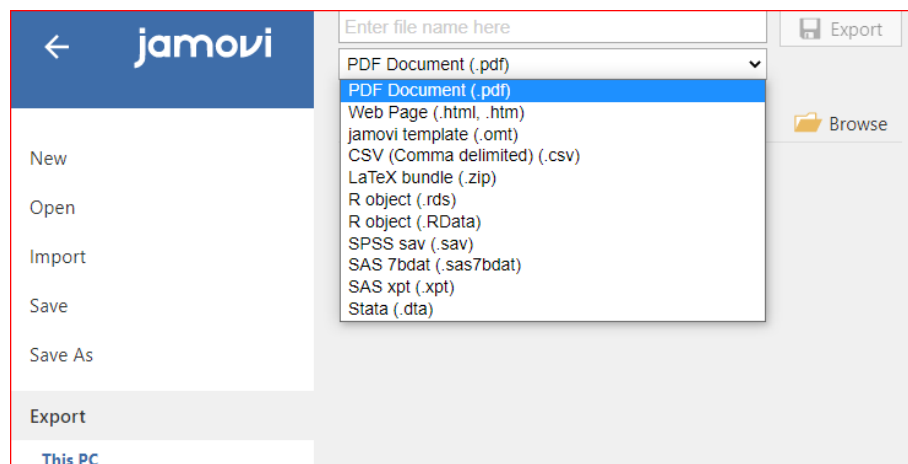
El contingut de la finestra de resultats es compon d'un conjunt d'objectes, generats per *jamovi* a partir de les "demandes" de l'anàlisi. Cada anàlisi demanat és un objecte-Analysis, que conté diversos objectes-Table. Mitjançant diversos procediments és possible exportar els resultats, inclosos els comentaris afegits:

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

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: "Export"
2. Selecciona, al desplegable corresponent, com a format d'exportació "PDF Document (.pdf)". Clica "Browse" per obrir la finestra que permet triar el nom i la ubicació de l'arxiu exportat.



Imatge 50. Exportació de resultats a PDF

3. Ara busca, al teu sistema local d'arxius, el que acabes d'exportar i fes-hi doble-clic per a obrir-lo amb la teva aplicació de lectura de PDF.
4. 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 / Export / ... a format HTML, que es pot obrir fàcilment amb editors de text, per a modificar, editar i complementar amb comentaris.

(B) Per a exportar **cada anàlisi 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 (LibreOfficeWriter, 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 "Analysis". Comprova que s'ha seleccionat (color blau) el que t'interessa exportar. Selecciona "Copy" per transferir tots els objectes 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, els 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 "Table" / "Copy" 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 "Image" / "Copy" 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, conjuntament, el seu títol i comentaris de text, al menú contextual cal seleccionar l'opció "Group" / "Copy".

ALTERNATIVA: Es pot desar una imatge particular (no "grup", 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 "Export..." i al desplegable de format d'arxiu no triar PDF sinó PNG o SVG (recomanat).

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 Analyses / Exploration / Descriptives (excepte el "gràfic de Pareto", que forma part d'un mòdul específic que cal instal·lar a part, anomenat "scatr").

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

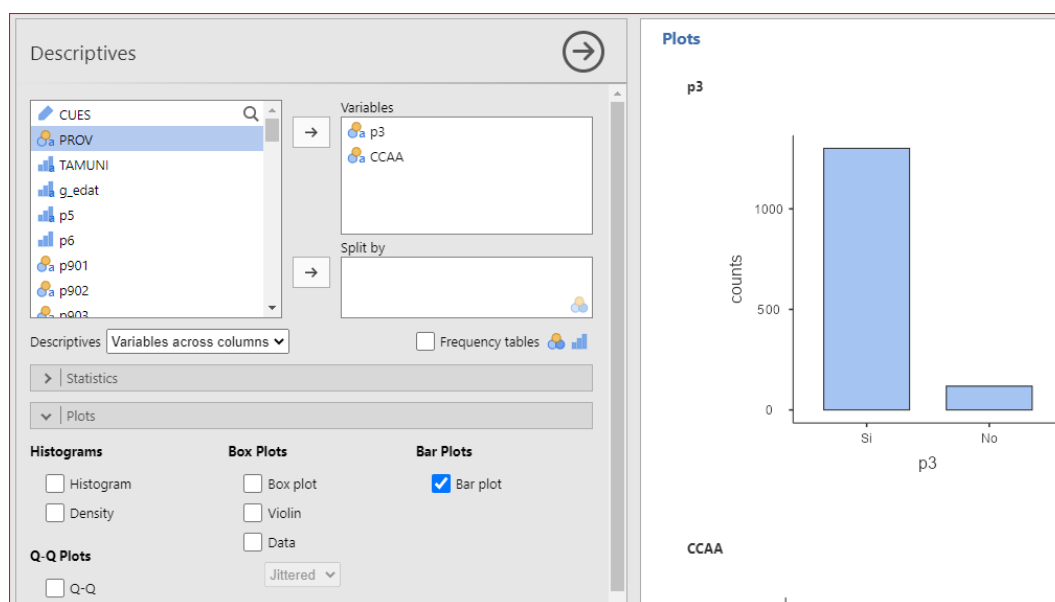
- Variables nominals (en *jamovi* Nominal/Text). El més habitual és fer gràfics de sectors i gràfics de barres. Si el nombre de categories és bastant gran, però, el diagrama de sectors és poc clar. El sistema de menús de *jamovi* no té incorporat el diagrama de sectors, que haurem de programar amb la 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'ordre.
- Variables quantitatives (en *jamovi* Ordinal/Integer). Són habituals tant els histogrames com els diagrames de caixa.

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).

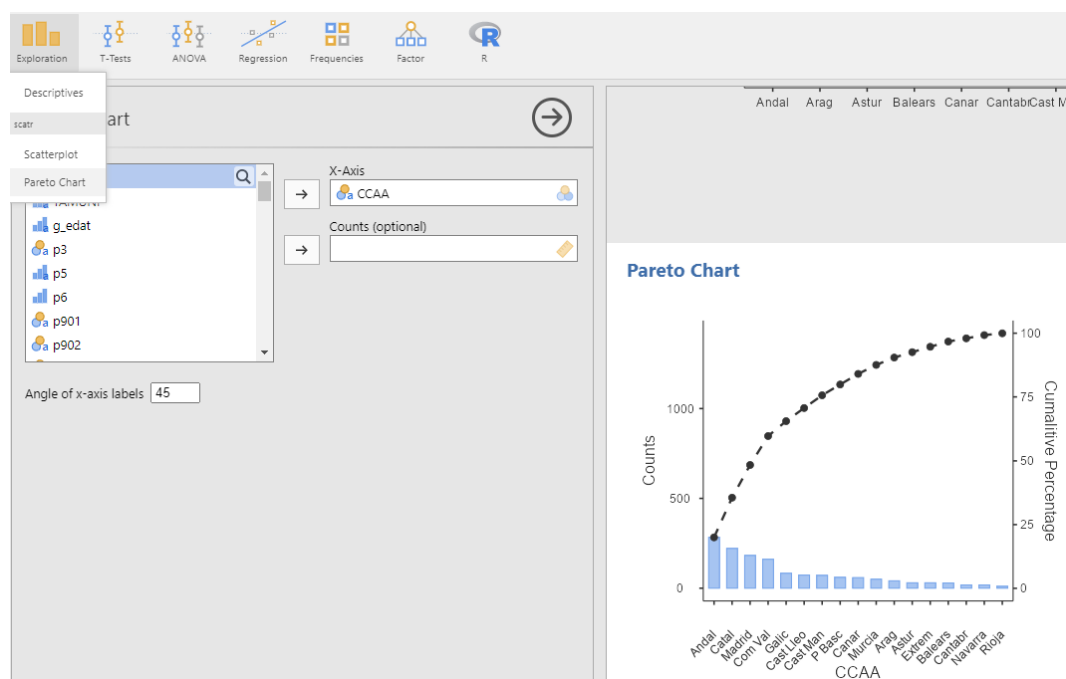
Per a graficar les variables p3 i CCAA, farem el següent:

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: Analyses / Exploration / 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 "Plots" per a desplegar les opcions gràfiques. A l'apartat "Bar Plots" activa l'opció "Bar Plot".



Imatge 51. Generar un diagrama de barres amb jamovi

3. Observa que algunes etiquetes de la variable CCAA queden encavalcades, un problema que tindrem sovint quan el nombre de categories sigui gran, o el text d'aquestes molt llarg.
4. 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.
5. Clica al menú Analyses / Exploration / Pareto Chart, 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 "X-Axis". Pots comprovar que les etiquetes també s'encavalquen. Ara bé, en aquest cas pots marcar, a la part inferior, la secció "Angle of X-axis labels" i seleccionar l'opció que permet girar 45 graus les etiquetes de l'eix horitzontal per tal de millorar la llegibilitat del gràfic.



Imatge 52. Exemple de Gràfic de Pareto generat amb jamovi

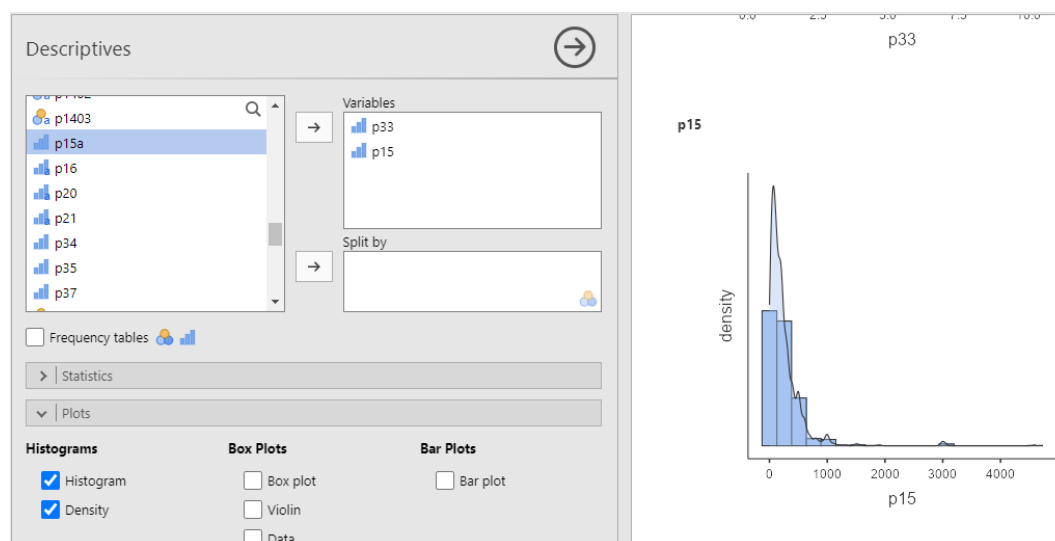


EXERCICI

- Tal i com hem 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 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: Analyses / Exploration / 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 "Plots" per a desplegar les opcions de gràfiques. A l'apartat de l'esquerra "Histograms" activa les opcions "Histogram" i "Density".



Imatge 53. 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 "Histogram" i "Density" i activa les opcions "Box-plot" i, si vols que ens dibuixi també la mitjana, la pestanya "Mean".
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 pots comprovar, per exemple, 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 atípic ("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 que coincideix amb el rang interquartil·lic RIQ ($Q_3 - Q_1$) i conté la Mediana, representada per una LÍNIA horitzontal. Els valors atípics s'identifiquen per PUNTS, i s'utilitzen com a criteri la seva 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 queden representats per "PATILLES" en forma de línies verticals.

Codi R per a generar representacions gràfiques



A continuació el tutorial indica alguns exemples de codi R que pots executar per a generar diversos tipus de gràfics, alguns 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 en variables nominals de poques categories). Anem-lo a fer 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$p3))`
2. Una versió més completa podria contenir el següent:
`pie(table(data$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 veure'l amb la variable *Freqüència d'ús d'Internet* (p5).

1. Per a generar la versió simple només cal escriure i executar:
`barplot(table(data$p5))`
2. Si, per exemple, volem incloure un títol al gràfic, canviar els límits de l'eix de les Y, i modificar l'orientació de les etiquetes de l'eix de les X, li direm:
`barplot(table(data$p5), main="Freqüència d'ús d'Internet", ylim=c(0,1000), las=2)`
3. I 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)`

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, generem la versió numèrica de p15 amb la següent instrucció:
`data$p15num<-as.numeric (as.character (data$p15))`
2. La versió més senzilla de l'histograma és:
`hist(data$p15num)`
3. Si volem canviar els títols (del gràfic i dels dos eixos), volem donar-li color, i volem que ens generi, per exemple, 20 categories diferents, li direm:
`hist(data$p15num, breaks = 20, main = "Histograma de Contactes a Xarxes Socials", xlab="Nombre de contactes", ylab="Freqüència absoluta", col="green")`

Per veure el 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 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:

```
boxplot(data$p15num, ylim=c(0,900),
main="Nombre d'amics a les xarxes socials",
col=c("magenta"))
```



EXERCICI

- 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 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.⁸

Tot i que, en la seqüència de treball habitual, 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 disposeu de les eines, es treballarà en la depuració de les dades.

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 2011.⁹ 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 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 (“*Levels*”), 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 de variables original i final:

⁸ 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.

⁹ 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.

		Variable FINAL (<i>transformada</i>)	
		Qualitativa (Nominal o Ordinal)	Quantitativa (Entera o Decimal)
Variable ORIGEN (inicial)	Qualitativa (Nominal o Ordinal)	[SE / Prim. / Sec. / FP / Grau / Doctorat] <i>es transforma en</i> [Prim o - / Sec. / Univ.]	[Si / No] <i>es transforma en</i> [1 / 0]
	Quantitativa (Entera o Decimal)	[Edat: 0 - 90] <i>es transforma en</i> [Joves / Adults / Vells]	[Ideologia escala 1 a 10] <i>es transforma en</i> [Ideologia escala 1 a 5]

En els quatre casos, per “recodificar” una variable amb *jamovi*, cal utilitzar el 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 recodifica en codi R? Com s'ha vist anteriorment, *jamovi* utilitza el panell de matriu de dades i els menús propis (o de mòduls addicionals) per a realitzar la gestió i anàlisi estadístiques, però també es pot fer servir la consola Rj per a executar operacions directament amb codi R. En principi, fins i tot es pot programar i executar recodificacions mitjançant codi R.

No es tractarà aquí aquesta qüestió perquè:

- (1) a data d'octubre de 2021 *jamovi* pot recodificar i crear noves variables en consola PERÒ aquestes noves columnes **NO** passen a estar disponibles a la matriu de dades, i
- (2) totes les recodificacions requerides en aquest curs es poden realitzar amb el sistema de menús i transformacions de variables actualment implementat.

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.

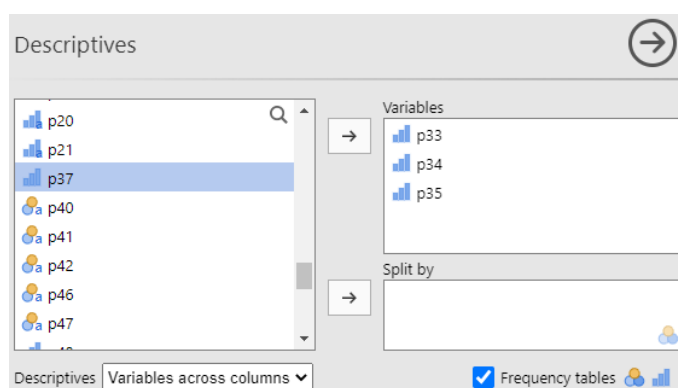
Tal i com pots veure al *MJDS_QüestionariCIS.pdf* aquestes variables són quantitatives, de manera que les hem definit en *jamovi* com a ORDINAL/INTEGER, tenint com a valors perduts 96, 97, 98 i 99.

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 INTEGER 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 “Analyses”. I a la icona “Exploration” 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 “Frequency tables”:



Imatge 54. Selecció de diverses variables

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

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:

Frequencies of p33				Frequencies of p34				Frequencies of p35			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
1	55	6.1 %	6.1 %	1	42	5.3 %	5.3 %	1	44	5.5 %	5.5 %
2	60	6.6 %	12.7 %	2	49	6.1 %	11.4 %	2	44	5.5 %	10.9 %
3	157	17.3 %	30.0 %	3	128	16.0 %	27.4 %	3	117	14.5 %	25.4 %
4	131	14.4 %	44.4 %	4	102	12.8 %	40.2 %	4	127	15.8 %	41.2 %
5	264	29.1 %	73.5 %	5	172	21.6 %	61.8 %	5	204	25.3 %	66.5 %
6	98	10.8 %	84.3 %	6	91	11.4 %	73.2 %	6	99	12.3 %	78.8 %
7	83	9.2 %	93.5 %	7	101	12.7 %	85.8 %	7	87	10.8 %	89.6 %
8	38	4.2 %	97.7 %	8	72	9.0 %	94.9 %	8	52	6.5 %	96.0 %
9	6	0.7 %	98.3 %	9	13	1.6 %	96.5 %	9	17	2.1 %	98.1 %
10	15	1.7 %	100.0 %	10	28	3.5 %	100.0 %	10	15	1.9 %	100.0 %

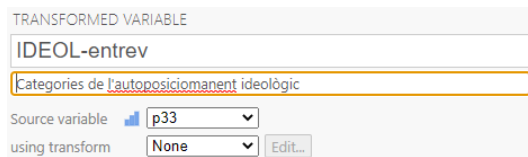
Imatge 55. Exemples de resultat de taules de freqüències

Observem que les variables es troben ben definides, de manera que podem aplicar la regla de transformació que assigni, per exemple, el valor *Esquerra* als valors originals 1-4, el valor *Centre* al valor original 5, i el valor *Dreta*, als originals 6-10.

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

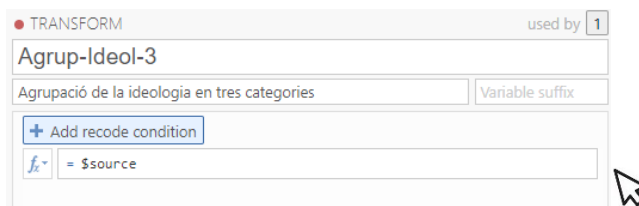
1. Situa’t sobre la variable p33 i clica, del menú “Data” la pestanya “Transform”.

2. Es crea una nova columna a la dreta de la p33, per defecte anomenada “p33 (2)” i basada en la variable d’origen p33 (observa que, de moment, és una còpia).
3. Canvia el nom (p. ex., IDEOL-entrev). Pots incloure una descripció més expressiva.



Imatge 56. 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 a el desplegable que actualment indica [None] per a “using transform”, i selecciona l’opció [Create New Transform...].
6. S’obrirà la finestra de regles de transformació, amb una transformació per defecte, “Transform 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 variable automàticament dependent de la variable original.
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 el sufix** amb el qual s’identificaran les noves variables a les que apliquem 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 “recode conditions”: 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 [Add recode condition] per afegir una condició de recodificació.

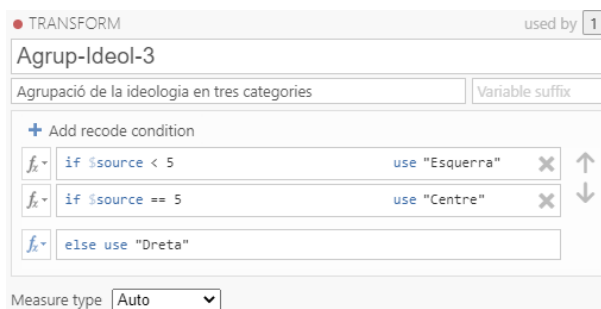


Imatge 57. Botó per afegir condicions de recodificació

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:
IF \$source < 5 use "Esquerra"
12. Un cop la tinguis escrita, torna a marcar “Add recode condition” per afegir la segona condició. Aquesta, que fa referència al posicionament de centre, es pot expressar en llenguatge de programació com:

IF \$source == 5 use "Centre"

13. Torna a marcar "Add recode condition" i afegeix la condició final, que serà pels que encara no has canviat (els majors de 5), de manera que es pot expressar programàticament com:
else use "Dreta"



Imatge 58. 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). És molt important que es respectin estrictament les convencions de codificació, per exemple, la **dobla 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 tindrà una variable ODINAL/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  / Save.

Quan estàs recodificant a partir d'una transformació de variables, els operadors que pots fer són:

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

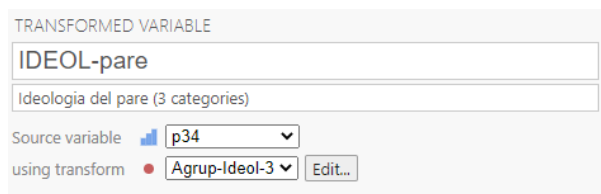
Ara pots visualitzar els canvis fets a la matriu de dades. Comprova, fent freqüències de la variable antiga i nova, que s'han aplicat correctament les tres regles de transformació.

Si tot ha funcionat correctament amb aquesta primera variable, ara ja pots procedir a aplicar LES MATEIXES transformacions a les altres dues variables, per a recodificar-les en les noves variables ordinals "IDEOL-pare" i "IDEOL-mare" (respectivament, p34 i p35).

Per a recodificar la variable origen p34 en "IDEOL-pare", fes:

1. A *jamovi*, fes doble-clic al nom de la variable p34, que encapçala la columna de dades. S'obre el panell superior de definició de la variable, que mostra la informació de la variable original p34.

2. Clica l'etiqueta "Data" del menú superior i marca la icona "Transform" del submenú de procediments amb dades. Es crea una nova columna a la dreta de la p34, per defecte anomenada "p34 (2)" i basada en la variable d'origen p34.
3. Clica el desplegable corresponent a "using transform" (que per defecte indica "None").
4. Selecciona la transformació **que recodifica en tres intervals** ordinals, i que hem anomenat "Agrup-Ideol-3".
5. Comprova com ha creat la variable "p34-Agrup-Ideol-3" tal i com la volíem. Canvia el nom (p. ex., IDEOL-pare) i inclou una descripció més expressiva, per exemple, *Ideologia del pare (3 categories)*.



Imatge 59. Exemple de transformació

6. 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.
7. Per a tancar la finestra de regles de transformació, clica la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra.
8. Per a tancar la finestra de definició de variables, clica a la icona circular FLETXA-AMUNT de la part superior dreta d'aquesta finestra.
9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

Fes el mateix que en l'anterior. Comprova, fent freqüències de la variable antiga i nova, que s'han aplicat correctament les tres 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 quatre variables que a hores d'ara tens (dues per la ideologia de la persona entrevistada i dues per a la del pare), pots fer:

Prerequisit: Tenir instal·lat a *jamovi* el mòdul d'extensió anomenat Rj.¹⁰

1. A l'apartat "Analyses", 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 quatre línies diferents:

```
summary(data$p34)
summary(data$IDEOL-entrev')
summary(data$p35)
summary(data$IDEOL-pare')
```

 Com has pogut veure, *jamovi* t'ajuda proposant ordres i variables disponibles.

¹⁰ A l'anterior apartat "Exploració de variables amb codi R" es descriu en detall com instal·lar aquesta funcionalitat de *jamovi*.

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”.

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.

5. Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.

The screenshot shows the Rj Editor interface. On the left, the console contains the following R code:

```
1 # summary(data[1:3])
2
3
4 summary(data$p33)
5 summary(data$`IDEOL-entrev`)
6 summary(data$p34)
7 summary(data$`IDEOL-pare`)
8
9
10
11
```

On the right, the output is displayed in two tables. The top table shows the summary of the first three variables:

variables	count	count %	count %
Una dia	244	18.8 %	82.0 %
3-5 setm	123	9.5 %	91.5 %
1-2 setm	69	5.3 %	96.8 %
Ménys 1 setm	30	2.3 %	99.1 %
Gaireb mai	12	0.9 %	100.0 %

The bottom table shows the summary of the variable 'IDEOL-entrev' (Ideology):

	1	2	3	4	5	6	7	8	9	10	NA's
Esquerra	55	60	157	131	264	98	83	38	6	15	512
Centre											
Dreta											
NA's											

Imatge 60. Exemple de resultats d'exploració mitjançant codi R

6. 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.
7. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.



EXERCICI

1. Utilitzant les mateixes regles de transformació, genera la variable IDEOL-mare.
2. 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 correcta.

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: used by 1, a la part superior dreta.

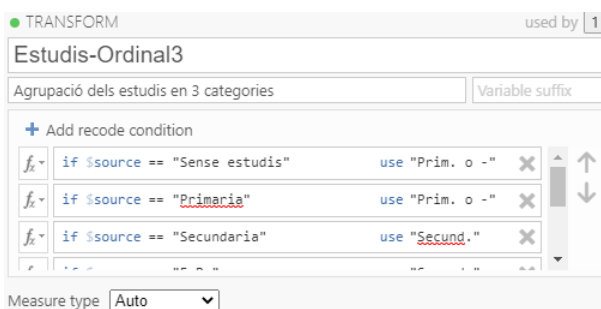
Recodificació (II). De variables qualitatives a qualitatives

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 correcta 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.¹¹
3. Clica a sobre de l'etiqueta "Data" del menú superior i a la icona "Transform" 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 [None] per a "using transform", i selecciona l'opció [Create New Transform...].
6. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada "Transform 3", 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 3" per "Estudis Ordinal3". Si vols pots introduir una descripció més detallada, com ara "Agrupació dels estudis en 3 categories".
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 -*, a les primeres dues categories; *Secund.*, a les dues següents i *Univers.* a les altres dues.
10. Clica el botó [+ Add recode condition] 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 -'. En el llenguatge de programació:
`IF $source == 'Sense estudis' use 'Prim. o -'`
12. Cal introduir **cada** condició en una **nova línia**. Clica el botó [+ Add recode condition] per a crear la següent condició.
13. Escriu la segona condició:
`IF $source == 'Primaria' use 'Prim. o -'`
14. Introdueix la resta de condicions en llenguatge de programació, clicant [+ Add recode condition] tantes vegades com calgui:
`IF $source == 'Secundaria' use 'Secund.'`
`IF $source == 'F.P.' use 'Secund.'`
`IF $source == 'Univers mitjos' use 'Univers.'`
`IF $source == 'Univers sup' use 'Univers.'`

¹¹ Pots consultar apartats anteriors d'aquesta mateixa pràctica, on s'ha vist com fer-ho.



Imatge 61. 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:

```
IF $source == 'Sense estudis' use 'Prim. o -'
```

```
IF $source == 'Primaria' use 'Prim. o -'
```

```
IF $source == 'Secundaria' use 'Secund.'
```

```
IF $source == 'F.P.' use 'Secund.'
```

```
else use '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  / Save.

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.

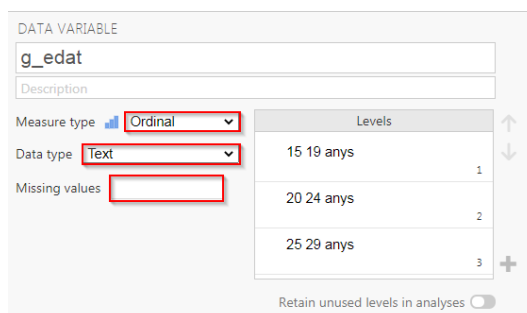
Recodificació (III). De variables qualitatives a quantitatives

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: INTEGER o DECIMAL.

Per a transformar la variable *g_edat* (que és ordinal/text) a quantitativa cal assignar a cada valor el corresponent a la seva "marca de classe". Així, la transformació es resumeix a passar de "15 19 anys" a 17,5; de "20 24 anys" a 22,5; i de "25 29 anys" a 27,5. A cadascuna de les tres etiquetes o nivells, doncs, els assignem el valor (que serà continuu/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.¹²
2. En la mateixa finestra comprova, també, que no té valors perduts.



Imatge 62. Comprovació de definició de variable

3. Selecciona la variable g_edat clicant el nom que encapçala la columna. Clica l’etiqueta “Data” del menú superior i clica, també, la icona “Trasform” 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.
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 [None] per a “using transform”, i llavors selecciona l’opció [Create New Trasform].
6. S’obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada “Transform 4” (el nom depèn, òbviament, del nombre de transformacions fetes anteriorment). La regla que ens dona per defecte és molt “els mateixos valors que la variable origen” (= \$source), i un tipus de variable que depèn de la variable original.
7. Primer **modifica el nom** perquè sigui més explicatiu, canvia “Transform 4” 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”.¹³
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 [+ Add recode condition] 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:
`IF $source == '15 19 anys' use 17.5`

¹² Pots consultar apartats anteriors d’aquesta mateixa pràctica, on s’ha vist com fer-ho.

¹³ Documentar o descriure en detall les transformacions que construeixis ajuda molt quan es volen reutilitzar posteriorment.

Compte! Per escriure decimals en *jamovi* has de fer servir el punt i no la coma (perquè s'està utilitzant el llenguatge R).

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 [+ Add recode condition]. Les següents condicions en llenguatge de programació segueixen la mateixa lògica:
`IF $source == '20 24 anys' use 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:
`else use 27.5`
13. Canvia el tipus de mesura a CONTINUOUS per a definir la nova variable com a quantitativa.



Imatge 63. 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 *Continuous*. 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 "Analyses", 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  / Save.

Recodificació (I). 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 ("Compute"). Aquest procediment permet aplicar un conjunt de funcions matemàtiques i estadístiques.

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 INTEGER, i no tenen valors impossibles, ja que els valors superiors a 9.000 (9.997, 9.998 i 9.999) han estat definits com a *Missing values*:

Imatge 64. Exemple de variables quantitatives sobre contactes a xarxes socials

Si demanem descriptius d'ambdues obtenim:

Descriptives		
	p15	p15a
N	1081	1054
Missing	338	365
Mean	249	42.7
Median	175	20.0
Standard deviation	333	70.0
Minimum	1	1
Maximum	4600	1000

Imatge 65. 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 *Continuous/Integer*. 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'*Integer* (no la de *Continuous*).

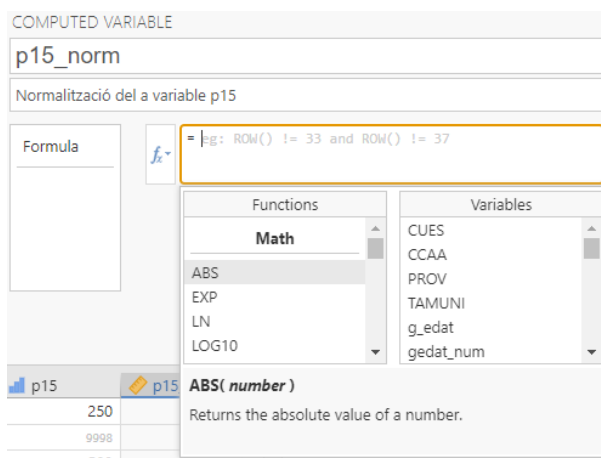
Operacions i càlculs utilitzant una única variable

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 qualsevol 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 passos:

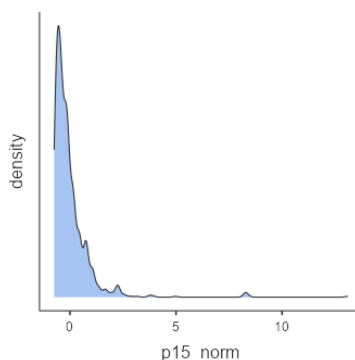
1. Comprova, fent doble-clic sobre el nom de la variable, que està definida com *Ordinal/Integer* i que els valors superiors a 9.000 no seran tinguts en compte (estan definits com a *Missing values*). Si no és el cas, **modifica'n la definició** fins que s'ajusti a aquests paràmetres.¹⁴
2. Clicant sobre la variable p15, selecciona-la. Clica l'etiqueta "Data" del menú superior i, a continuació, selecciona la icona "Compute" 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 66. Menú de càlcul amb variables

5. L'estandardització ve notada per la **funció Z**. Busca-la al menú de Funcions, marca-la amb clic-esquerra i, a continuació, busca la variable que vols estandarditzar (p15). En llenguatge de programació comprensible per jamovi li hauràs escrit:
z (p15)
6. Fes un 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. Comprova també que la desviació típica de la segona és 1 (i, en conseqüència, la variància ser 1).
7. 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.

¹⁴ Pots consultar apartats anteriors d'aquesta mateixa pràctica, on s'ha vist com fer-ho.



Imatge 67. Gràfica de densitat de la variable P15 normalitzada

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

El llenguatge de programació ofereix múltiples funcions per a realitzar els càlculs. Totes tenen la mateixa “syntaxi”: 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 entre variables diferents

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 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/INTEGER, i estan depurades, en el sentit que no tenen valors impossibles, ja que els valors superiors a 9.000 (9.997, 9.998 i 9.999) han estat definits com a *Missing values*.

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/Integer* i que els valors superiors a 9.000 no seran tinguts en compte (estan definits com a *Missing values*). Si no és el cas, **modifica'n la definició** fins que s'ajusti a aquests paràmetres.¹⁵

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

2. Clica sobre del nom de la variable p15a, que encapçala la columna de dades, per seleccionar-la
3. Clica l'etiqueta "Data" del menú superior i, a continuació, selecciona la icona "Compute" 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 *continuous*, sí que veiem que el mínim és 0,00109 i el màxim és 1:

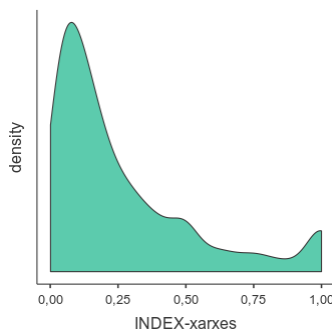
Descriptives			
	p15	p15a	CONFIA_xarxes
N	1081	1054	1051
Missing	338	365	368
Mean	249	42.7	0.267
Median	175	20.0	0.167
Standard deviation	333	70.0	0.271
Minimum	1	1	0.00109
Maximum	4600	1000	1.00

Imatge 68. 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 "Analyses", 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  / Save.

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 69. 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).

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 sumàriem directament les 14 variables. Seguint aquesta lògica el què fem és:

1. Transformem (TRANSFORM) 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 podem fer el càlcul (COMPUTE) 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 (CONTINUOUS/INTEGER), 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 “Data” del menú superior i la icona “Trasform” 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 càlcul), 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 [None] per a “using transform”, i llavors selecciona l'opció [Create New Transform].
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 Binària (0,1).

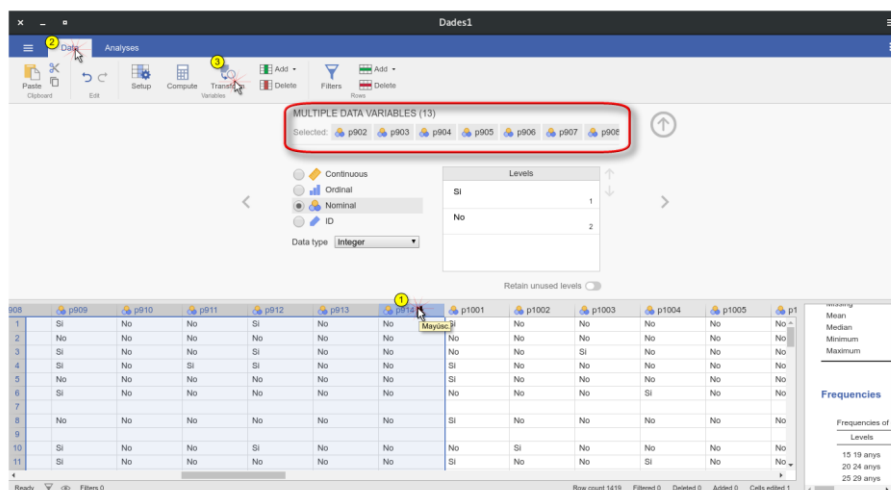
Observa que... aquest sufix sovint és important que l'utilitzem quan vulguem repetir la mateixa transformació varies vegades, com és el cas.

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 [+ Add recode condition] 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:
`IF $source == 'Sí' use 1`
9. Cal introduir cada condició en una nova línia, clicant el botó [+ Add recode condition] i seguir la mateixa lògica:
`IF $source == 'No' use 0`

- En la icona que apareix al costat del nom de la variable podràs comprovar que la nova variable és Nominal/Integer. Als nostres efectes això és positiu, ja que en ser numèrica (Integer) ens permetrà els càlculs que posteriorment haurem de fer.
- 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.
- 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.
- Fes freqüències de les dues variables (p901 i p901_bin) i comprova que la transformació ha estat correcta.
- Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

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 a un conjunt de variables és la següent:

- Cal seleccionar les variables p902 fins a la p914. Per fer-ho, a la matriu de dades, clics sobre el nom de la primera variable a transformar (p902). Queda marcada tota la columna.
- 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).
- Prem la tecla de [Majúscules] al teclat i conjuntament fes clic sobre el nom de "darrera" variable (p914).



Imatge 70. Selecció de conjunts de variables (per aplicar la mateixa transformació)

- Ja les tens totes seleccionades. Per a aplicar la mateixa transformació a totes, ves al menú principal superior fes clic a l'apartat "Data" i clics l'opció "Transform". Al panell de transformacions múltiples, clics el desplegable de "using transform" (que indica [None]) i selecciona SINOa10.
- S'han generat 13 noves variables quantitatives (amb 1 i 0) basades en les 13 variables nominals originals. Per tancar el panell de transformació clics la FLETXA-AMUNT encerclada de la part superior dreta.
- Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

La selecció múltiple de variables també es pot utilitzar per a definir [Setup] 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 té un petit punt (groc) 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 (COMPUTE) 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 / Add / Compute. 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 71. Exemple de la funció SUMATORI amb un conjunt de variables

11. Quan la fórmula 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 bé, 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  / Save.

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 de les dades de les columnes originals que intervenen en l'operació té el valor perdut.

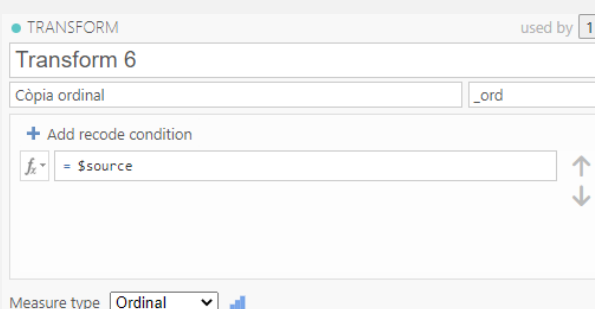
El resultat d'aquesta operació és una nova variable que mostra que hi ha 39 persones que donen un únic ús a Internet, 93 persones que usen Internet per a 2 funcionalitats, etc. fins arribar a 24 persones (un 1,8%) que utilitzen Internet per a gairebé tot (12 usos diferents). Sembla que la major part 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 (INTEGER) 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 podem 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 còpia d'ella mateixa (= \$source) tot modificant el tipus de mesura a la part inferior de la pantalla de transformació.



Imatge 72. 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*.

Transformacions complexes (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 (TRANSFORM) 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 algebraïques 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:¹⁶

Contingency Tables			
p41	p42		Total
	Si	No	
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 estarem dient a *jamovi* és alguna cosa com: *Si la p42 és 'Si' i la p41 és 'Casat/da', la nova variable ha de tenir el valor de 'Matrimoni', mentre que si la p42 és 'No' 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 la finestra de fórmula (al costat del signe =) cal incloure l'operació algebraica que ha de ser realitzada. Tant per un pas com per l'altre farem servir el condicional IF.

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

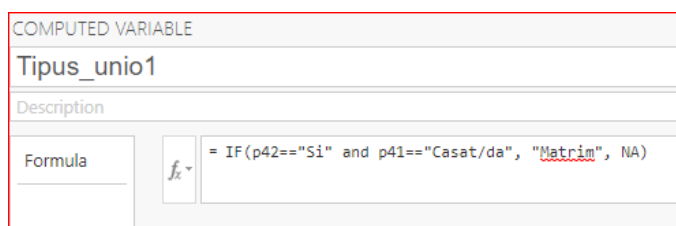
¹⁶ Per generar aquesta taula ves al menú i clica sobre Analyses / Freqüències / Independent samples. Un cop ho hakis fet, en la finestra de diàleg situa la p41 a les files i la p42 a les columnes. Per obtenir més informació sobre aquest procediment, vegeu la generació d'aquest tipus de taules a l'apartat "10. Anàlisi de taules de contingència mitjançant *jamovi*" d'aquest tutorial.

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

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

- 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.
- 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.
- 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 nova variable a continuació de la p42, situa't a sobre i clica al menú Data/Compute. Apareix la nova columna de dades.
- Posa-li un nom, com per exemple, Tipus_unio1, i si vols una descripció més extensa.
- A la finestra de fórmula (al costat del signe =) cal incloure la condició que hem escrit més amunt,¹⁷ de manera que la finestra quedarà:



Imatge 73. Càlcul utilitzant el condicional IF i els valors d'altres variables (I)

- En clicar [Intro] ja veuràs la variable construïda. En alguns casos puntuals ha escrit *Matrim*, mentre que la majoria de casos els ha deixat en blanc. Comprova que els *Matrim* es corresponen amb la definició que hem fet.

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 popi de *jamovi*, li haurem de dir:

```
IF (p42=="Si" and p41!="Casat/da", "Par de Fet", Tipus_unio1)
```

¹⁷ 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 caràcters " i " .

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 *Matrim*, que ja havíem definit a la columna.

5. Per crear, doncs, la nova variable, situa't a sobre de *Tipus_unio1* i clica al menú Data/Compute. 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:

COMPUTED VARIABLE

Tipus_unio

Tipus d'unió, per a les persones que viuen en parella

Formula

= IF(p42=="Si" and p41!="Casat/da", "Par de Fet", Tipus_unio1)

Imatge 74. 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 *Matrim*, quan ja estava definida, afegeix alguns *Par de Fet*, mentre que encara en deixa molts en blanc. Comprova que els *Par de Fet* es corresponen amb la definició que hem fet, i que els blancs són persones que no viuen en parella (p42=No).
9. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

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

Descriptives		Frequencies of Tipus_unio			
	Tipus_unio	Levels	Counts	% of Total	Cumulative %
N	369	Par de Fet	220	59.6 %	59.6 %
Missing	1050	Matrim	149	40.4 %	100.0 %

Imatge 75. 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 com hem fet, sinó que podem crear directament la variable final. Per fer-ho, la condició que hauríem d'escriure substituïria, 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", "Matrim",
IF (p42=="Si" and p41 != "Casat/da", "Par 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 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*:

Levels	Counts	% of Total	Cumulative %
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 76. Descriptiva de la variable ordinal nivell educatiu

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 heterosexuales, 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... Analyses/Freqüències/Independent Samples) 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 ($M>P$):

Contingency Tables

p54	p55				Total
	Menys prim	Primaria	Secundaria	Universit	
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 Data/Compute. 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 Data/Compute. 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 Data/Compute. 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  / Save.

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

Frequencies of Homogamia			
Levels	Counts	% of Total	Cumulative %
P>M	194	14.7 %	14.7 %
M>P	190	14.4 %	29.1 %
P=M	934	70.9 %	100.0 %

Imatge 77. 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 en les tres variables. Volem crear una variable que digui, per exemple, "Si utilitza Facebook o no". Tal i com està construïda la matriu de dades, s'utilitzarà Facebook si aquesta xarxa social és esmentada 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 Data/Compute. 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 Data/Compute. 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 Data/Compute. 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:

Frequencies of Facebook1				Frequencies of Facebook2				Frequencies of Facebook3			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
Si	928	77.9 %	77.9 %	No	491	34.6 %	34.6 %	No xarxes	248	17.5 %	17.5 %
No	263	22.1 %	100.0 %	Si	928	65.4 %	100.0 %	No	243	17.1 %	34.6 %
								Si	928	65.4 %	100.0 %

Imatge 78. Descriptiva de tres variables amb modificacions acumulatives consecutives

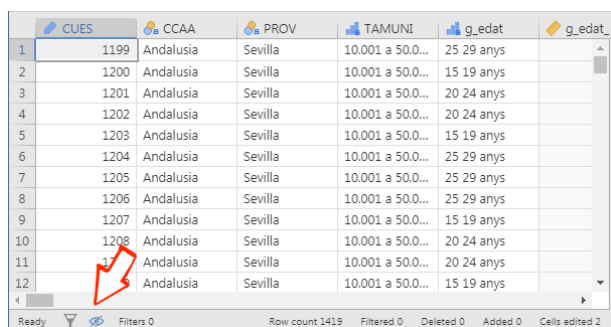
7. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

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 només 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.

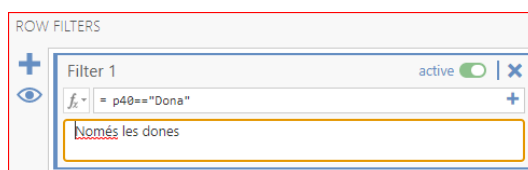


	CUES	CCAA	PROV	TAMUNI	g_edat	g_edat
1	1199	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	
2	1200	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	
3	1201	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
4	1202	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
5	1203	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	
6	1204	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	
7	1205	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	
8	1206	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	
9	1207	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	
10	1208	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
11	1209	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
12	1210	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	

Imatge 79. 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 pot activar o desactivar independentment. Tots queden desats en l'arxiu de *jamovi* *.omv



Imatge 80. Pantalla de control de filtres de casos

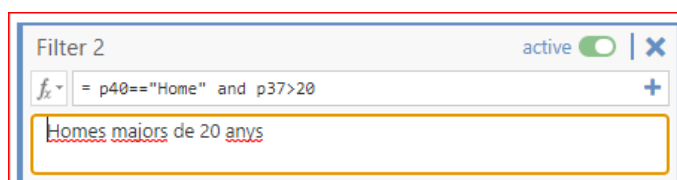
4. Qualsevol modificació sobre la fórmula d'un filtre **triga un temps en fer-se efectiva**, atès que *jamovi* recalcula totes les transformacions i anàlisis de l'arxiu per ajustar-los a la submostra seleccionada.
5. Quan un o més filtres estan activats, s'indica a la part inferior esquerra (1) i es 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.

	p37	p40	p41	p42	p41_a (2)	p42_a (2)
1	29	Mujer	Casado/a	Si	Casado/a	Si
2	19	Mujer	Soltero/a	No	Soltero/a	No
4	20	Mujer	Soltero/a	No	Soltero/a	No
6	25	Mujer	Soltero/a	Si	Soltero/a	Si
9	17	Mujer	Soltero/a	No	Soltero/a	No
11	23	Mujer	Casado/a	Si	Casado/a	Si
13	29	Mujer	Soltero/a	No	Soltero/a	No
14	19	Mujer	Soltero/a	No	Soltero/a	No
15	29	Mujer	Soltero/a	Si	Soltero/a	Si
19	20	Mujer	Soltero/a	No	Soltero/a	No
20	26	Mujer	Soltero/a	No	Soltero/a	No
23	20	Mujer	Soltero/a	Si	Soltero/a	Si

Imatge 81. 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.¹⁸ Per desactivar un filtre (no per esborrar-lo) has de clicar sobre la petita icona situada a dal a la dreta, deixant el filtre inactiu.

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**.



Imatge 82. Exemple de condició combinada de filtratge de casos

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. En aquest cas, com que volem seleccionar els homes majors de 20 anys, volem que compleixi dues condicions, de manera que utilitzarem el connector **and**, de la següent manera:
`p40 == "Home" and p37>20`

¹⁸ Similarment al “robloqueo” que pateixen els robots als relats d'Isaac Asimov (1950).

3. Sempre és útil afegir una descripció especificant la utilitat del filtre.

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.

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

- Crea un filtre amb el següent criteri:

$$-1.5 < (p37 - \text{VMEAN}(p37)) / \text{VSTDEV}(p37) < 1.5$$
- 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 univariats cal estudiar les freqüències de totes les variables, tant de les quantitatives (ORDINAL - INTEGER) 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 Analyses / 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 Analyses / 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 entrat els codis "SI" i "si", en comptes del correcte "Sí".

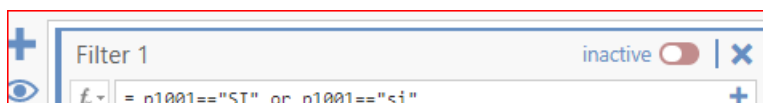
Si	No	SI	si
592	707	1	1

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 83. 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 "Sí").



Anem a fer el mateix per a la **variable quantitativa p6** (edat que tenia en començar a utilitzar Internet):

1. Clica Analyses / 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.

Errors en la congruència entre variables

Per avaluar possibles errors que afecten creuaments de variables haurem de determinar, en primer lloc, quins creuaments poden dur-nos a incongruències entre variables. 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...

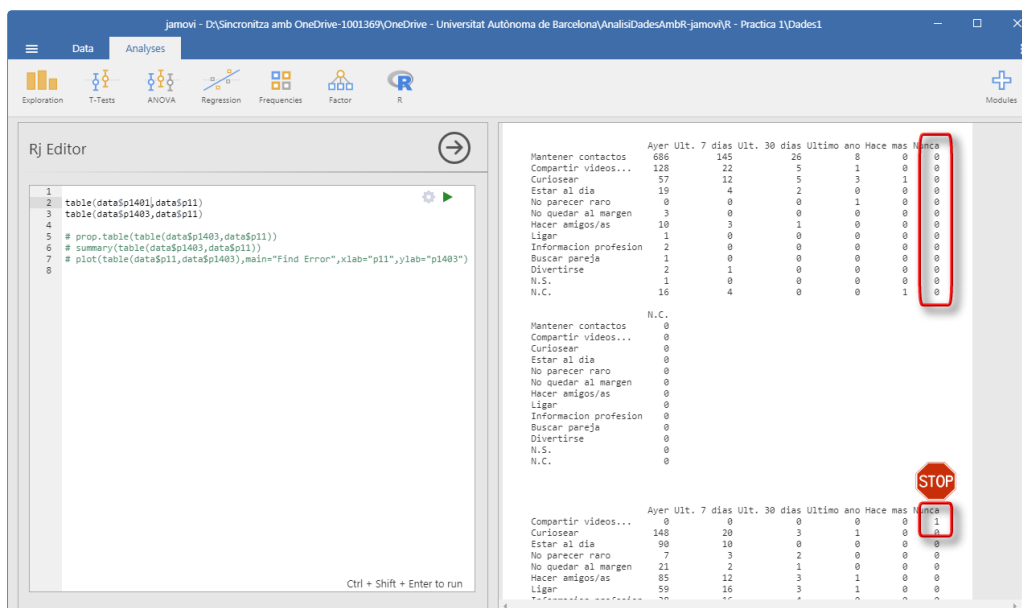
Per detectar possibles incongruències ho solem fer mitjançant "taules de doble entrada" (taules de contingència). Per exemple, si creues la p11 amb la p1401 veuràs que no hi ha cap incongruència, és a dir, ningú dels que han contestat que no fan servir mai les xarxes socials ha respost aquesta pregunta, 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 Analyses / 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 84. 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 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.

Exemple 1. En el cas 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 les dues coses, de manera que hem hagut d'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 les dues condicions, de manera que hem hagut d'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 "7. Transformació i depuració de dades mitjançant *jamovi*", obre a *jamovi* el fitxer resultant *MJDS_Dades_Exercicis.omv* i respon 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 pareja*”. 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 freqüència.

Combinant ambdues, construeix una variable que tingui tres categories: 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	1418		
Missing	1		

Frequencies			
Frequencies of Internet			
Levels	Counts	% of Total	Cumulative %
No fa servir Internet	118	8.3 %	8.3 %
Internet, no a diari	234	16.5 %	24.8 %
Internet a diari	1066	75.2 %	100.0 %

Imatge 85. 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 (TAMUNI i PROV). Un cop hakis 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'esquerreres que la mare i (3) La mare és més d'esquerreres 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.

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'interval 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 nominals/ordinals, estimació de mitjanes per a 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 2011.¹⁹ 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).

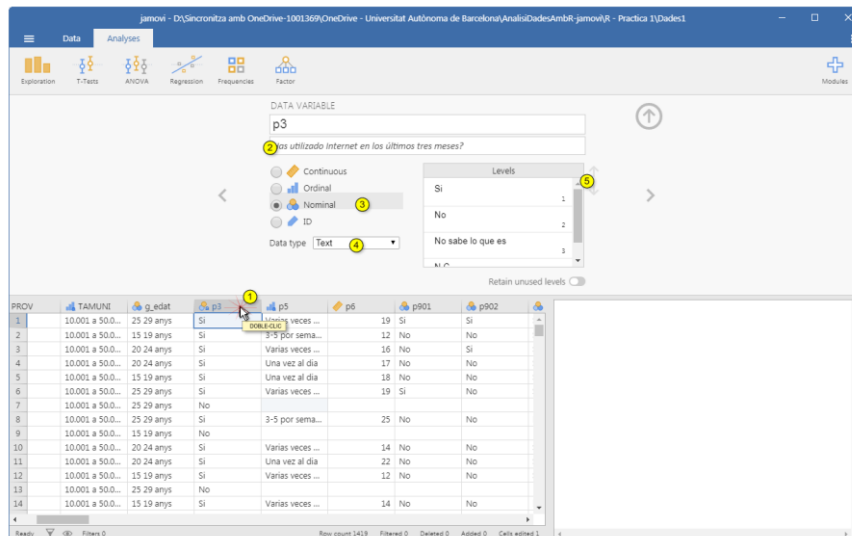
Construcció d'un interval de confiança per a una proporció (π).

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 gent 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...” (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 “Test binomial per a 2 resultats” (2 Outcomes Binomial Test).

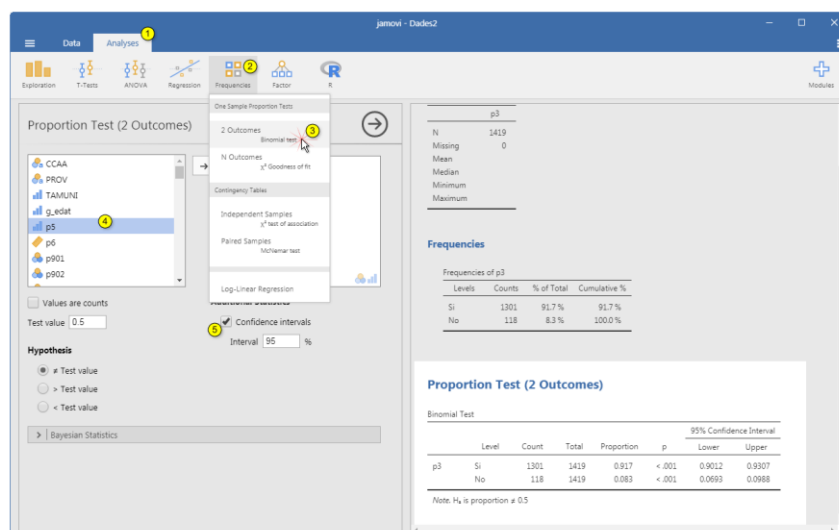
1. Comprova que la variable p3 està ben definida en *jamovi*. Fes doble-clic a sobre del nom que encapçala la seva columna. S'obre el panell de definició, que hauria d'especificar que és de tipus NOMINAL / TEXT. Si no és així, modifica la definició de la variable p3.

¹⁹ 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.



Imatge 86. Definició correcta de la variable nominal p3

2. Clica l'etiqueta "Analyses" del menú superior de *jamovi*.
3. Selecciona l'opció "Frequencies", cap a la dreta del submenú de procediments estadístics.



Imatge 87. Càlcul d'interval de confiança per a una proporció.

4. Al menú que es desplega, selecciona l'opció "2 Outcomes Binomial Test". 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 (Count), casos totals a la mostra (Total), proporció de casos respecte al total (Proportion), 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 Additional Statistics. Clica el requadre per a activar "Confidence intervals".

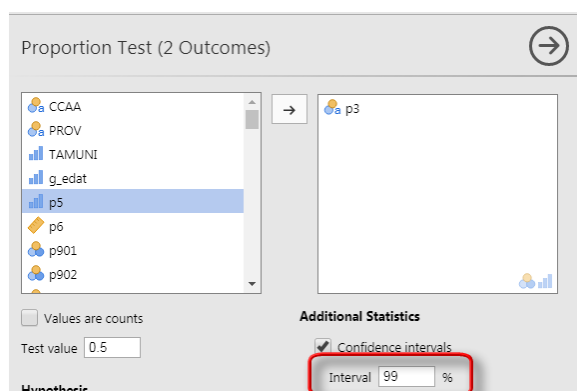
Els intervals de confiança (les dues columnes de la dreta de la taula) estan expressats en “tant per 1”, i si es multipliquen per 100 passen a ser percentatges.

Binomial Test						95% Confidence Interval	
	Level	Count	Total	Proportion	p	Lower	Upper
p3	Si	1301	1419	0.917	< .001	0.9012	0.9307
	No	118	1419	0.083	< .001	0.0693	0.0988

Note. H_a is proportion $\neq$ 0.5

El resultat que obtenim ens diu que podem assegurar, amb un 95% de confiança, que la proporció de persones (de 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 Outcomes Binomial Test” 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 88. Control del nivell de significativitat d'intervals de confiança



EXERCICI

- 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 CONTINUU (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 (Setup) com a *Data type=INTEGER*. 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 *Analyses / T-Tests / One Sample T-Test...* . S'obre el panell de definició del procediment.
4. Situa la variable de la que vols calcular l'interval a l'espai *Dependent Variables*
5. Per defecte no es calcula l'interval de confiança. Cal que activis l'opció "Confident interval" a l'apartat **Additional Statistics**.
6. Per visualitzar els estadístics descriptius cal activar l'opció "Descriptives" a l'apartat **Additional Statistics**.
7. Per defecte, el nivell de confiança és 95%, però, si cal, es pot canviar a la "finestra" corresponent.

The screenshot shows the Jamovi interface with the 'Analyses' menu open, selecting 'One Sample T-Test'. The variable 'p15' is moved to the 'Dependent Variables' box. In the 'Additional Statistics' section, 'Confidence interval' is checked and set to 95%, and 'Descriptives' is also checked. The 'Results' panel displays the following table:

One Sample T-Test						
	Statistic	df	p	Mean difference	95% Confidence Interval	
					Lower	Upper
p15_up	Student's t	24.6	1080	< .001	249	229 269

Imatge 89. Càlcul d'interval de confiança per a una mitjana

El resultat és:

One Sample T-Test

						95% Confidence Interval	
		Statistic	df	p	Mean difference	Lower	Upper
p15_sp	Student's t	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 com a *Measure type=ORDINAL*.

Codi R per a la construcció d'interval 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 Analyses / R / Rj Editor...
2. Tot i que la instrucció que cal teclejar, és la següent (*amb p15, o el nom de la variable inferida*):

```
t.test(data$p15, conf.level=0.95)
```
3. Aquesta instrucció ens donarà un problema, ja que per R la variable p15 no és numèrica. Per tant, en primer lloc l'hauré de reconvertir. Una manera de fer-ho utilitzant R és escrivint les dues instruccions següents (la primera genera, a partir de p15, la variable p15_2 que serà numèrica, i la segona demana directament l'interval de confiança):

```
data$p15_2<-as.numeric(as.character(data$p15))
t.test(data$p15_2, 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, per exemple al 99% caldrà escriure i executar (en vermell el canvi introduït):

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

De manera que ara el resultat serà:

One Sample t-test

```
data: data$p15_2
t = 25, df = 1080, p-value <2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 222.6 274.8
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 ?



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 mateixa fitxer, *MJDS_Dades_def.omv*, respon les següents qüestions.

(1) 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%.

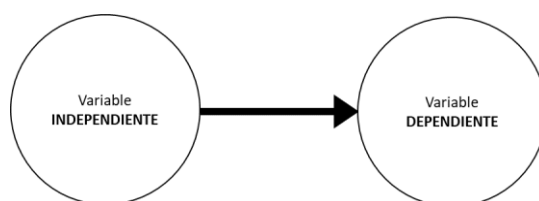
(2) A partir de la pregunta 6 del qüestionari, digues quin percentatge dels entrevistats que fan servir Internet, va començar a fer-ne ús abans dels 12 anys? Tingues en compte, abans de construir l'interval, que cal depurar la variable, atès que l'enquesta s'ha fet a joves de 15 a 29 anys, i que no hi 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.

(3) 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

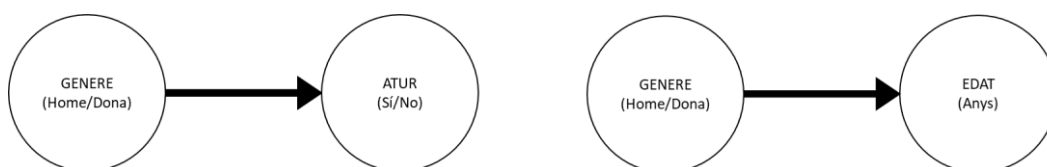
Les seccions anteriors del present Tutorial pràctic estan dedicats a la descripció i la inferència en relació a UNA variable. Aquesta és una part fonamental dels estudis socials: el coneixement de la distribució quantitativa dels valors d'un indicador concret o característica entre la població. Correspon a hipòtesis DESCRIPTIVES respecte a la població, del tipus “més d'un 20% es troba en situació d'atur”, o “la edat mitjana es 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 un 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 90. 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”.

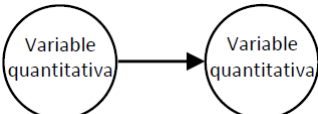
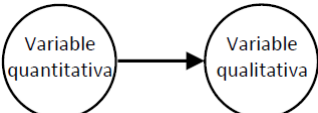


Imatge 91. Exemples d'anàlisi de la relació entre 2 variables.

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

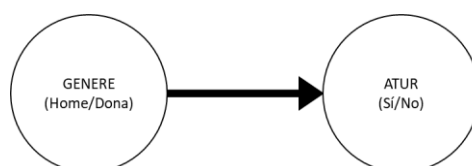
En analitzar la relació entre dues variables, en funció de la mètrica de cadascuna d'elles, tenim quatre opcions diferents, que es poden expressar de la següent forma:

COMBINACIÓ DE VARIABLES	OPERACIONS ESTADÍSTIQUES
	Taula de contingència. Test de la khi quadrat
	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 “Contingency Tables”, 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 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 92. 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 perquè 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 bivariants, 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 a la mostra 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 dependent és la que ens generarà els grups de població comparables.

En aquesta secció es treballarà amb el fitxer de dades originals “Dades2.rdata”. Es tracta d’un fitxer de dades en format *R*, semblant al que teníem en anteriors apartats i que es basa en el mateix qüestionari. L’únic que s’ha fet ha estat depurar-lo, 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. 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*” <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, les que podeu consultar en el document *MJDS_QüestionariCIS.pdf*, que necessitareu per anar seguint el Tutorial.

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 TAULES DE CONTINGÈNCIA.
2. Descarrega al teu PC els arxius necessaris: “MJDS_QüestionariCIS.pdf” i “MJDS_Dades.rdata”.
3. Executa l’aplicació *jamovi* al teu ordinador.
4. Fes clic al menú general (dalt, esquerra).

5. Selecciona l'opció Open / This PC. Al panell central i clica [Browse] (dalt, dreta).
6. S'obre una finestra per seleccionar l'arxiu de dades a obrir. Tria l'arxiu "MJDS_Dades.rdata".

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.

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 de "endevinar" el tipus de la variable a partir de les dades contingudes a l'arxiu. Aquesta identificació automàtica no sempre és l'òptima.

Sempre cal revisar la definició de cadascuna de les variables quan s'obre (o s'importa) amb *jamovi* un arxiu de dades que no està en el seu propi format (.omv).

Com veuràs més endavant, caldrà repassar les variables de l'arxiu "Dades2.rdata" utilitzades en aquest apartat, comprovant que estan correctament definides a *jamovi*.

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

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

L'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 "Contingency Tables" al menú *Analyses / Frequencies*. 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.

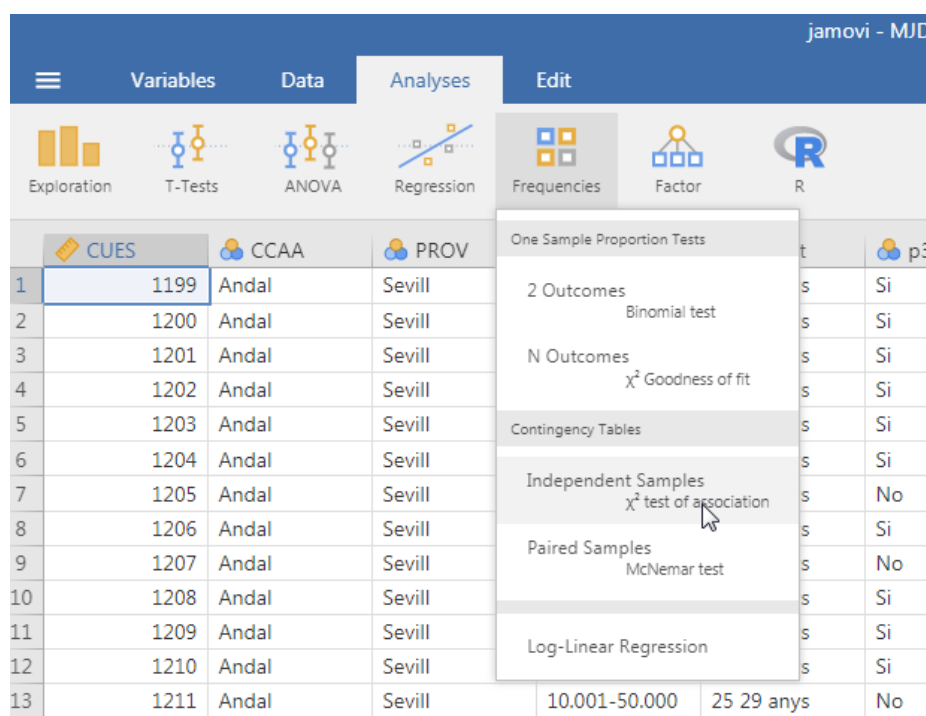
En aquest exemple anem a explorar la possible relació entre la utilització d'Internet (p3) i l'estatus socioeconòmic (ESTATUS). Una taula de contingència (TdC) o taula de freqüències creuades, té dues dimensions, una variable se situa en horitzontal amb les seves categories a les columnes de la taula, 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", anomenades així perquè és impossible que un mateix cas/persona estigui 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 sub-grups d'una població que es volen comparar, té un significat TOTALMENT DIFERENT a quan la mateixa terminologia es refereix al paper que juga 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/INTEGER. Si no és així, modifica la seva definició ("Setup"). Utilitza el que has après anteriorment al Tutorial: pots ajudar-te de l'arxiu "MJDs_QüestionariCIS.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 "Analyses" / "Freqüències" / "Contingency Tables" / "Independent Samples – χ^2 test of association".



Imatge 93. 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).

Contingency Tables

Rows: ESTATUS
Columns: p3

Results

Contingency Tables

ESTATUS	p3		Total
	Si	No	
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

χ^2 Tests

	Value	df	p
χ^2	38.2	4	< .001
N	1343		

Imatge 94. Taula de contingència simple, mostres independents

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").

Contingency Tables

ESTATUS	p3		Total
	Si	No	
C.A/C.M-A	218	3	221
Nueva C.M	223	12	235
Antigua C.M	209	15	224
Obreros cualif	433	47	480
Obreros no cualif	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 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:

- Observed. Freqüència absoluta i realment observada per a cada casella.
- Expected. Freqüència "simulada" o calculada "en condicions d'independència". Ens fa una estimació dels efectius que seria esperable trobar a cada cel·la ES CAS QUE NO HI hagués cap relació entre les dues variables.
- % within row. 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 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%.
- % within column. 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%.
- % of 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.

Observació! Els percentatges que millor expliquen la relació entre dues variables són sempre els percentatges respecte de la variable independent, que en aquest cas és l'estatus. 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.

A grans trets, en l'exemple anterior:

1. Fes clic a la taula. S'obre el panell de definició de la taula de contingència.
2. Clica el desplegable de les opcions > Cells.
3. Activa l'opció de percentatges [] Row.
4. Desactiva l'opció de recomptes [] Observed counts.

Contingency Tables

Rows: ESTATUS
Columns: p3

Counts (optional):
Layers:

Cells

Counts:
☐ Observed counts
☐ Expected counts

Percentages:
☒ Row
☐ Column
☐ Total

Results

Contingency Tables

Contingency Tables

ESTATUS		p3		Total
		Si	No	
Alta/mitj-alta	% within row	98.6 %	1.4 %	100.0 %
Nova classe mitj	% within row	94.9 %	5.1 %	100.0 %
Vella classe mitj	% within row	93.3 %	6.7 %	100.0 %
Obr qualif	% within row	90.2 %	9.8 %	100.0 %
Obr no qualif	% within row	83.1 %	16.9 %	100.0 %
Total	% within row	92.0 %	8.0 %	100.0 %

χ^2 Tests

	Value	df	p
χ^2	38.2	4	< .001
N	1343		

Imatge 95. Percentatges per fila (V.I.) a una Taula de Contingència.

Si llegim els percentatges respecte de l'estatus (*within row*) 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 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 un 83,1% fa servir Internet.



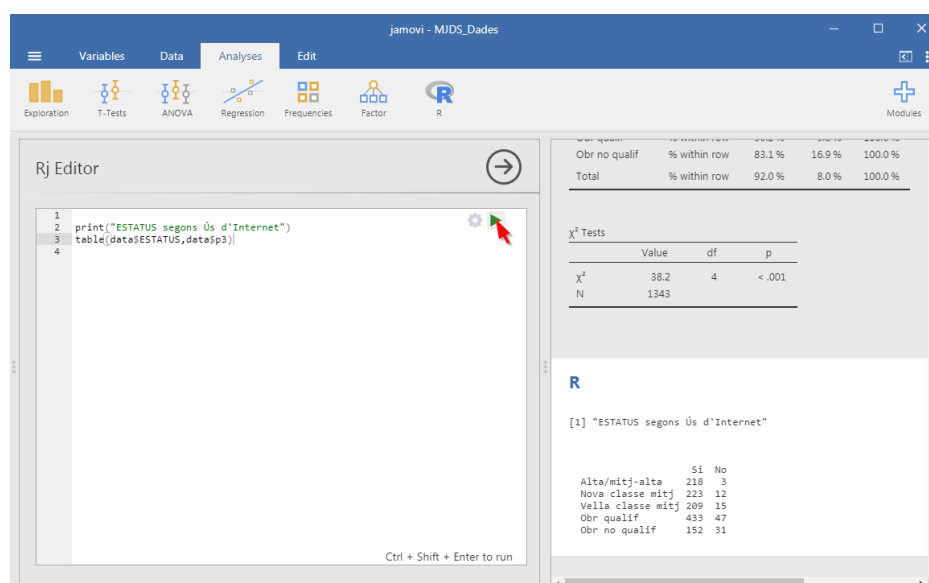
Per a obtenir el resultat d'una **taula de contingència directament executant codi R** amb l'extensió (Rj) de *jamovi* cal:

1. Tenir instal·lat el mòdul o extensió de *jamovi* anomenat Rj Editor.²⁰
2. A l'apartat "Analyses", clicar a 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:

```
print("ESTATUS segons Ús d'Internet")
table(data$ESTATUS, data$p3)
```

²⁰ 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. 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: ‘ESTATUS segons Ús d'Internet’ ”, 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 96. Taula de contingència generada amb codi R

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  / Save.



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 de “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.

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

Aquestes ordres que has escrit en llenguatge R signifiquen:

“genera l’objecte anomenat TITOL que serà ‘ESTATUS segons Ús d'Internet (%)’ ”,

“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 decimal (,2)”

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.

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. Escribeu i executa, una a una, les següents ordres:

```
TITOL
TPerc
TP2dec
```

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.

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

Atenció! Com has pogut veure, el codi R permet més control sobre el contingut i el disseny de cada taula, i alhora has pogut veure com una taula definitiva es pot construir 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 un **gràfic de barres apilades al 100%**. Cada barra representa el 100% de cada categoria

de la variable independent situada a les columnes, mentre que cadascuna d'aquestes barres està dividida proporcionalment a la distribució de la variable dependent situada a les files.

Per poder generar amb *jamovi* aquest tipus de gràfics, a dia d'avui cal utilitzar la consola Rj Editor; no hi ha cap menú *jamovi* habilitat per fer-ho.

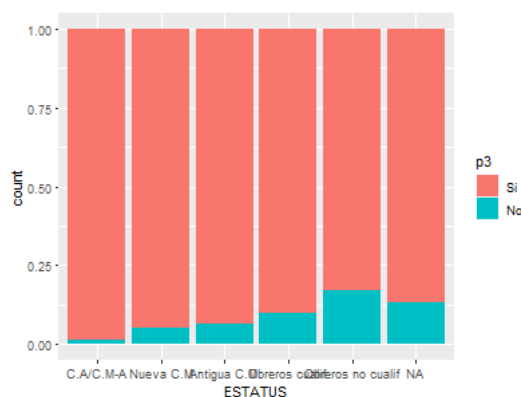
Cal tenir en compte que *jamovi* és un projecte en desenvolupament i que incorpora la potència de R. D'aquí a un temps possiblement s'incorpori un menú o un mòdul per facilitar la construcció i modificació d'aquest i d'altres gràfics.

La programació R que genera 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:²¹



VERSIÓ SIMPLE

```
library(ggplot2)
ggplot() + geom_bar(aes(y=..count.., x=ESTATUS, fill=p3), data=data, position = position_fill())
```



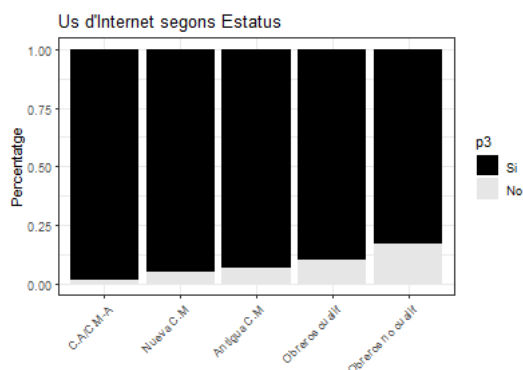
Imatge 97. Representació gràfica de taula de contingència (SIMPLE)



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, end = .9) +
  theme_bw() +
  theme(axis.text.x=element_text(angle=45, hjust=1))
```

²¹ 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 una altra vegada, tot i que aquí la incloquem a cada gràfic.



Imatge 98. 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, end = .9) + theme_bw()`
- S'ajusta l'angle de les etiquetes de l'eix horitzontal, perquè no se solapin i siguin totes visibles, amb `theme(axis.text.x=element_text(angle=45, hjust=1))`

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.

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, donat que les dades han estat obtingudes d'una mostra, de manera que les conclusions derivades de l'anàlisi no poden extrapolar-se al conjunt de la població.

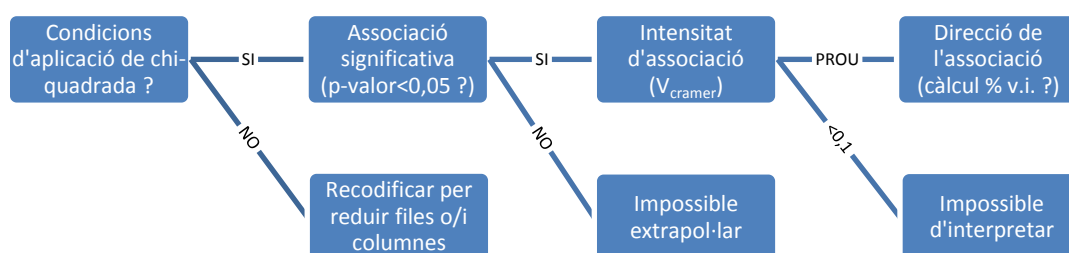
L'anàlisi complet de la relació entre dues variables qualitatives (nominal o ordinal) segueix els clàssics QUATRE passos successius i necessaris de qualsevol test d'hipòtesis:

- 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.
- Hi ha **associació estadística significativa** entre ambdues variables? Les diferències observades, són significatives? Cal comprovar que el p-valor sigui $\leq 0,05$ (*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 99. 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 ($I \times J$, número de files multiplicat pel número 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*). I és que aquestes **freqüències esperades no poden ser 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 ("Expected"):

1. Fes clic a 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 > Cells.
2. Activa l'opció de percentatges [] Expected counts.
3. Desactiva la resta d'opcions de recomptes.
4. El resultat es mostra al panell de la dreta:

Contingency Tables		p3	
ESTATUS		Si	No
C.A/C.M-A	Expected	203	17.8
Nueva C.M	Expected	216	18.9
Antigua C.M	Expected	206	18.0
Obreros cualif	Expected	441	38.6

Contingency Tables

ESTATUS		p3	
		Si	No
Obreros no cualif	Expected	168	14.7
Total	Expected	1235	108.0

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 extrapolables a la població,
- mentre que si la significació de χ^2 és **petita** és que les diferències difícilment seran degudes a l'atzar, i per tant 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 ($p\text{-valor} \leq 0,05$) 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 $\leq 0,05$ per a considerar les conclusions "extrapolables".

El n.c. 95% és acceptable a Ciències Socials en general. En altres disciplines (més restrictives en quant als requeriments d'evidència empírica) només és acceptable un n.c. del 99%; cosa que significa que s'exigeix un p-valor $\leq 0,01$ per poder extrapolare 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.

χ^2 Tests			
	Value	df	p
χ^2	38.2	4	< .001
N	1343		

Si no vols que ens doni aquesta informació, al panell de definició del procediment “Contingency table” cal desplegar l’apartat “Statistics” i desmarcar l’opció ☐ χ^2 .

Recorda! Als resultats de 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 número decimal. Així 0,001 apareix com .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 només 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).

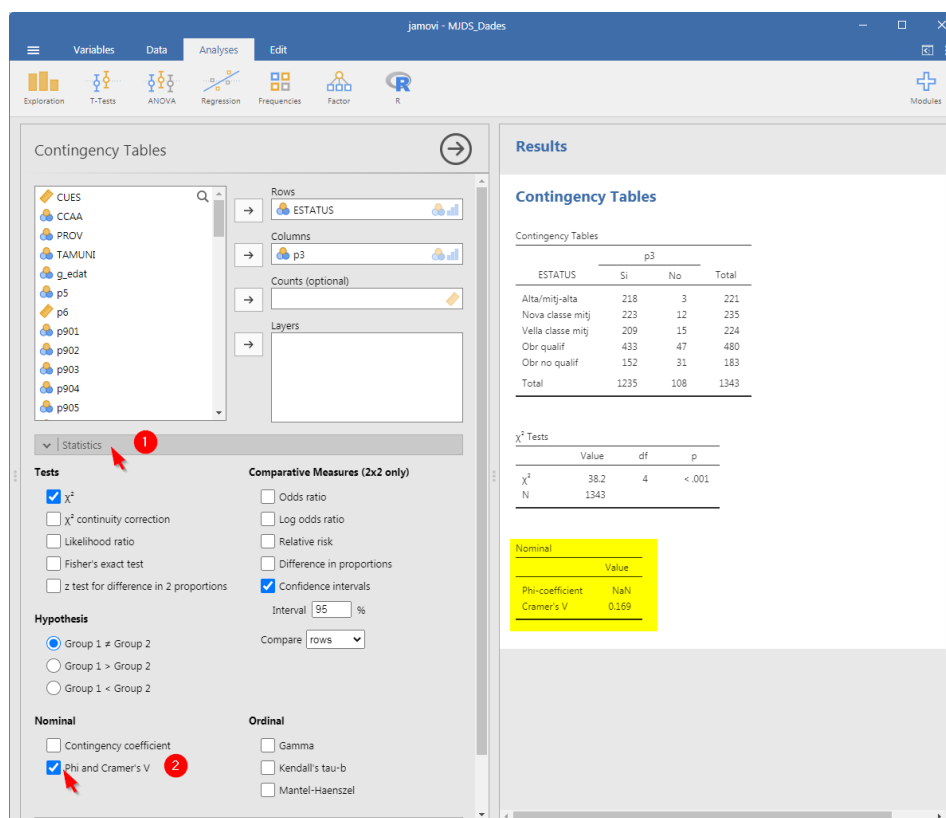
Un cop establert que hi ha relació significativa entre ambdues variables se sol calcular una mesura per avaluar la intensitat d’aquesta relació. Es habitual fer-ho a partir de l’indicador d’associació bilateral **V de Cramer**, que varia entre 0 i 1: Un valor de 0 equivaldria a dir que no hi ha cap relació (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 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. Fes clic a 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 > Statistics.
2. A l’apartat “Nominal”, activa l’opció “☐ Phi and Cramer’s V”.



Imatge 100. 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 relació de determinació entre elles. La diferencia és que l’associació no pressuposa que una de les variables determini o sigui causa de l’altra, contràriament la determinació pressuposa un “sentit” o una “direcció” concreta de l’explicació.

Si una de les variables determina l’altra, en aquest cas, voldria dir que en algunes categories socioeconòmiques tothom fa servir Internet i en altres no hi ha ningú que es connecti, per tant es pot suposar que la posició d’ESTATUS determina l’Ús d’Internet (i no al contrari, el fet de començar o deixar d’usar Internet no “determina” 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 bi-variable.

En cas que existeixi justificació teòrica, com la que s’acaba d’apuntar, es pot procedir a analitzar amb *jamovi* el sentit de la determinació. Per això es pot utilitzar qualsevol de dues alternatives:

1. una anàlisi del percentatges (disponible als menús de *jamovi*), o
2. una anàlisi de residuals (disponible executant R a la consola de *jamovi*).

(1) Fer una **anàlisi dels percentatges** d’una taula de contingència a la qual s’ha identificat una variable com a dependent i una altra com a independent implica:

- Demanar una taula encreuada 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 encreuada de percentatges amb *jamovi*, cal:

1. Selecciona el procediment de Taules per a mostres Independents, a menú “Analyses” / “Freqüències” / “Contingency Tables” / “Independent Samples – χ^2 test of association”.
2. Selecciona la variable DEPENDENT (p3) i mou-la a l’espai dret per a “Rows” (files).
3. Selecciona la variable INDEPENDENT (ESTATUS) i mou-la a l’espai dret per a “Columns” (columnes).
4. Clica per a desplegar l’apartat > Cells.
5. Com que la variable independent (ESTATUS) està en el sentit de les COLUMNES, cal activar l’opció de percentatges per [] Columns.
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  / Save.

Contingency Tables

		ESTATUS				
p3						
		C.A/C.M-A	Nueva C.M	Antigua C.M	Obreros cualif	Obreros no cualif
Si	% within column	98.6 %	94.9 %	93.3 %	90.2 %	83.1 %
No	% within column	1.4 %	5.1 %	6.7 %	9.8 %	16.9 %
Total	% within column	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %

Recorda la regla del “creuament”! Si la variable independent (causal) està a les columnes els percentatges es calculen per columnes i es comparen per a cada fila de la variable dependent (efecte). Si la variable independent (causal) 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 (Clase Alta / Mitja-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.

(2) 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:

Poder executar codi R amb l'extensió (Rj) de *jamovi* que implica tenir instal·lat el mòdul o extensió de *jamovi* "Rj Editor".²²

A l'apartat "Analyses", clica a la icona "R". Al desplegable, seleccionar l'opció "Rj Editor",

Escriu, 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 la ordre:

```
taula1<-table(data$p3,data$ESTATUS)
taula2<-chisq.test(taula1)
taula3<-taula2$stdres
taula3
```

Aquestes simples ordres que has escrit en llenguatge R signifiquen:

"Genera un objecte anomenat taula1 què conté la taula de freqüències observades del creuament de la variable p3 a les files i la 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".

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.

Apareixerà el resultat al panell de la dreta, integrat amb la resta de resultats de l'arxiu.

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.

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

²² 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*.

El resultat es la següent taula de residus ajustats (o “corregits”).

Us	WWW	C.A/C.M-A	Nueva	C.M	Antigua	C.M	Obreros	cualif	Obreros	no cualif
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 més gran ($>1,96$) o més baix ($<1,96$) que d'esperats en condicions d'independència de les variables. Això permet identificar la 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 (C.A/C.M-A) 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 què 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 determina el nivell d'ús d'Internet.

(3) 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:

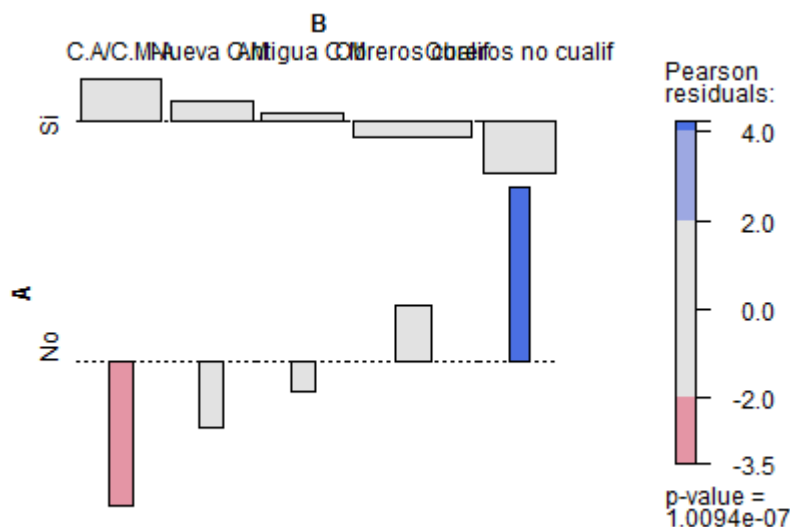
1. Poder executar codi R amb l'extensió (Rj) de *jamovi* que implica tenir instal·lat el mòdul o extensió de *jamovi* “Rj Editor”.²⁴
2. A l'apartat “Analyses”, clica a la icona “R”. Al desplegable, seleccionar l'opció “Rj Editor”,
3. Escriu, poc a poc, la ordre per a generar el gràfic d'associació amb 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)
```

²³ 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.

²⁴ 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. Aquestes simples ordres que has escrit en llenguatge R signifiquen:
 - a. “Activa a R la possibilitat de fer gràfics i càlculs per analitzar variables categòriques”, i
 - b. “Mostra un gràfic d’associació amb la variables p3 com a dependent i la variable ESTATUS com a independent o causa, coloreix els valors residuals segons la seva desviació a la freqüència esperada”.
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’anàlisi 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  / Save.



Imatge 101. Representació gràfica de residuals d'una taula de contingència.

Al gràfic d’associació es pot veure, en horitzontal, les dues categories de la variable dependent (Sí i No, ús d’Internet) amb una línia discontinua 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**.²⁵

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 la existència de categories que “contribueixen” en una mesura elevada a la diferencia 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**

²⁵ La ordre R “assoc” només suporta el càlcul dels residus de Pearson (no estandarditzats, ni tipificats, ni coregits).

ús d'Internet que el conjunt. Correlativament, a la part superior del gràfic es pot veure què 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 determina el nivell d'ús d'Internet.

Solució als problemes de compliment 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 (ESTUDIOS) 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:

Contingency Tables

		ESTUDIOS				
p20		Sin est	Prim	F.P.	Second	Univers
Molt imp.	Expected	0.4056	222.7	102.2	110.7	86.0
Bastant imp.	Expected	0.4406	241.9	111.0	120.3	93.4
Poc imp.	Expected	0.0963	52.9	24.3	26.3	20.4
Gens imp.	Expected	0.0575	31.6	14.5	15.7	12.2
Total	Expected	1.0000	549.0	252.0	273.0	212.0

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 podem analitzar la taula tal i com hem fet en l'anterior apartat.



EXERCICI

1. Crea una nova variable, basada en la variable original ESTUDIOS, mitjançant una **transformació** que recodifiqui els valors de forma que "Sin Est" i "Prim" s'unifiquin en una

nova categoria “Prim. o menys”, mentre que la resta de categories d’ESTUDIS es mantinguin igual a la nova variable.²⁶

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 continuity correction.

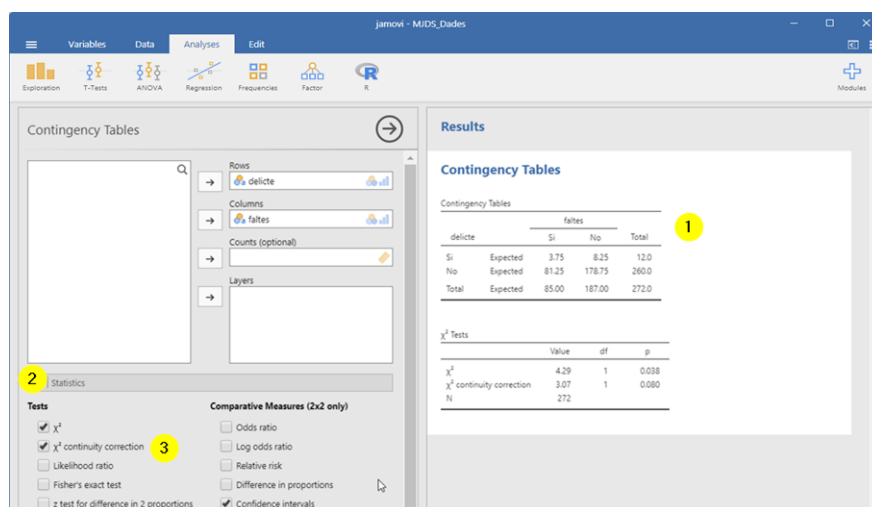
Anem a veure-ho amb un exemple extret d’un fitxer provinent del món de la criminologia, que només conté dues variables, el fitxer *reincid.rdata*. En aquest fitxer trobaràs les variables *faltes* i *delicte*. La primera fa referència a haver tingut faltes de comportament durant la condemna (Si/No), mentre 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 “Analyses” / “Freqüències” / “Contingency Tables” / “Independent Samples – χ^2 test of association”.
3. Selecciona la variable “faltes” (independent) i mou-la a l’espai dret per a “Columns” (columnes).
4. Selecciona la variable “delicte” (dependent) i mou-la a l’espai dret per a “Rows” (files).
5. Clica per a desplegar l’apartat > Cells.
6. Activa l’opció de recompte [] Expected counts.
7. Per tenir una visió més clara, desactiva tota la resta d’opcions de > Cells.
8. Recorda desar l’arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

Contingency Tables			
		faltes	
delicte		Si	No
Si	Expected	3.75	8.25
No	Expected	81.25	178.75
Total	Expected	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.

²⁶ Les transformacions de variables i recodificació es van veure a a la pàgina núm 60 d’aquest tutorial.



Imatge 102. Chi-quadrat amb correcció de Yates per a taules 2x2.

Per a demanar la correcció de Yates o correcció per continuïtat:

1. Fes clic a la taula generada, al panell dret de resultats.
2. Al panell de definició del procediment estadístic (esquerra) desplega l'apartat > Statistics".
3. Activa l'opció de tests l'opció [] χ^2 continuity correction.
4. No desactives la resta d'opcions de > Statistics.
5. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

χ^2 Tests

	Value	df	P
χ^2	4.29	1	0.038
χ^2 continuity correction	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.

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 95% n.c., entre les variables. Compte amb casos com aquest, on el criteri general del test khi quadrat (p-valor $\leq 0,05$) 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)



EXERCICIS PROPOSATS: TAULES DE CONTINGÈNCIA

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

Obre amb *jamovi* el fitxer anterior *Dades1.omv* i respon les següents qüestions.

(1) A partir del creuament de la variable p16 i el sexe (p40), contesta aquestes tres preguntes.

1. Quin percentatge del total d'entrevistats és dona i considera molt probable que quedés amb algú que ha conegut a través d'Internet?
2. Quin percentatge dels que consideren gens probable que quedessin amb algú que han conegut a Internet, és dona?
3. 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.

(2) A partir del creuament gràfic de la variable freqüència d'ús d'Internet (p5) i la variable estudis de l'entrevistat (ESTUDIOS) 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 dues variables.

(3) Comprova a la mostra les diferències entre homes i dones en l'ús d'Internet. Avaluja amb un 95% de confiança, si aquestes diferències 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

(4) 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, representa-les gràficament, calcula la V de Cramer i analitza els residuals ajustats. Enganxa els principals resultats en un processador de textos i comenta'ls.

(5) Fes el mateix utilitzant com a variable independent la mida del municipi (TAMUNI) i estudia la seva relació amb l'ús d'Internet. A l'hora d'interpretar els resultats tingues en compte que la variable mida del municipi està en milers d'habitants. Observa com en aquest cas l'agrupació no és tant necessària. Calcula la V de Cramer i analitza els residuals. Enganxa els principals resultats en un processador de textos i comenta'ls.

(6) 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.

(7) 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.²⁷ Per fer-ho, basa't en la pregunta 14 del qüestionari. Arriba al màxim de conclusions possibles en la relació entre aquesta variable i el sexe (p40).

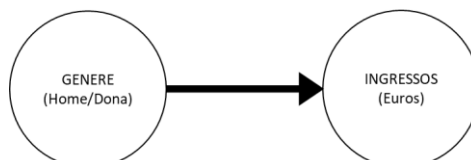
(8) Tal i com mostra la pregunta 21 del qüestionari un dels usos d'Internet és el d'obtenir informació sobre l'actualitat política. Digues si hi ha alguna relació entre aquesta variable i la ideologia política (p33), arribant a totes les conclusions possibles. Fes el test corresponent amb un 99% de confiança.

(9) 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 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.

²⁷ 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 "Ligar" i "Buscar pareja". Crea una variable que ens informi de quanta gent fa servir les xarxes socials amb aquesta finalitat. Fes-ho canviant els perduts inicials de les tres variables per algun valor conegut, per exemple NR.

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 es pot calcular la mitjana) mitjançant els procediments “T-tests” i “One-way ANOVA”, 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’extrapol·lar-la al conjunt de la població (aspecte inferencial de l’anàlisi).



Imatge 103. Exemple de model d’anàlisi de comparació de mitjanes.

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), en el context de les anàlisis bivariades, quan parlem de diferències ens referim a diferències entre els grups que genera una variable respecte d’una altra. Si disposem d’una mostra de dades de la variable “gènere” i de la variable “ingressos”, quan parlem de diferències ens referim a les possibles diferències en 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 és mostrar com 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), i una altra de quantitativa, la variable dependent (en *jamovi*, de tipus CONTINUOUS, amb dades INTEGER o DECIMAL). El tipus de variables és, de fet, la diferència entre el test de la khi quadrat i la comparació de mitjanes.

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 T-test de comparació de dues mitjanes independents, mentre que en el segon ens trobem davant d’una anàlisi de la variància.²⁸

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:

²⁸ La comparació de dues mitjanes independents (T-test) i l’anàlisi de la variància s’expliquen més endavant en aquesta mateixa Pràctica.

1. Tests paramètrics
2. Tests no paramètrics

Tests paramètrics. Són els més recomanats, en tant que treballen amb la informació “completa”, 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 les mostres 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.²⁹

A banda d'aquesta qüestió, i entrant directament dintre de la família de tests paramètrics, disposem de dues possibilitats de test, en funció de com són les variàncies dels grups considerats:

- Variàncies iguals (“homoscedasticitat”): Si en tots els grups hi ha la mateixa variància.
- Variàncies diferents (“heteroscedasticitat”): Si algun dels grups presenta una variància diferent dels altres.

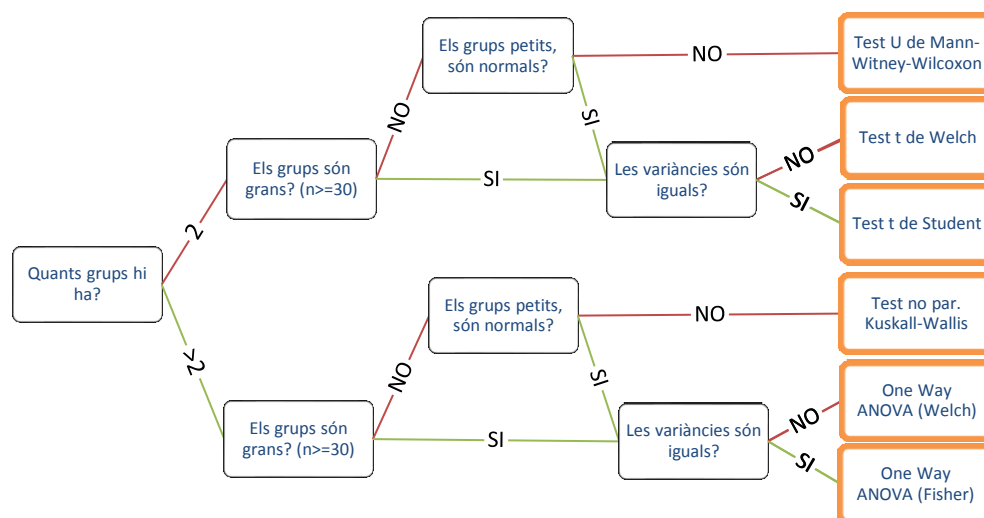
Tests no paramètrics. Tenen l'inconvenient que en la seva aplicació es perd molta informació, en tant que no treballen amb els valors quantitativs sinó, habitualment, amb el rang, amb la posició que ocupen les observacions. En aquest sentit només els farem servir quan, per qüestions de mostra o normalitat, no sigui possible fer un test paramètric.

Observació! La majoria d'autors considera que si un grup té més de 30 casos ja es pot considerar gran, de manera que si tots els grups tenen més de 30 observacions 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-ne la normalitat d'aquest/s grups/s; i, sinó es compleix en un d'ells, fer un test no paramètric.

Quin test cal fer?

Esquemàticament, les diferents situacions es poden expressar seguint aquest arbre de decisions.

²⁹ Per avaluar la normalitat en una distribució vegeu més endavant l'apartat corresponent d'aquesta Pràctica 3.



Imatge 104. 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?

Per poder seguir el Tutorial pas a pas, en aquesta secció es treballarà amb el fitxer de dades "MJDS_Dades3.omv" que és pràcticament el mateix que en la pràctica anterior, i que es basa en el qüestionari "MJDS_QüestionariCIS.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_QüestionariCIS.pdf" i "Dades3.omv".
3. Executa l'aplicació *jamovi* al teu ordinador.
4. Fes clic a el menú general (dalt, esquerra).
5. Selecciona l'opció Open / This PC. Al panell central Fes clic a [Browse] (dalt, dreta).
6. S'obre una finestra per seleccionar l'arxiu de dades a obrir. Tria l'arxiu "Dades3.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.

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 de "endevinar" el tipus de la variable a partir de les dades contingudes a l'arxiu. Aquesta identificació automàtica no sempre resulta en la forma més adient de 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 que no està en el seu propi format (.omv).

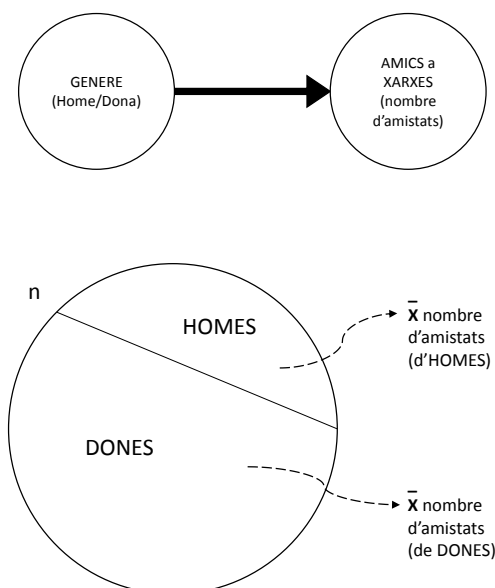
Com veuràs més endavant, caldrà repassa les variables de l'arxiu "Dades2.rdata" utilitzades a aquest apartat, comprovant que estan correctament definides a *jamovi*.

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

- Utilitza el menú general / Save As ... i desa'l en una ubicació que puguis recuperar en un altre moment l'arxiu en format propi de *jamovi* ("Dades2.omv").

L'estadística descriptiva en la comparació de mitjanes.

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).

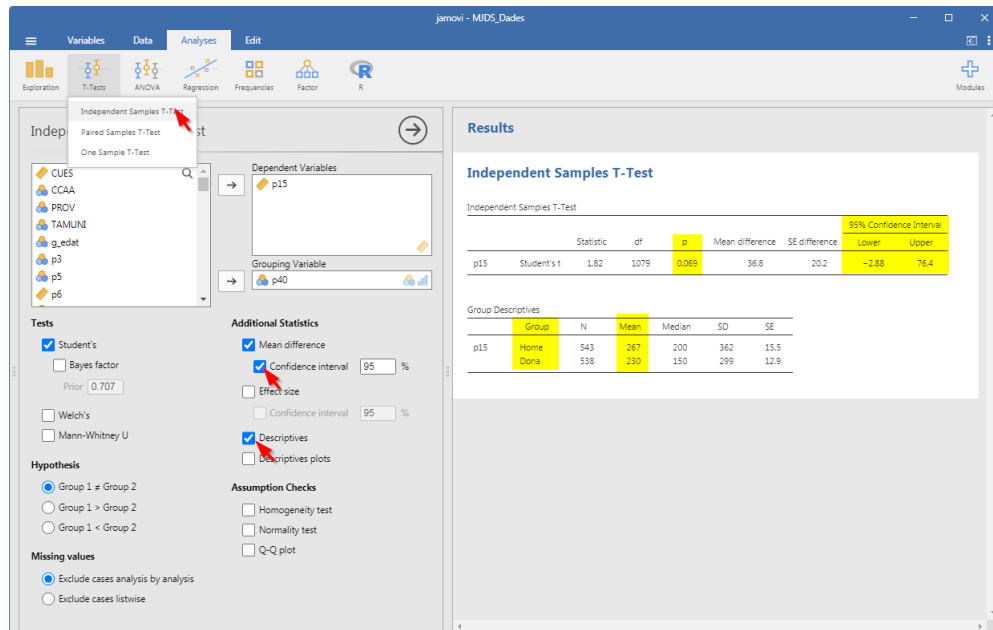


Imatge 105. Exemple d'anàlisi de comparació de 2 mitjanes, mostres independents.

Per descriure aquesta relació:

1. Comprovar el tipus de les variables i de dades: Contactes (p15) ha de ser CONTINUOUS i INTEGER, i Sexe (p40) ha de ser NOMINAL i TEXT.
2. Clic a menú Analyses / T-Test / Independent Samples T-Test...
3. La variable quantitativa dependent (p15) la situarem sempre a l'espai superior "Dependent Variables".
4. La variable qualitativa independent (p40) a l'espai inferior *Grouping Variable*.
5. Per defecte no es calcula l'interval de confiança. Cal que activis l'opció "Confident interval" a l'apartat **Additional Statistics**.
6. Per visualitzar els estadístics descriptius cal activar l'opció "Descriptives" a l'apartat **Additional Statistics**.

7. Per defecte, el nivell de confiança és 95%, però, si cal, es pot canviar a la “finestra” corresponent.

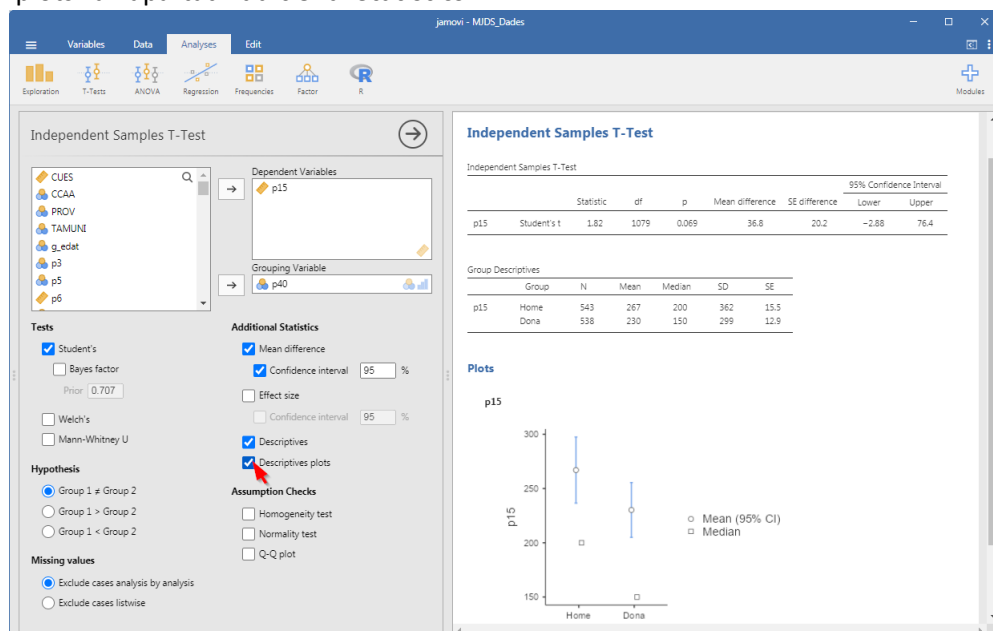


Imatge 106. Anàlisi paramètrica de comparació de dues mitjanes de mostres independents.

La interpretació d'aquests resultats és que la mitjana de contactes a xarxes socials dels homes (267 amitats) és, significativament, més alta que la de les dones (230 amitats).

Gràficament aquesta informació se sol resumir a partir de la comparació de la posició de les mitjanes, les medianes i els intervals de confiança de cada grup sobre la mateixa escala:

1. Clic a la taula resultat del procediment “Independent Samples T-Test”
2. Per defecte no es generen representacions gràfiques. Cal que activis l'opció “Descriptive plots” a l'apartat **Additional Statistics**.

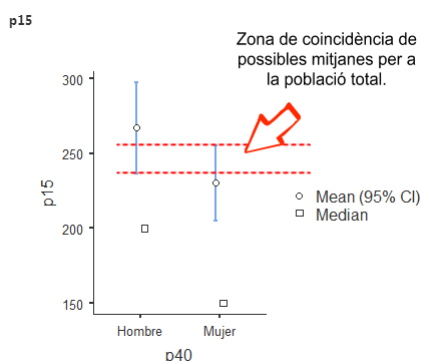


Imatge 107. Gràfic de BARRES D'ERROR per a comparació de dues mitjanes de mostres independents.

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 (“error bars”) s’estenen amunt i avall fins a una desviació estàndard (SD, o *sigma*) de distància a partir de les mitjanes.

La interpretació d’aquesta visualització gràfica es:

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 a la població en general, perquè els nivells de confiança “coincideixen” en part.



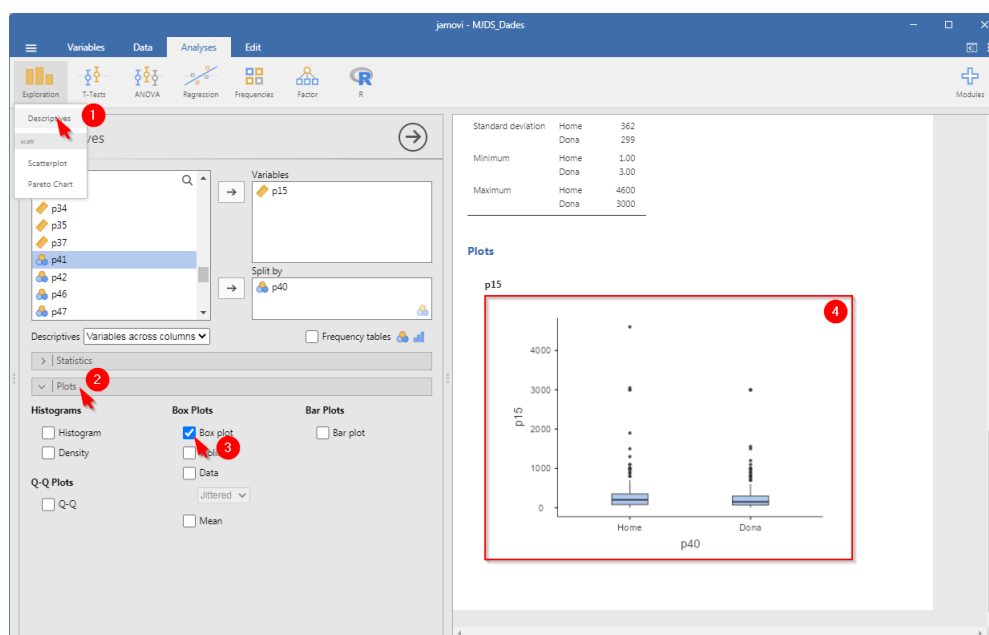
Imatge 108. Solapament d’interval·ls de confiança.

A la vista dels resultats d’aquest gràfic, es podria donar el cas que la mitjana “real” de tota la població d’homes fos 248 i la mitjana “real” de tota la població de dones fos 252.

1. Identifica aquests dos punts al gràfic, tots dos situats dins dels respectius interval·ls de confiança.
2. Comprova com aquesta “realitat”, possible dins del marge d’error, és contrària al resultat que s’ha obtingut a la mostra (*que la mitjana de contactes a xarxes socials dels homes és més alta que la de les dones*).

Una representació alternativa pot ser un gràfic de caixes múltiples, amb una caixa per a cada grup de la variable qualitativa.

1. Clic a menú Analyses / Exploration / Descriptives...
2. La variable quantitativa dependent (p15) la situarem sempre a l’espai superior “Variables”.
3. La variable qualitativa independent (p40) a l’espai inferior *Split by*.
4. Per defecte no es mostren resultats gràfics. Cal que facis clic a > Plots. Es desplega el menú d’opcions gràfiques.
5. Activa l’opció “Box plot” a l’apartat **Box Plots**.



Imatge 109. Gràfic de CAIXA per a comparació de dues mitjanes de mostres independents.

Per a interpretar els resultats cal recordar com es construeixen els diagrames de caixes: el rectangle de la caixa està delimitat en alçada (eix Y) pel quartils 1er i 3er de la variable quantitativa (percentils 25% i 75%) i delimita el 50% central de la distribució de casos, la línia horitzontal interior indica la mediana (50%), i els punts indiquen valors “desviats” (*outliers*) que queden fora de la centralitat de la distribució.

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 els gràfics de barres d'error concentren l'atenció en la part central de la distribució i per tant 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 un dels grups comparats supera els 30 casos ja és suficient. Així, només avaluem 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:

- H_0 : la distribució és normal
- H_1 : 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 situació laboral (p49) i el temps (en minuts) dedicat a les xarxes socials (minuts_XS), el primer que fariem seria estimar la **mida de la mostra per a cada grup** que genera la variable situació laboral. L'opció més raonable és fer-ho "dividint" la variable p49 amb les categories de la variable minuts_XS a partir del procediment "Descriptives", tal i com hem vist en l'apartat anterior:

1. Clic a menú Analyses / Exploration / Descriptives...
2. La variable quantitativa dependent (minuts_XS) la situarem sempre a l'espai superior "Variables".
3. La variable qualitativa independent (p49) a l'espai inferior *Split by*.
4. Ens interessa només resultats sobre la mida dels grups (N). Si vols una taula més "neta" que la que apareix per defecte pots desplegar el menú d'opcions estadístiques [> Statistics] i des-marcar tots els resultats que no siguin N a l'apartat **Sample Size**.

Descriptives

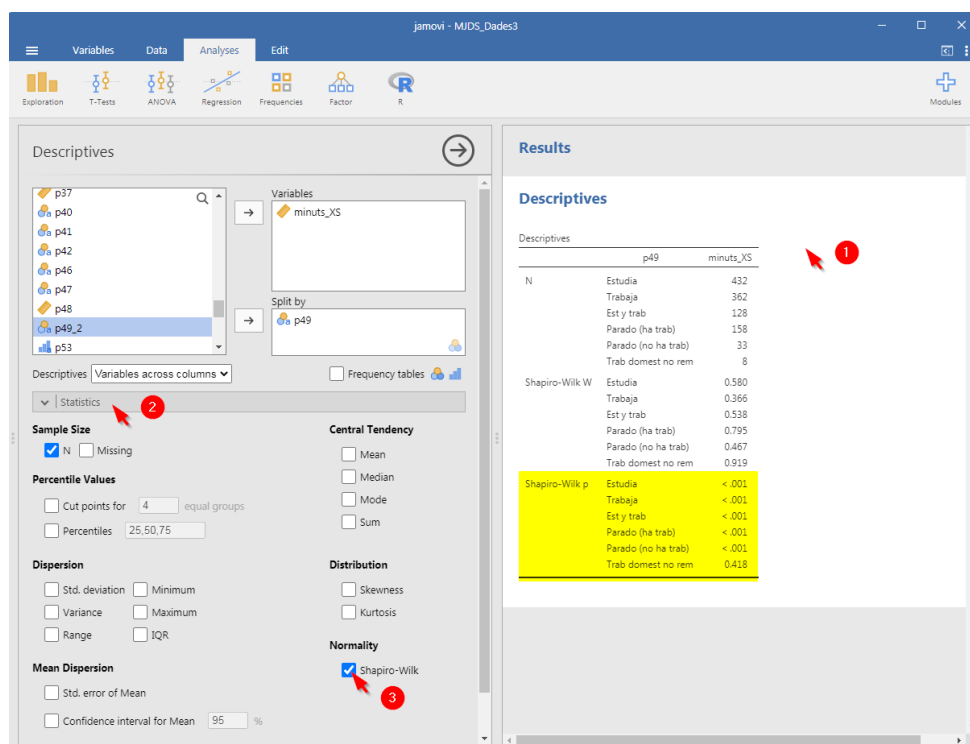
	p49	minuts_XS
N	Estudia	432
	Trabaja	362
	Est y trab	128
	Parado (ha trab)	158
	Parado (no ha trab)	33
	Trab domest no rem	8

Observa que hi ha un grup molt poc nombrós, que són les persones que manifesten que la seva situació laboral és la de treball domèstic no remunerat ("*Trab domest no rem*"), dels quals només disposem de 8 observacions vàlides per a la variable minuts_XS.³⁰

Si es vol fer un test que avalui si les diferències observades a la mostra són significatives a nivell de tota la població, primer hem de saber si podem assumir que **aquest grup** segueix una distribució normal. Per fer-ho:

1. Clic a la taula resultat que s'acaba de generar.
2. Desplega el menú d'opcions fent clic a > Statistics
3. Troba, a la part inferior dreta, l'apartat **Normality**, i activa l'opció [] Shapiro-Wilk.

³⁰ En realitat hi ha 26 persones en aquesta situació, però 18 tenen valors perduts per a la variable "minuts_XS".



Imatge 110. Test de Normalitat Shapiro-Wilk

La taula de resultats queda ampliada amb el p-valor de la prova de normalitat Shapiro-Wilk per a cadascun dels grups o categories de la variable qualitativa. En el cas del grup que “donava problemes” (Trab domest no rem, amb menys de 30 casos vàlids) el p-valor és 0,418; **superior** a 0,05, de manera que **podem assumir** el supòsit de normalitat, i en conseqüència fer el test paramètric corresponent.

En resum! En els test habituals (amb un nivell de confiança del 95% de confiança), si en la prova de Shapiro-Wilk observem un $p\text{-valor} \leq 0,05$ per a cada grup amb menys de 30 casos vàlids, concloem que no hi ha normalitat, mentre que si el $p\text{-valor} > 0,05$ assumim la normalitat.

ATENCIÓ! Pot ser has comprovat que cap dels altres grups no ha superat la prova de normalitat de Shapiro-Wilk. És un problema? No, perquè tots aquests grups tenen 30 o més efectius vàlids, i per tant s'assumeix que són “prou grans” com per no haver d'aplicar una prova de normalitat per assegurar la realització de test de tipus paramètrics.

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? 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 com per no ser degudes a l'atzar. Si es passa 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.

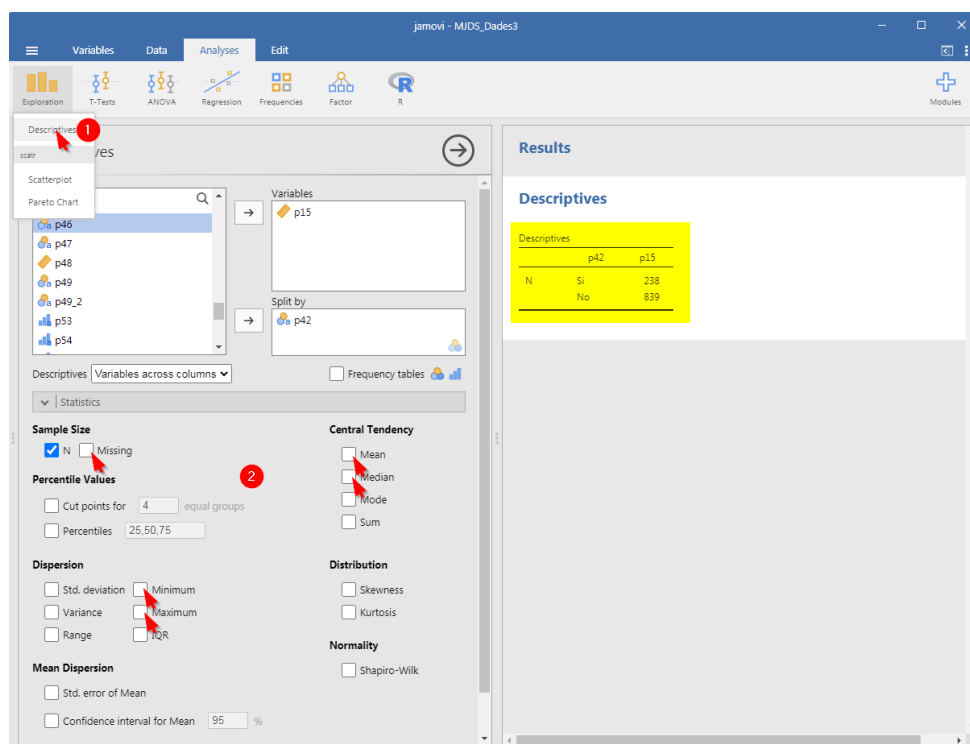
Per a comparar la mitjana d'una variable resposta (el nombre d'amics a xarxes, p15) entre dos grups independents (conviuen o no en parella, p42) es pot fer servir la prova anomenada "T d'Student per a mostres independents". En aquest cas el que volem contrastar s'expressa en un test com:

- $H_0: \mu_{si} = \mu_{no}$ (la mitjana de nombre d'amics a la xarxa dels que conviuen en parella és la mateixa que la mitjana dels que no hi conviuen)
- $H_1: \mu_{si} \neq \mu_{no}$ (la mitjana de nombre d'amics a la xarxa dels que conviuen en parella és diferent que la mitjana dels que no hi conviuen)

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. Si hi ha evidència estadística per a la segona hipòtesis (H_1) és que la situació de convivència **si** 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 a fer un test paramètric o cal comprovar la normalitat d'alguns dels dos grups. Per avaluar-ho hi ha diversos procediments a *jamovi* que generen els resultats adients (un descriptiu de la variable dependent en funció de la variable independent), però es convenient utilitzar Analyses / Exploration / Descriptives tal i com s'ha vist anteriorment.

1. Clic a menú Analyses / Exploration / Descriptives...
2. La variable quantitativa dependent (p15) la situarem sempre a l'espai superior "Variables".
3. La variable qualitativa independent (p42) a l'espai inferior *Split by*.
4. Per defecte es mostren més resultats que la mida dels grups (N). Si vols una taula més "neta" pots desplegar el menú d'opcions estadístiques [> Statistics] i des-marcar tots els resultats que no siguin N a l'apartat **Sample Size**.



Imatge 111. Estimació de la mida dels grups o mostres per comparar 2 mitjanes.

On 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.

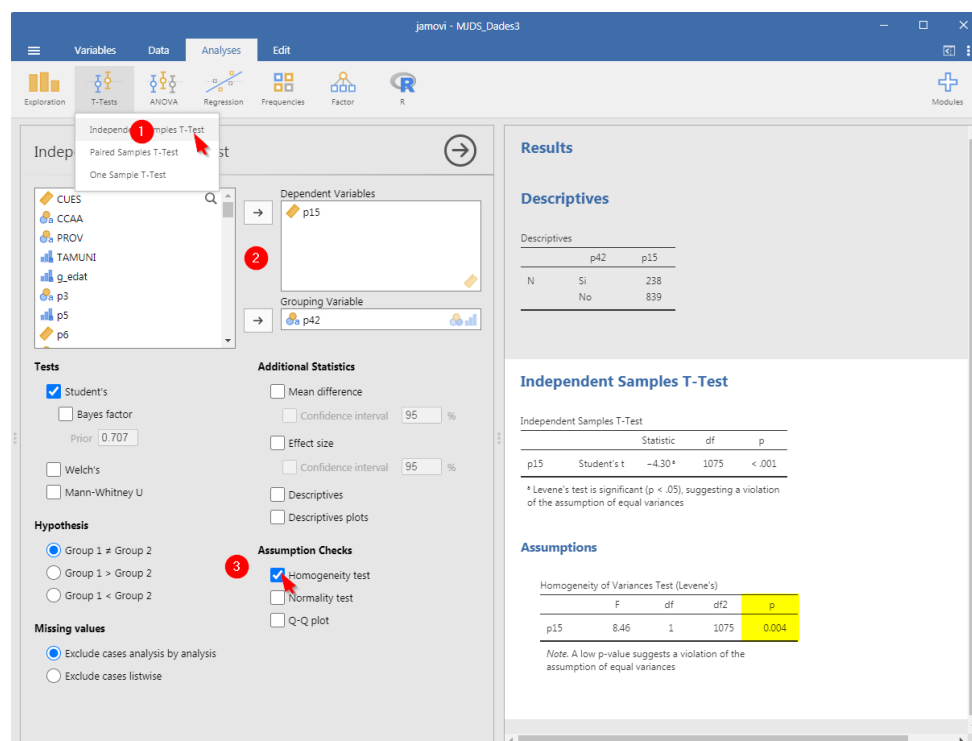
Tal i com s'esmenta en la introducció, hi ha dues maneres diferents de fer aquesta comparació paramètrica: (1) quan les variàncies són iguals i (2) quan no ho són; de manera que en primer lloc haurem de comprovar si es dona la condició d'aquesta igualtat entre variàncies (homoscedasticitat). Per fer-ho duem 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:

- $H_0: \sigma^2_{si} = \sigma^2_{no}$ (les dues variàncies són iguals)
- $H_0: \sigma^2_{si} \neq \sigma^2_{no}$ (les dues variàncies són diferents)

Per fer el test de comparació de dues variàncies amb *jamovi*:

1. Clic a menú Analyses / T-Tests / Independent Samples T-Test...
2. La variable quantitativa dependent (p15) la situarem sempre a l'espai superior "Dependent Variables".
3. La variable qualitativa independent (p42) a l'espai inferior "Grouping Variable".
4. Per defecte es mostren resultats del test t de Student. A l'apartat inferior dreta **Assumption Checks** cal activar l'opció "[] Homogeneity test".



Imatge 112. Test de Levene, per a comprovar igualtat de variàncies.

Al resultat cal interpretar el p-valor. En aquest cas prenem el valor de 0,004, de manera que hi ha prou evidència estadística per a concloure que **les variàncies són 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 homocedasticitat) 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 assumida (1) la normalitat, i (2) la diferència entre les variàncies (*homocedasticitat*) ja podem fer el test de comparació de dues mitjanes independents corresponent: el Test de Welch.³¹

1. Clic a menú Analyses / T-Tests / Independent Samples T-Test...
2. Situa la variable dependent (p15) a l'espai superior "Dependent Variables".
3. La variable independent (p42) a l'espai inferior "Grouping Variable".
4. Per defecte es mostren resultats del test t de Student. A l'apartat superior esquerra **Tests** cal activar l'opció "[] Welch's". Per a obtenir una taula més "neta" es pot desactivar l'opció "Student's", o deixar-la activada donat que genera informació rellevant sobre el test de Levene.
5. Es convenient activar també l'opció "Descriptives" a l'apartat **Additional Statistics**.

³¹ Per saber quin test cal fer en cada cas, consulta l'arbre de decisió Quin test cal fer?.

Independent Samples T-Test						
		Statistic	df	p		
p15	Student's t	-4.30*	1075	< .001		
	Welch's t	-4.94	481	< .001		
* Levene's test is significant (p < .05), suggesting a violation of the assumption of equal variances						
Group Descriptives						
	Group	N	Mean	Median	SD	SE
p15	Si	238	167	100	268	17.4
	No	839	271	200	345	11.9

Imatge 113. Test de Welch per a comprovar igualtat entre dues mitjanes, amb la normalitat i diferència entre les variàncies dels grups.

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 diferència 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 conclusió haurà de ser que hi ha suficient evidència estadística ($p\text{-valor} < 0,001$) per afirmar que el nombre d'amics **és diferent** en funció de si es viu en parella o no, i 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% de confiança, si observem un $p\text{-valor} \leq 0,05$ ens quedem amb la hipòtesi alternativa, mentre si el $p\text{-valor} > 0,05$ 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 **Tests** del panell de definició del procediment "Test T per a Mostres Independents" com 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 en cas que algun grup fos petit i no normal. S'escau utilitzar-ho en cas que algun dels grups amb menys de 30 persones no hagi superat el test de Normalitat (Shapiro-Wilk).
- **t de Student:** Test paramètric en cas de normalitat i variàncies iguals (per defecte a *jamovi*). S'escau utilitzar-ho en cas que es superi el test d'homoscedasticitat (Levene).
- **t de Welch:** Test paramètric en cas de normalitat i variàncies diferents. S'escau utilitzar-ho en cas que no es superi el test d'homoscedasticitat (Levene).

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 té més de dos categories o valors, 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 variància³² (que utilitza les sumes de quadrats, que en definitiva són els variàncies, per a comparar mitjanes). En aquest cas el que volem contrastar se sol expressar en un test d'hipòtesis com:

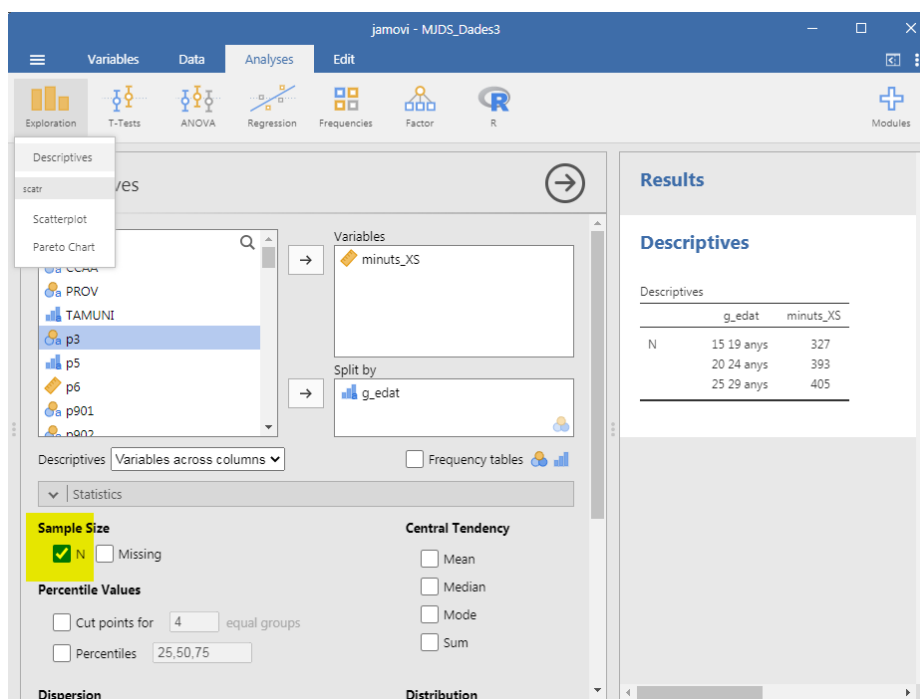
- $H_0: \mu_{15-19} = \mu_{20-24} = \mu_{25-29}$ (la mitjana de temps dedicat a les xarxes socials és la mateixa en els tres grups d'edat)
- $H_1: \mu_i \neq \mu_j$, per algun i, j (la mitjana de temps dedicat a les xarxes socials no és la mateixa en els tres grups d'edat, alguna mitjana difereix de les altres)

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, també aquí tenim dues famílies de tests, el paramètric i el no paramètric. Anàlogament, farem el test paramètric si tots els grups són grans o petits però normals, mentre que en cas contrari (algun grup 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 INTEGER, i amb les categories ben ordenades.
2. Demana a *jamovi* estadística Descriptiva de la variable "minuts_XS" agrupada per "g_edat".

³² La denominació habitual ANOVA es, òbviament, un acrònim anglès de *analysis of variance*.

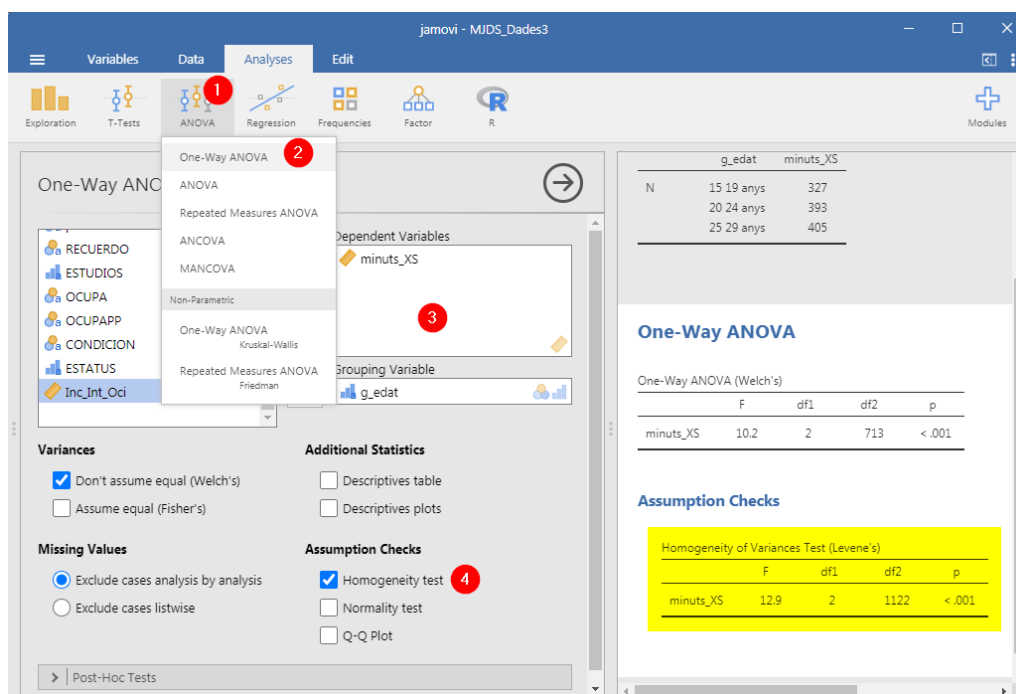


Imatge 114. Estimació de la mida dels grups o mostres per comparar DIVERSES mitjanes.

Pots veure que els tres grups tornen a ser grans ($n_i > 30$), de manera que podem fer directament un test paramètric. En aquest cas també hi ha dos tests paramètrics, 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 per decidir quin és el test t adient per aquest cas.

Per a fer test de Levene en el cas d'ANOVA:

1. Clic al menú Analyses / ANOVA / One Way ANOVA... S'obre el panell de definició del procediment "Anàlisi de Variàncies simple".
2. Situa la variable dependent (minuts_XS) a l'espai superior "Dependent Variable", i la variable independent (g_edat) a l'espai inferior "Fixed Factors".
3. Per defecte es mostren resultats del test t per a variàncies diferents (Welch). Però cal comprovar si aquesta *heterocedasticitat* es certa o no.
4. Per a realitzar el test de Levene, a l'apartat inferior dreta **Assumption Checks** activa l'opció "[] Homogeneity test".



Imatge 115. Anàlisi de Variàncies simple (one-way).

El resultat ens el torna a donar en forma de p-valor. En aquest cas pren un valor menor a 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 es pot comprovar que hi ha DOS tests possibles:

- el test paramètric en cas de variàncies diferents: “Don’t assume equal (Welch’s)”, i
- el test paramètric en cas de variàncies iguals: “Assume equal (Fisher’s)”.

Per defecte, *jamovi* marca el primer, que és precisament el adient en aquest cas.

One-Way ANOVA (Welch's)

	F	df1	df2	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 (p-valor < 0,05) que hi ha relació** entre l’edat i el temps que es dedica a les xarxes socials.

Dir que hi ha relació entre edat i temps és equivalent a dir que el temps dedicat a les xarxes socials és diferent en funció de l’edat. Ara bé, aquesta anàlisi no ens diu entre quins grups es donen les diferències. Per veure-ho hem de comparar cada possible parella de grups, el que s’anomena comparacions múltiples a posteriori, o comparacions 2 a 2 (post-hoc Tests). Aquestes comparacions són equivalents a plantejar:

- $H_0: \mu_{15-19} = \mu_{20-24}$ (el grup d’edat 15-19 i el 20-24 es comporten igual)
- $H_1: \mu_{15-19} \neq \mu_{20-24}$ (el grup d’edat 15-19 i el 20-24 es comporten diferent)
- $H_0: \mu_{20-24} = \mu_{25-29}$ (el grup d’edat 20-24 i el 25-29 es comporten igual)
- $H_1: \mu_{20-24} \neq \mu_{25-29}$ (el grup d’edat 20-24 i el 25-29 es comporten diferent)

- $H_0: \mu_{15-19} = \mu_{25-29}$ (el grup d'edat 15-19 i el 25-29 es comporten igual)
- $H_1: \mu_{15-19} \neq \mu_{25-29}$ (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. Per avaluar-les cal:

1. Al mateix panell de definició del procediment “Anàlisi de Variàncies simple”, fes clic a > Post-Hoc Test per activar les comparacions múltiples a posteriori, o comparacions 2 a 2,
2. Per defecte no es calcula cap comparació. Cal activar el tipus adient, segons les conclusions del test de Levene previ: “Games-Howell” en aquest cas perquè hi ha variàncies diferents.

The screenshot shows the jamovi software interface for a One-Way ANOVA analysis. The 'Analyses' menu is open, and 'One-Way ANOVA' is selected. The 'Dependent Variables' list contains 'minuts_XS' and the 'Grouping Variable' is 'g_edat'. In the 'Post-Hoc Tests' section, 'Games-Howell (unequal variances)' is selected. The right panel displays the results of the One-Way ANOVA (Welch's), including the F-statistic (10.2), df1 (2), df2 (713), and p-value (<.001). It also shows the Homogeneity of Variances Test (Levene's) and the Games-Howell Post-Hoc Test results for 'minuts_XS'.

	F	df1	df2	p
minuts_XS	10.2	2	713	<.001

	F	df1	df2	p
minuts_XS	12.9	2	1122	<.001

		15 19 anys	20 24 anys	25 29 anys
15 19 anys	Mean difference	—	-11.1	22.0
	p-value	—	0.477	<.001
20 24 anys	Mean difference		—	33.1
	p-value		—	0.002
25 29 anys	Mean difference			—
	p-value			—

Imatge 116. Comparacions entre grups en un ANOVA simple.

Games-Howell Post-Hoc Test – minuts_XS

		15 19 anys	20 24 anys	25 29 anys
15 19 anys	Mean difference	—	-11.1	22.0
	p-value	—	0.477	<.001
20 24 anys	Mean difference		—	33.1
	p-value		—	0.002
25 29 anys	Mean difference			—
	p-value			—

Els resultats els mostra en format de 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. 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 grups ($p\text{-valor} > 0,05$).

Contràriament, 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.

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, d'edat més elevada.

Si n'avaluem les mitjanes per a cada grup (que podem obtenir en la taula generada per *Analysis / Exploration / 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.

Observació 1! El procediment de demanar les comparacions grup a grup **només té sentit** en el cas que puguem assegurar que hi ha diferències entre els grups.

Observació 2! Aquest procediment el farem exactament igual tant si es tracta d'un test paramètric com no paramètric.

Recordatori! 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 evidència estadística per a afirmar-ho. És en aquest sentit que sovint diem que hem observat diferències no significatives.

La intensitat de la relació: El coeficient de determinació.

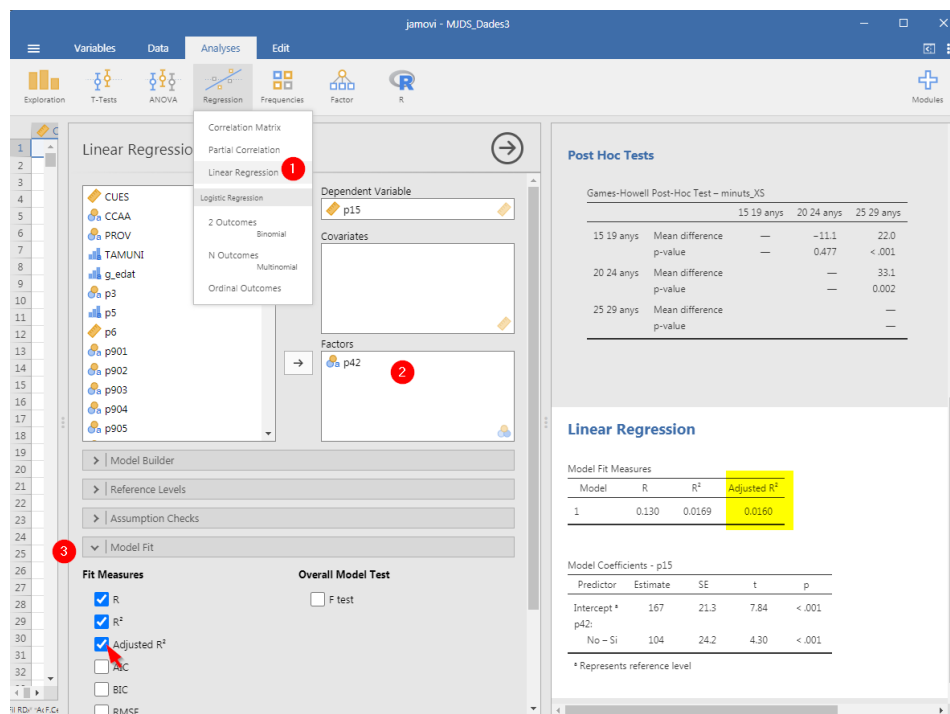
Mentre que en les taules de contingència s'avalua la intensitat de la relació a partir de la *V de Cramer*, quan comparem mitjanes ho fem mitjançant l'estadístic R^2 , anomenat "coeficient de determinació". La diferència més important respecte de la *V de Cramer* (que varia entre 0 i 1) és que R^2 es pot expressar en percentatge, i ens diu **quina part de la variable dependent és explicada** per la variable independent.

Una vegada comprovat (mitjançant test de comparació de mitjanes) que dues variables estan relacionades, per conèixer la intensitat cal demanar R^2 amb *jamovi*, utilitzant una opció d'un procediment diferent al ANOVA, anomenat "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. Clic al menú *Analyses / Regression / Linear Regression...* S'obre el panell de definició del procediment "Regressió lineal".

2. Situa la variable dependent (p15) a l'espai superior "Dependent Variable", 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 fes clic a el desplegable ">" Model fit" i activa l'opció [] Adjusted R^2 .



Imatge 117. Intensitat de la relació (R^2) per a comparació de mitjanes.

Model Fit Measures			
Model	R	R^2	Adjusted R^2
1	0.130	0.0169	0.0160

El resultat ara afegit a la taula indica un valor del coeficient de determinació ajustat del 0,0160; que és el mateix que dir una variància explicada del 1,6%. Aquesta intensitat és molt baixa: la convivència explica només un 1,6% de la variabilitat del total d'amics que es tenen a les xarxes socials.



EXERCICI

- Pots analitzar si altres variables tenen una relació explicativa major del total de amics a xarxes socials? Prova amb diverses variables i determina quina o quines expliquen més.

Exercicis proposats: Comparació de mitjanes

Després de seguir completament el tutorial en l'apartat "Comparació de mitjanes", obre en *jamovi* el fitxer resultant *Dades3.omv* i respon les següents qüestions.

(1) Construeix, amb un 95% de confiança, un interval per al temps que els individus passen a les xarxes socials (minuts_XS). Enganxa els resultats en un processador de textos i interpreta'ls. Fes-ho també amb un 99% de confiança i interpreta el resultat i analitza la diferència en l'amplitud dels dos intervals.

(2) Resumeix gràficament la relació existent entre el nivell d'estudis (ESTUDIOS) i el temps diari dedicat a les xarxes socials (minuts_XS). Fes-ho de manera que en el gràfic es visualitzi clarament si hi ha diferències entre els diferents nivells d'estudi en el temps dedicat a les xarxes socials. A partir de la sintaxi, canvia el nom dels eixos de manera que aparegui el significat real de cada variable. Procura eliminar els valors perduts del gràfic i enganxa'l en un processador de textos.

(3) 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.

(4) A partir de la pregunta 10 s'ha construït un indicador comprès entre 0 i 8 sobre la incidència d'Internet en l'oci dels joves (Inc_Int_Oci). L'indicador s'ha construït comptant en quants ítems relacionats amb l'oci consideraven els joves que Internet els havia comportat un descens de les activitats (no fer res, passejar, veure la televisió, escoltar la ràdio, anar al cinema, llegir, practicar algun esport, anar a biblioteques), de manera que els valors més alts signifiquen que Internet els ha canviat en gran mesura els hàbits d'oci, mentre que els valors baixos que no. Planteja les hipòtesis corresponents per tal de valorar si hi ha prou evidència estadística per afirmar que els efectes d'Internet en l'oci són diferents en funció del nivell d'estudis (ESTUDIOS). En cas afirmatiu avalua en quins grups podem considerar que es donen diferències significatives, així com la intensitat de la relació. Enganxa en un processador de textos tots els resultats que consideris rellevants.

(5) Quina de les següents variables explica en major mesura el nombre d'amics a les xarxes socials (p15): la convivència (p42), el sexe (p40), el grup d'edat (g_edat) o el nivell d'estudis (ESTUDIOS). Enganxa en un processador de textos tots els resultats que consideris rellevants.

(6) 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.

(7) 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).³³ Per tal de comprovar si l'has fet bé, aquí tens les freqüències d'aquest indicador, que hauries de llegir de la següent manera: 118 entrevistats no fan servir Internet per a res, 39 persones per a una sola cosa, 93 per a dues... mentre que 24 persones han fet servir Internet per a 12 coses diferents:

³³ Per construir-lo basat en la primera pràctica.

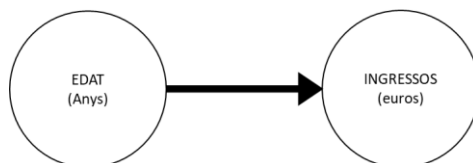
# de usos	Recompte
0	118
1	39
2	93
3	164
4	176
5	191

# de usos	Recompte
6	179
7	129
8	134
9	79
10	60
11	33
12	24

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.

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 “Regression” corresponents de càlcul de correlació i 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'extrapol·lar-la al conjunt de la població (aspecte inferencial de l'anàlisi).



Imatge 118. Exemple de model d'anàlisi de regressió lineal simple.

L'objectiu principal de la practica és comparar les distribucions de **DUES variables quantitatives**. Ens trobem davant la possible relació entre variables de mètrica quantitativa (en *jamovi*, de tipus CONTINUOUS i INTEGER o DECIMAL). S'utilitzarà el procediment de estimació de recta de regressió lineal simple per introduir la lògica de la modelització.

En aquest apartat es treballarà amb el fitxer de dades originals Dades3.omv que és el mateix que en la pràctica anterior, i que es basa en el mateix qüestionari, que necessitareu per anar seguint la pràctica. Descarrega al teu PC els arxius necessaris: “MJDS_QüestionariCIS.pdf” i “Dades3.omv”.

Recorda ! Sempre cal revisar la definició de cadascuna de les variables quan s'obre (o s'importa) amb *jamovi* un arxiu de dades que no està en el seu propi format (.omv).

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

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 que té el seu fonament en la recta que millor ajusta amb les dades, anomenada “*recta de regressió*” dibuixada 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).

Compte! Tot i que aquí s'explicarà primer l'anàlisi descriptiva i després la inferència, la lògica analítica és la contrària. És a dir:

(1er) Assegurem si hi ha o no relació significativa **per a tota la població** (inferència, p-valors).

(2on) **En cas que hi hagi** relació, interpretem tots els paràmetres que siguin interpretables. En cas que no s'asseguri la relació, no cal interpretar descriptivament els paràmetres.

Per descriure la relació entre dues variables quantitatives se soleu utilitzar tres estratègies:

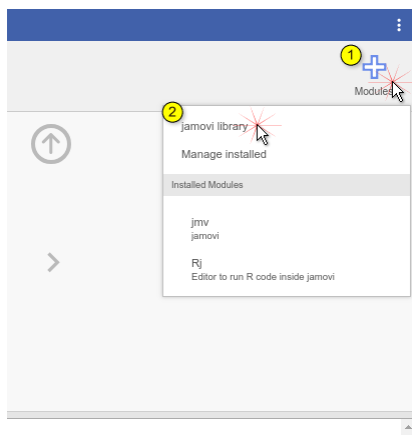
1. Dibuixar gràficament el “núvol de punts” que representa la relació, utilitzant l’anomenat diagrama de dispersió o diagrama XY.
2. Avaluar el nivell o intensitat de la relació entre les dues variables a partir de l’anomenat coeficient de correlació de Pearson (r).
3. Estimar els coeficients que determinen la “recta de regressió”, així com l’indicador d’ajustament del núvol de punts a la recta: el coeficient de determinació (r^2).

El diagrama de dispersió

Per dibuixar els diagrames XY, sempre representarem la variable independent a l’eix de les X (horitzontal) i la variable dependent al de les Y (vertical) que en el nostre cas 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, en el nostre cas, “lineal”.

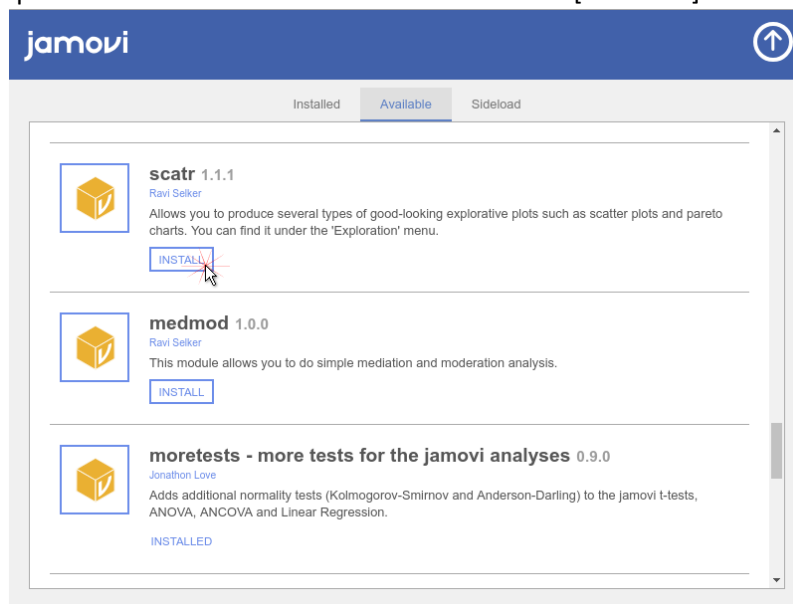
Per a generar un diagrama de dispersió, o diagrama XY (“*scatter plot*”) a *jamovi*, si és que no s’ha fet prèviament, cal instal·lar un mòdul addicional anomenat “*scatr*”.

1. Assegurat que l’ordinador disposa de connexió a Internet.
2. Per a carregar el mòdul “*scatr*”, a l’extrem dret del menú fes clic a la icona “ + Modules ”. Es desplega el llistat d’opcions, selecciona “*jamovi library*”.



Imatge 119. Instal·lació de mòduls addicionals.

3. S'obre la finestra d'instal·lació de mòduls o extensions disponibles. Baixa pel llistat fins que trobis el mòdul *scatr*. Fes clic al botó blau [**INSTALL**]



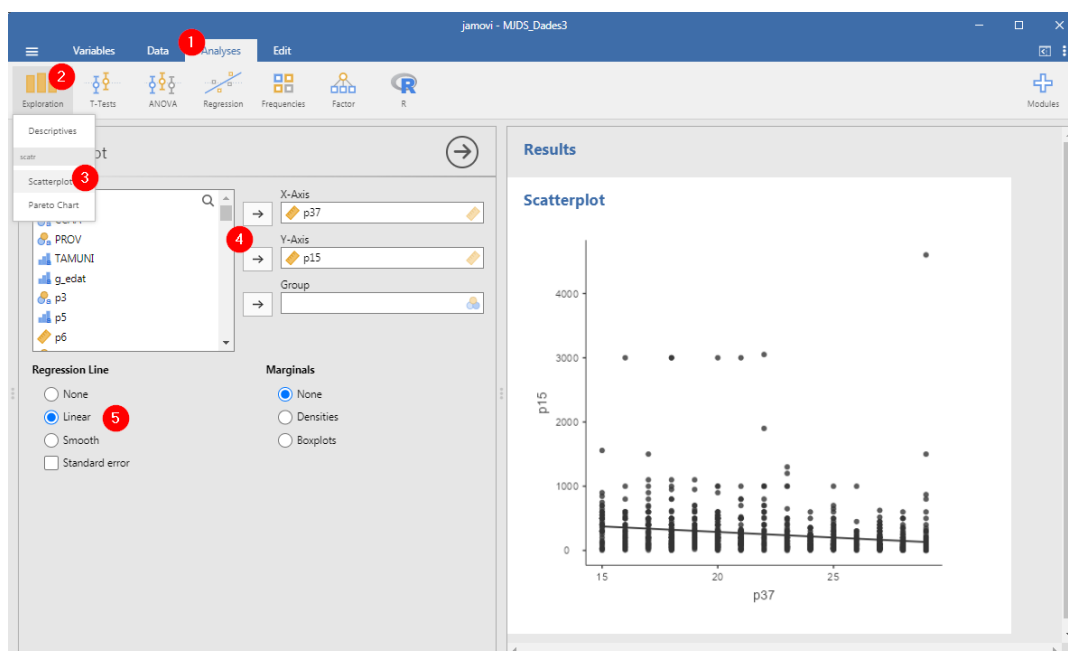
Imatge 120. Instal·lació de mòdul per a generar diagrames XY

Els mòduls o extensions de *jamovi* s'instal·len descarregant-los des d'Internet, cal tenir connexió per poder instal·lar-los. Una vegada instal·lat un mòdul apareixeran noves opcions als menús de *jamovi*.

Compte! Si l'arxiu *.omv s'obre a un altre ordinador es requereix que aquest ordinador 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 "Dades3.omv" cal comprovar que les dues variables estiguin ben definides. Fes doble-clic al nom de p15, que encapçala la columna de dades: cal que sigui de tipus CONTINUOUS i les dades INTEGER. Fes el mateix per a la p37.
2. Fes clic a la FLETXA-AMUNT ENCERCLADA de la part superior dreta per a tancar la finestra de definició de variables.
3. Ara fes clic al menú Analyses / Exploration / Scatterplot... S'obre la finestra de definició del diagrama de dispersió.
4. Situa la variable p37 a l'espai superior dret "X-Axis" i la variable p15 a l'espai central dret "Y-Axis".
5. Per defecte, només es dibuixa el núvol de punts. Per a afegir una línia d'ajustament, selecciona l'opció " () Linear " a l'apartat inferior esquerre **Regression Line**. Es genera el resultat gràfic següent:



Imatge 121. Diagrama de dispersió o XY amb jamovi.

Diagrama de dispersió amb codi R



Si cal modificar el resultat gràfic, per exemple incloent-hi títols, etc..., o controlar la escala dels eixos, és preferible utilitzar codi R per executar a la consola de *jamovi*. A més, per generar el diagrama de dispersió directament amb codi R no cal tenir instal·lat el mòdul d'extensió “scatr” (però sí el mòdul “Rj Editor” que dona accés a la consola per escriure codi R).

1. Fes clic a menú Analyses / 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:

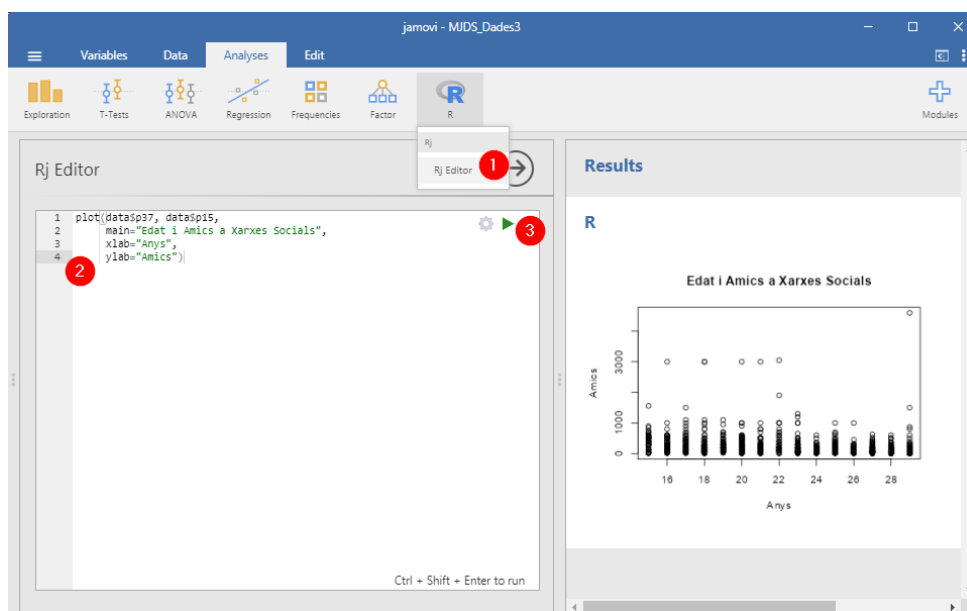
```
plot(data$p37, data$p15,
      main="Edat i Amics a Xarxes Socials",
      xlab="Anys",
      ylab="Amics")
```

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:

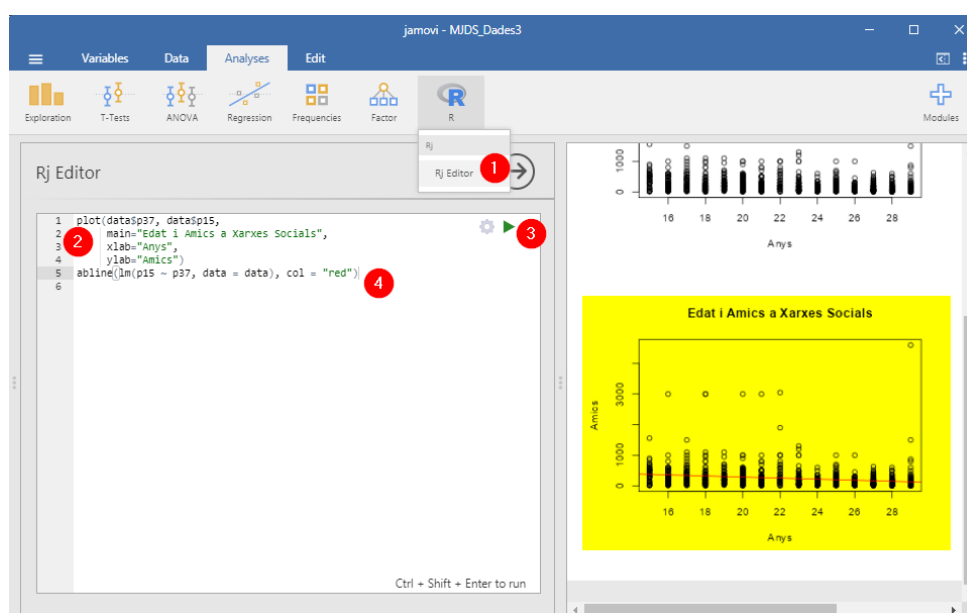
```
plot(data$p37, data$p15, main="Edat i Amics a Xarxes Socials", xlab="Anys", ylab="Amics")
```

3. Executa el codi R: clica al tringle verd (o combinació de teclat [Ctrl]+[Mayús]+[Intro])



Imatge 122. Diagrama de dispersió o XY amb codi R

4. Per afegir una línia de regressió cal afegir una segona instrucció a la consola i executar:³⁴
`abline(lm(p15 ~ p37, data = data), col = "red")`



Imatge 123. Diagrama XY i recta amb codi R

5. Fes clic a la FLETXA-DRETA ENCERCLADA per a tancar la finestra de consola.

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 que aquests van descendent a mesura que augmenta l'edat. A més, observem una forta dispersió de les dades, de manera que la correlació no serà massa intensa.

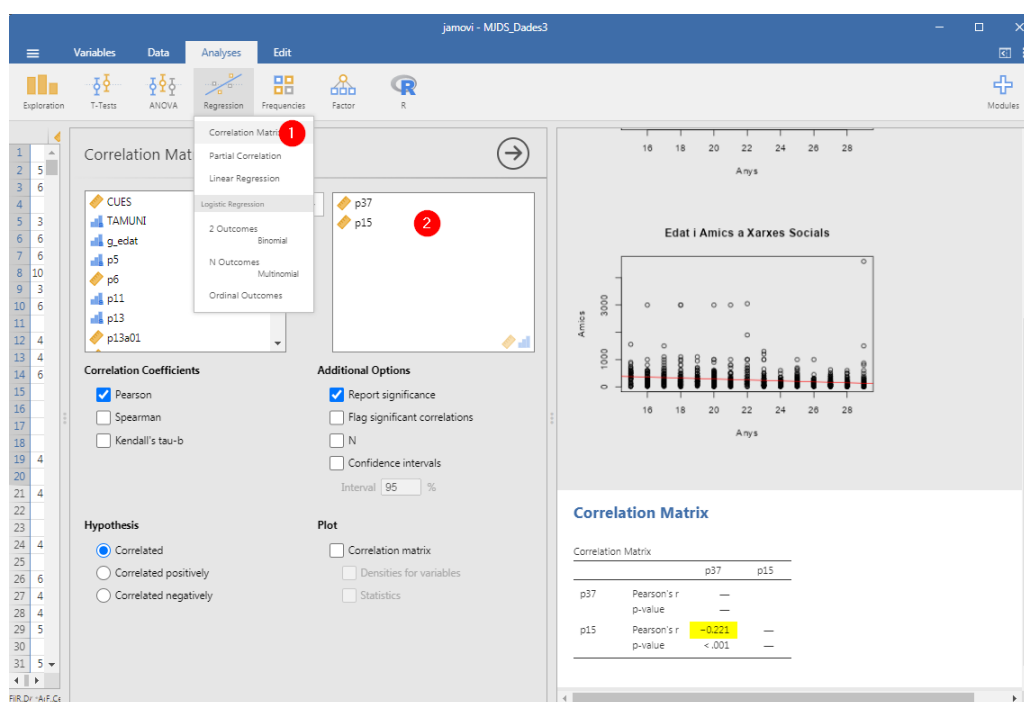
³⁴ El símbol ~ es pot aconseguir amb la combinació de teclat [Alt Gr] + [4] i a continuació [Espai].

La correlació entre ambdues 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. Fes clic a el menú Analyses / Regression / Correlation Matrix... S'obre el panell de definició del procediment matriu de correlacions,
2. Situa les dues variables, p37 i p15, a l'espai superior dret "Variables".



Imatge 124. Coeficient de correlació (r) amb jamovi.

El resultat 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.

Observació! No confonguis els conceptes i tingues en compte que l'existència de correlació entre dues variables no implica causalitat.
Hi ha perfecta correlació estadística entre la longitud del braç dret i l'esquerre.

El model de regressió lineal simple

La correlació entre dues variables quantitatives —el coeficient r de Pearson— es pot avaluar sempre, no comporta que hi hagi una variable dependent i una d'independent, simplement ens mostra si

les dues variables canvien (correlacionen) 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 ($Y = \text{efecte o resposta}$) i una d'independent ($X = \text{causa o tractament}$).

En aquests casos, la intensitat de la relació ens la dona el “coeficient de determinació”, un estadístic que ens diu quina part de la variabilitat total és explicada per la variable independent (r^2). Es tracta d'un coeficient que varia entre 0 i 1, i que en percentatge ens indica precisament això, la proporció de la variabilitat total de Y explicada per X .

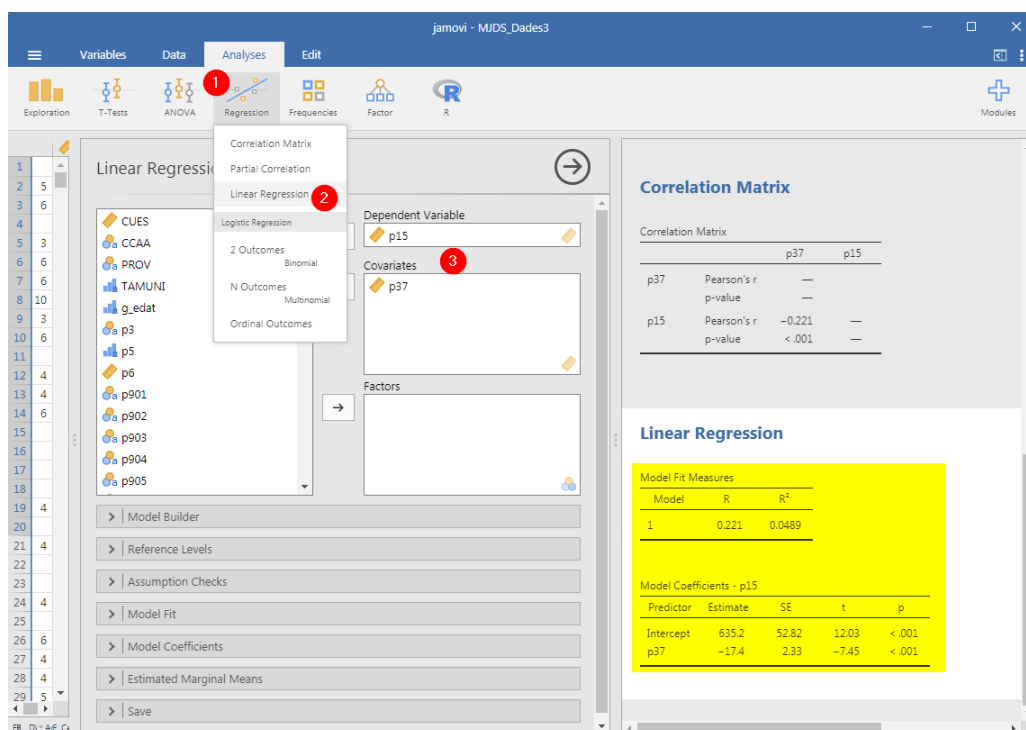
A més, sol ser habitual que ens interressi quina és la funció de la recta de regressió (la recta que *jamovi* ha dibuixat en el gràfic de dispersió). Aquesta funció ens permetrà predir, per a cada valor de la variable independent, el valor que pren, en mitjana, la variable dependent.

En general, el model funcional que correspon a la recta de regressió lineal simple es:

$$Y = a + b.X$$

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 “Linear Regression”.

1. Fes clic a el menú Analyses / Regression / Linear Regression... S'obre el panell de definició.
2. Selecciona la variable dependent amics p15 (Y , efecte) i situa-la a l'espai superior dret “Dependent Variable”.
3. Selecciona la variable independent edat p37 (X , causa) i situa-la a l'espai central dret “Covariates” (que admet únicament variables numèriques).
4. Per defecte es generen dues taules de resultats.



Imatge 125. Model de regressió lineal simple, 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 (R) entre ambdues variables, i el coeficient de determinació, que notem com a R^2 . El valor 0,0489 de R^2 , el podem llegir com que la variable edat explica fins a un 4,9% de la variabilitat donada en el nombre d'amics a les xarxes socials.

Per altra banda, els paràmetres per construir la recta de regressió es poden obtenir a la columna "Estimate" de la taula inferior: la "b" o "pendent" (p37) i la "a" o "constant" (Intercept). Així es pot escriure la recta d'aquesta forma:

$$\text{Nombre d'amics} = 635,2 - 17,4 \times \text{Edat}$$

Aquesta funció permet estimar el nombre d'amics en les xarxes socials en funció de la 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 la "constant" no sigui interpretable. De fet, només ho serà si en la mostra tenim valors mesurats en $X=0$. En aquest cas voldria dir que tenim persones d'edat 0 anys que han dit quants amics tenen a les xarxes socials (*la qual cosa indica baixa qualitat de les dades disponibles...*).

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.

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 el mateix que el 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:

$$\text{Nombre d'amics} = 635,2 - (17,4 \times \text{Edat}) = 635,2 - (17,4 \times 23) = 635,2 - 400,2 = \mathbf{235 \text{ amics}}$$

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 sinó disposàvem de valors propers a edat=0, tampoc no podem 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 per a edats inferiors als 15 anys ni superiors als 29.

Com pots comprovar al panell corresponent de *jamovi*, el procediment “Linear Regression” permet moltes més opcions, per exemple, incorporar més d’una variable quantitativa explicativa (*Covariates*), 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

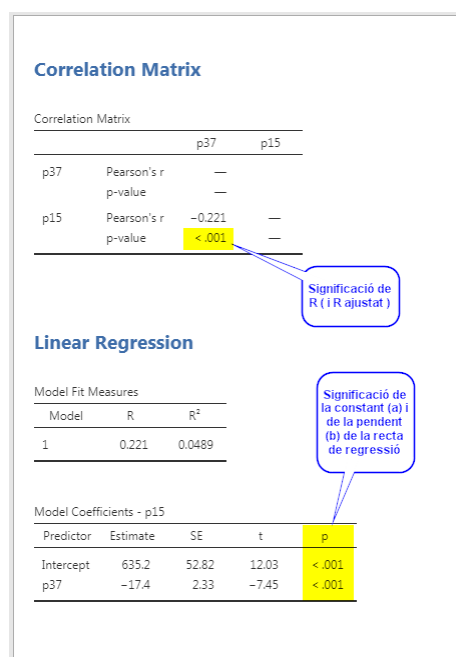
Pots comprovar com en els anteriors apartats hem estimat quatre paràmetres: coeficient de correlació (R), coeficient de determinació (R²), constant (a) i pendent (b).

La seva estimació (per a tota la població) comporta quatre tests d’hipòtesis, que serien:

- Test sobre el coeficient de correlació:
 - $H_0: R = 0 \rightarrow$ La correlació entre X i Y és 0. No hi ha relació entre X i Y
 - $H_1: R \neq 0 \rightarrow$ La correlació entre X i Y és diferent de 0. Hi ha relació entre X i Y
- Test sobre el coeficient de determinació:
 - $H_0: R^2 = 0 \rightarrow$ La X explica un 0% la Y. No hi ha relació entre X i Y
 - $H_1: R^2 \neq 0 \rightarrow$ La X explica més d’un 0% la Y. Hi ha relació entre X i Y
- Test sobre la constant:
 - $H_0: a = 0 \rightarrow$ Quan X=0, la Y també és 0
 - $H_1: a \neq 0 \rightarrow$ Quan la X=0, la Y no és 0
- Test sobre el pendent:
 - $H_0: b = 0 \rightarrow$ Els canvis en X no comporten canvi en Y. No hi ha relació entre X i Y
 - $H_1: b \neq 0 \rightarrow$ Els canvis en X comporten algun canvi en Y. Hi ha relació entre X i Y

Observa que la conseqüència final de tres d’aquests tests és la mateixa: dir “no hi ha correlació”, “una variable explica el 0% de l’altra”, “si una variable canvia l’altra no ho fa”, són, en el fons, el mateix. Aquests tres casos comporten, simplement, que no hi hagi relació entre les variables. Així, **els tests sobre R, R² i B són exactament el mateix**. I aquest és el resultat que ens importa a

nosaltres, en tant que el nostre objectiu és, precisament, avaluar la possible relació entre dues variables.



Imatge 126. Estimació poblacional de coeficients de correlació i model de regressió lineal.

Si observes el resultat dels tres tests veuràs que la significació és molt petita, gairebé 0; de manera que podem assegurar, amb un 99% de confiança (p -valor $< 0,01$) **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 "a" o constant (*Intercept*), 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 $X=0$ també $Y=0$ 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.

Els supòsits del model: L'anàlisi de residuals.

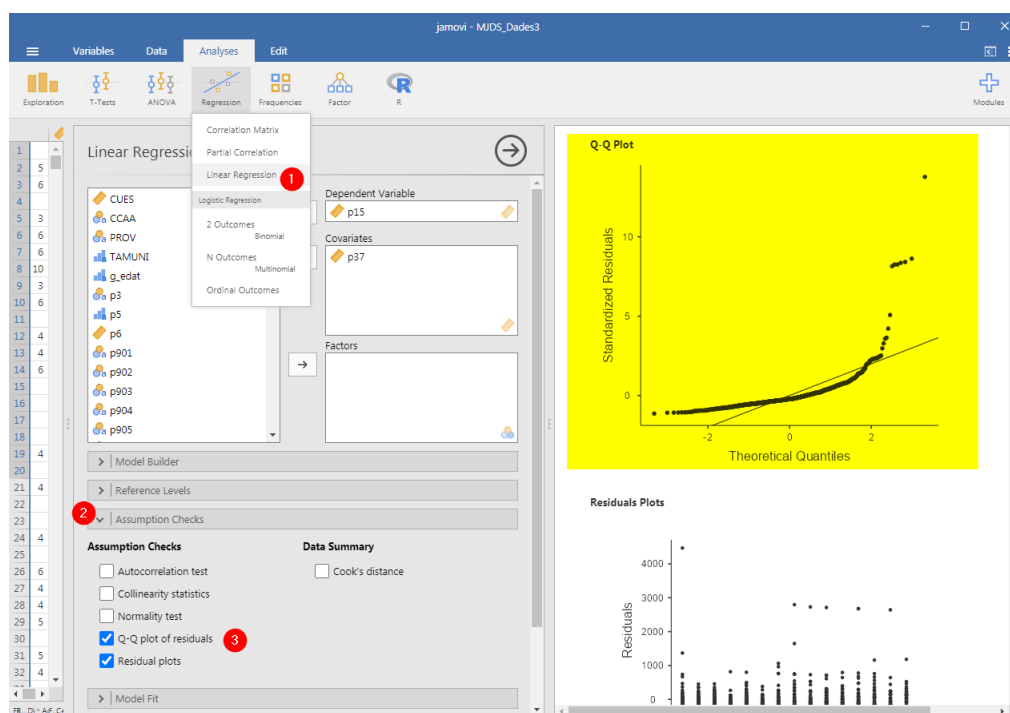
Recorda de quan es va tractar l'anàlisi de les taules de contingència que els "residuals" són les diferències entre el valor observat i el valor "esperat" o "teòric". En el cas de la Regressió Lineal són les diferències entre el valor observat i el valor predit per la recta de regressió. Pot ser considerat com "l'error aleatori" observat. L'estudi d'aquests es fa, bàsicament, a partir del diagrames de dispersió XY entre els valors residuals i els valors predits (en *jamovi*, "Residual plots"). En termes

generals per a validar el model de regressió mitjançant 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:

1. El model està correctament especificat.
2. Els residus són **independents** els uns dels altres (*Colinearity*, molt rellevant en anàlisi multivariable, però no tant en aquest cas de regressió simple amb només una variable independent).
3. En mostres petites, hi ha **normalitat** en els residus (*Q-Q plot of residuals*)
4. Els residus tenen una mitjana de zero i una variància constant (*Residual plots*).
5. No hi ha valors anòmals.

El sistema de menús de *jamovi* actualment permet controlar només la Normalitat i la Homoscedasticitat dels residuals. Per fer això:

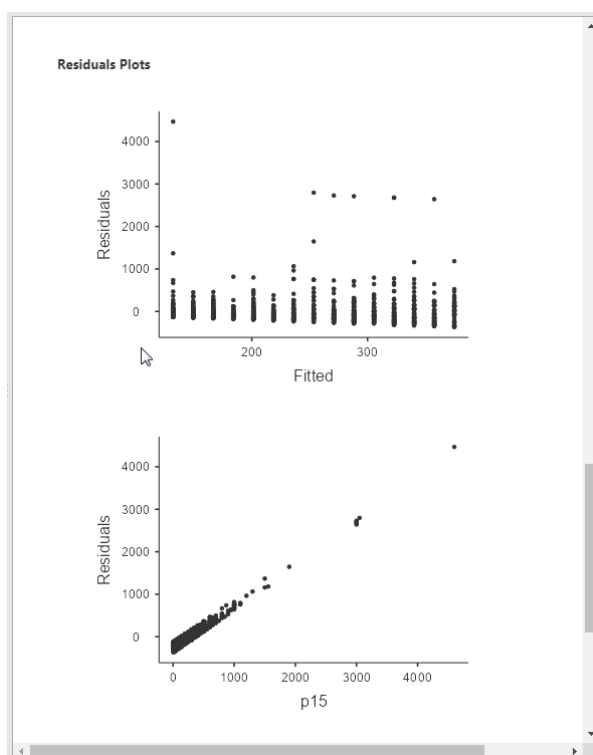
1. Fes clic a la taula de resultats de la Regressió Lineal corresponent. S'obrirà el panell de definició del procediment.
2. Desplega l'apartat “> Assumption Checks” i activa les opcions [] Q-Q plot of residuals, i [] Residual plots.



Imatge 127. Anàlisi de residuals al model RLS. Gràfic Q-Q.

Si la distribució fos normal els punts es distribuïrien exactament sobre la recta del gràfic “*Q-Q plot of residuals*”. En aquest cas, pots veure que els valors més petits, i –sobre tot- 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 és més gran.

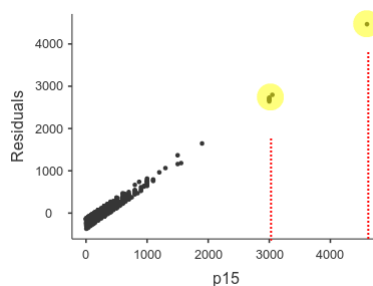
El gràfic sembla indicar que es pot resoldre el problema de la no-normalitat si s'eliminen de l'anàlisi els casos identificats com a “extrems” (*outliers*), situats més enllà de 2 desviacions típiques del punt central.



Imatge 128. Anàlisi de residuals al model RLS. Gràfic de Residuals.

El gràfic “*Residuals plots*” representa els residuals (diferència entre el valor real i el valor ajustat/predit per la recta) respecte dels valors ajustats. Les condicions del model ens diuen que (1) els residus han de tenir una mitjana de zero i (2) variància constant. 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, l’homocedasticitat.

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 129. Identificació de casos extremadament desviats.

En concret, respecte a la variable p15 es pot identificar un cas proper al valor 4500 i diversos casos propers al valor 3000 (amics o contactes a xarxes socials) que presenten la màxima desviació. Es tracta de valors possibles, però que ens condiciona 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 $p15 = 635,2 - 17,4 \cdot p37$. Per a calcular la regressió lineal sense tenir en compte els casos que es desviïn 3 desviacions estàndard de la mitjana caldrà crear noves columnes a la matriu de dades. L'objectiu és IDENTIFICAR els casos (files) que estan fora del rang de desviació recomanat (2 desviacions típiques).



Per generar amb *jamovi* una nova “versió” de la variable original p15 on aquests casos siguin marcats com a valors perduts (NA):

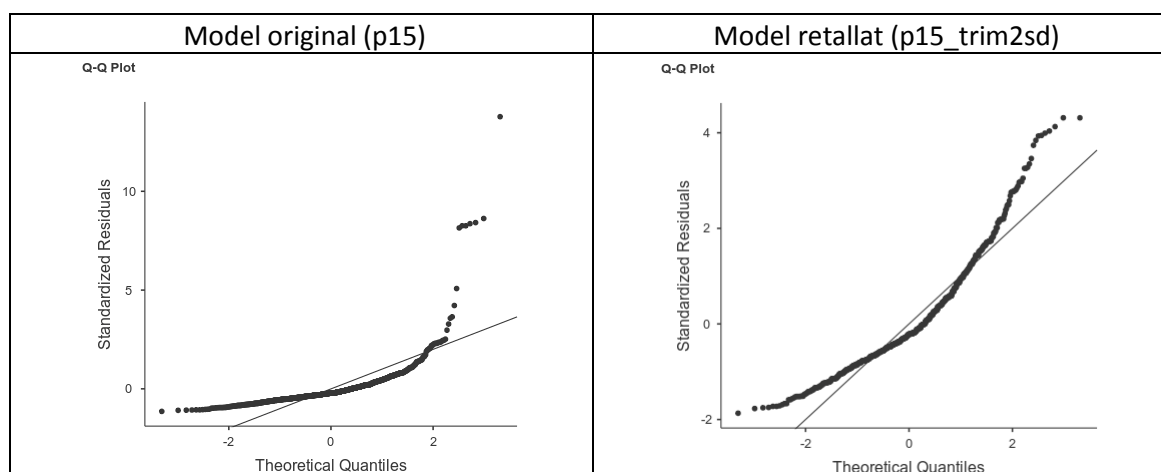
1. Calcula la predicció d'amics a xarxes a partir de la edat i aplicant la funció. Clic al nom de la variable p15, clic a menú Data / Compute... Al panell de càlcul, anomena la **nova variable “p15_pred”** i a la finestra de fórmula escriu el codi corresponent a *“Si el valor de la variable amics (p15) és perdut, llavors a aquesta variable també ho serà, i si no ho es, 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. Calcular la desviació o diferencia del nombre d'amics “reals” respecte a “predits”. Clic al nom de la variable p15_pred, clic a menú Data / Compute... 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. Calcular la estandardització de la desviació anterior. Clic al nom de la variable p15_desv, clic a menú Data / Compute... 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. Crear la variable final, transformant p15_desvEst per que incorpori només els valors de p15 dins del rang de desviació acceptable (2 SD). Clic al nom de la variable p15_desvEst, clic a menú Data / Transform... Al panell de càlcul, anomena la **nova variable “p15_trim2SD”** i al desplegable “using transform” selecciona l'opció “Create New Transform”. Pots anomenar-la “trim 2sd” i a la finestra de regles de transformació, fes clic a **+ Add recode condition** per escriure el codi corresponent a *“Si el valor en la variable origen és major que 2 llavors el valor d'aquesta variable serà perdut (NA) i en qualsevol altre cas el valor serà el nombre d'amics reals (p15)”*, en codi R:
`if $source > 2 use NA
else use 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..

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.

- Repeteix el procediment que has fet abans per analitzar ara la influència de la variable edat (p37) sobre el nombre d'amics, amb la versió “retallada” (p15_trim2sd).

Model	R	R ²	Variància explicada	Recta de regressió
Original	0,221	0,0489	4'9%	Amics = 635,2 – 17,4 x Edat
Retallat	0,364	0,133	13'3%	Amics = 570,5 – 16 x Edat



Imatge 130. Comparació de gràfic Q-Q abans i després de “treure” els casos extremadament desviats identificats mitjançant el gràfic de residus.

Es preferible i més acceptable el model amb la variable “retallada” perquè és més explicatiu, i perquè compleix millor els supòsits del model de regressió lineal simple. La diferència entre els dos models és la excussió o no de 25 casos (un 2,3% de la mostra).

Exercicis proposats: Regressió lineal simple

Després de seguir completament el tutorial d'aquest apartat dedicat a “Regressió lineal simple”, obre amb *jamovi* el fitxer resultant *Dades4.omv* i respon les següents qüestions.

(1) 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). Respon a les següents preguntes:

(1.1) Quina és la variable dependent i quina la independent?

(1.2) En termes generals, el test que ens plantejem és el següent:

H_0 : L'edat al connectar-se per primer cop a Internet, no ha canviat en el temps

H_1 : 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.

(1.3) Escribe la funció que explica una variable a partir de l'altra, i interpreta els elements de la funció.

(1.4) Representa gràficament la relació entre ambdues variables.

(1.5) Digues quant val el coeficient de determinació i interpreta'l.

(1.6) 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.

(1.7) Comprova gràficament els supòsits del model. Quants individus tenim amb un residu estudentitzat fora del rang $[-3, 3]$. Elimina'ls i repeteix el model. S'ha produït algun canvi important?

(2) 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

(2.1) Entra les dades al *jamovi* i dibuixa el núvol de punts. Digues si et suggereix una relació directa o inversa.

(2.2) Planteja el test d'hipòtesis que creguis més oportú i contrasta'l amb un 95% de confiança.

(2.3) 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?

(2.4) Quin percentatge de la variabilitat d'una variable és explicada per la variabilitat de l'altra? Com s'anomena a aquest valor?

(2.5) Calcula la recta de regressió i interpreta'n els resultats.

...

ANNEX. Del paper a la matriu de dades: Flux de treball amb dades de qüestionaris

Eines requerides

1. Text col·laboratiu: Per poder redactar conjuntament amb la resta de persones del grup de treball.
=> **Etherpad** <<https://etherpad.net/>>
=> **Google Documents** <<https://www.google.es/intl/es/docs/about/>>
2. Formularis on-line: Prepara qüestionaris i administra'ls *online*
=> **Google FORMS** <<https://www.google.es/intl/es/forms/about/>>
3. Gestió de respostes: Organitza, modifica i prepara les dades recollides *online* en format "matriu de dades".
=> **LibreOffice-Calc** <<https://es.libreoffice.org/descarga/libreoffice/>>
=> **MicroSoft-Excel**
4. Analitza respostes a qüestionaris: Realitza les operacions de transformació i anàlisi de la matriu de dades pertinents per a justificar conclusions amb justificació estadística.
=> **JAMOVİ** <<https://www.jamovi.org/download.html>>

Google FORMULARIS (GF)

GForms permet (1) construir un "formulari" per facilitar la recopilació de respostes, i (2) generar un "full de càlcul" de Google Drive, vinculat al formulari corresponent, on es van emmagatzemant totes les respostes. Amb GF pots aplicar el teu qüestionari fàcilment, mitjançant una unitat de telefonia portàtil o una tablet-PC, a les persones enquestades personalment.

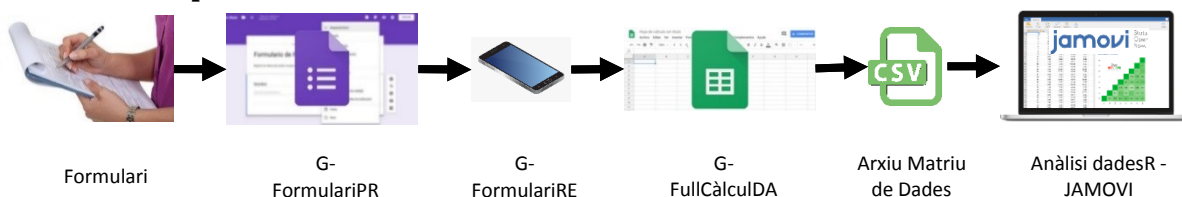
Pre-requisits

1. Disposar d'un conjunt de preguntes (amb respostes) per construir el qüestionari. Cal preparar un document previ, amb qualsevol processador de text, per poder estalviar temps fent retalla i enganxa cap al GF.
2. Disposar de connexió WWW.
3. Disposar d'un compte d'usuari/a de Google (si tens un "*smartphone*" amb SO Android tens un compte de Google).

Operació (5 fases)

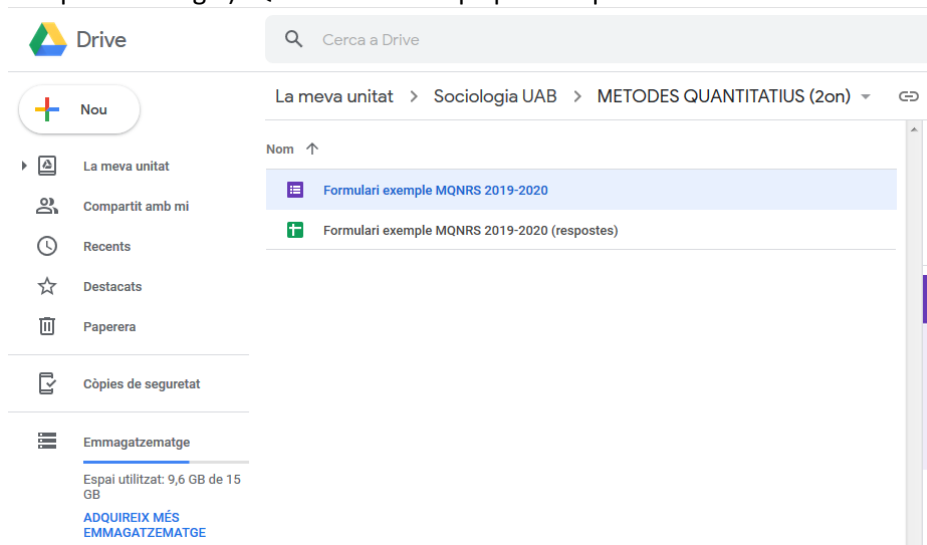
- A. Primer, cal construir el formulari (amb textos, preguntes i respostes). Es convenient utilitzar un PC d'escriptori.
- B. Segon, cal accedir al formulari en mode "visualitza", per poder aplicar-ho a les persones seleccionades. Es convenient utilitzar un dispositiu portàtil, telefonia mòbil o tauleta amb connexió WWW, el que permet fer el treball de camp.
- C. Tercer, cal exportar el full de càlcul G-Drive amb les respostes cap a un arxiu compatible per la posterior importació des de JAMOVİ (la aplicació d'anàlisi de les dades).
- D. Quart, cal importar les dades dins de JAMOVİ.
- E. Cinquè, cal ajustar la descripció i la definició de les variables, i deixar la matriu de dades preparada per a l'anàlisi estadística, en el format propi de JAMOVİ (*.omv).

Flux de Tasques

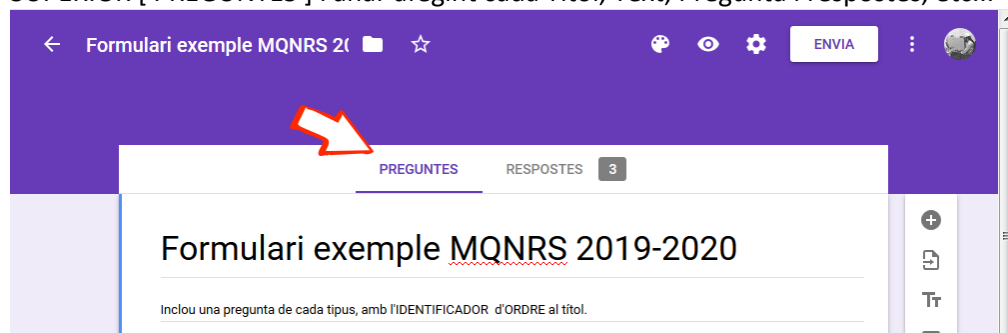


(A) Construcció del Qüestionari

1. Navegar a Google Drive <https://www.google.com/intl/es_ALL/drive/>
Es recomana crear una carpeta especial pel projecte del qüestionari. Aquesta carpeta es pot COMPARTIR amb la resta de membres del grup (cal que també disposin d'un compte de Google). Qui l'ha creat és qui pot compartir-la.



2. Fer clic al botó per crear un [Nou] formulari. Al menú inicial cal indicar "Més" per accedir als G-Formularis.
3. Cada formulari disposa de 2 pestanyes: "PREGUNTES" i "RESPOSTES". Per a introduir els continguts de la enquesta, segons disseny del qüestionari, cal utilitzar la PESTANYA SUPERIOR [PREGUNTES] i anar afegint cada Títol, Text, Pregunta i respostes, etc...



4. **ATENCIÓ:** Els "tipus" de preguntes -en la terminologia pròpia de Google- NO CORRESPON amb els tipus segons el Manual L&F i la resta del material del curs.
 - Preguntes DICOTÓMIQUES: Tria el format "**Selecció múltiple**", i afegir les opcions corresponents.
 - Preguntes OBERTES: Tria el format "**Paràgraf**".

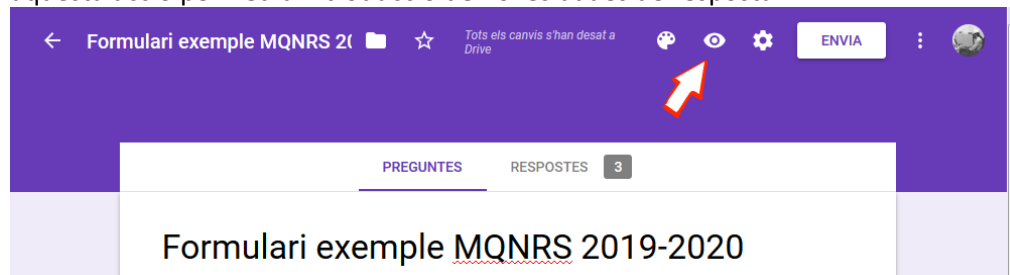
- Preguntes TANCADES: Tria el format “**Desplegable**”, i afegir les opcions corresponents. Pots fer clic-i-moure a la icona lateral esquerra de cada resposta { Sis-Punts } -fins que el punter tingui forma de creu blanca amb 4 fletxes- per a DESPLAÇAR la resposta amunt i avall dins la pregunta.
- Preguntes MULTIRESPOTA: Tria el format “**Graella amb caselles de selecció**”, redacta cadascuna de les diferents opcions de resposta com a "Files" -a la part esquerra- i cadascuna de les possibles respostes com a "Columnes" -a la part dreta-.
- Preguntes ESCALA-NUMERICA: Tria el format “**Escala lineal**”, selecciona els límits numèrics inferior i superior de l'escala, i especifica amb etiqueta textual el significat de cada extrem.
- Preguntes ESCALA-ORDINAL: Tria el format “**Graella de selecció múltiple**”, redacta cadascuna de les diferents "estímuls" o preguntes com a "Files" -a la part esquerra- i cadascuna de les possibles respostes com a "Columnes" -a la part dreta-. Assegura't que les opcions de resposta de l'escala estan correctament ordenades.
- Preguntes CONTINUES (amb unitat): Tria el format “**Resposta breu**” i no oblidis indicar clarament la unitat de mesura (hores?, euros?, nombre de germans?,...)

5. Per a TOTES les preguntes:

- No oblidis incloure l'opció de resposta NS/NC.
 - Es convenient anomenar les variables amb un identificador curt (p2, B17) separat amb un punt del text descriptiu de la pregunta.
 - Les “descripcions” -de Google Form- a cada pregunta es perdran en el procés d'exportació. NO CAL INCLOURE-LES, és millor incorporar la ID i la descripció pregunta a l'espai “Pregunta” de GF.
 - Fes clic a cada pregunta per a editar-la.
 - Fes clic a la icona lateral { Afegir } + per afegir noves preguntes a continuació.
 - Fes clic a la icona { Paperera } per a eliminar la pregunta.
 - *Molt útil per estalviar temps*: fes clic a la icona { Doble-Pàgina } per fer un duplicat de la pregunta, que posteriorment es pot modificar.
 - Fes clic-i-moure a la icona superior { Sis-Punts } -fins que el punter tingui forma de creu blanca amb 4 fletxes - per a DESPLAÇAR la pregunta amunt o avall dins del qüestionari.
 - Fes clic a la icona inferior { Obligatòria } per activar que sigui requerida al formulari.
6. NO CAL, però si t'interessa separar les preguntes en blocs diferents, pots clicar la última pregunta d'un bloc i clicar a la icona lateral { Quadrat-Línia-Blanca-Mig } per afegir una nova "secció" del qüestionari.

(B) Aplicació del Qüestionari

1. En qualsevol moment, pots clicar a la icona { ULL }  per a pre-visualitzar el formulari, aquesta acció permet la introducció de noves dades de resposta.

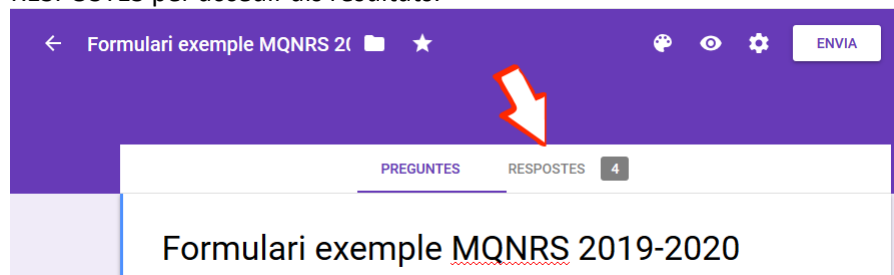


2. Una altra opció és aconseguir el vincle URL associat al formulari de respostes. Qualsevol persona que disposi d'aquesta adreça URL pot respondre al qüestionari, per la qual cosa és important mantenir **estrictament CONTROLADA** la seva difusió.
 - Per aconseguir la URL del formulari: A G-Formularis, clicar al botó superior dret [ENVIA] , veieu que per defecte està activada l'opció icona { Sobre }  que lliura per correu-e l'adreça URL, seleccioneu l'opció a la seva dreta, icona { Cadena }  i veureu l'enllaç URL del formulari, copieu-lo per utilitzar-ho en qualsevol navegador WWW per poder aplicar el qüestionari al subjectes seleccionats com a mostra.
3. En acabar de respondre, clic botó [ENVIA] . Es rep una nota confirmació.
4. Continuar enviant noves respostes (hi ha un vincle) o tancar definitivament el formulari.



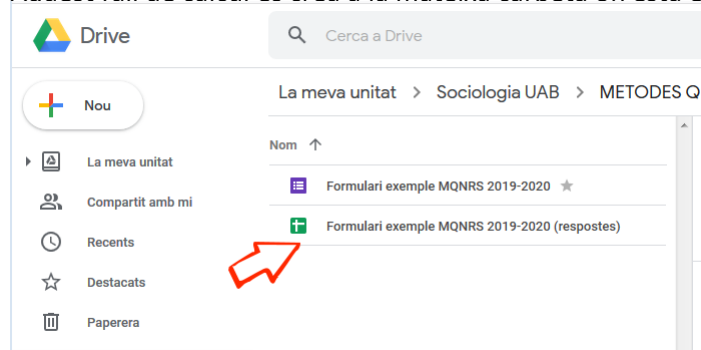
(C) Gestió de les Respostes

1. Cada G-Formulari disposa de 2 pestanyes: "PREGUNTES" i "RESPOSTES". Fes clic a RESPONDES per accedir als resultats.



2. A la part superior dreta i ha una icona per accedir a les dades en format full de càlcul {quadre-verd-amb-creu-blanca} . Clic per a CREAR un nou full de càlcul a Google Drive amb el contingut de les respostes al qüestionari.

3. Aquest full de càlcul es crea a la mateixa carpeta on està el G-Formulari.



4. S'obre el full de càlcul a G-Drive.

- Primera columna és la marca de temps de cada resposta.
- Següents columnes, encapçalades pel TITOL de cada pregunta.
- SI al GF hi ha descripcions, es perden en aquest pas.
- Les preguntes "graella amb caselles de selecció" generen una columna per a cada resposta, correctament identificades a l'encapçalament amb el text de l'opció.
- Les preguntes "escala lineal" generen correctes Escales Numèriques (C).
- Les preguntes "graella de selecció múltiple" generen una columna per a cada resposta, correctament identificades a l'encapçalament amb el text de l'opció.

Formulari exemple MQNRS 2019-2020 (respostes)						
Fitxer Editar Mostra Inserir Format Dades Eines Formulari Complementos Ajuda						
Marca de temps						
	A	B	C	D	E	F
1	Marca de temps	DICOTOMICA p1. Estas	OBERTA p2. Què opines	TANCADA p3. Amb quin	MULTIRESPOSTA p4. A	MULTIRESPOSTA p4.
2	08/10/2019 12:10:01	Sí	dfbxxdfgndfdg	Cotxe	Sí	No
3	08/10/2019 12:10:24	Sí		Bus	Sí	Sí
4	08/10/2019 12:17:00	Sí	ghmncmncv	Cotxe	Sí	No
5	08/10/2019 13:10:59	No	xdfbxxcb	Bus	NS(NC)	NS(NC)
6	14/10/2019 12:21:05	No	sdgsdfgdf	Cotxe	Sí	Sí
7	14/10/2019 12:21:25	No	shdghfdgh	Moto	Sí	No
8	14/10/2019 16:51:20	No sap / No contesta	dfhdfghdfhdfg, aaoi oaishj	Bus	Sí	Sí
9	15/10/2019 12:47:24	No sap / No contesta	hdfgjdfdfhjd	Tren	No	Sí

(D) Exportació de les Respostes

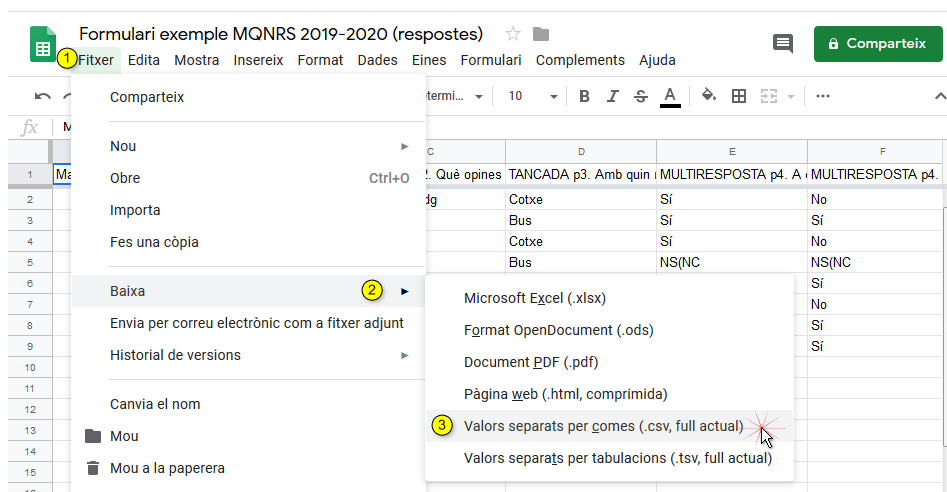
Un fitxer de tipus CSV conté les dades corresponents a les files i columnes d'un full de càlcul, organitzades per "registres". Cada fila és un registre, on cada columna està separada per comes. La primera fila inclou els encapçalaments de cada columna, igualment separats per comes.

Exemple d'arxiu de dades CSV (*Comma Separated Values*):

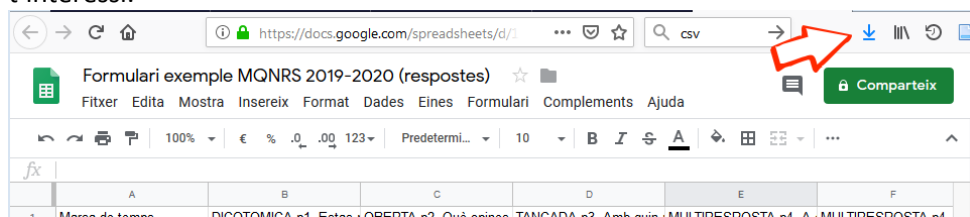
```
Data,Grup,Prova,NIU,Aprovat?,"Curs, si s'escau"
12/02/2019,A,1234567,Parcial,No,1
12/03/2019,B,2222222,Parcial,Si,1
24/10/2019,A,7654321,Final,Si,2
```

Aquest format s'utilitza habitualment per a transferir informació de bases i matrius de dades d'unes aplicacions a altres. Es possible descarregar una versió CSV de totes les dades contingudes al full de càlcul Google Drive.

- Menú / Fitxer / Baixa / Valors separats per comes (.csv, valors actuals). Desar el fitxer al sistema local (descàrrega per defecte del navegador).



- Segons la configuració del navegador utilitzat, caldrà clicar a la icona de descarregues {fletxa-avall-sobre-línia} , accedir a la carpeta on es troba i moure-la al lloc que t'interessi.



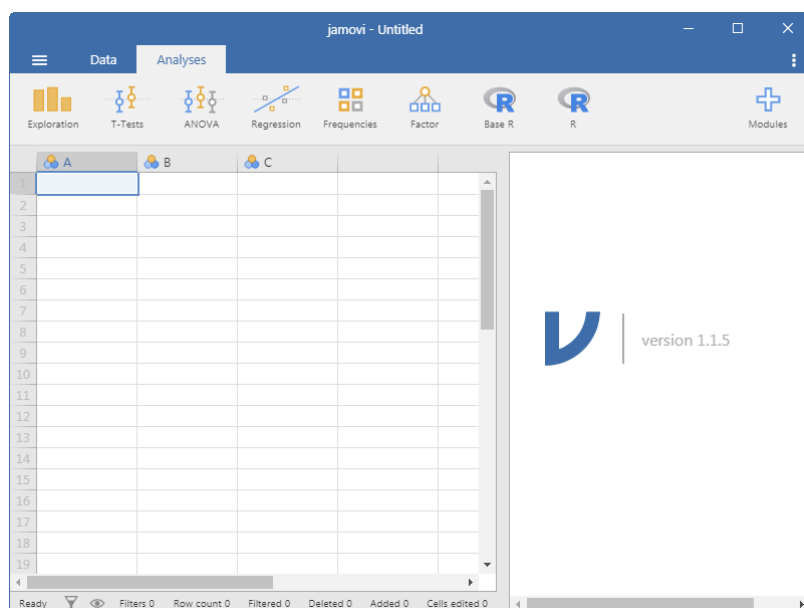
- Ja tens les dades de totes les respostes a un arxiu compatible de tipus CSV.

Gairebé totes les aplicacions d'anàlisi estadística de dades poden importar, o "llegir", dades que estiguin en format CSV.

(E) Importació de les Respostes a *jamovi*

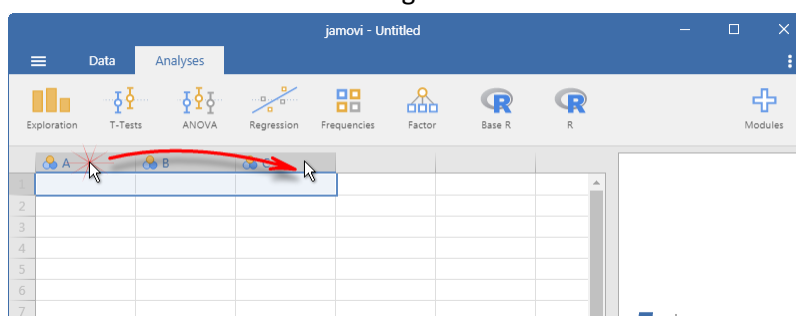
PREREQUISITS

- Disposar d'una matriu de dades.
- Disposar d'una instal·lació de JAMOMI al vostre PC. (A <https://www.jamovi.org/download.html>> trobareu com instal·lar-ho en sistemes Windows, Mac i Linux)

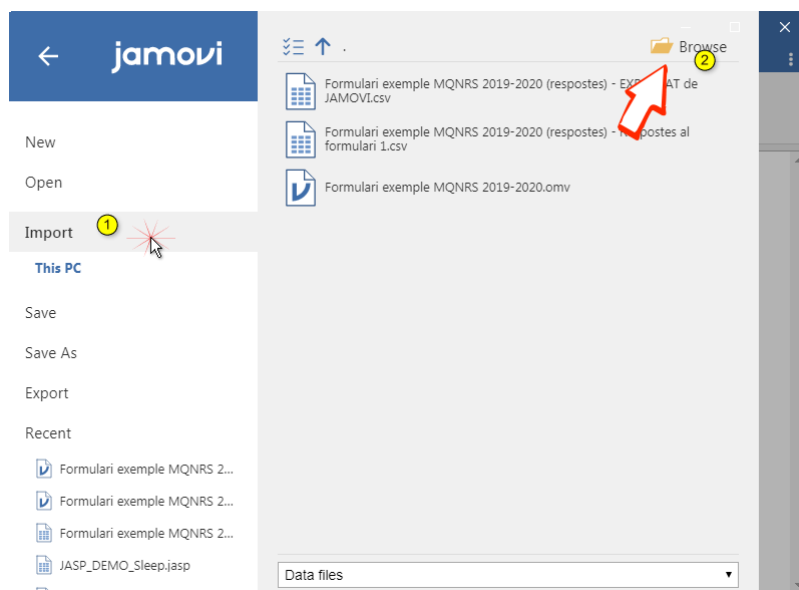


L'aplicació de gestió i anàlisi de dades JAMOVI permet importar arxius de dades tipus CSV. No pot importar "Descripcions", ni tampoc la "ordinalitat" d'algunes variables (O). Aquests són dos aspectes que CALDRÀ AJUSTAR després de fer la importació de les dades.

1. Obrir l'aplicació JAMOVI. Per defecte hi ha TRES variables, anomenades A, B i C, que no necessitem per res i cal eliminar. Selecciona totes tres, fes clic-arrossegat a la primera fila encapçalament de "A", i mou-te fins a la columna "C". Les tres columnes estan seleccionades: marcades amb un gris més fosc.

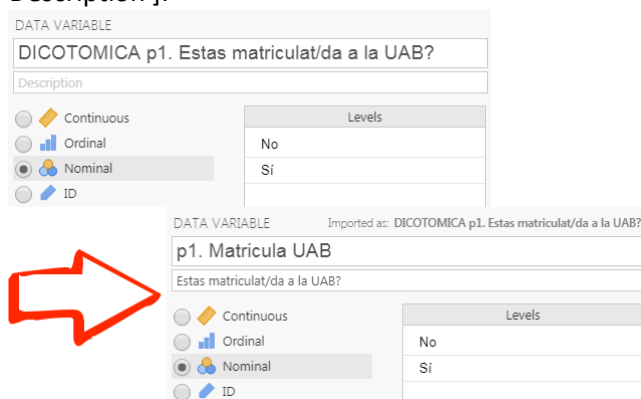


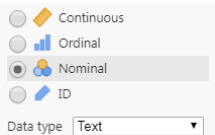
2. Fes clic al menú Data. Fes clic a l'opció "Delete" o icona { Columna-Vermell }  Delete. Confirma per a eliminar les 3 columnes. Ara la matriu està buida, preparada per a importar les teves dades CSV.
3. Fes clic a la icona de menú general { Tres-Línies-Horitzontals } , a la dreta del menú Data.
4. Selecciona Import / This PC / Browse { Carpeta-Groga }  Browse, i selecciona l'arxiu CSV amb les dades. Confirma i comprova com s'han importat les respostes a la matriu de dades.



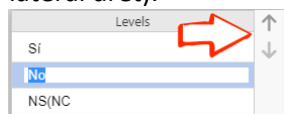
5. CAL MODIFICAR la definició de les variables, fes clic a icona { Setup }

- JAMOVÍ importa els encapçalaments de cada columna com a "nom de la variable", però és **CONVENIENT** que les variables tinguin un Nom curt acompanyat d'una Descripció complerta. Fes retalla/enganxa per aconseguir que el nom contingui només l'identificador (o un abreviatura clara) i enganxa el text complet de la pregunta a [Description].

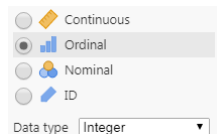


- La variable "marca de temps", cal canviar-la a tipus ID.
- Per passar a la definició de la següent variable fes clic a la icona lateral dreta { > } >, o directament un clic a la columna corresponent. La icona { Fletxa-Amunt-Encerclada }  serveix per tancar la zona de definició de variables (columnes).
- Per a variables de tipus DICOTÓMIC: Comprova que són Nominal i Text.
 
- Per a variables de tipus OBERTA: Comprova que són Nominal i Text.
- Per a variables de tipus TANCADA: Comprova que són Nominal i Text.
- Per a variables de tipus MULTIRESPONSA: Comprova que són Nominal i Text. Re-numera la ID de les respostes per distingir-les, tot i ser resposta múltiple a la mateixa pregunta (p.e., p41., p42., p43.,...). Si cal, mou els nivells amunt o avall, fins aconseguir el **mateix ordre** per a totes les multi-respostes (clic al nivell + fletxes

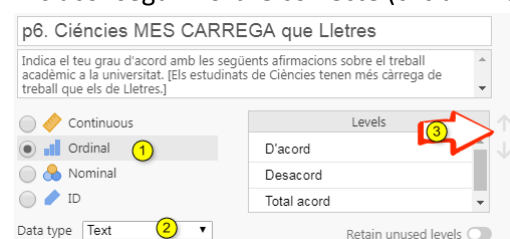
lateral dret).



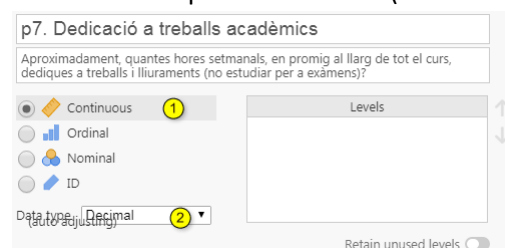
- Per a variables de tipus ESCALA NUMERICA: Comprova que són Ordinal i Integer.



- Per a variables de tipus ESCALA ORDINAL: Comprova que són Ordinal i Text. Re-numera la ID de les respostes per distingir-les, tot i ser aplicable la mateixa escala a diferents "preguntes" (p.e., p41., p42., p43.,...). Si cal, mou els nivells amunt o avall, fins aconseguir l'ordre correcte (clic al nivell + fletxes lateral dret).



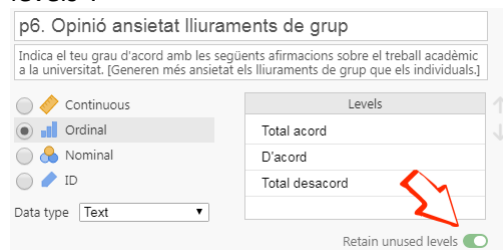
- Per a variables de tipus CONTINUA: Comprova que són Continuous i Decimal (o Integer si no tenen decimals). ATENCIÓ, com que JAMOMI és una aplicació en anglès utilitza el PUNT com a separador decimal (i no la coma).



6. JAMOMI només importa els "nivells" (Levels) o etiquetes de categoria EXISTENTS a la base de dades. Per exemple, si ningú ha contestat "No" aquesta opció no apareixerà com a "nivell" de la variable; això NO ES CAP PROBLEMA, però si es vol que l'anàlisi consideri també les categories buides CAL afegir-les expressament.

Si voleu afegir a la variable una etiqueta, categoria, valor, o nivell que NO CONSTA entre les dades (perquè no hi ha cap resposta a qüestionari que la inclogui) cal utilitzar el següent mètode (indirecte):

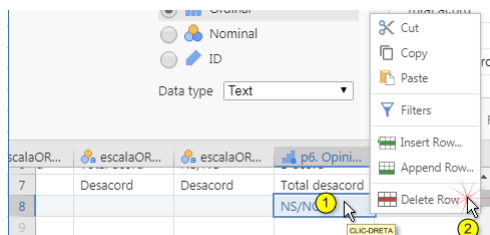
- Fes clic a la columna de la variable que vols modificar. Doble-clic a l'encapçalament de columna si està "tancada" la zona de modificació de la variable.
- A sota del llistat de "Nivells" (Levels) de la variable, activa l'opció "Retain unused levels".



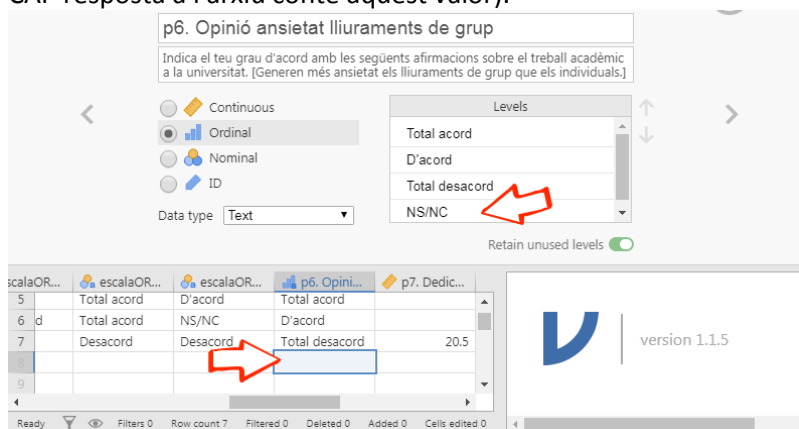
- Fes clic a una fila de la variable que estigui per sota de la última dada importada des de l'arxiu CSV. Escriu el text pel Nivell que vulguis afegir. Pica la tecla [Intro] per confirmar.

	Desacord	Desacord	Total desacord	
7				20.5
8			NS/NC	
9				

- Com veuràs, ha aparegut al final de la llista de Nivells de la variable.
- Fes clic al text que acabes d'introduir. Elimina'l picant la tecla [Supr].



- Com veuràs, la llista de Nivells de la variable ara inclou el valor desitjat (tot i que CAP resposta a l'arxiu conté aquest valor).



- Desa la matriu de dades en el format propi de JAMOMI (*.omv). Fes Menú / Save As..., fes clic a [Browse] per triar la destinació on vols emmagatzemar la teva matriu de dades. Confirma amb el botó [Desa].

JAMOMI 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. Es software lliure, multiplataforma (Windows, Linux, Mac) i és ampliable mitjançant “moduls”.

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.