



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

MUJADES v1.1



2017-2021

AJENJO COSP, M. (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	5
2. Instal·lació del programari <i>jamovi</i>	8
Mòduls d'extensió de <i>jamovi</i>	9
Com afegir el mòdul per a operar directament amb codi R ?	9
3. Interfície gràfica de <i>jamovi</i>	12
Menús d'opcions	13
1. Menú d'arxius	13
2. Modalitat de gestió de dades	15
3. Modalitat d'operacions estadístiques.....	16
4. Modalitat de comentaris de text	17
5. Menú de configuració de <i>jamovi</i>	18
4. Lectura i escriptura d'arxius de dades amb <i>jamovi</i>	21
Introducció directa de dades.	21
Copia de dades des d'altres aplicacions.....	23
Carrega de les dades d'arxius.	25
5. Definició de dades amb <i>jamovi</i> : Tipus de variables.....	29
6. Estadística descriptiva i exploratòria bàsica amb <i>jamovi</i>	34
Exploració bàsica descriptiva de les variables.	34
Exploració de variables amb codi R.....	38
Resultats de la exploració de variables: Afegir comentaris i exportar.....	43
Modalitat comentari de text	45
Exportació de resultats cap a altres aplicacions	46
Descripció gràfica de variables.....	48
7. Transformació i depuració de dades mitjançant <i>jamovi</i>	55
El tractament de les "no resposta": Els valors perduts.....	56
Recodificació de variables.	63
Recodificació de variables quantitatives a qualitatives	64
Recodificació de variables qualitatives a qualitatives	70
Recodificació de variables qualitatives a quantitatives	71
Recodificació de variables quantitatives a quantitatives.....	74
Càlculs amb variables quantitatives.....	74
Operacions i càlculs entre diverses variables	76
Transformacions complexes (1): Combinació de recodificacions i càlculs	79
Transformacions complexes (2): Construcció de tipologies.....	83
Filtrar casos	90
Detecció i depuració d'errors.....	92
Errors en variables individuals	92
Errors en creuaments de variables	93
Exercicis proposats: Transformació de dades	94
8. Inferència amb una variable utilitzant <i>jamovi</i>	96
Construcció d'un interval de confiança per a una proporció (π).	96
Construcció d'un interval de confiança per a la mitjana (μ).	99
Codi R per a la construcció d'interval de confiança (mitjana).	100
9. Inferència de relació entre variables	102

10. Anàlisi de taules de contingència mitjançant <i>jamovi</i>	103
L'estadística descriptiva en les taules de contingència.	104
La representació gràfica de les taules de contingència	109
La inferència estadística amb dues variables no quantitatives: El test de la khi quadrat de Pearson.	111
A) Les condicions d'aplicació en el test de la khi quadrat de Pearson.....	111
B) Test de significació estadística khi quadrat de Pearson	112
C) Intensitat de l'associació entre variables nominals (V de Cramer).	113
D) Avaluació del sentit de l'associació	114
Solució als problemes de compliment de les condicions d'aplicació.	117
La taula de contingència 2*2 (la correcció de Yates).....	118
11. Comparació de mitjanes amb <i>jamovi</i>	122
Quin test cal fer?	123
L'estadística descriptiva en la comparació de mitjanes.....	125
El test de Shapiro-Wilk per a la Normalitat de la variable.	128
La comparació de dues mitjanes independents.	130
La comparació de més de dues mitjanes independents: L'anàlisi de la variància (ANOVA). ...	134
La intensitat de la relació: El coeficient de determinació.	138
Exercicis proposats: Comparació de mitjanes	139
12. Regressió lineal simple mitjançant <i>jamovi</i>	141
Descripció de la relació entre dues variables quantitatives.	141
El diagrama de dispersió	142
La correlació entre ambdues variables	145
El model de regressió lineal simple.....	146
Inferència estadística en els models de regressió	148
Els supòsits del model: L'anàlisi de residuals.....	149
Exercicis proposats: Regressió lineal simple	153
ANNEX. Del paper a la matriu de dades: Flux de treball amb dades de qüestionaris	154
Eines	154
Google FORMULARIS (GF).....	154
Prerequisits	154
Operació (5 fases)	154
Flux de Tasques	155
(A) Construcció del Qüestionari	155
(B) Aplicació del Qüestionari.....	157
(C) Gestió de les Respostes	157
(D) Exportació de les Respostes	158
(E) Importació de les Respostes a <i>jamovi</i>	159

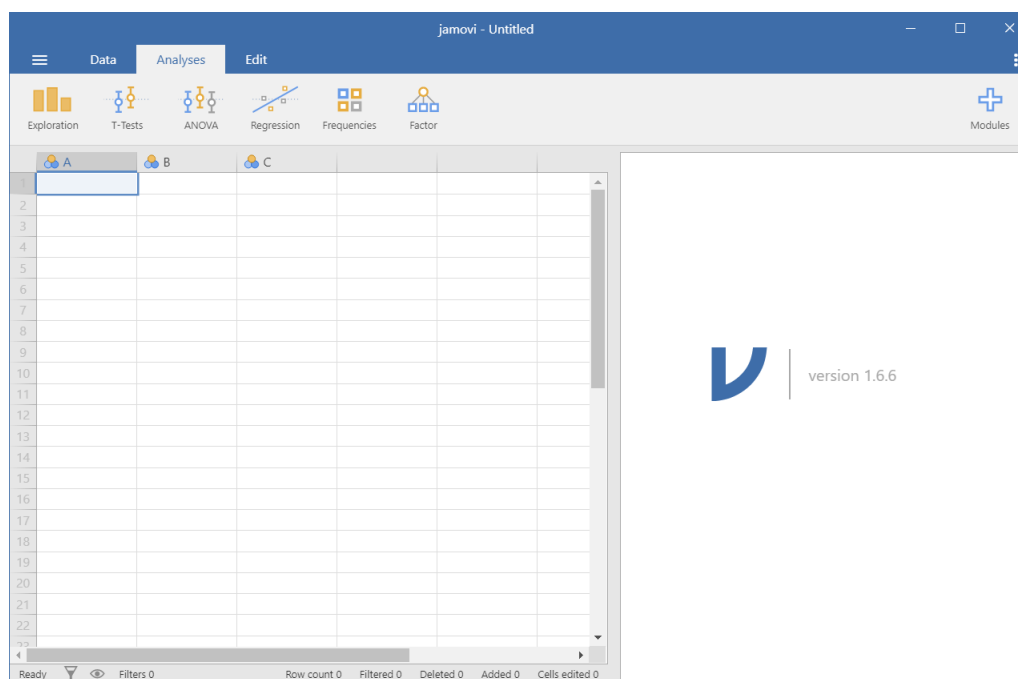


1. *jamovi*: Tractament i anàlisi de dades quantitatives

*jamovi*¹ és un sistema de tractament de dades i anàlisi estadístic que proporciona totes les funcionalitats bàsiques requerides per a la recerca social amb mètodes quantitius. 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 permet 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.

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 1. Pantalla principal de jamovi

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



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 (be sigui al mòdul Rj de *jamovi*, o directament a R). D'aquesta forma, *jamovi* resulta una eina per l'anàlisi de dades quantitatives tant pel seu ús directe com per la seva utilitat per l'aprenentatge de programació amb R.

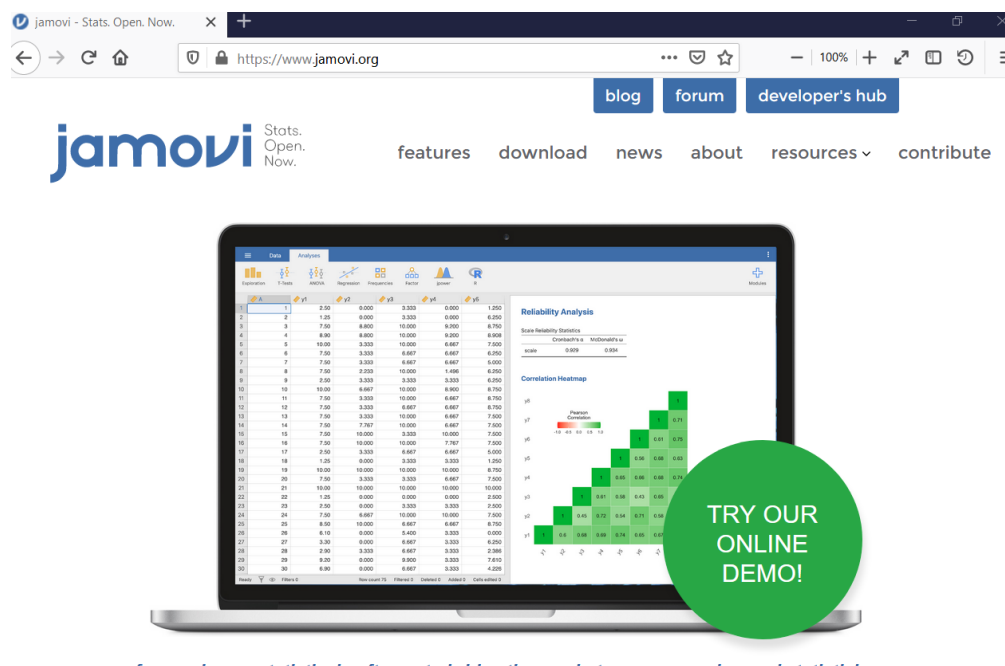
Aquesta versió del tutorial MUJADES utilitza imatges de dues versions diferents de *jamovi*., la 1.6.6. i la 1.1.5, amb algunes diferències en quant als menús.

L'espai web del projecte *jamovi* inclou un apartat molt interessants per a l'autoaprenentatge d'aquest sistema d'anàlisi estadístic. Es recomana complementar el present 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 i instal·lar al PC local.²



Imatge 3. Website de *jamovi* (6 nov. 2020)

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. Fes clic a la teva opció. Espera que el 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.

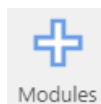
Mòduls d'extensió de *jamovi*

L'aplicació *jamovi* inclou un conjunt de procediments d'anàlisi estadístic inicial que es pot ampliar amb els anomenats "modules". Relacionat amb la seva relació (i dependència) del llenguatge R, *jamovi* segueix la mateixa filosofia:

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

Els 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*). Aquests 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 descarrega 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. Es possible descarregar-los 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 descarrega si es disposa de connexió WWW.



Imatge 5. Accés a gestió d'extensions

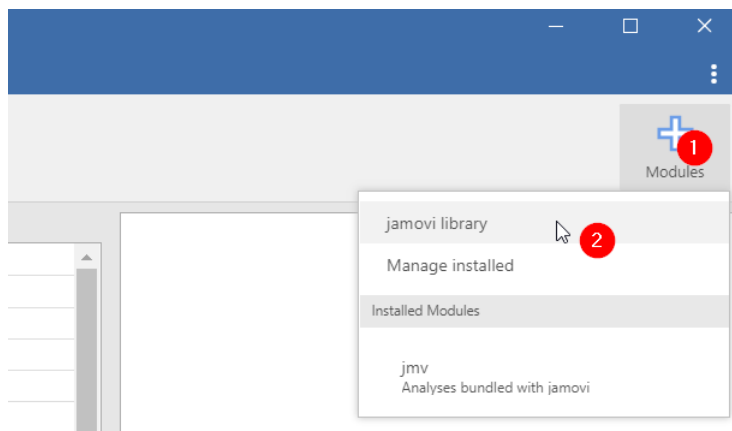
A continuació s'exemplifica aquest procediment d'instal·lació de mòduls des de dins de *jamovi*.

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

Al llarg del present tutorial es mostrarà com realitzar determinades operacions estadístiques directament executant codi en llenguatge R. En ocasions es tractarà d'una forma alternativa al sistema de menús que aconsegueix els mateixos resultats, en altres ocasions s'utilitzarà codi R per aconseguir superar determinades limitacions del sistema de menús bàsic. D'aquesta forma, al llarg de tot el tutorial MUJADES, s'aniran introduint elements bàsics del llenguatge de programació R.

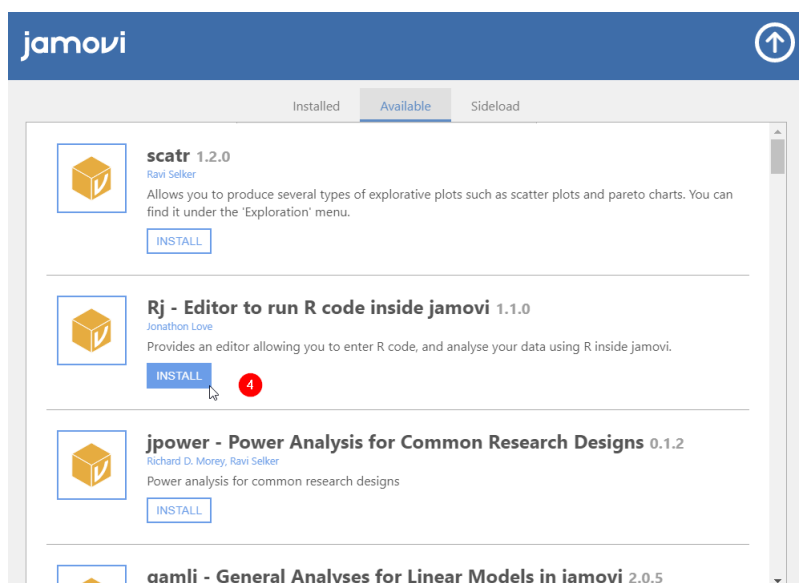
Per a poder executar codi R directament amb *jamovi*, sense el sistema de menús, cal prèviament instal·lar el mòdul anomenat "**Rj – Editor to Run R code inside jamovi**". Per fer-ho, pas a pas:

1. A la pantalla principal de *jamovi*, a la modalitat ANALISI DE DADES ("Analyses"), al menú gràfic, extrem dret, clic a la icona "Modules".



Imatge 6. Accés al sistema de gestió d'extensions

2. El sistema connecta amb el repositori WWW de mòduls i mostra el catàleg de les extensions disponibles, juntament amb una descripció i botons per procedir a la seva instal·lació. Busca el mòdul "Rj – Editor".



Imatge 7. Catàleg d'extensions (jamovi library)

3. Fes clic el botó [INSTALL]. Apareix un missatge mentre s'està fent la descàrrega. Quan culmina el procés, al catàleg s'indica "INSTALLED".
4. Tanca la finestra del catàleg de mòduls (jamovi library) fent clic a la icona fletxa-amunt-encerclada de la part superior dreta.
5. Al menú gràfic de *jamovi* ara 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.



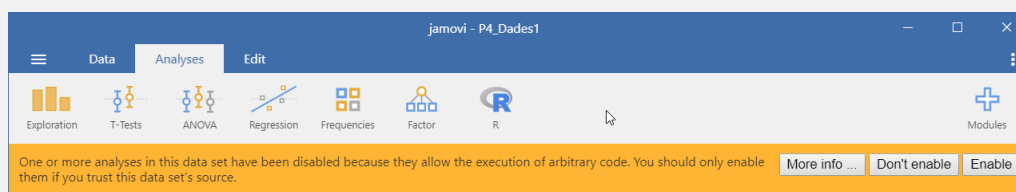
Imatge 8. Menú gràfic ampliat amb mòdul Rj-Edit

6. Ara el mòdul corresponent ha quedat instal·lat al PC.

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

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

En cas de missatge d'error o advertiment, respondre [Enable]



Imatge 9. Missatge d'error associat a l'ús de codi R directe (mòdul Rj)

Per a gestionar els diferents mòduls instal·lats:

1. A la pantalla principal de *jamovi*, al menú gràfic superior, extrem dret, clic a 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].

Pot comprovar com hi ha una extensió en concret 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, clic a la icona "Exploration". Comprova que la 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 opcions s'utilitzaran al llarg del present tutorial)

Tingues en compte que si estàs utilitzant un PC on hi ha *jamovi* instal·lat (p.e., sales d'informàtica, o accés remot 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*

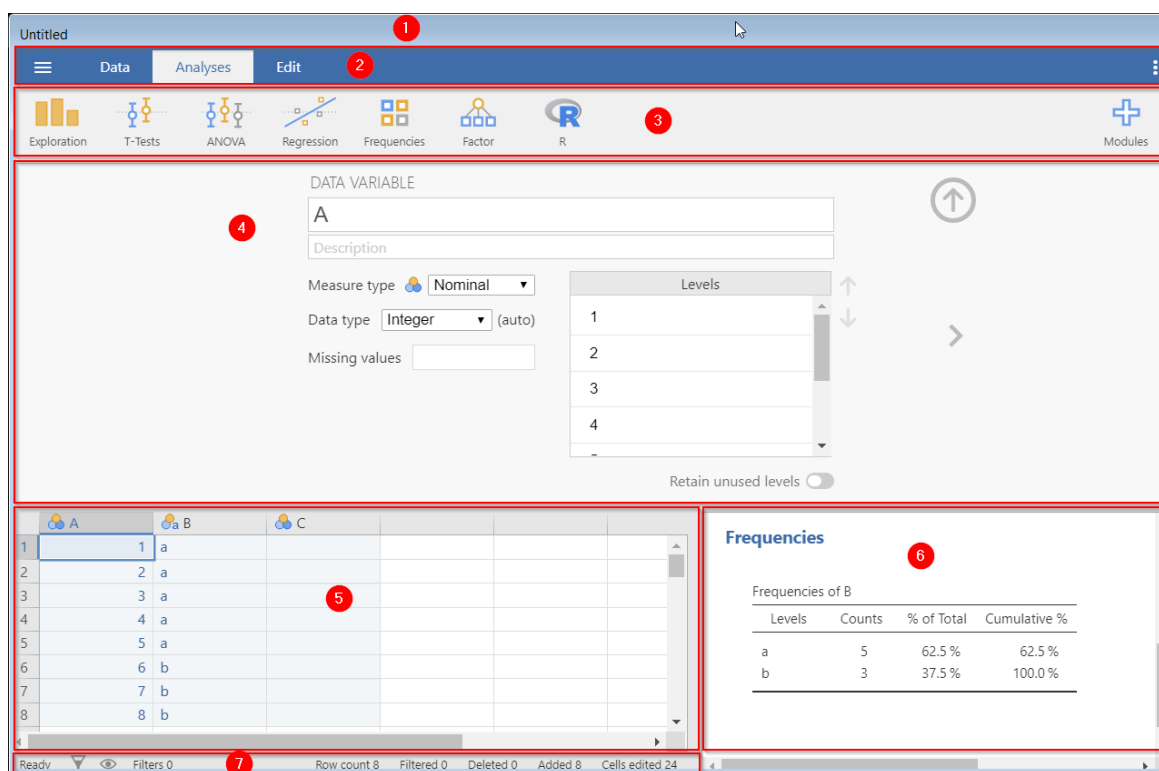
La aplicació *jamovi* és una interfície gràfica d'usuari (GUI) per executar operacions en el llenguatge de programació R especialitzat en l'anàlisi de dades. Com a tal, el seu sistema de menús inclou una part (mínima) de totes les potencialitats de R, però ho fa d'una forma amigable i de màxima facilitat d'ús sense haver d'aprendre programació.

En aquesta secció es presenten sistemàticament els diferents elements d'aquesta interfície gràfica.

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

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

1. Modalitat GESTIÓ de DADES,
2. Modalitat ANÀLISI de DADES, i
3. Modalitat COMENTARI de RESULTATS.



Imatge 10. 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 tres modalitats de funcionament: gestió de dades, operacions estadístiques, i comentaris de text).
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 amb casos per files i variables a les columnes) i de configuració de les anàlisis (depèn de la modalitat activa).
6. Finestra de resultats (amb les taules, gràfiques i textos generats).
7. Barra d'informació i sistema de filtratge de casos.

La finestra de definició de variables (4) es mostra o 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) és substituïda 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 d'altres “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 d'opcions

A la part superior de la pantalla principal es troba l'accés als CINC menús d'opcions.³ La barra de color blau conté els activadors d'aquest menú, respectivament, d'esquerra a dreta:



Imatge 11. Menús d'opcions generals de *jamovi*

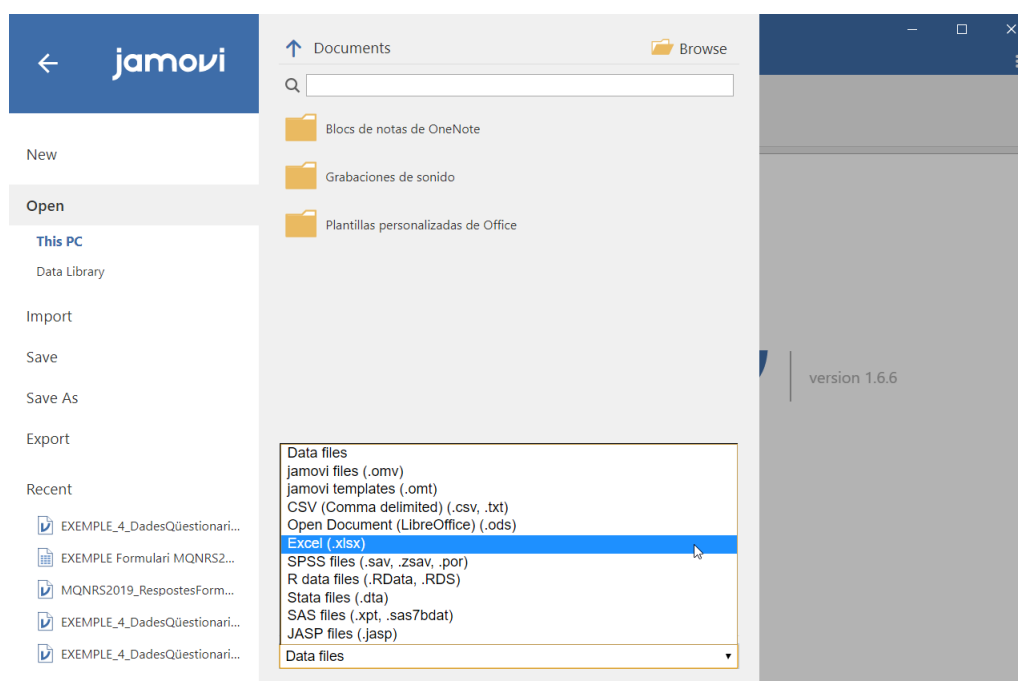
1. [Icona-tres-línies-horitzontals] Menú d'arxius: Gestiona arxius de *jamovi* i d'altres formats.
2. “Data” Activa la modalitat de GESTIO de DADES: Per a treballar amb les dades i variables.
3. “Analyses” Activa la modalitat de ANÀLISI de DADES: Per a demanar resultats d'operacions.
4. “Edit” Activa la modalitat de EDICIÓ de TEXT: Per a afegir comentaris propis als resultats.
5. [Icona-columns-tres-punts] Menú de configuració: Controla aspectes de visualització.

1. Menú d'arxius

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

Permet crear nous arxius de dades propis de *jamovi* (amb un format anomenat OMV), obrir-ne d'existents, desar l'actual, així com importar i exportar dades i resultats (segons el format) de i a arxius d'altres formats.

³ El nombre de menús pot variar per a versions diverses de *jamovi*. En aquest cas es tracta de la versió 1.6.



Imatge 12. Menú d'arxius

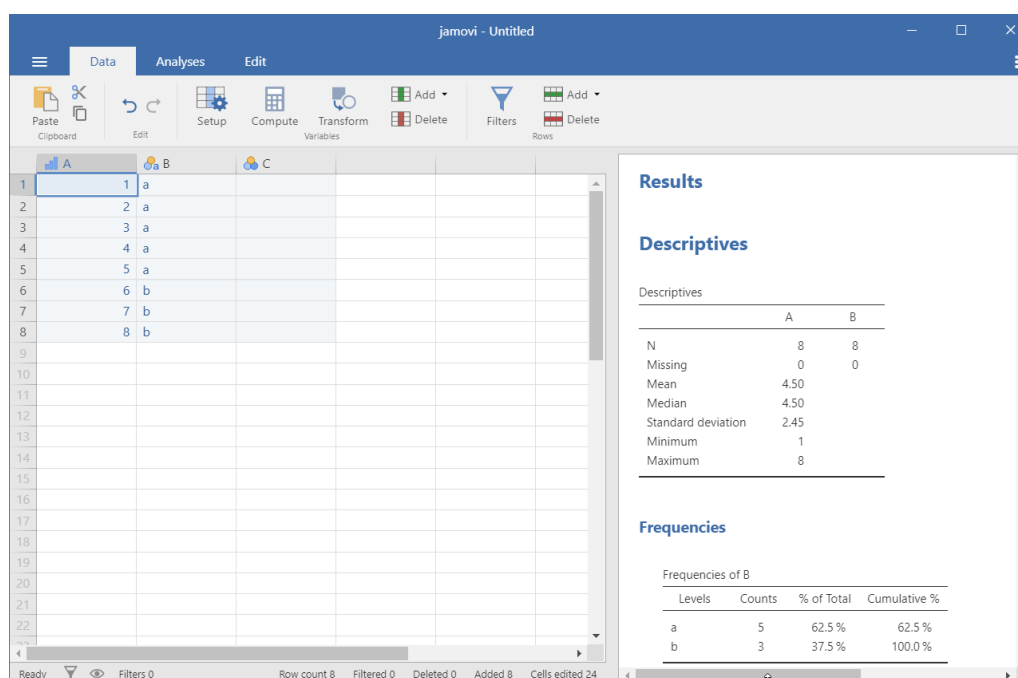
- **New:** Obre una nova instància de *jamovi*, sense tancar l'actual.
- **Open:** Mostra una finestra per seleccionar un arxiu de dades per a carregar-les dins de *jamovi*. Funciona amb diversitat d'arxius d'aplicacions habitualment utilitzades per a l'anàlisi de dades quantitatives (SPSS, Stata, SAS, JASP) i de Fulls de càlcul (LibreOffice Calc, M\$-Excel), així com arxius de R, o el format genèric per intercanvi de dades CSV. 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 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.
- **Import:** Mostra una finestra per seleccionar un més arxiu de dades per a afegir-los dins de *jamovi*. Funciona amb arxius de tipus propi OMV. 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 assignat un nom a un arxiu "Save" ho desa amb aquest nom a la mateixa ubicació.
- **Export:** Genera diferents tipus d'arxius amb continguts específics, segons el format seleccionat per a la exportació.
 - Resultats: Es poden generar pàgines amb totes les taules i gràfics dels resultats i amb els textos adjunts, en el cas d'exportacions cap a PDF, a HTML, o a LaTeX (bundle zip).
 - Dades: En el cas d'exportacions cap a formats d'aplicacions d'anàlisi (SPSS, Stata, SAS, CSV, o R) el que 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 la obertura, però resulta en una còpia exacta de la tasca realitzada anteriorment.

2. Modalitat de gestió de dades

S'utilitzarà quan es vulgui modificar les dades existents, bé sigui a nivell de casos (filtrant-los segons criteris, afegint o eliminant, enganxant d'altres aplicacions,...) i a nivell de variables (creant-ne de noves a partir d'existents, eliminant, ...).

Mostra el sub-menú gràfic amb les opcions disponibles per a editar, eliminar, moure, modificar i filtrar les dades actuals, organitzades en una matriu de files de casos/individus i columnes de variables.



Imatge 13. Modalitat gestió de dades

- **Clipboard:** Controls per a intercanviar dades entre diverses zones de la matriu, o amb d'altres aplicacions (per exemple, intercanvi amb fulls de càlcul, mitjançant operacions de retalla/enganxa del porta-papera).
- **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 possibilitat d'ajustar el seu nom, la descripció, el tipus de variable (nominal, ordinal, contínua), el tipus de les dades (text, decimal, nombre sencer), els valors perduts, i les categories o etiquetes ("levels").
 - **Compute:** Activa i desactiva la finestra de CÀLCUL de variables, amb possibilitat de crear NOVES variables a partir d'una o més existents, tot aplicant funcions estadístiques, lògiques, o matemàtiques (p.e., 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).

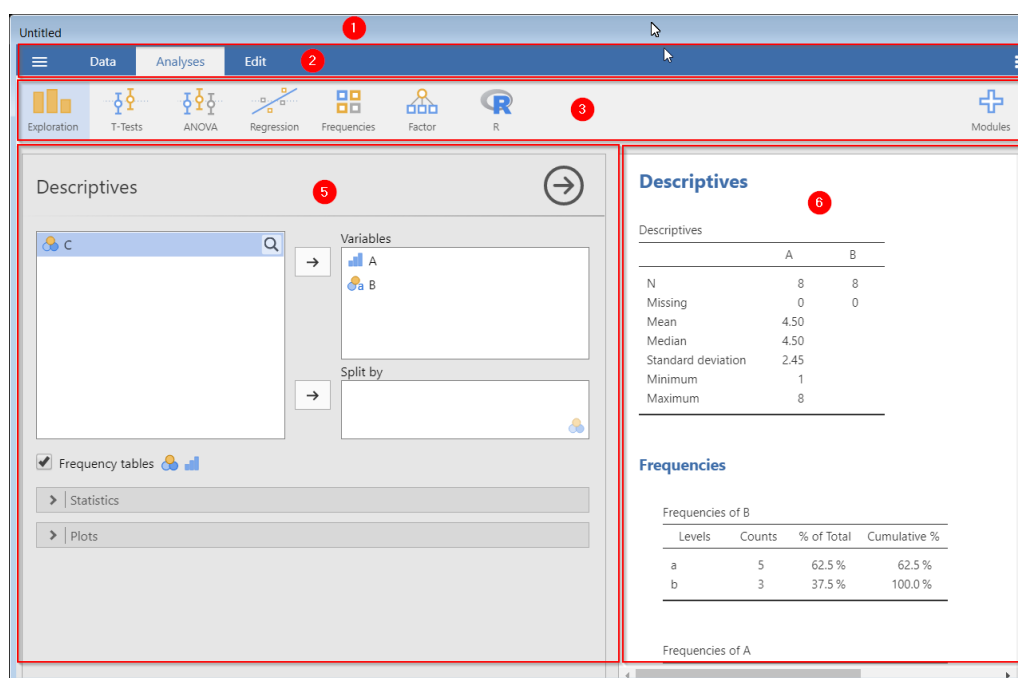
- **Transform:** Activa i desactiva la finestra de TRANSFORMACIÓ de variables, amb possibilitat de crear NOVES variables a partir d'una existent, tot aplicant criteris condicionals (del tipus IF-THEN-ELSE) fonamentats en els valors o etiquetes de la variable original (*p.e., la recodificació de la variable 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 nova columna en blanc.
- **Rows (files, casos)**
 - **Filters:** Activa i desactiva la finestra de FILTRATGE de casos, amb 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. La 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.e., 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, a continuació de la selecció 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 encapçalament de les columnes, o dels nombres identificadors de les files de la part esquerra de la matriu.

3. Modalitat d'operacions estadístiques

S'utilitzarà quan es vulgui 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 sub-menú gràfic amb els procediments disponibles per a realitzar les operacions disponibles per la instal·lació local de *jamovi*. Conté tots els procediments bàsics, més l'accés als mòduls d'expansió que han estat instal·lats.



Imatge 14. Disposició en modalitat d'operacions estadístiques

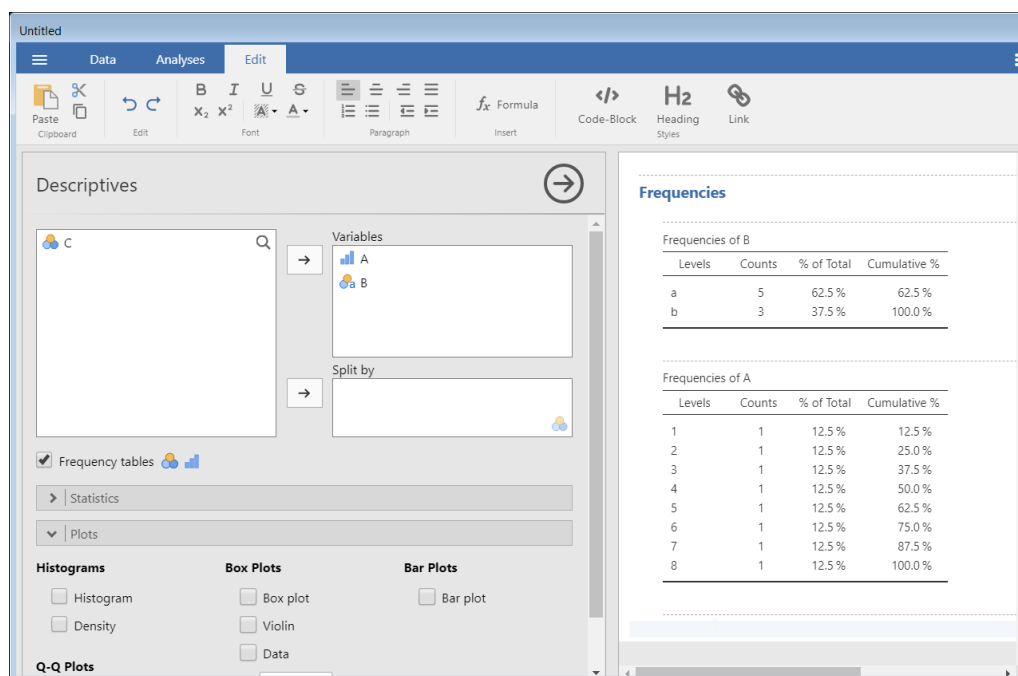
Fent clic a les icones del sub-menú gràfic desplega un menú amb els procediments estadístics disponibles, que han de ser utilitzats específicament per a condicions concretes de les dades. Correspon a qui realitza l'anàlisi conèixer les condicions d'aplicabilitat de cadascú dels procediments: *jamovi* 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 inadequat.

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

4. Modalitat de comentaris de text

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

Mostra el sub-menú 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 15. Modalitat comentari de text

Les diferents icones d'edició de text del sub-menú gràfic funcionen de forma similar a les que disposa qualsevol aplicació de processament de textos (LibreOffice Writer, M\$-Word, MacOS-Pages, etc.).

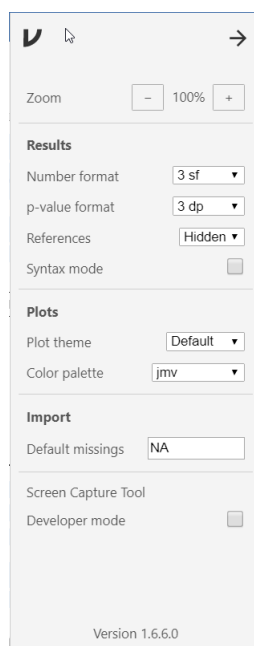
Quan s'activa amb clic a "Edit" aquesta modalitat a la finestra de resultats queden destacats (en blau clar) les àrees on es pot afegir text, associades als diferents objectes de resultats generats per les operacions. Clicar a aquestes àrees permet la seva edició, de forma que -típicament- es poden fer comentaris adjunts a cada taula, així com redactar extensions de text de presentació i de conclusions que conformen un INFORME complet dins de *jamovi*, amb l'avantatge que es poden desar i compartir en format OMV, o publicar 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 conjuntament les dades originals, més totes les transformacions i càlculs realitzades, més tots els resultats generats, més tot el text que acompanya i comenta als 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 de cursos de la Facultat.

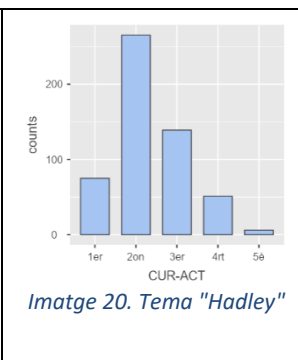
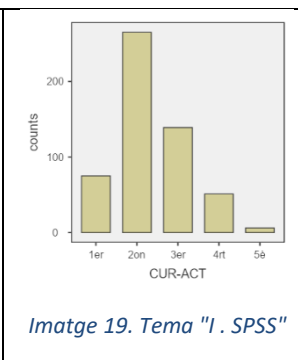
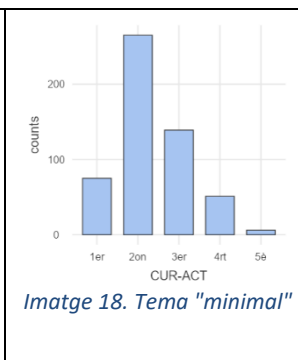
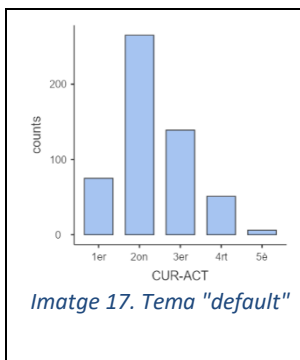
5. Menú de configuració de *jamovi*

Desplega un llistat d'opcions que permet controlar algunes configuracions de l'aplicació, així com activar l'eina de captura de pantalla.



Imatge 16. 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, be en forma "relativa" de dígitos significatius (sense tenir en compte els zeros a l'esquerra, així el format 3sf genera 0,123 i 0,0045) 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 p-valors de significativitat, be en forma "relativa" o en forma "absoluta".
 - **References:** Permet ocultar o visibilitzar automàticament al final del resultats les referències bibliogràfiques pertinents al projecte *jamovi* i al llenguatge **R**, en les versions utilitzades per les anàlisis realitzades.
 - **Syntax mode:** Activa / desactiva el mode en que els resultats no es formaten per a publicació sinó que apareixen tal i com els calcula el llenguatge **R**, e inclou el codi **R** corresponent a cada resultat generat. Habitualment desactivat, per a aconseguir resultats més presentables i 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.
- **Screen Capture tool:** Activa una eina per a obtenir captures d'imatge del contingut del conjunt de finestres d'una instància de *jamovi*. Útil per a disposar de material gràfic associat a les anàlisis; per exemple, s'ha utilitzat per a realitzar aquest tutorial.
- **Developer mode:** Activa un mode especial d'utilitat per a qui programa nous mòduls i expansions de *jamovi*. En general, per a analistes usuàries no te cap més efecte que disposar de més informació 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 ocasions interessa analitzar amb *jamovi* dades que es troben en formats no digital, en qüestionaris, o en publicacions impreses. Caldrà introduir, teclejant manualment, les dades.

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

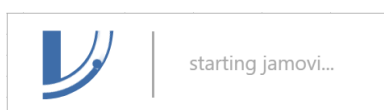
- HI ha unes dades publicades que es volen integrar dins d'una matriu de dades de *jamovi*.

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:

1. Executa l'aplicació *jamovi* al teu ordinador.



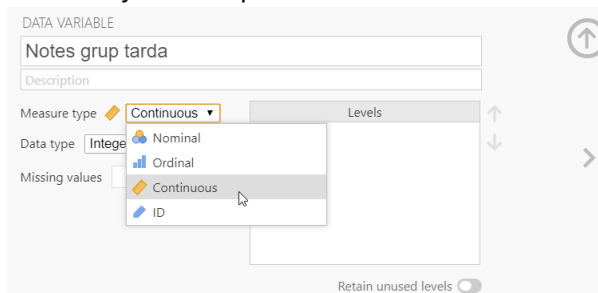
Imatge 21. Indicador d'inici

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 números sencers i sense decimals). Per fer-ho...
4. Clic al menú / Data / Setup
(alternativament, doble-clic a l'encapçalament de la primera columna (lletra A))
5. S'obre l'àrea de definició de variables.
6. Tecleja el nom desitjat, substituint les lletres assignades automàticament pel sistema, per exemple: Notes torn de tarda.
7. Pots incloure qualsevol caràcter (fins i tot accentuats, 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.
8. A "Measure type" (tipus de variable) fes clic a desplegable i selecciona "Continuous".⁴
9. A "Data type" (nivell de mesura de les dades) selecciona "Integer" (és l'opció per defecte)
10. Pots continuar definint d'altres variables, fent clic a les icones laterals > i <
(alternativament, fent clic a l'encapçalament de la columna corresponent)
11. Pots tancar l'àrea de definició de variable amb clic a la fletxa-amunt-encerclada.

En el cas de variables de tipus nominal, o ordinal, caldrà comprovar les categories de la variable ("Levels"). Veure en detall el capítol posterior "

⁴ Per a més informació sobre "Measure Type" i "Data Type", consulta l'apartat posterior "

5. Definició de dades amb *jamovi*: Tipus de variables".

12. 5. Definició de dades amb *jamovi*: Tipus de variables”.

Imatge 22. Definició de variable numèrica sense decimals

13. Comprovar com l'encapçalament de la primera columna ara inclou el nom, i la petita icona que identifica variables quantitatives (una regla de mesura de color taronja).
14. Clic al 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'equivocació, només cal clicar a sobre de la casella que es vulgui 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.

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: Introduir les dades de qüestionaris en paper en un full de càlcul, que permet la introducció contínua per files, i posteriorment utilitzar 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, s'avança aquí com procedir per a DESAR les dades ara 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. Fer clic a el Menú / Arxius [tres línies horitzontals] 
3. Al desplegable, selecciona l'opció "Save". S'obrirà una finestra d'emmagatzematge.
4. Teclejar el nom desitjat, p.e., "P3DadesManuais" (recomanable sempre sense espais)
5. Clic a "Browse", per a seleccionar la ubicació d'aquest arxiu *jamovi*.
6. S'obre una finestra de 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 23. Missatge indicació procés de desar arxius

8. Comprova que la barra superior de l'aplicació ara indica "jamovi - P3DadesManuais".

Copia de 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 el línia. Per exemple, les dades poden estar a un full de càlcul, o 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 "P3_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 24. Exemple de dades disponibles a un full de càlcul

Observa com 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 tarda de l'assignatura.

A continuació es mostrarà com AFEGER aquestes dades 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). Tot plegat s'utilitzarà per a practicar les operacions disponibles amb *jamovi* i el porta-papers del sistema operatiu.

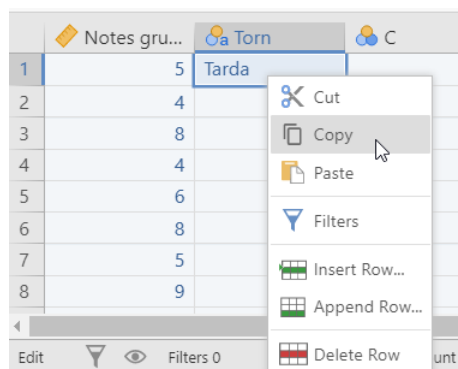


EXERCICI

1. Si està tancat, obre l'arxiu de *jamovi* que vas desar anteriorment amb les dades introduïdes manualment. Si ho ja ho tens obert només cal que continuïs treballant a aquell arxiu.
2. Afegeix una nova variable a la matriu de dades de *jamovi*, a la segona columna (que automàticament es va anomenar B en crear la matriu de nou). Per fer-ho, pots consultar la secció anterior.
3. Anomena-la "Torn", i configura-la de tipus de variable "nominal", i amb tipus de dades "Text".

Pots tancar, o no, la finestra d'identificació de variables. A continuació introduïrem les dades restants fins aconseguir una matriu de dades amb dos variables.

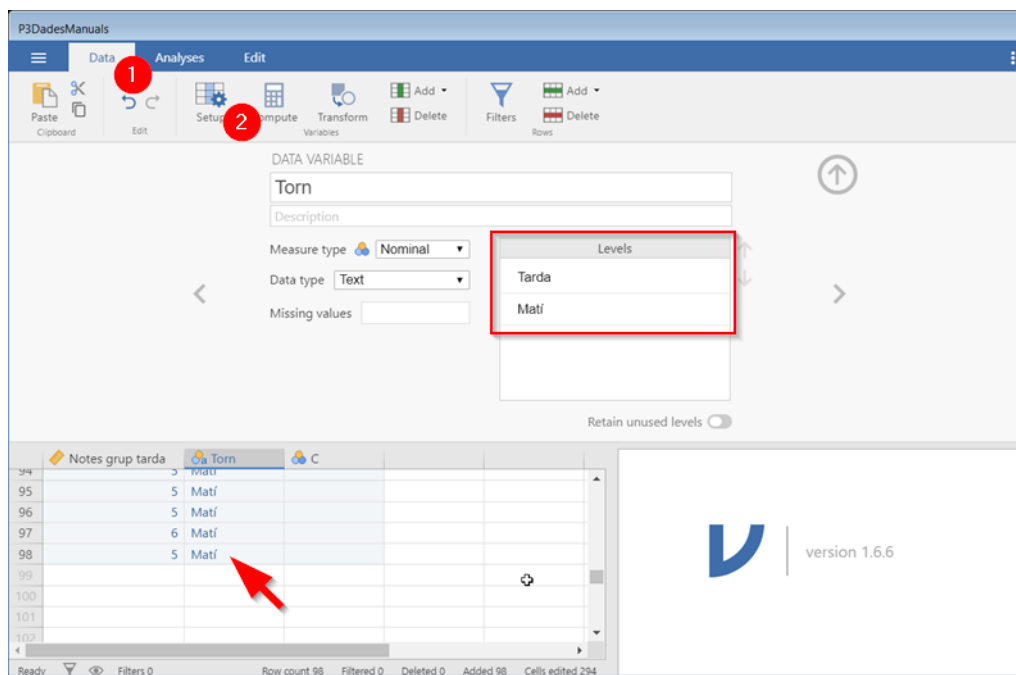
1. A la finestra de dades, fes clic a la casella (buida) del primer individu per a aquesta segona variable que has creat.
2. Introdueix manualment el valor de la categoria corresponent: Tarda.
3. Podries fer-ho manualment pels 41 individus restants, però millor utilitzar el procediment de retallar/enganxar: 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 teclat [Ctrl]+[C]).



Imatge 25. Menú contextual de dades

5. Amb el ratolí, fes clic a la casella corresponent al Torn del segon individu, i MANTENIN 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. Clic a l'opció "Paste" per a enganxar-hi el porta-papers. Ara tots ens 42 primers individus consta com que són del torn de tarda.
7. Ves al full de càlcul prèviament obert amb les dades de l'arxiu extern "P3_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*. Fes clic a 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 (clic casella + “Copy” + seleccionar múltiples caseles + “Paste”) per acabar d’introduir ràpidament les dades dels 98 individus.



Imatge 26. Retalla/Enganxa dades matriu

Si ara obres la finestra d’identificació de variables, amb Menú / Data / Setup (o doble-clic a encapçalament d’una columna) pots comprovar com la variable nominal té dos categories.

Com s’ha pogut comprovar, 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 de errors de teclat. Compte, que per a *jamovi* “Matí” és un valor (“Level”) totalment diferent de “matí” o de “Mati”.

Carrega de les dades d’arxius.

En ocasions interessa analitzar amb *jamovi* dades que es troben emmagatzemades en format digital a arxius en diferents 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, be siguin fulls de càlcul genèrics o aplicacions d’anàlisi (p.e., 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 és una variable (atribut o resposta dels individus).

	CUES	CCAA	PROV	TAMUNI	g_edat	p3	p5	p6	p901	p902	p903
1	1199	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	Si	Varias veces ...	19	Si	Si	Si
2	1200	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	Si	3-5 por sema...	12	No	No	Si
3	1201	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	Si	Varias veces ...	16	No	Si	Si
4	1202	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	Si	Una vez al dia	17	No	No	Si
5	1203	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	Si	Una vez al dia	18	No	No	Si
6	1204	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	Si	Varias veces ...	19	Si	No	Si
7	1205	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	No					
8	1206	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	Si	3-5 por sema...	25	No	No	No
9	1207	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	No					
10	1208	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	Si	Varias veces ...	14	No	No	Si
11	1209	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	Si	Una vez al dia	22	No	No	Si
12	1210	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	Si	Varias veces ...	12	No	No	Si
13	1211	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	No					
14	1212	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	Si	Varias veces ...	14	No	No	No
15	1213	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	No					
16	1214	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	Si	Varias veces ...	21	No	No	Si
17	1215	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	Si	Varias veces ...	16	No	No	Si
18	1216	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	Si	Una vez al dia	14	No	No	Si
19	1217	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	Si	3-5 por sema...	17	No	No	Si
20	1218	Andalusia	Sevilla	400.001 a 1.0...	25 29 anys	Si	Varias veces ...	13	No	No	Si
21	1219	Andalusia	Sevilla	400.001 a 1.0...	15 19 anys	Si	Varias veces ...	12	No	No	Si
22	1220	Andalusia	Sevilla	400.001 a 1.0...	25 29 anys	Si	Varias veces ...	17	No	No	Si
23	1221	Andalusia	Sevilla	400.001 a 1.0...	20 24 anys	Si	Varias veces ...	11	No	No	Si
24	1222	Andalusia	Sevilla	400.001 a 1.0...	20 24 anys	Si	Varias veces ...	17	No	No	Si
25	1223	Andalusia	Sevilla	400.001 a 1.0...	15 19 anys	Si	1-2 por sema...	12	No	No	Si
26	1224	Andalusia	Sevilla	400.001 a 1.0...	20 24 anys	Si	Varias veces ...	15	No	No	Si
27	1225	Andalusia	Sevilla	400.001 a 1.0...	20 24 anys	Si	Varias veces ...	13	Si	Si	Si
28	1226	Andalusia	Sevilla	400.001 a 1.0...	25 29 anys	Si	Varias veces ...	18	No	No	Si

Imatge 27. Matriu de dades a *jamovi*

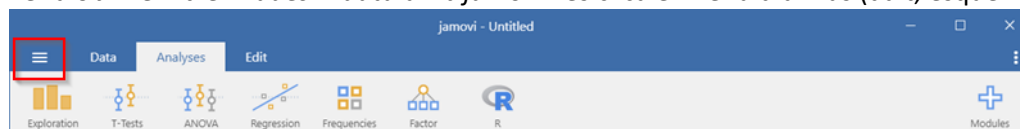
L'opció del menú d'arxius "Import" permet AFEGIR dades noves a les ja existents a l'espai de treball. D'aquesta forma es poden afegir casos o variables per a poder AJUNTAR diferents arxius "parcials" de dades fent servir l'opció d'importació i seleccionant (amb Ctrl+Clic) els diferents arxius que es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

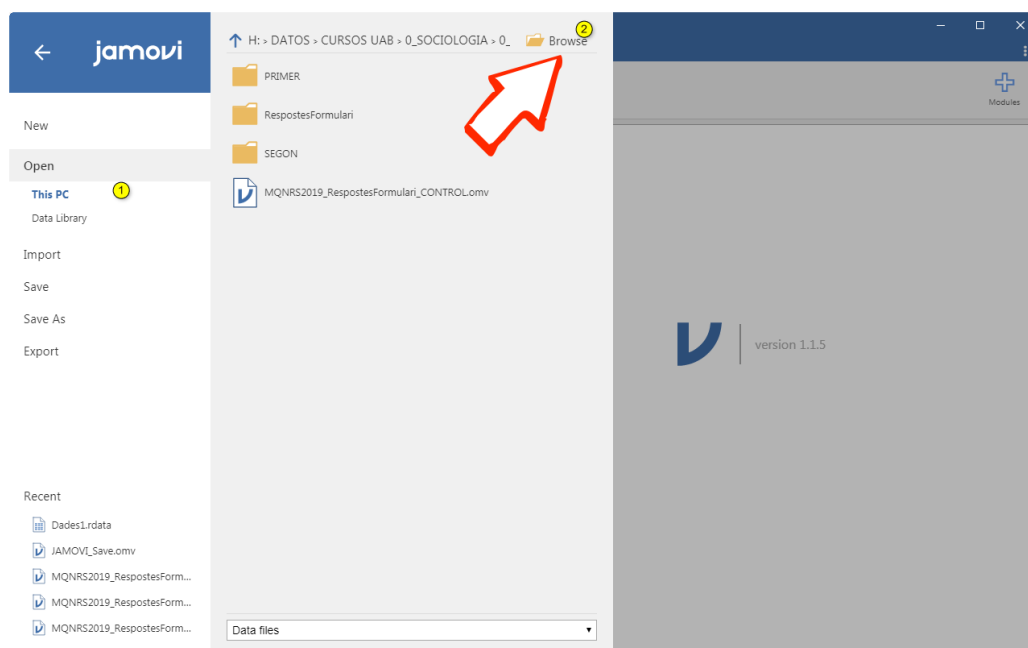
A l'espai docent Campus Virtual (CV) de l'assignatura hi pots trobar l'arxiu "P3_Dades1.rdata", un arxiu de dades corresponents a una enquesta del CIS .

1. Descarrega aquest arxiu format .XLSX 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 *Dades1.rdata* amb *jamovi* Fes clic a el menú d'arxius (dalt, esquerra).



Imatge 28. Ubicació menú d'arxius

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



Imatge 29. Obrir un nou arxiu de dades

6. S'obre una finestra per seleccionar l'arxiu de dades a obrir. Tria l'arxiu "*P3_ades1.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, amb 1419 individus entrevistats i 70 variables.

L'aplicació *jamovi* pot obrir directament arxius de dades en diferents formats, corresponents amb altres aplicacions abastament utilitzades en l'anàlisi estadístic 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 de 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) ⁵



EXERCICI

- A l'espai Moodle corresponent a la PRACTICA 3 pots trobar dos arxius format *jamovi* que contenen dades de la mateixa enquesta del CIS: "*P3_Dades1_A.omv*" i "*P3_Dades1_B.omv*" cadascú amb un conjunt per a diferents províncies.
- Pot ser es va repartir la tasca d'introduir les dades entre dos persones o dos empreses de servei, per accelerar el procediment, i ara cada arxiu té les MATEIXES variables i una part dels casos.
- Descarrega els dos arxius al teu PC local.
- Obre cadascú dels arxius per separat. En quin d'ells es troben les dades de les enquestes realitzades a Barcelona? Quantes entrevistes en total conté aquest arxiu?

⁵ *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 la estadística bayesiana. Es pot descarregar, per a diferents plataformes, a <http://jasp-stats.org>.

- Prova d'unificar totes les dades en un nou i únic arxiu de *jamovi*. Utilitza l'opció "Import" del menú d'arxius, tot seleccionant alhora els dos arxius. Comprova que el total sigui de 1419 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). Pot ser en versions posteriors aquesta restricció desaparegui.

5. Definició de dades amb *jamovi*: Tipus de variables

En tota anàlisi estadística el primer que cal és demanar-se, des del punt de vista del nivell de mesura, de quin tipus de variables es tracta. A efectes estadístics hi ha dos tipus de variables, d'una banda les qualitatives o categòriques i, d'altra les quantitatives o numèriques. Malgrat això, dintre de les qualitatives sovint interessa distingir entre nominals i ordinals.

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

- Existeix un qüestionari realitzat pel CIS, i es vol preparar una matriu de dades de *jamovi* per a poder introduir les dades corresponents a les respostes d'aquest qüestionari.
- Existeix 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 “P3_Qüestionari.pdf” (que reproduïx una mostra d'un qüestionari del CIS amb diverses preguntes), i “P4_Dades1.omv” (que inclou les respostes per a una mostra de n=1419 persones enrevistades).

1. Descarrega tots dos arxius a una carpeta local del teu PC.
2. Obre l'arxiu PDF, per tenir-ho disponible al llarg de tot el tutorial.
3. Obre amb *jamovi* l'arxiu amb la matriu de dades (respostes),
4. Comprova com a en aquest qüestionari del CIS hi trobaràs exemples de variables dels diferents nivells de mesura:
 - Variables **nominals**: Utilització o no d'Internet (p3), sexe (p40), estat civil (p41).
 - Variables **ordinals**: Freqüència 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)⁶.

Encara que aquests són els 3 tipus més habituals en anàlisi de dades socials, no és la terminologia que utilitza *jamovi*, que distingeix entre dotze classes 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

Segons el tipus de la variable:

- **“Continuous”**: Per a variables quantitatives (numèriques) que disposen d'una unitat d'escala, com ara pot ser número 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 pot ser nivell educatiu assolit o opinió respecte a un tema (D'acord, Neutral, En desacord).
- **“Nominal”**: Per a variables nominals que permeten classificar tots els casos en grups mútuament excloents, com ara pot ser religió professada o la participació en un esdeveniment (Sí o No).

⁶ Les escales d'opinió o d'auto-posicionament són en realitat variables de tipus ordinal, però hi ha un ampli consens (poc fonamentat) en considerar-les com a numèriques o quantitatives a efectes estadístics si tenen més de 7 opcions de resposta.

- “ID”: Per a identificar de forma única a 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.

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.

Segons el tipus **de les dades**:

- “Integer”: Dades numèriques sense decimals, o variables ordinals identificades per un factor numèric però amb una “etiqueta” (p.e., intervals).
- “Decimal”: Només per a variables contínues, tractades com a números amb decimals.⁷
- “Text”: Només per a nominals o ordinals, tractades com a dades textuales, cadenes de caràcters alfabètics i/o numèrics.

En funció del tipus de dades *jamovi* utilitza determinats formats o altres per a la seva forma interna d'emmagatzemar les dades i en la generació dels resultats dels processos estadístics. Resulta rellevant en algunes anàlisis, perquè permet el tractament “dual” d'una mateixa variable (nominal o ordinals) be com a factor d'agrupació o be com a quantitatives.

En referència a les variables corresponents a preguntes del qüestionari i que anteriorment han estat presentades com a exemples:

- Utilització o no d'Internet (p3),
Sexe (p40) i
Estat civil (p41)
... cal que siguin definides com a “Nominal” & “Text”.
- Freqüència utilització d'Internet (p5),
Probabilitat de quedar amb algú que hagi conegut per Internet (p16) i
Nivell d'estudis del pare (p54)
... cal que siguin definides com a “Ordinal” & “Text”.
- Edat que tenia en començar a utilitzar Internet (P6),
Nombre de contactes a les xarxes socials (p15) i
Escala ideològica (p33)
... cal que siguin definides com a “Continuous” & “Integer”.

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** resulta en la forma més adient de definir les variables.

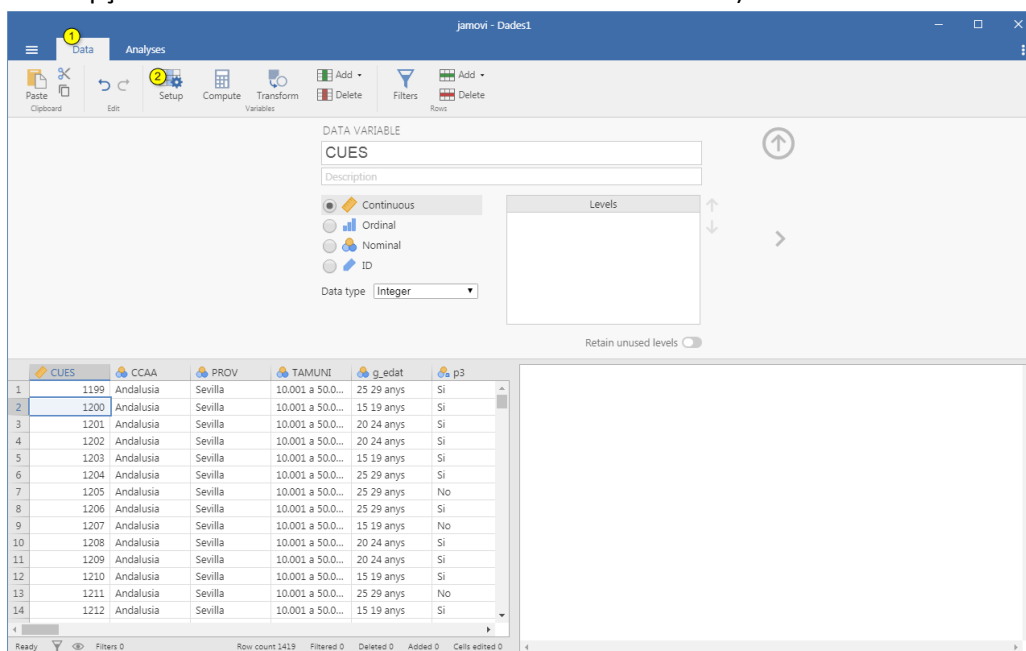
Sempre cal revisar la definició de cadascuna de les variables quan s'obre (o s'importa) amb *jamovi* un arxiu de dades que no està en el seu propi format (.omv).

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

1. Obre dins *jamovi* l'arxiu de dades R anomenat “P4_Dades1.rdata”.

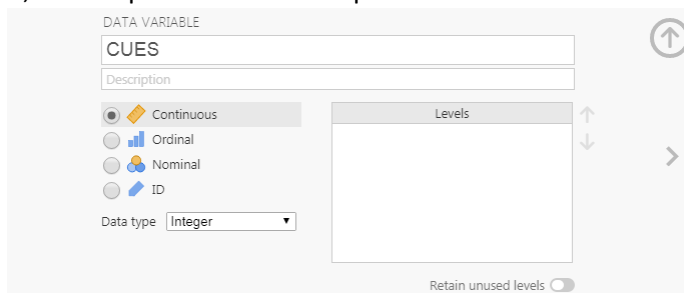
⁷ Els nombres decimals a *jamovi* utilitzen un PUNT (no una coma) per a separar la part sencera de la decimal, seguint l'estàndard de notació anglosaxó.

- Per accedir al panell de definició de les variables, cal clicar a la pestanya “Data” del menú superior, i a continuació clic a la icona “Setup”. Comprova com clicar a aquesta icona serveix per a obrir o tancar aquest panel.
(Una drecera per aconseguir el mateix efecte és fer doble-clic al nom de la variable que encapçala cada una de les columnes de la matriu de dades.)



Imatge 30. Accés a l'àrea d'identificació de variables

- 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.
- Es pot passar seqüencialment d'una variable a la següent o anterior fent clic a les fletxes > < que es mostren a ambdós laterals del panel.
- En aquest cas, obre el panell de definició per a la variable anomenada “CUES”.



Imatge 31. Identificació automàtica (Open) de variable CUES

Com pots veure, *jamovi* ha identificat automàticament la “CUES” com a variable numèrica a partir de les dades (números 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:

- Clic sobre ID, per canviar el tipus de variable,
- No cal canviar el tipus de dades, “integer” ja està bé perquè es tracta de números sencers,
- 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, “Número consecutiu que identifica a cada cuestionario”.

DATA VARIABLE

CUES

Número consecutivo que identifica a cada cuestionario

Measure type ID

Data type Text

Missing values

Levels

Retain unused levels ☐

Imatge 32. Correcta identificació de variable CUES

En el cas de les següents dos variables: CCAA i PROV, totes dues han estat definides automàticament com de tipus “Nominal”. Això és correcte, però el tipus de dades ha estat establert com a “Integer” de forma **incorrecta** perquè totes les dades són cadenes textuales o literals. Per tant, per a definir correctament d’aquesta variable fes el següent:

1. No cal canviar el tipus de variable,
2. Clic sobre el selector desplegable de “Data type”, i canvia el tipus a “text” perquè es tracta de caràcters alfabètics,
3. A la descripció (sota del nom de la variable) pot ser útil afegir respectivament “Comunitat Autònoma de residència” i “Província de residència”.

En el cas de la següent variable de la matriu de dades TAMUNI, ha estat definida automàticament com de tipus “Nominal”. Això és **incorrecte** perquè es tracta d’una variable “Ordinal” amb 7 categories que representen intervals de població del municipi perfectament ordenables de menor a major, per tant cal canviar el tipus de variable. Igualment cal canviar el tipus de dades, que ha estat establert com a “Integer” de forma incorrecta: totes les dades són cadenes textuales o literals que defineixen o anomenen l’interval de població corresponent. Per tant, per a definir correctament d’aquesta variable fes el següent:

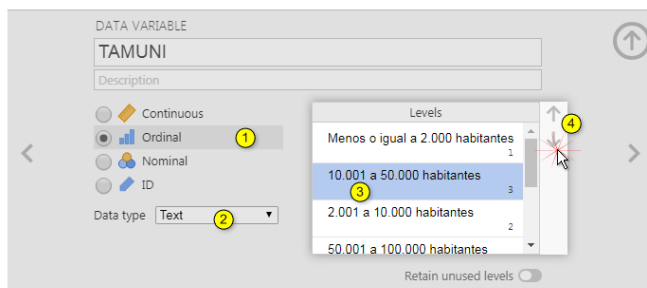
1. Clic sobre Ordinal, per canviar el tipus de variable,
2. Clic sobre el selector desplegable de “Data type”, i canvia el tipus a “text” perquè es tracta de caràcters alfabètics,⁸
3. A la descripció (sota del nom de la variable) pot ser útil afegir “Tamaño del municipio”, i
4. En el cas concret de les 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 la ordenació de les diferents categories sigui incorrecta. Sovint succeeix que quan es fa una importació de dades *jamovi* ordena les categories alfabèticament i no segons la jerarquia intrínseca de la variable.

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

- Fes clic a sobre de cadascuna de les categories, i quan queda seleccionada, fer servir les fletxes que apareixen a la part dreta de la finestra de categories per completar la ordenació cap amunt o avall de la jerarquia.

⁸ Tot i que representen intervals numèrics, les “etiquetes” de cada categoria són en realitat textuales.



Imatge 33. 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"*).



EXERCICI

1. Repassa **TOTES** les variables de l'arxiu "P3_Dades1" per comprovar que estan correctament definides a *jamovi*.
2. Pots ajudar-te de l'arxiu "Questionari.pdf" per a aquesta tasca.
3. Igualment, el qüestionari et servirà per a introduir descripcions per a cada variable, basades en les preguntes literals del mateix qüestionari.
4. 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* ("P3_Dades1.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* ("P3_Dades1.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 “visualitzar” de forma simple y sintètica com es distribueixen els valors de 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 poden ser calculats 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** o la **mediana** i la **desviació típica**.

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

- Existeix un qüestionari realitzat pel CIS, i es vol preparar una matriu de dades de *jamovi* per a poder introduir les dades corresponents a les respostes d’aquest qüestionari.
- Existeix 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.
- Es vol elaborar un informe amb una descripció exploratòria de variables d’aquesta enquesta.

Com a resultat de la teva tasca pràctica anterior (si has realitzat aquest tutorial pas a pas) disposes ara d’un arxiu, amb TOTES les variables correctament identificades, amb les dades d’un estudi concret del CIS que has anomenat “P3_Dades1.omv”. Pots utilitzar aquest arxiu teu pel seguiment del tutorial a partir d’ara.

Si no has realitzat encara la pràctica anterior (identificació de variables) però vols continuar -immediatament- amb el tutorial, pots obrir l’arxiu “P4_Dades1.omv” disponible a l’espai CV de l’assignatura.

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

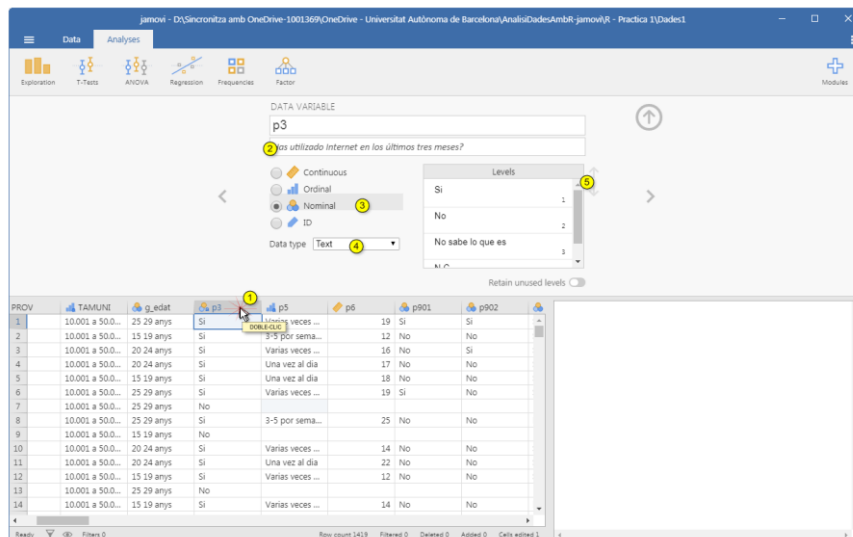
A *jamovi* existeix un procediment per a seleccionar un conjunt de variables i demanar la generació d’aquests indicadors de distribució descriptiva que es considera Estadística Descriptiva.

Per fer-ho, per exemple, amb les següents variables:

- p3 (“has fet servir Internet en els darrers tres mesos” NOMINAL/TEXTO),
- p5 (“en cas afirmatiu, amb quina freqüència utilitzes Internet”, ORDINAL/TEXTO), i
- p33 (“Auto-posicionament polític en una escala esquerra (1) a dreta (10)”, ORDINAL/TEXTO)

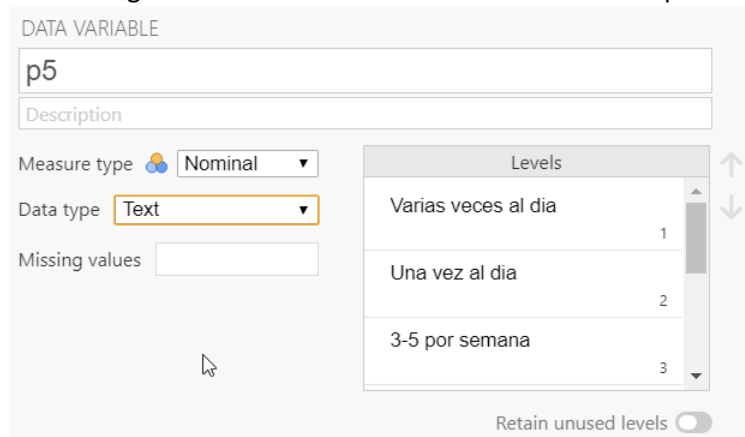
...segueix els passos indicats a continuació:

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.



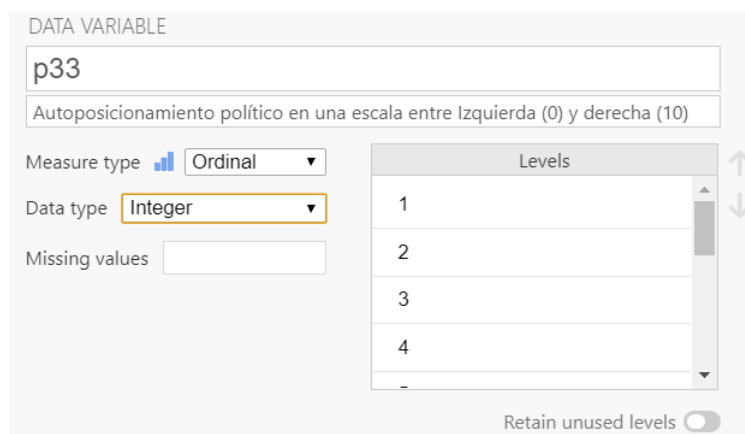
Imatge 34. Identificació correcta de variable P3

2. Comprova que la variable p5 està ben definida en *jamovi*: de tipus ORDINAL / TEXT, i amb les categories ("levels") de la variable en l'ordre adient. Si cal modificar l'ordre, cal clicar a cada etiqueta de categoria i utilitzar les fletxes amunt i avall corresponents.



Imatge 35 Identificació correcta de variable P5

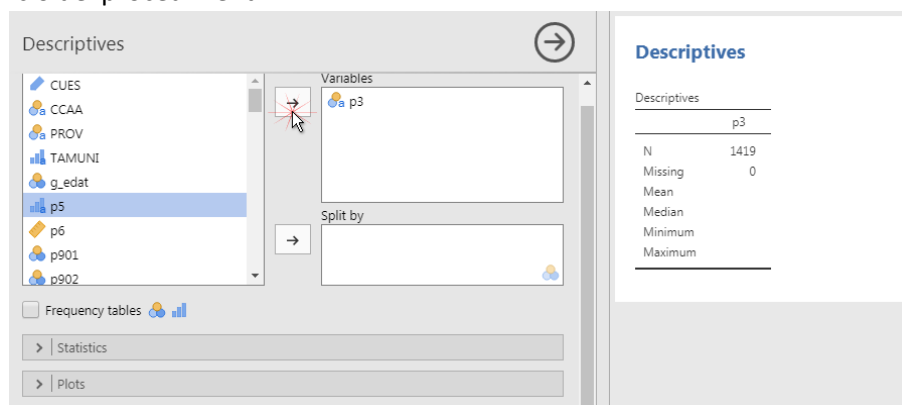
3. Comprova que la variable p33 està ben definida en *jamovi*. Fes doble-clic a sobre del nom que encapçala la seva columna. S'obre el panell de definició, que hauria d'especificar que és de tipus ORDINAL / INTEGER.



Imatge 36 Identificació correcta de variable P33

RECORDA ! Es molt important que totes les variables estiguin ben definides en quant a la seu nivell de mesura, 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.

4. Ja pots tancar la finestra d'identificació de variables, amb la fletxa-amunt-encerclada.
5. Si no ho has fet abans, fes ara clic a la etiqueta "Analyses" del menú superior de *jamovi*.
6. Selecciona l'opció "Exploration", a l'extrem esquerra del sub-menú de procediments estadístics.
7. Al menú desplegable, selecciona l'opció "Descriptives". S'obre un panell central de definició del procediment.



Imatge 37. panell típic de definició de procediments estadístics

8. 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.
9. Fes el mateix amb les variables p5 i p33.
10. 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).
11. Per defecte aquest procediment només genera 6 resultats: recompte de casos totals (N), recompte de casos perduts (Missing), mitjana aritmètica (Mean), mediana (Median), valor mínim i valor màxim de la distribució. Més endavant veuràs com afegir o treure d'altres resultats.

12. 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.
13. Si vols amagar el panell central, pots clicar a la FLETXA-EN-CERCLE de la part superior dreta.
14. Si vols visualitzar el panell central de definició d’un procediment, fes doble-clic a sobre de qualsevol resultat al panell dret de resultats.

Quina és la mitjana de la variable p33? Te cap sentit això? Perquè?

Recorda que la variable recull l’auto-posicionament polític de les persones enquestades a una escala esquerra-dreta, que va de 1 a 10. A properes seccions del tutorial veuràs com solucionar aquest important problema.



EXERCICI

1. Repassa les variables de l’arxiu (definides a *jamovi*) que vas generar en l’apartat anterior fins que trobis algunes (4 o 5) que siguin descriptives de la població espanyola i que facin referència a aspectes que t’interessin.
2. Pots ajudar-te de l’arxiu “Questionari.pdf” per a aquesta tasca, si és que els noms de les variables no aporten prou informació.
3. 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 ?
4. Desa l’arxiu actualitzat en format propi de *jamovi* (.omv), posant-li un nom que t’ajudi a identificar-ho (p.e., “P4_DadesCIS_Miguel.omv”).
5. Tanca totalment *jamovi*.
6. Torna a obrir-lo amb l’arxiu que acabes de desar. Comprova com s’han conservat les dades JUNTAMENT amb les taules que has generat.

En resum, per a generar amb *jamovi*:

- Una taula amb **freqüències absolutes** (f_i), només variables nominals o ordinals.
- Una taula amb freqüències **relatives percentuals** (p_i), només variables nominals o ordinals.
- Una taula amb freqüències **percentuals acumulades** (P_i), només variables nominals o ordinals.
...fes clic a Menú / Analyses / Exploration / Descriptives, i activa l’opció “Frequency tables”
- La **moda** (Mo), de tot tipus de variables.
...fes clic a Menú / Analyses / Exploration / Descriptives, desplega la secció “Statistics” (clic) i activa l’opció “Mode”
- La **mediana** (Me , Md , Q_2), només variables ordinals o continues.
...fes clic a Menú / Analyses / Exploration / Descriptives, desplega la secció “Statistics” (clic) i activa l’opció “Median” (està per defecte).
- La **mitjana** ($\bar{x}$), només variables continues.

...fes clic a Menú / Analyses / Exploration / Descriptives, desplega la secció “Statistics” (clic) i activa l'opció “Mean” (està per defecte).

- Els **quartils** (Q_1 , Q_2 , Q_3), només variables ordinals o contínues.
...fes clic a Menú / Analyses / Exploration / Descriptives, desplega la secció “Statistics” (clic) i activa l'opció “Quartiles”.
- Els **tercils** (P_{33} , P_{66}), només variables ordinals o contínues.
...fes clic a Menú / Analyses / Exploration / Descriptives, desplega la secció “Statistics” (clic) i activa l'opció “Cut points for 3 equal groups”.
- La **mitjana** ($\bar{x}$), només variables contínues.
...fes clic a Menú / Analyses / Exploration / Descriptives, desplega la secció “Statistics” (clic) i activa l'opció “Mean” (està per defecte).

Exploració de variables amb codi R.

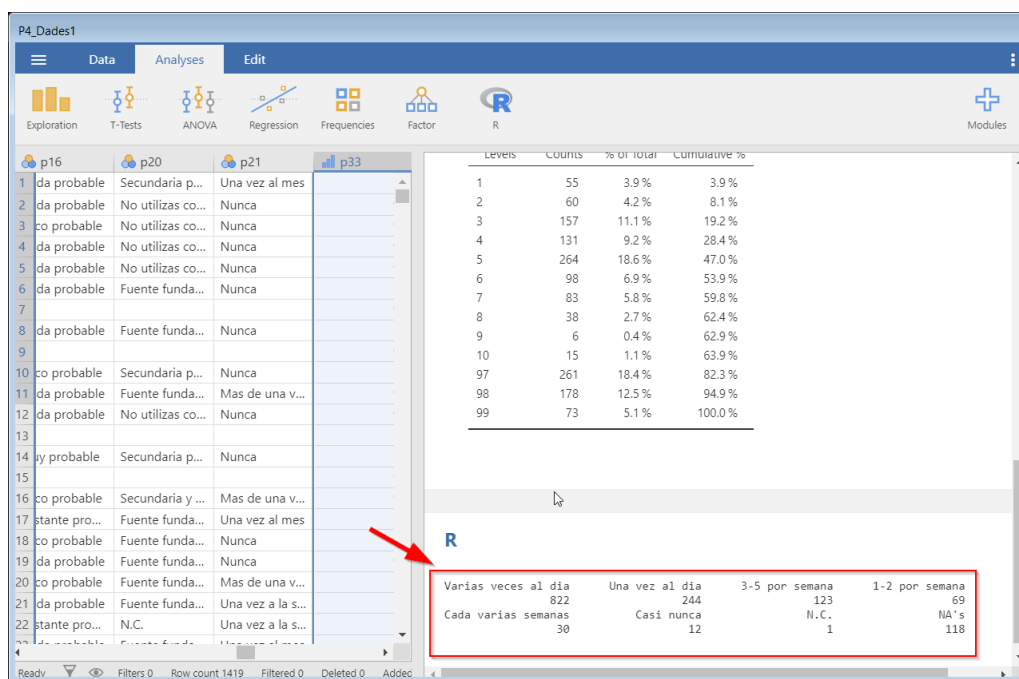
Com a alternativa a l'ús dels menús de *jamovi*, es pot fer servir directament el llenguatge de programació orientat a l'anàlisi de dades “R” per a explorar la distribució de variables de la matriu de dades. Es tracta, en aquest cas, d'introduir una ordre R molt simple (“*summary(data)*”) dins del sistema perquè *jamovi* la executi. Per fer això cal:

1. afegir a *jamovi* la capacitat addicional, que no està implementada per defecte al sistema, de executar directament ordres R mitjançant l'anomenada “consola”,
2. escriure la ordre R a la “consola”, i
3. executar la consola R per generar el resultats.

Per a instal·lar el mòdul o extensió (Rj) que permet la execució, des de *jamovi*, de codi R cal disposar de connexió a Internet, i seguir els passos indicats 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 ho 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 els següents passos:

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])
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 ordre R:
summary(data\$p5)
Com has pogut veure, *jamovi* t'ajuda proposant ordres i variables disponibles.
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 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. 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 38. Exemple resultat d'execució codi R

- Per tancar la consola R cal clicar a la fletxa-encercada 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 la 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 a les funcions que actualment pot desenvolupar *jamovi* per la via dels seus menús i les seves extensions. Per exemple, utilitzant la ordre de 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.

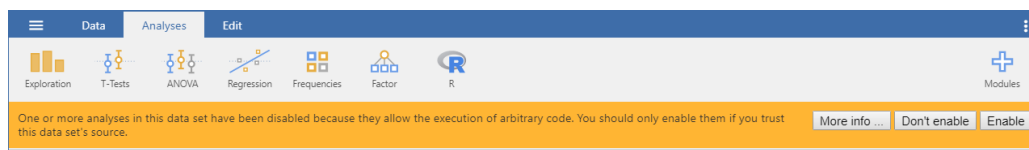
Generar taules de contingència amb codi R

Hi ha diverses ordres del codi R que permeten generar taules de freqüències bàsiques, o recompte dels casos per a cada categoria de les variables. 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*.

El següent apartat del tutorial t'acompanyarà en la utilització pràctica d'ordres R que són directament accessibles.

A l'espai docent Campus Virtual (CV) de l'assignatura hi pots trobar els arxius "P3_Qüestionari.pdf" (que reproduïx una mostra d'un qüestionari del CIS amb diverses preguntes), i "P4_Dades1.omv" (que inclou les respostes per a una mostra de n=1419 persones entrevistades).

1. Si és que no els tens actualment oberts a *jamovi*, descarrega ara de l'espai docent Campus Virtual els arxius "P3_Qüestionari.pdf" i "P4_Dades1.omv" a una carpeta local del teu PC.
2. Comprova que la instal·lació de *jamovi* que estàs utilitzant té instal·lat el mòdul d'extensió "Rj – Editor". Si no és així, recorda instal·lar-ho abans de continuar.
3. Obre tots dos arxius. Si reps un missatge sobre si vols REALMENT obrir l'arxiu OMV (color taronja, a la part superior) confirma-ho fent clic a el botó de l'extrem dret [Enable].



Imatge 39. Confirmació d'obertura d'arxiu OMV amb codi R

4. Accedeix a la consola de R, amb Menú / Analyses / R / Rj Editor... Apareix una consola amb una ordre per defecte, que està desactivada (comença amb #) i que pots esborrar per a escriure el teu codi.
5. Per a generar una taula de freqüències simple de la variable NOMINAL "Has utilizado Internet en los últimos tres meses?" (P3), escriu el següent codi a la consola:
poc a poc, la teva ordre R:
table(data\$p3)
6. 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".
7. Executa el codi fent clic a el petit triangle verd de la part superior dreta de la consola, o amb la combinació de teclat [Ctrl] + [Majúscules] + [Intro].
8. Comprova com a la finestra de resultats s'ha generat un recompte de les categories de la variable:

Si	No
1301	118
9. Aquesta taula –bàsica- no disposa ni de títol ni d'altres càlculs com ara freq. relatives o acumulats.



EXERCICI

- Genera, amb codi R, una taula de freqüències simple de la variable ORDINAL "Auto-posicionamiento político en una escala de izquierda (1) a derecha (10)" (P33).
 - Quin problema hi ha amb aquesta taula ? Què són els valors "fora" del rang de la escala ?
1. Per a generar una taula de freqüències simple de la variable ORDINAL P33, però **excloent** del recompte un conjunt de valors, escriu i executa el següent codi a la consola:
table(data\$p33, exclude = c(97, 98, 99))
 2. Observa com pots incloure o excloure, segons et convingui, els valors perduts d'una variable.
 3. Per a generar una taula de freqüències simple de la variable CONTINUA P15 excloent els valors perduts, escriu i executa el següent codi a la consola:
ftable(data\$p15, exclude = c(9998, 9999))
 4. Observa la diferencia entre executar el codi ftable() o table().
Quina de les dues disposicions de la taula és la millor per a exportar-la cap a altres aplicacions?

Ampliant aquest codi R es pot variar el contingut i disposició de la taula. Observa a continuació l'efecte d'aplicar noves funcions al codi anterior, que consisteix en incloure entre parèntesi resultats anteriors.

1. Per a generar una taula de freqüències EN VERTICAL de la variable d'autoposicionament ideològic (P33), excloent valors perduts (fora del rang 0-10), escriu i executa el següent codi a la consola:
`cbind(table(data$p33, exclude = c(97, 98, 99)))`
2. Observa l'efecte d'incloure el codi que genera com a resultat una taula DINS de la funció `cbind()`. Aquesta funció s'utilitza per a canviar la orientació dels resultats generats.
3. Per a generar una taula EN VERTICAL de la variable P3 que contingui freqüències RELATIVES, escriu i executa el següent codi a la consola:
`cbind(prop.table(table(data$p3)))`
4. Observa l'efecte d'incloure el codi que genera la taula **`table(data$p3)`** dins de la funció **`prop.table()`** que és inclosa dins de la funció **`cbind()`**. El resultat final es: la taula de freqüències "original", convertida a freqüències relatives, i tot convertit a orientació vertical.
5. Per a generar una taula EN VERTICAL de la variable P3 que contingui freqüències ABSOLUTES i també freqüències RELATIVES, escriu i executa el següent codi a la consola:
`cbind(table(data$p3), prop.table(table(data$p3)))`
6. Observa l'efecte de cadascun dels elements del codi. Aquesta ordre significa en llenguatge R: *"Fes una taula vertical amb una primera columna que sigui una taula de freq. absolutes de la variable P3 del conjunt de dades (data) i una segona columna que sigui una versió freq. relatives de la taula de freq. absolutes de la variable P3 del conjunt de dades (data)".*

Una de les més útils potencialitats del llenguatge R és poder "posar nom" a qualsevol dels resultats obtinguts, de forma que es poden re-utilitzar per a posteriors operacions. Per exemple, el codi `7 + 2` genera com a resultat 7, i el codi `suma <- 7 + 2` anomena "suma" al valor final d'aquesta operació, per tant el codi `suma + 1` dona com a resultat 10. Per anomenar qualsevol resultat a R cal indicar el nom desitjat, seguit de `<-` seguit de la operació corresponent.

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

1. Per a generar una taula de freqüències de la variable P3 més completa, escriu el següent codi a la consola, a línies diferents i sense executar-ho fins al final:
`dades = data$p33`
`n_i <- table(dades)`
`f_i <- prop.table(n_i)`
`F_i <- cumsum(f_i)`
`cat("Taula de freqüències: variable p33")`
`cbind(n_i, f_i, F_i)`
2. Observa com les primeres línies de codi van generant i anomenant objectes, i com s'utilitzen els objectes per a posteriors operacions.
3. 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.
4. Aquesta seqüència d'ordres significa, en llenguatge R:
*"A partir d'ara anomenarem **`dades`** a la columna de valors de la variable p33 de la matriu de dades de jamovi", i*

“Anomenarem **n_i** a la taula de freqüències absolutes d’allò anomenat **dades**”, i
 “Anomenarem **f_i** a la taula de freqüències relatives d’allò anomenat **n_i** (es a dir, l’anterior taula de freqüències absolutes)”, i
 “Anomenarem **F_i** a la taula de freqüències acumulades d’allò anomenat **f_i** (es a dir, l’anterior taula de freqüències relatives)”, i
 “Escriu a la finestra de resultats el text següent: Taula de freqüències: variable p33”, i
 “Mostra, amb orientació vertical, les columnes de les taules que s’anomenen **n_i**, **f_i**, **F_i**”.

5. Executa el codi. S’ha generat la següent taula a la finestra de resultats.

Taula de freqüències: variable p33

	n_i	f_i	F_i
1	55	0.038760	0.03876
2	60	0.042283	0.08104
3	157	0.110641	0.19168
4	131	0.092319	0.28400
5	264	0.186047	0.47005
6	98	0.069063	0.53911
7	83	0.058492	0.59760
8	38	0.026779	0.62438
9	6	0.004228	0.62861
10	15	0.010571	0.63918
97	261	0.183932	0.82311
98	178	0.125440	0.94856
99	73	0.051445	1.00000



EXERCICI

- Prova ara de generar la taula de freqüències de p33, però amb el sistema de menús de *jamovi*.
- Quines columnes NO es poden generar amb el sistema de menús ?

Estadística bàsica descriptiva amb codi R

Tots els estadístics bàsics es poden demanar amb una ordre de codi R, per exemple:

- La **mitjana** ($\bar{x}$), només variables contínues.
 ...fes clic a Menú / Analyses / R / Rj Editor, escriu a la consola el codi R següent:
mean(as.numeric(data\$p33, na.rm = TRUE))
 i executa-ho (clic al petit triangle verd superior dreta)
- Aquesta ordre vol dir: “Demana a R que calculi la mitjana aritmètica, de la variable anomenada p33 pertanyent al la matriu de dades de *jamovi*, però considerant-la com si fos variable numèrica, i sense tenir en compte els valors perduts”.
 ...mira d’identificar els diferents elements al codi R.
- Es molt important incloure la part del codi **as.numeric()**, perquè aquesta variable està actualment definida a *jamovi* com de tipus ORDINAL i per tant és impossible calcular la seva mitjana. Afegir aquesta porció de codi assegura que es pugui calcular la mitjana.
- Prova d’executar una copia de la ordre sense modificar el tipus de la variable, i comprova què fa R quan se li demana calcula mitjana de variables que no són numèriques

- Els **quartils** (Q_1 , Q_2 , Q_3), només variables ordinals o contínues.
...fes clic a Menú / Analyses / R / Rj Editor, escriu a la consola el codi R següent:
`quantile(data$p15, na.rm = TRUE)`
i executa-ho (clic al petit triangle verd superior dreta)
- El coeficient de variació (desviació típica dividida per la mitjana), només per a variables contínues.
...fes clic a Menú / Analyses / R / Rj Editor, escriu a la consola el codi R següent:
`sd(data$p15, na.rm = TRUE) / mean(data$p15, na.rm = TRUE)`
i executa-ho (clic al petit triangle verd superior dreta)

Alguns estadístics bàsics NO estan disponibles als menús de *jamovi*, i és imprescindible utilitzar codi R per aconseguir-los com a resultats:

- El **rang inter-quartil** (*IRQ*), només variables ordinals o contínues.
...fes clic a Menú / Analyses / R / Rj Editor, escriu a la consola el codi R següent:
`IRQ(data$p15, na.rm = TRUE)`
i executa-ho (clic al petit triangle verd superior dreta)
- Comprova que el resultat és la diferència entre el 3er. i el 1er. quartil, ja calculats prèviament a altres ordres anteriors del tutorial.

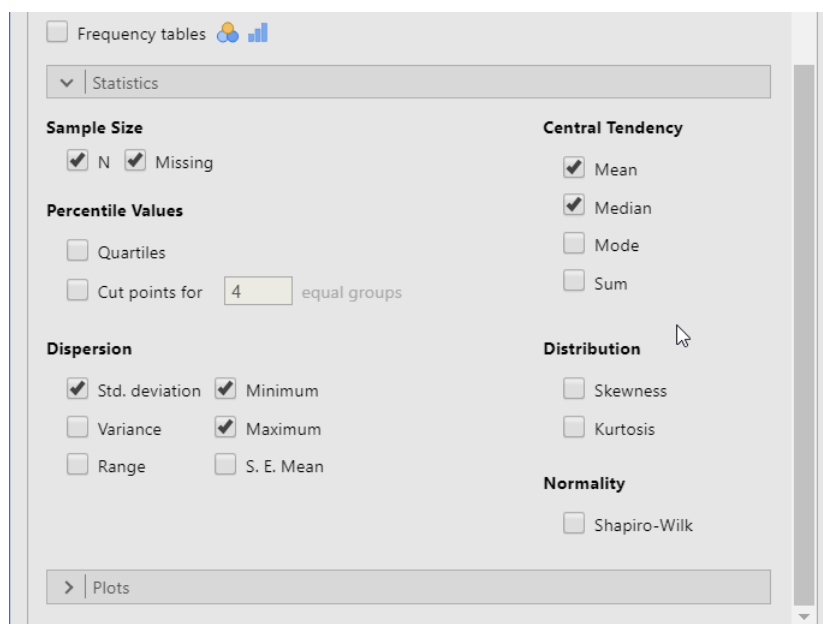
Algunes persones prefereixen treballar amb el sistema de menús (restringit) i d'altres prefereixen escriure codi R i aprofitar tota la potencia. En tot cas *jamovi* permet utilitzar totes dues estratègies, combinar-les, i introduir-se de forma simple a la programació en R.

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 –per motius de seguretat- *jamovi* sempre demana la confirmació per a obrir aquests arxius.

Resultats de la 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 comentaran els resultats generats pel procediment d'Estadística Descriptiva i dos interessants possibilitats de treball a partir d'aquests resultats: afegir comentaris personals, i exportar cap a altres aplicacions.

En relació amb la configuració per defecte del sistema de generació d'anàlisis **estadístics descriptius univariats** per a conjunts de variables cal tenir en compte que :



Imatge 40. Configuració per defecte d'Estadística Descriptiva

- Si vols generar taules de freqüències s'ha d'activar expressament l'opció. **Només** ho fa amb variables de tipus NOMINAL o ORDINAL.
- Sempre s'indica la mitjana i la mediana, però no la moda. Per a variables qualitatives cal desactivar la mitjana i activar la moda, expressament.
- Per a variables de tipus ordinal o quantitatiu, cal activar expressament els Quartils, a l'apartat "Percentile Values". Si vols generar càlcul de decils, cal activar i indicar "Cut points for 10 equal points" (el mateix s'aplica a la generació de centils, tercils, quintils, etc...).
- Sempre s'indica la desviació típica, el màxim i mínim, però no el rang. Per a variables ordinals NO es calcula el "rang inter-quartil" (IQR), que només es pot demanar mitjançant codi R directament executat.
- Per a variables de tipus quantitatiu, es pot activar expressament un indicador d'asimetria ("Skewness") i un altre de Kurtosi, així com un indicador per a comprovar si la distribució dels valors segueixen una forma "normal" (similar a la popular i coneguda campana de Gauss).

La operació menú / Analyses / Exploration / Descriptive genera:

1. Una taula resumen per al conjunt de variables seleccionades, amb els valors possibles dels estadístics configurats, calculats només per a les variables del tipus adient.

Descriptives

	p3	p5	p33
N	1419	1301	1419
Missing	0	118	0
Mean			38.2
Median			6
Standard deviation			44.7
Minimum			1
Maximum			99

Imatge 41. Exemple resultat taula descriptiva

2. Si s'ha activat l'opció, una taula de freqüències per a cada variable NOMINAL o ORDINAL seleccionada.

Frequencies

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 %
Varias veces al dia	822	63.2 %	63.2 %
Una vez al dia	244	18.8 %	81.9 %
3-5 por semana	123	9.5 %	91.4 %
1-2 por semana	69	5.3 %	96.7 %
Cada varias semanas	30	2.3 %	99.0 %
Casi nunca	12	0.9 %	99.9 %
N.C.	1	0.1 %	100.0 %

Imatge 42. Exemple taules de freqüències

Les taules de freqüències es generen ordenades segons les categories ("Levels").

- Si vols canviar l'ordre d'aparició de categories a una taula, cal que re-ordenis les categories a la identificació de la variable.

Les taules de freqüències generen tres columnes: freqüència absoluta ("Counts"), freqüència relativa percentual ("% of Total") i freqüència relativa percentual acumulada ("Cumulative %").

No es pot canviar aquesta configuració bàsica, sense executar codi R.

SOLUCIÓ 1: Què passa si demanes una taula de freqüència d'una variable de tipus numèric? Prova de fer-ho amb la variable P.15 (*Podrias decirme cuántos contactos, amigos/as, personas agregadas, tienes, en total, en tu/s red/es sociales?, amb dos valors perduts: No ho sap 9998 i No contesta 9999*). Amb el menú de *jamovi* aquesta opció és impossible, però es pot fer una copia de la variable (pots veure com, més endavant, l'apartat de transformacions de variables) i fer que aquesta copia sigui de tipus ORDINAL per poder demanar la taula de freqüències amb menús de *jamovi*.

SOLUCIÓ 2: Es pot utilitzar codi R per demanar una taula de freqüències de qualsevol variable, fins i tot les que són de tipus numèric (Continuous) fent servir el codi R per a demanar taules de freqüències (fins i tot, es pot excloure un conjunt de valors que siguin considerats perduts):
`fable(data$p15, exclude = c(9998, 9999))`

Modalitat comentari de text

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 text mitjançant el menú corresponent.

Alternativament, estant a la modalitat ANÀLISI es pot fer clic-esquerre 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. Fes clic el menú / Edit
2. A la finestra de resultats queda destacat en blau els espais on es pot incloure text propi: al inici i final de cada procediment estadístic demanat, i al peu de cada taula generada. Inclou el títols dels procediments, que poden ser modificats lliurement.

3. Fes clic a qualsevol àrea editable. Escriu el text desitjat i dona-li format bàsic amb les diferents opcions del sub-menú d'icones (negreta, cursiva, subratllat, sub/super-index, color de lletra o de fons, justificació, llistes...). L'opció "Heading" genera sub-títols.
4. Busca la taula de freqüències de la variable P3. Comprova que no resulta gaire 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 frequency table on the right. The frequency table is for variable p3, titled "Taula de freqüència de la variable 'has fet servir Internet en els darrers tres mesos' (p3)". The table has columns for Levels, Counts, % of Total, and Cumulative %. The data is as follows:

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

Below the table, a comment is added: "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 43. Exemple de comentaris afegits a resultats estadístics

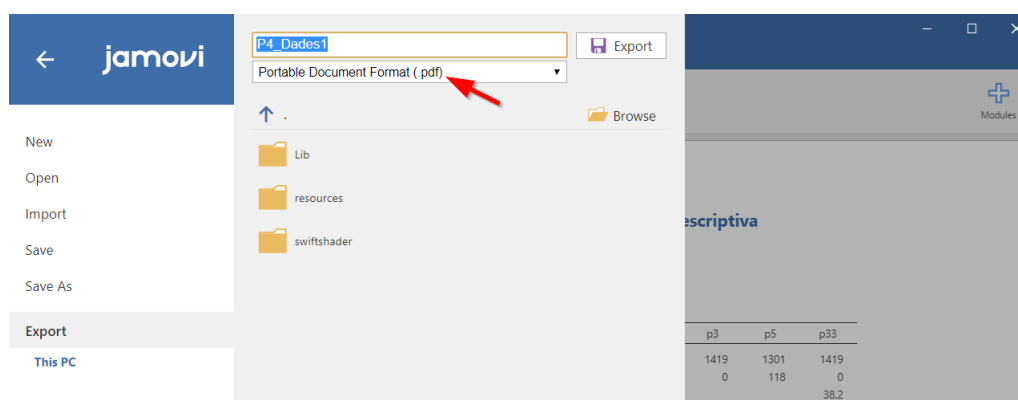
Exportació de resultats cap a altres aplicacions

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: 1) del conjunt de tots resultats, 2) de cada anàlisi per separat, incloent totes les seves taules i gràfiques, o 3) 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.

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

1. Fes clic al menú d'arxius / Export
2. Selecciona, al desplegable corresponent, com a format d'exportació "Portable Documents Format (pdf)". Clic a "Browse" per obrir la finestra que permet triar el nom i la ubicació de l'arxiu exportat.



Imatge 44. 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. Aquest arxiu PDF no és fàcil de modificar una vegada generat (es 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.

Per a exportar cada anàlisi per separat, **incloent totes les seves taules, gràfiques 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 (LibreOfficeWriter, MacOS Pages, M\$-Word,...). Comença a redactar el teu informe, al que 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 "Pegar" (amb clic-dreta + opció Pegar, o alternativament 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 amplies possibilitats per a modificar, editar, seleccionar, eliminar o moure qualsevol part, i integrar-ho dins del redactat general.

Per a exportar **només una taula** en particular:

1. Obre un document al teu processador de text amb el teu informe, al que incorporaràs 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 "Pegar" (amb clic-dreta + opció Pegar, o alternativament amb la combinació de teclat [Ctrl]+[v]).
4. En aquest cas, la taula s'han enganxat al processador de text com a taules, amb amplies 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*.

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 que 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 "Pegar" (amb clic-dreta + opció Pegar, o alternativament amb la combinació de teclat [Ctrl]+[v]).
4. En aquest cas, la imatge s'han enganxat al processador de text (o presentació, o disseny gràfic) com a imatges, 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.

Amb el mateix sentit o objectiu que las estadística descriptiva bàsica, existeixen representacions gràfiques de dades que permeten una exploració de les variables i són 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 estadístic corresponent.

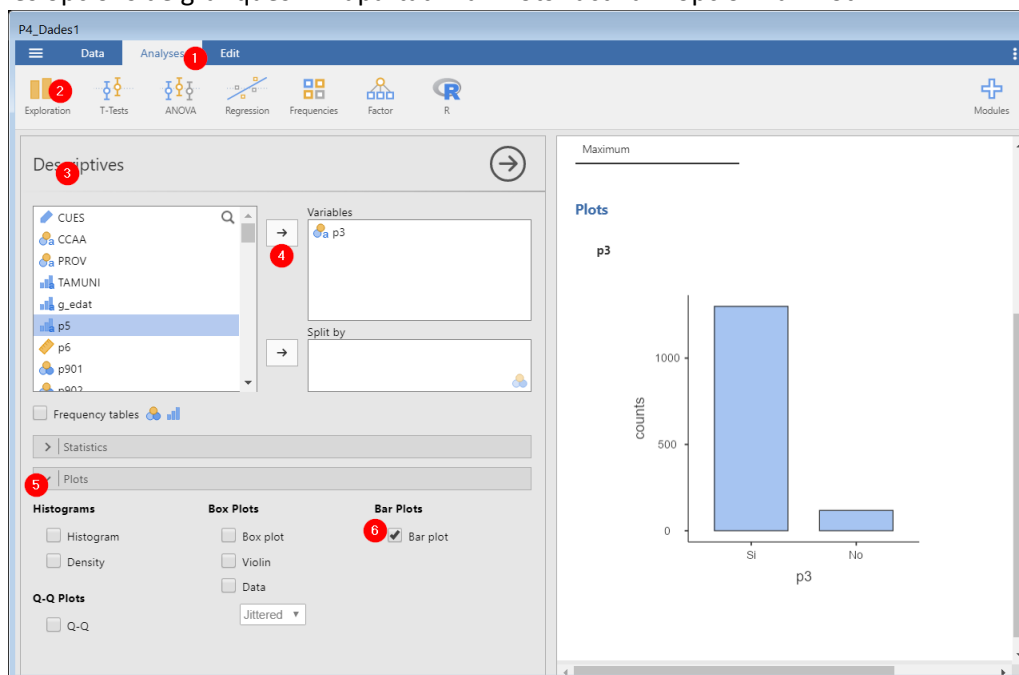
Respecte a la estadística descriptiva bàsica, l'opció per a activar representacions gràfiques està disponible al menú Analyses / Exploration / Descriptives (excepte el "gràfic de Pareto", que forma part d'un mòdul d'expansió que cal instal·lar específicament).

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

- Existeix un qüestionari realitzat pel CIS "P3_Qüestionari.pdf", amb les dades disponibles a "P4_Dades1.omv".
- Es volen afegir representacions gràfiques per a un informe exploratori d'algunes variables d'aquesta enquesta.
- Les variables són de tipus diferents, i per tant, caldrà realitzar la representació més adient per a cadascuna d'elles:
 - P3: "Has utilizado Internet en los últimos tres meses?"
 - g_edat: "Grup d'edat: adolescent (15-19), jove (20-24), adult (25-29)"
 - p33: "Auto-posicionament polític en una escala esquerra (1) a dreta (10)"
 - P15: "Nombre de contactes a xarxes socials"

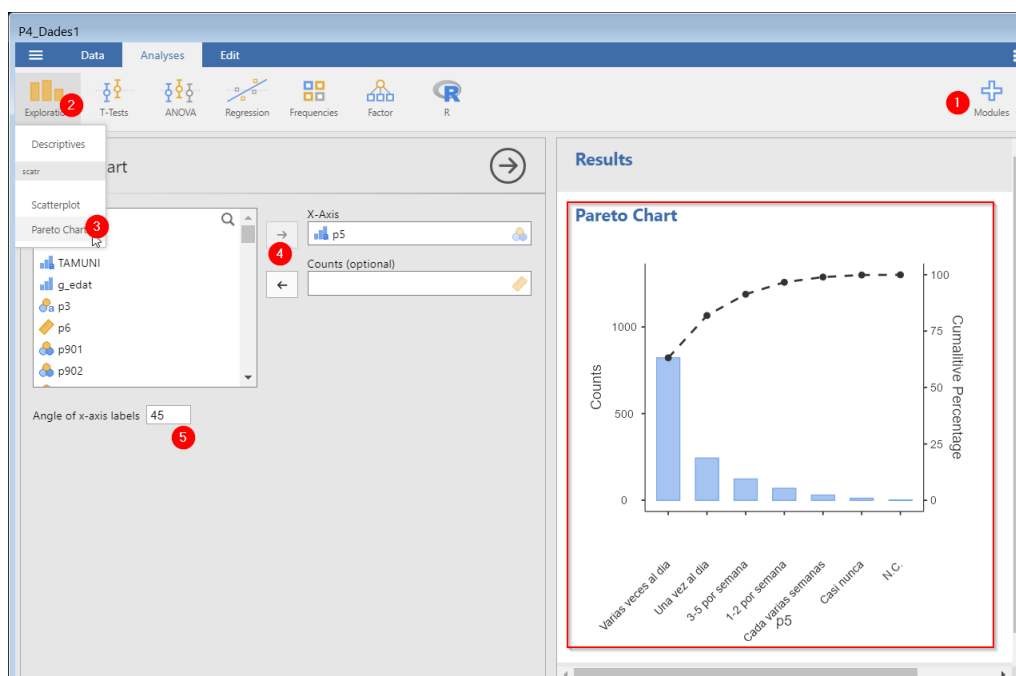
Per a generar els diagrames gràfics adients amb el sistema de menú de *jamovi*, segueix els següents passos:

1. Per a la variable P3 (“*Has utilizado Internet en los últimos tres meses?*”), que és **NOMINAL**, es pot utilitzar Diagrama de Barres, Diagrama de Sectors, i Diagrama de Pareto.
2. El recomanat és el de BARRES, que es pot fer amb el sistema de menús de *jamovi*.
3. Fes clic a Menú / Analyses / Exploration / Descriptives, a la configuració del procediment (pantalla esquerra) busca i selecciona la variable p3, fes clic al botó central per a incloure la variable p3 al conjunt d’anàlisi, i a la part inferior fes clic a la secció “Plots” per a desplegar les opcions de gràfiques. A l’apartat “Bar Plots” activar l’opció “Bar Plot”.



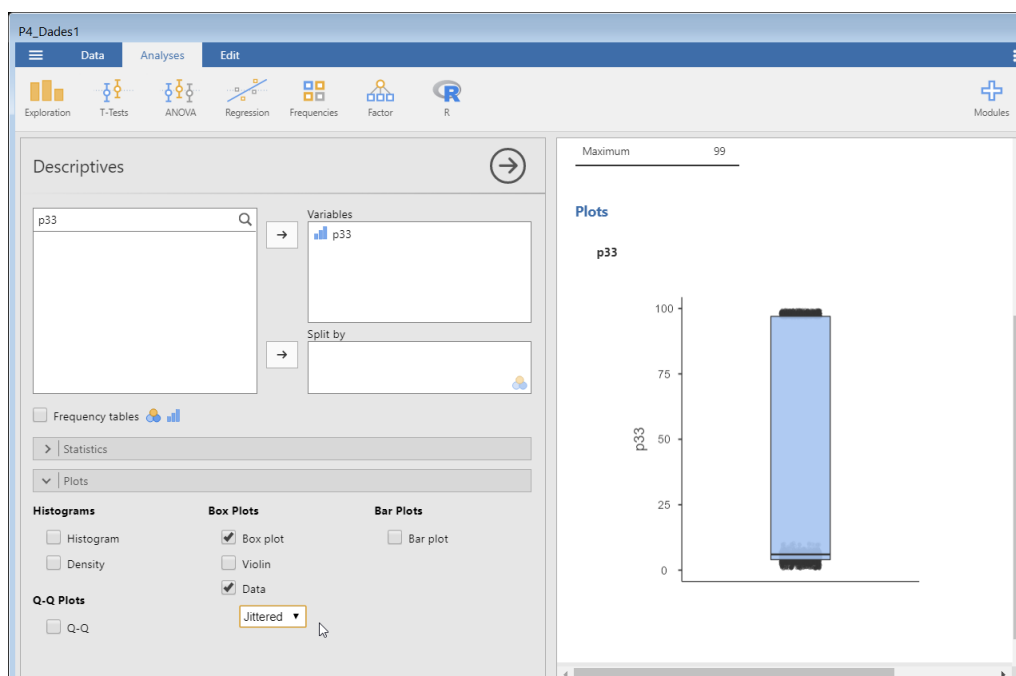
Imatge 45. Generar un diagrama de barres generat amb jamovi

4. Si s’instal·la el mòdul d’extensió anomenat “scatr” es pot generar un gràfic de Pareto per a variables de qualsevol tipus, especialment adient per a les NOMINALS.
5. Fes clic a Menú / Analyses / Exploration / Pareto Chart, a la configuració del procediment (pantalla esquerra) busca i selecciona la variable p5, fes clic al botó central per a incloure la variable al conjunt d’anàlisi “X-Axis”, i a la part inferior fes clic a la secció “Angle of X-Axis labels” per a desplegar l’opció que permet girar 45 graus les etiquetes de l’eix horitzontal per tal de millorar la llegibilitat de la gràfica.



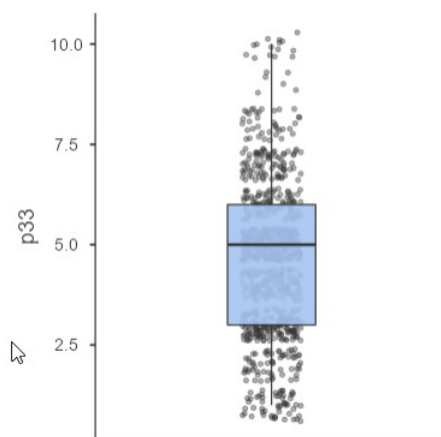
Imatge 46. Exemple de Gràfica de Pareto generada amb *jamovi*

6. Per a la variable *g_edat* ("Grup d'edat: adolescent (15-19), jove (20-24), adult (25-29)"), que és **ORDINAL/TEXT**O, es poden utilitzar tots els diagrames adjacents per a variables nominals.
7. El recomanat és el de CAIXA, però NO es pot fer amb el sistema de menús de *jamovi* amb aquesta identificació de la variable com a ORDINAL/TEXT. Per aquest tipus de variables resulta millor generar diagrama de BARRES, com s'ha vist anteriorment.
8. Per a la variable *p33* ("Auto-posicionament polític en una escala esquerra (1) a dreta (10)"), que és **ORDINAL/SENCER(integer)**, es poden utilitzar tots els diagrames adjacents per a variables nominals, més el Diagrama de Caixa. Si la escala té més de 7 punts, com en aquest cas, s'accepta la representació utilitzant Histograma o Diagrama de densitat.
9. El recomanat és el de CAIXA, que es pot fer amb el sistema de menús de *jamovi*.
10. Fes clic a Menú / Analyses / Exploration / Descriptives, a la configuració del procediment (pantalla esquerra) busca i selecciona la variable *p33*, fes clic al botó central per a incloure la variable al conjunt d'anàlisi, i a la part inferior fes clic a la secció "Plots" per a desplegar les opcions de gràfiques. A l'apartat central "Box Plots" activar les opcions "Box Plot" i també "Data".



Imatge 47. Generar un diagrama de caixa generat amb jamovi

11. Com pots veure, aquesta representació gràfica mostra un problema: les dades (que són els punts representats “a sota” de la caixa) es reparteixen en dos grans grups, els que estan entre 0 i 10, i els que estan a prop de 99. Per a aquesta variable, que és una escala 0-10, el diagrama de caixa permet detectar que hi ha força “valors perduts”. Caldrà depurar-los per a deixar només visibles només entre 0 i 10 (tal com veuràs a una secció posterior del tutorial).
12. Una vegada depurades les dades *jamovi* genera la següent representació gràfica:



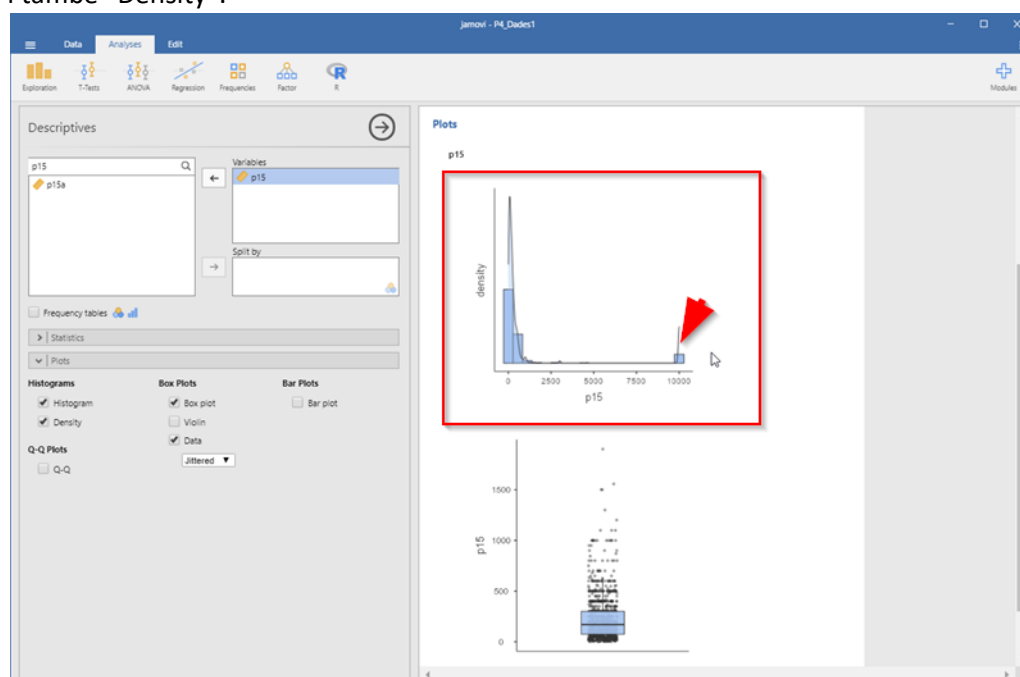
Imatge 48. Exemple de diagrama de caixa

13. Comprova com la mediana, o segon quartil, (Md , Q_2) correspon a una 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.

RECORDA: Els diagrames o gràfiques de caixes són molt útils i representen gràficament la distribució d'una variable ordinal o numèrica 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"). 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.

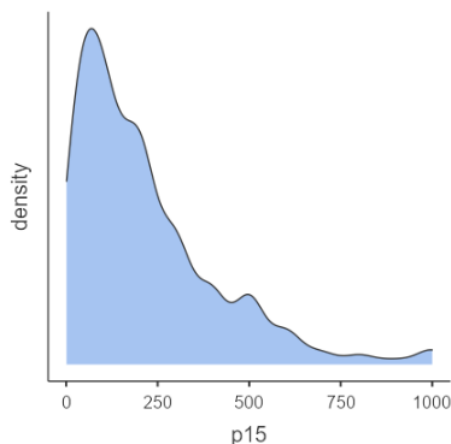
La part central de la distribució, que correspon al 50% central de la població ordenada segons la variable, queda representada per la CAIXA que coincideix amb el rang intercuartil RIQ ($Q_3 - Q_1$) i conté la Mediana, representada per una LÍNIA horitzontal. Els valors atípics s'identifiquen per PUNTS, i s'utilitza 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 $Q_3 + 1,5 \cdot \text{IQR}$, o inferiors a $Q_1 - 1,5 \cdot \text{IQR}$. Aquests límits queden representats per "PATILLES" en forma de línies verticals.

14. Per a la variable P15 ("Nombre de contactes a xarxes socials"), que és **NUMÈRICA** (*Continuous & Integer*) es poden utilitzar tots els diagrames.
15. El recomanat és l'HISTOGRAMA, que es pot fer amb el sistema de menús de *jamovi*, i també és molt útil el de CAIXA.
16. Fes clic a Menú / Analyses / Exploration / Descriptives, a la configuració del procediment (pantalla esquerra) busca i selecciona la variable p15, fes clic al botó central per a incloure la variable al conjunt d'anàlisi, i a la part inferior fes clic a la secció "Plots" per a desplegar les opcions de gràfiques. A l'apartat esquerre "Histograms" activa les opcions "Histogram" i també "Density".



Imatge 49. Exemple d'Histograma generat amb jamovi

17. Comprova com, de nou, la representació gràfica ajuda a descobrir problemes a la distribució (en aquest cas hi ha valors perduts 9998 i 9999, que caldrà depurar).
18. Quan un histograma s'acompanya de la gràfica de densitat, resulta 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 al punt central.



Imatge 50. Exemple de diagrama de densitat

Si el teu sistema *jamovi* te instal·lat el mòdul d'extensió anomenat "Rj – Editor" el Menu / Analyses disposarà d'una opció amb la que podràs accedir a una consola per a escriure i executar 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 d'ells impossibles de generar directament amb els menús de *jamovi*.

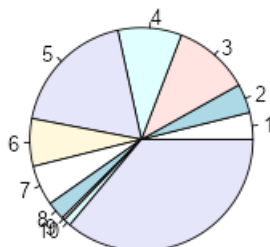
1. Per a generar diagrames de BARRES de la variable d'auto-posicionament ideològic P33, escriu i executa a la consola de R:

```
cat("Barres freq. absolutes")
barplot(table(data$p33, exclude = c(97, 98, 99)))
```
2. Comprova com s'ha adjuntat un títol identificatiu al gràfic, i s'ha exclòs un conjunt de valors considerats perduts.
3. Per a generar diagrames de SECTORS de la variable P3 "Has utilizado Internet en los últimos tres meses?", escriu i executa a la consola de R:

```
pie(table(data$p3), main = "P3. Has utilizado Internet en los últimos tres meses?")
```
4. Comprova com s'ha adjuntat un títol identificatiu al gràfic.
5. Per a generar diagrames de SECTORS de la variable P33 "Autoposició ideològica", escriu i executa a la consola de R:

```
pie(table(data$p33, exclude = c(97, 98, 99)), main = "P33. Autoposició ideològica")
```
6. Compara la capacitat descriptiva (baixa...) d'aquest tipus de gràfic amb el diagrama de barres de la mateixa variable generat anteriorment.

P33. Autoposició ideològica



Imatge 51. Exemple de diagrama de sectors

7. Per a generar diagrama de CAIXA de la variable d'auto-posicionament ideològic P33, sense valors perduts i amb un títol personalitzat, escriu i executa a la consola de R:
`boxplot(data$p33, na.omit=TRUE, main = "P33. Autoposició ideològica")`
8. Comprova com s'ha adjuntat un títol identificatiu al gràfic. Tot i que aquesta variable és ordinal amb una escala numèrica (integer), és possible millorar la seva representació afegint, a més, etiquetes orientatives a l'eix vertical.
9. Escriu les següents dos línies, i després executa a la consola de R:
`boxplot(data$p33, main = "P33. Autoposició ideològica", yaxt = "n")`
`axis(2, at = c(1, 5, 10), labels = c("Esquerra", "Centre", "Dreta"))`
10. Comprova com ara ha millorat la interpretabilitat de la representació gràfica, amb l'ús de la ordre axis() que serveix per a modificar els eixos del darrer gràfic que s'acaba de generar.
11. Per a generar l'HISTOGRAMA de la variable P15 Contactes a les Xarxes Socials, amb exclusió dels valors perduts 9998 i 9999, escriu i executa a la consola de R:
`hist(data$p15, exclude = c(9998, 9999))`
12. Comprova com s'ha adjuntat un títol automàtic, poc explicatiu.
13. Quants blocs te aquest histograma ? Com fer si vols representar una variable numèrica mitjançant un histograma amb un nombre determinat d'interval, o blocs d'alçada ?
14. Per a generar l'HISTOGRAMA de la mateixa variable P15, amb 5 interval i un títol personalitzat, escriu i executa a la consola de R:
`hist(data$p15, exclude = c(9998, 9999), breaks = 5, main = "Contactes a Xarxes Socials")`
15. Comprova com l'opció **main** adjunta un títol, i l'opció **breaks** controla els intervals.
16. Per a generar una gràfica de DENSITAT de la variable d'auto-posicionament ideològic P33, amb exclusió dels valors perduts 97, 98 i 99, escriu i executa a la consola de R:
`plot(density(data$p33, na.rm=TRUE))`
17. Compara-la amb l'histograma anterior corresponent.



EXERCICI

- De la matriu de dades "P4_Dades1.omv", selecciona tres variables (diferents de les ja utilitzades p3, p5, g_edat, p15, i p33), una NOMINAL, una altra ORDINAL, i una tercera QUANTITATIVA.
- Ajuda't amb el qüestionari del CIS "P3_Qüestionari.pdf" disponible al CV.
- Genera amb *jamovi*, la representació gràfica que millor representi a cada variable.
- Redacta un document en format DOCX amb les tres gràfiques importades al processador de text, cadascuna acompanyada d'un breu comentari respecte a 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: els tractament dels **valors perduts**, la **depuració**, detecció i correcció de possibles errors en les dades, així com la **transformació** de variables.

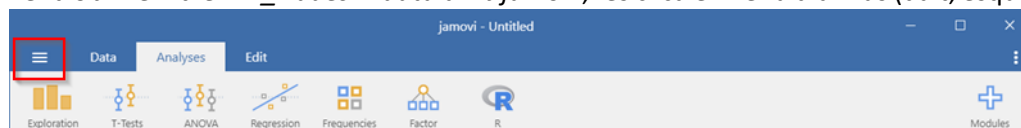
Aquestes tasques, que impliquen típicament la generació de “nova” informació a partir de les dades originals, no formen part pròpiament de les operacions estadístiques. Per aquest motiu algunes de les aplicacions informàtiques disponibles per a l'anàlisi de dades qualitatives no inclouen eines prou eficients 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.⁹

Tot i que, en la seqüència de treball habitual, primer es depura el fitxer i després es fan les transformacions de variables pertinents, aquí -per 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 “*Dades2.rdata*”. 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 2011.¹⁰ De totes les variables de l'estudi només se n'han conservat algunes, les que podeu consultar en el document “*Qüestionari.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: “*P4_Dades2.rdata*”, i “*Qüestionari.pdf*” corresponents a l'enquesta del CIS.

1. Descarrega aquests dos arxius a una carpeta local del teu PC.
2. Executa l'aplicació *jamovi* al teu ordinador.
3. Per a obrir el fitxer *P4_Dades2.rdata* amb *jamovi*, fes clic al menú d'arxius (dalt, esquerra).



Imatge 52. Ubicació menú d'arxius

4. Selecciona l'opció Open / This PC. Al panell central, fes clic a [Browse] (dalt, dreta)



Imatge 53. Obrir un nou arxiu de dades

5. S'obre una finestra per seleccionar l'arxiu de dades a obrir. Tria l'arxiu “*P4_Dades2.rdata*” que has descarregat al teu PC local.
6. A la part inferior dreta de la pantalla principal apareix una notificació del procés d'obertura d'arxiu i càrrega de dades.

⁹ 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.

¹⁰ 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.

- Una vegada completat el procés, *jamovi* disposa de la matriu de dades, amb 1419 individus entrevistats i 70 variables.

El tractament de les “no resposta”: Els valors perduts.

Sovint algunes persones entrevistades deixen sense 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 deixar **en blanc**, sense anotar res, els que no procedeix (*no cal fer la pregunta, degut a la seva resposta a preguntes anteriors*), i assignar algun **codi especial** a la resta (per exemple, “NS/NC” en variables de text; i 98, o 99, o 999 en variables numèriques).

La existència de valors “perduts” a la matriu de dades pot suposar dos tipus de problemes al moment de fer l'anàlisi:

- Tot i que la mostra de l'estudi sigui, en general, prou gran com per assegurar representativitat estadística (*permet assegurar que els resultats i conclusions fonamentats en el conjunt de casos de la mostra es poden “aplicar” a tot el conjunt de la població*) és possible que alguna variable concreta contingui molts valors perduts i per tant no se'n puguin treure conclusions vàlides d'aquesta part de la informació.
- En el cas d'utilització de codis o valors numèrics per a expressar valors perduts de la variable, si es tenen en compte i s'utilitzen en càlculs estadístics s'estarà “deformant” la realitat social mesurada.

Quan *jamovi* troba un **espai en blanc** a les dades li assigna 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**. A més es pot definir, per a cada variable, una quantitat de codis especials, com a valor perdut **d'usuari**, que *jamovi* tampoc no tindrà en compte en els procediments de càlcul corresponents.

- Demana freqüències per a les variables p3 (“*has fet servir Internet en els darrers tres mesos*”), i p5 (en cas afirmatiu, “*amb quina freqüència utilitzes Internet*”).

La primera és una variable NOMINAL i la segona ORDINAL, totes dues amb dades de tipus TEXT. Revisa la descripció d'aquestes dues variables a Menú / Data / Setup. No et refiïs mai de que estigui perfecta la descripció d'una variable si l'has carregat des d'un arxiu extern.

- Quan demanis la taula de freqüències, recorda desplegar (clic) l'apartat “Statistics” per **desactivar** les opcions que no necessites ara. Deixa només “N” i “Missing”.

Descriptives

	p3	p5
N	1419	1301
Missing	0	118

Frequencies of p3

Descriptives

	p3	p5	
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 %
Varias veces al dia	822	63.2 %	63.2 %
Una vez al dia	244	18.8 %	81.9 %
3-5 por semana	123	9.5 %	91.4 %
1-2 por semana	69	5.3 %	96.7 %
Cada varias semanas	30	2.3 %	99.0 %
Casi nunca	12	0.9 %	99.9 %
N.C.	1	0.1 %	100.0 %

Observa que a p3 no hi ha 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 quant no era procedent (degut a la resposta negativa a p3), mentre que si no contestava a p5 hi ha assignat el codi N.C.

Arribats aquest punt cal plantejar que fer amb les “no respostes”, és a dir, amb els que han respost N.C. en la p5. La qüestió és establir si ens interessa o no distingir diferents tipus de no resposta.

1. Demana freqüències per a la variable p33 (“Auto-posicionament polític”).
2. A l’arxiu original és una variable CONTINUA, per tant, recorda que si vols visualitzar la taula de freqüències has de canviar la seva definició per que sigui ORDINAL / INTEGER.

Frequencies of p33

Levels	Counts	% of Total	Cumulative %
1	55	3.9 %	3.9 %
2	60	4.2 %	8.1 %
3	157	11.1 %	19.2 %
4	131	9.2 %	28.4 %
5	264	18.6 %	47.0 %
6	98	6.9 %	53.9 %
7	83	5.8 %	59.8 %
8	38	2.7 %	62.4 %
9	6	0.4 %	62.9 %

Frequencies of p33

Levels	Counts	% of Total	Cumulative %
10	15	1.1 %	63.9 %
97	261	18.4 %	82.3 %
98	178	12.5 %	94.9 %
99	73	5.1 %	100.0 %

En el cas de la variable ideologia pròpia (p33), per exemple, hi ha tres tipus de no resposta:

- no té cap ideologia “*Ninguna*”, codi 97
- no sap quina és la seva ideologia “*NS*”, codi 98, i
- no vol contestar “*NC*”, codi 99.

Per tant, cal considerar com a valors perduts la categoria “N.C.” de la variable ordinal p5, i les categories 97, 98 i 99 de la variable numèrica p33.

Tot i que la versió actual de *jamovi* permet definir “valors perduts” d’usuari¹¹ a les variables (i ja queden així identificats i s’exclouen de les anàlisis) és més pràctic fer una còpia de la variable **que no contingui els valors** considerats “perduts” (crear una nova columna). El nom d’aquesta variable còpia pot ser el mateix de la original amb un sufix que denota la seva condició de variable amb dades no perdudes o disponibles, per exemple, “_a” (en anglès, “available”).

Es important fer una còpia de la variable, per poder mantenir sempre les dades originals intactes. D’aquesta forma es manté la possibilitat de fer anàlisis tenint en compte o no els valors “perduts”.

En general, es procedeix en dos passos, per a cada variable:

1. Primer cal utilitzar la funció de *jamovi* “Transform” per tal de crear la còpia de la variable, i
2. llavors cal crear una nova “regla de transformació” que expliqui com canviar els valors de la variable original per a omplir la nova variable.

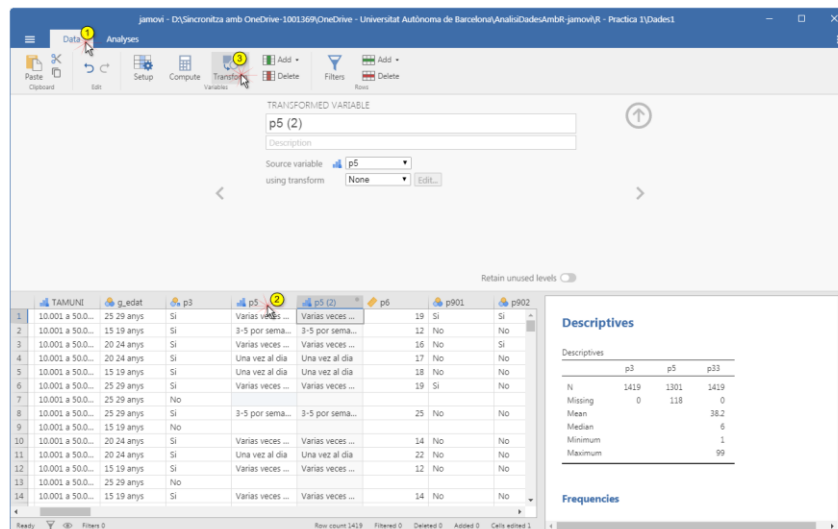
En el cas de la variable NOMINAL o ORDINAL: per exemple, la p5 “¿Con qué frecuencia utilizas Internet?”, l’objectiu és eliminar de la variable original tots els valors que continguin la etiqueta textual “N.C.”. Per tant, la regla de transformació que cal aplicar a cada fila o individu es pot expressar LITERALMENT com: **“si a la variable origen el teu valor és diferent a ‘N.C.’ llavors la variable transformada manté aquest valor”**. Com pots comprovar, l’aplicació d’aquesta regla de transformació a totes les files de la variable p5 farà que la nova variable sigui una còpia exacta EXCEPTE en el cas dels ‘N.C.’ que quedaran en blanc, per tant *jamovi* els tractarà com a valors perduts.

Per generar la nova variable “p5_a”, a partir de la p5 original, fes:

1. Clic a la etiqueta “Data” del menú superior.
2. Selecciona la variable origen, fent clic el nom de la columna corresponent a la variable p5.

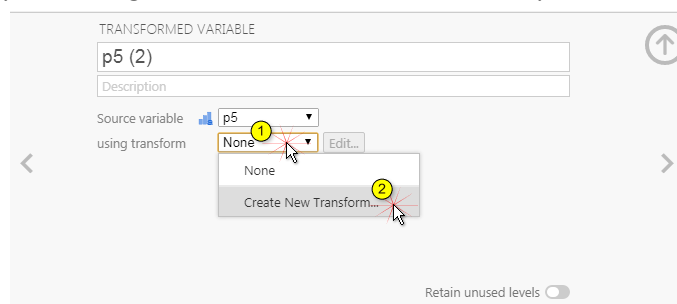
¹¹ Per a definir valors perduts d’usuari cal introduir condicions a l’apartat “Missing values” de Menú / Data / Setup. La sintaxi de condicions és la mateixa que la utilitzada pels filtres (consultar l’apartat corresponents del Tutorial).

3. Fes clic a la icona “Trasform” del sub-menú de procediments amb dades.
4. Es crea una nova columna a la dreta de la p5, per defecte anomenada “p5 (2)” i basada en la variable d’origen p5, sense transformacions per ara.



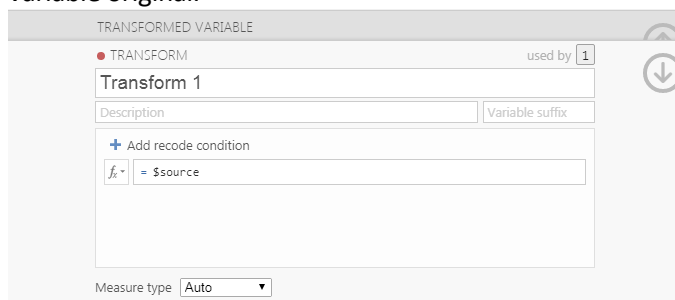
Imatge 54. Mètode per a transformar una variable

5. A continuació crearem i aplicarem la regla de transformació requerida per aquest cas. Per a crear una nova transformació, Fes clic a el desplegable que actualment indica [None] per a “using transform”, i llavors selecciona l’opció [Create New Transform].



Imatge 55. Mètode per a crear nova regla de transformació

6. S’obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada “Transform 1”, sense descripció, amb una regla molt simple “els mateixos valors que la variable origen” (= \$source), i un tipus de variable automàticament dependent de la variable original.

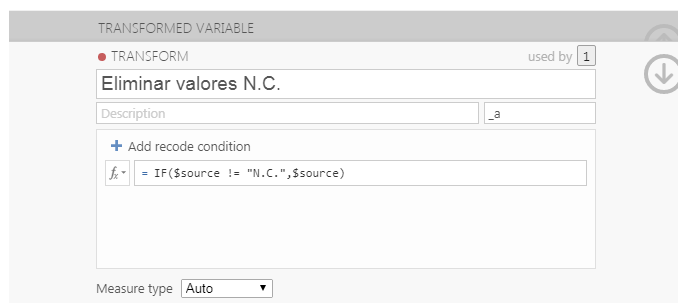


Imatge 56. Finestra de definició de regles de transformació

7. Primer **modifica el nom** per que sigui més significatiu, canvia “Transform 1” per “Eliminar valores N.C.”. Si vols pots introduir a sota una descripció més detallada.
8. A continuació, a la dreta de la descripció cal **introduir el sufix** amb el que s’identificaran les noves variables que siguin transformades per aquesta regla: es pot utilitzar “_a” (sense les cometes) per destacar que la variable transformada conté només valors DISPONIBLES.

Qualsevol altra notació que et serveixi per distingir de la variable original que té “perduts” pot estar bé, p.e. “_np”, o “_sp”.

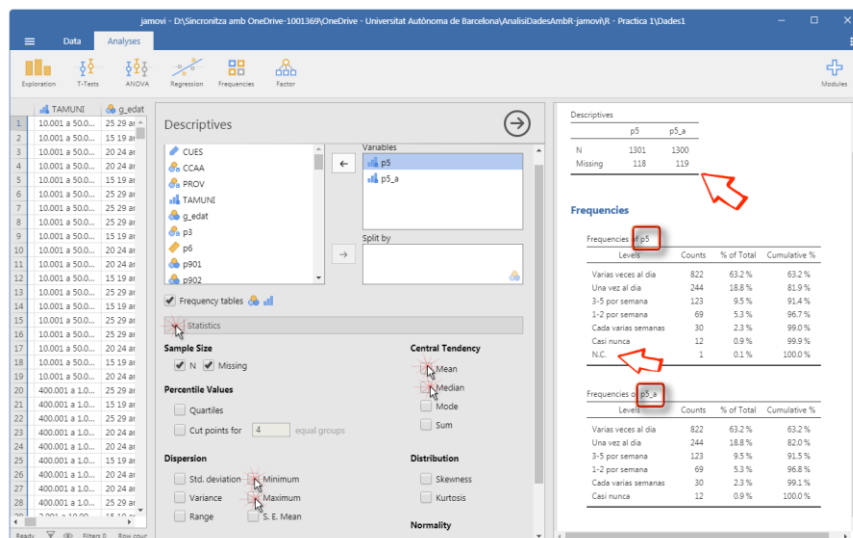
9. A l'espai corresponent a la condició de recodificació cal **introduir la regla** de transformació que fa desaparèixer els valors si són “N.C.”, i que ha d'estar escrita en un llenguatge de programació comprensible per *jamovi* (que com ja saps, és R). Òbviament, hi ha diverses formes d'aconseguir aquest efecte, una de les més simples pot ser: “*si el valor de la variable original no és igual a 'N.C.' llavors el valor a la variable transformada serà el mateix*”. En el llenguatge de programació que *jamovi* utilitza aquesta regla es pot expressar com:
IF(\$source != "N.C.", \$source)



Imatge 57. Exemple de regla de transformació

Si ...	IF(
... el valor de la variable original ...	\$source
... no ...	!
... és igual ...	=
... a 'N.C.' ...	"N.C."
... llavors ...	,
... el valor a la variable transformada serà el mateix que el valor de la original.	\$source
=> Si a la variable original és “N.C.” a la transformada no hi ha cap valor (missing)	)

10. Deixa el tipus de mesura com a AUTO per a “copiar” el mateix de la variable original.
11. Per a tancar la finestra de regles de transformació, cal clicar a la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra. Fes clic per tancar-la.
12. Com pots veure ara al panell de transformació, la nova variable s'anomena “P5_a” (aplicant el sufix automàtic), està basada en els valors originals de la p5, i utilitza la transformació anomenada “Eliminar valores N.C.” que acabes de crear.
13. Per a veure l'efecte de la transformació cal incloure la nova variable p5_a a la llista del procediment “Descriptives”. Fes clic a sobre de qualsevol dels resultats d'aquest procediment. S'obre el panell central de definició de resultats.
14. Per afegir p5_a als resultats, selecciona aquesta nova variable fent clic a la llista de l'esquerra. “Passa-la” cap la llista de la dreta, fent clic el botó FLETXA superior central.



Imatge 58. Resultat aplicació transformació de variables

15. S'ha actualitzat automàticament el resultat, on pots veure com p5 te **118** valors perduts mentre que p5_a te **119** valors perduts (incloïa una persona que ha contestat "N.C." i que ha estat "eliminat" en aplicació de la regla de transformació que genera la nova variable).
16. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

En el cas de la variable CONTINUA: per exemple, la p33 "Auto-posicionamiento político de las personas encuestadas en una escala izquierda-derecha que va de 1 a 10", cal eliminar tots els valors que continguin qualsevol dels tres codis corresponents a "perduts": 97, 98 i 99. En realitat, tractant-se d'una escala de 1 a 10 el que cal és assegurar que la nova variable (p33_a) és una copia de la original però **NOMÉS** si els valors estan dins d'aquest rang que limita la escala d'auto-posicionament, val a dir, "son mes petits que 11".

Per tant, la regla de transformació simple aplicable en aquest cas pot ser **"si a la variable origen el teu valor és menor de 11 llavors la variable transformada manté aquest mateix valor"**. Com pots comprovar, l'aplicació d'aquesta regla de transformació a totes les files de la variable p33 farà que a la nova variable els valors originals 97, 98 i 99 quedaran en blanc, i per tant *jamovi* els tractarà com a valors perduts.

Per generar la nova variable "p33_a", a partir de la p33 original, segueix els mateixos passos que en el cas anterior, però adaptant la regla de transformació. Assegura't que la definició de la variable és correcta, cal que sigui CONTINUA/INTEGER (també pot ser ORDINAL/INTEGER)-

9. A l'espai corresponent a la condició de recodificació cal **introduir la regla** de transformació que fa desaparèixer els valors si són majors que 10, escrita en un llenguatge de programació comprensible per *jamovi*. Per aconseguir aquest efecte, una de les regles més simples pot ser: *"si el valor de la variable original és menor a 11, llavors el valor a la variable*

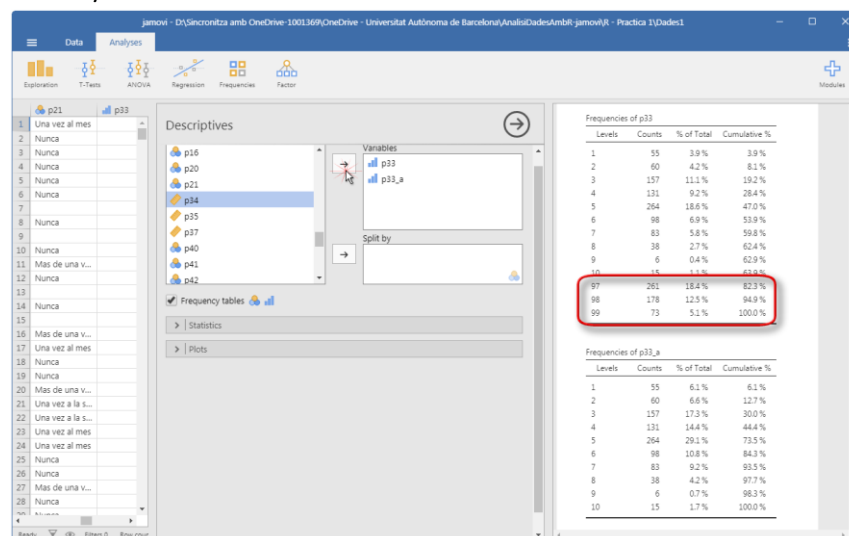
transformada serà el mateix". En el llenguatge *jamovi*:
IF(\$source < 11,\$source)



Imatge 59. Regla de transformació per excloure valors fora de rang

Si ...	IF(
... el valor de la variable original ...	\$source
... és més petit que ...	<
... 11 ...	11
... llavors ...	,
... el valor a la variable transformada serà el mateix que el valor de la original.	\$source
=> Si a la variable original és major que 10, a la transformada no hi ha cap valor (missing)	)

- Pots afegir ara la nova variable p33_a al procediment "Descriptive" per comprovar si la transformació realment elimina els valors no desitjats.
- Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general / Save.



Imatge 60. Resultat d'aplicar transformació per excloure fora de rang

Pots comprovar l'efecte de tenir o no en compte els valors perduts a la següent taula de freqüències de les dues variables p33. La mitjana de la variable sense perduts (P33_a) és 4,59 mentre que els seus valors varien entre 1 i 10. Compara aquesta situació amb els resultats descriptius de la variable original, que –incorrectament- varia entre 1 i 99 amb una mitjana de 38,2 ...**en una escala de 1 a 10 !**

Descriptives

p33 p33_a

N	1419	907
Missing	0	512
Mean	38.2	4.59
Median	6	5
Minimum	1	1
Maximum	99	10

Resulta fonamental “depurar” la matriu de dades con a tasca prèvia a l’anàlisi estadística realitzada amb les seves dades. La depuració consisteix bàsicament en detectar errades, en forma de valors fora del rang “teòric” de la variable, i gestionar les variables en que apareguin per tal d’obtenir variables amb els valors identificats com a “perduts” en blanc.

S’han vist les dues eines bàsiques per realitzar aquesta tasca: la descripció simple de variables, i la transformació per eliminar valors perduts.

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 (tal i com hem vist en el cas dels valors perduts).

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

		Variable FINAL	
		Qualitativa (Nominal o Ordinal)	Quantitativa (Interval o Raó)
Variable ORIGEN	Qualitativa (Nominal o Ordinal)	[SE / Prim. / Sec. / FP / Grau / Doctorat] es transforma en [Prim o - / Sec. / Univ.]	[Si / No] es transforma en [1 / 0]
	Quantitativa (Interval o Raó)	[Edat: 0 - 90] es transforma en [Joves / Adults / Vells]	[Ideologia 1 - 10] es transforma en [Ideologia 1 - 5]

En tots 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 a la base de dades, en funció dels valors d’altra o altres columnes existents.

$$\text{Variable ORIGEN} + \text{Regla de Transformació} = \text{Variable FINAL}$$

I com es recodifica amb codi R? Com s'ha vist anteriorment a *jamovi* s'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 novembre de 2020 *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 pel present curs es poder realitzar amb el sistema de menús i transformacions de variables actualment implementat.

Recodificació de variables quantitatives a qualitatives

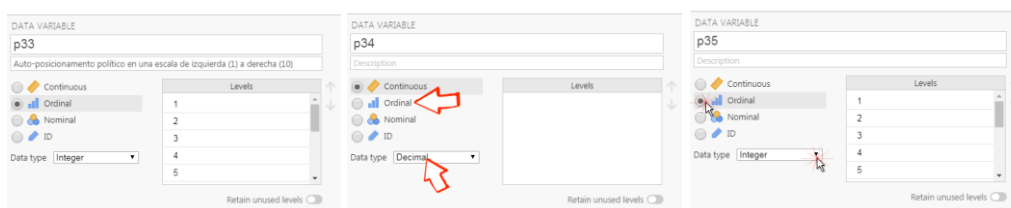
En el següent exemple (recodificació de quantitativa a qualitativa), a partir de les variables p33, p34 i p35 (respectivament, ideologia de l'individu, del seu pare i de la seva mare), es generen tres de noves, que s'anomenaran ideol-entrev, ideol-pare i ideol-mare, i que tenen, a més dels valors perduts,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

Segons es pot veure a Qüestionari.pdf, aquestes variables cal que siguin definides com a ORDINALS/INTEGER i amb els valors perduts 96, 97, 98 i 99.

Com que les dades han estat IMPORTADES dins de *jamovi* des d'un arxiu de dades extern és possible que la seva definició "automàtica" al moment de realitzar-se la importació no sigui la correcta.

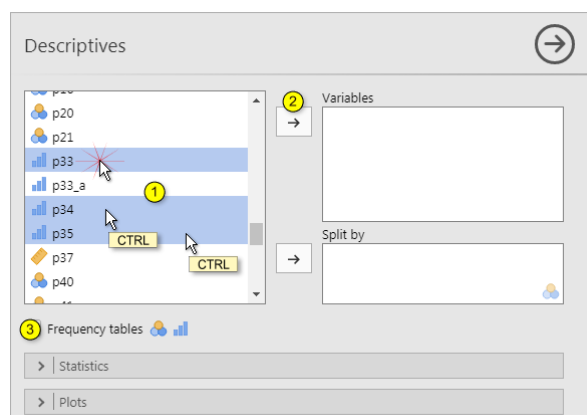
Per començar, cal comprovar el tipus de cada variable implicada.

1. A *jamovi*, fes doble-clic al nom de la variable p33, que encapçala la columna de dades. S'obre el panell superior de definició de la variable.
2. Comprova que és una variable de tipus ORDINAL i amb dades de tipus INTEGER.
3. Fes doble-clic al nom de la segona variable (p34), que encapçala la columna corresponent.
4. Cal que estigui definida igual que la p33. Fes-ho si no és el cas, i procedeix igual amb p35.



A continuació, és útil obtenir prèviament un resum descriptiu de cada variable implicada.

1. Amb *jamovi*, Fes clic al menú superior "Analyses",
2. Fes clic a la icona "Exploration" del sub-menú general de procediments analítics,
3. Al desplegable, Fes clic a l'opció "Descriptives". S'obre la finestra de definició del procediment.



Imatge 61. Forma de seleccionar diverses variables

4. 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 panel,
5. Fes clic a el quadre corresponent per activar "Frequency tables".
6. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

RECORDA: Com pots veure a les icones, les "taules de freqüències" només es poden 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 CONTINUES).

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	3.9 %	3.9 %	1	42	3.0 %	3.0 %	1	44	3.1 %	3.1 %
2	60	4.2 %	8.1 %	2	49	3.5 %	6.4 %	2	44	3.1 %	6.2 %
3	157	11.1 %	19.2 %	3	128	9.0 %	15.4 %	3	117	8.2 %	14.4 %
4	131	9.2 %	28.4 %	4	102	7.2 %	22.6 %	4	127	8.9 %	23.4 %
5	264	18.6 %	47.0 %	5	172	12.1 %	34.7 %	5	204	14.4 %	37.8 %
6	98	6.9 %	53.9 %	6	91	6.4 %	41.2 %	6	99	7.0 %	44.7 %
7	83	5.8 %	59.8 %	7	101	7.1 %	48.3 %	7	87	6.1 %	50.9 %
8	38	2.7 %	62.4 %	8	72	5.1 %	53.3 %	8	52	3.7 %	54.5 %
9	6	0.4 %	62.9 %	9	13	0.9 %	54.3 %	9	17	1.2 %	55.7 %
10	15	1.1 %	63.9 %	10	28	2.0 %	56.2 %	10	15	1.1 %	56.8 %
97	261	18.4 %	82.3 %	96	67	4.7 %	61.0 %	96	22	1.6 %	58.4 %
98	178	12.5 %	94.9 %	97	84	5.9 %	66.9 %	97	113	8.0 %	66.3 %
99	73	5.1 %	100.0 %	98	384	27.1 %	93.9 %	98	391	27.6 %	93.9 %
				99	86	6.1 %	100.0 %	99	87	6.1 %	100.0 %

Per recodificar (i depurar els valors perduts) d'aquestes tres variables cal aplicar DUES regles de transformació:

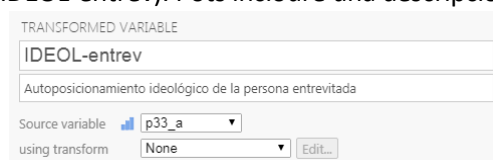
1. Regla per a eliminar els valors perduts "fora del rang 1-10", i
2. Regla per a assignar als valors origen 1 a 4 el nou valor "Izquierda", al valor origen 5 el nou valor "Centro", i als valors origen 6 a 10 el nou valor "Derecha".

Aquestes dues regles cal aplicar-les a les tres variables. Com podeu recordar de l'apartat anterior on es depuraven els valors perduts per a la variable p33, la regla que elimina valors fora del rang 1

a 10 ja existeix a l'arxiu *jamovi* (*.omv), i per tant es podria reutilitzar en aquesta nova transformació (i eventualment, en altres).

Comença per generar la nova variable recodificada a partir de la p33:

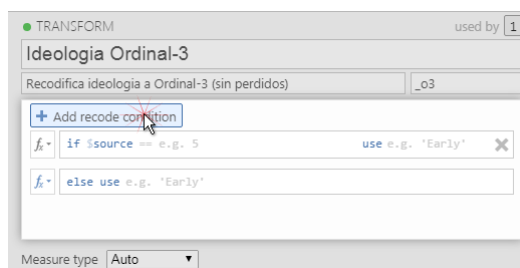
1. A *jamovi*, fes doble-clic al nom de la variable p33_a, que encapçala la columna de dades. S'obre el panell superior de definició de la variable. Aquesta columna incorpora dades de la variable original p33, amb una transformació ("Escala 1-10") tal que s'eliminen tots els valors "perduts" per a aquesta escala de 10 punts.¹²
2. Clic a la etiqueta "Data" del menú superior. Fes clic a la icona "Trasform" del sub-menú de procediments amb dades. Es crea una nova columna a la dreta de la p33_a, per defecte anomenada "p33_a (2)" i basada en la variable d'origen p33_.
3. Canvia el nom (p.e., IDEOL-entrev). Pots incloure una descripció més expressiva.



Imatge 62. Exemple de nova variable transformada

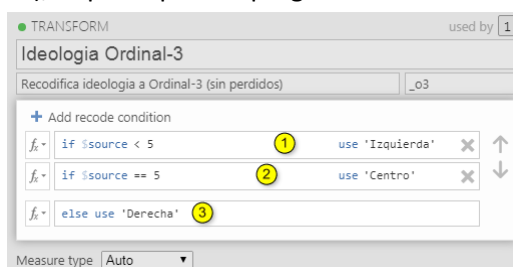
4. Fins aquí has creat una nova variable, o columna, copia exacta de la origen p33. Ara cal transformar-la per a eliminar els valors perduts.
5. A continuació crearem la nova transformació requerida per fer la recodificació de variable quantitativa (1-10) a variable ordinal (Izquierda-Centro-Derecha). Fes clic a el desplegable que actualment indica [None] per a "using transform", i llavors selecciona l'opció [Create New Transform].
6. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada "Transform 3", sense descripció, 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** per que sigui més significatiu, canvia "Transform 3" per "Ideologia Ordinal-3". Si vols pots introduir a sota una descripció més detallada, com ara "Recodifica ideologia a Ordinal-3 (sin perdidos)".
8. A continuació, a la dreta de la descripció cal **introduir el sufix** amb el que s'identificaran les noves variables que siguin transformades per aquesta regla: es pot utilitzar "_o3" (sense les cometes) per destacar que la variable transformada serà la versió ordinal amb 3 intervals.
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*.
10. En aquest cas concret es tracta d'un conjunt de 3 regles, o "recode conditions": Assignarem la etiqueta "Izquierda" als auto-posicionaments ubicats de 1 a 4, i "Centro" als ubicats a 5, de forma que la resta seran etiquetats com "Derecha" (ubicats de 6 a 10). Comença fent clic a el botó per afegir una condició de recodificació.

¹² Consultar l'apartat anterior per veure com crear aquesta transformació de "depuració" de valors perduts.



Imatge 63. Botó per afegir condicions de recodificació

11. La primera condició pot ser: *Si el valor de la variable original és menor que 5 llavors el valor a la variable transformada serà "Izquierda"*. En el llenguatge de programació que *jamovi* utilitza aquesta regla es pot expressar com: `IF $source < 5 use 'Izquierda'`
12. La segona condició, que fa referència als posicionaments de centre, es pot expressar en llenguatge de programació com: `IF $source == 5 use 'Centro'`
13. I la condició final, que afecta a tots els valors de la variable original que no han estat encara canviats (els majors de 5), es pot expressar programàticament com: `else use 'Derecha'`



Imatge 64. Combinació de 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, el **doblet-igual** (`==`), o les etiquetes textuais **entre cometes** (`'xxx'`).

14. Deixa el tipus de mesura com a AUTO per a "copiar" el mateix de la variable original (ordinal).
15. Per tancar aquesta finestra de regles de transformació, cal clicar a la icona circular FLETXA-AVALL de la part superior dreta d'aquesta finestra. Fes clic per tancar-la.
16. Com pots veure ara al panell de transformació, la nova variable està basada en els valors originals de la p33_a (per tant, ja està "neta" de valors perduts), i a més utilitza la transformació anomenada "Ideologia Ordinal-3" que acabes de crear.
17. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

Ara pots visualitzar els canvis fets a la matriu de dades. Comprova com han desaparegut els valors perduts (majors de 10) i s'han aplicat les noves TRES etiquetes en funció del valors de la variable original.

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

Per a recodificar la variable origen p34 en la nova variable "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. Aquesta columna incorpora dades de la variable original p34, sense eliminació de valors “perduts”.
2. Clic a la etiqueta “Data” del menú superior. Fes clic a la icona “Trasform” del sub-menú 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. Fes clic a el desplegable corresponent a “using transform” (que per defecte indica “None”).
4. Selecciona la transformació **que elimina els perduts** per sobre del valor 10: “Escala 1-10”.
5. Comprova com ara ja existeix la nova variable 34_a, transformada sense valors perduts.
6. Fes doble-clic al nom de la nova variable p34_a, que encapçala la columna de dades.
7. Clic a la etiqueta “Data” del menú superior. Fes clic a la icona “Trasform” del sub-menú de procediments amb dades. Es crea una nova columna a la dreta de la p34_a, per defecte anomenada “p34_a (2)” i basada en la variable d'origen p34_a.
8. Fes clic a el desplegable corresponent a “using transform” (que per defecte indica “None”).
9. Selecciona la transformació **que recodifica en tres intervals** ordinals: “Ideologia Ordinal-3”.
10. Comprova com ara ja existeix la nova variable 34_a_o3, transformada sense valors perduts i recodificada. Canvia el nom (p.e., IDEOL-pare) i inclou una descripció més expressiva.

Imatge 65. Exemple de transformació

11. Com pots veure ara al panell de transformació, la nova variable està basada en els valors originals de la p34_a (per tant, ja està “neta” de valors perduts), i utilitza la transformació anomenada “Ideologia Ordinal-3” que has creat anteriorment per a una altra variable.
12. 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. Fes clic per tancar-la.
13. Per a tancar la finestra de definició de variables, cal clicar a la icona circular FLETXA-AMUNT de la part superior dreta d'aquesta finestra. Fes clic per tancar-la.
14. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general / Save.

Pots comprovar les transformacions que has fet demanant a *jamovi* una descripció simple d'aquestes variables. Recorda que la forma ràpida de fer-ho és utilitzar la “consola R” per a escriure directament llenguatge de programació R amb la comanda específica.

En aquest cas, si vols una descripció sintètica de les tres variables que fan referència a la ideologia del pare de la persona entrevistada:

Pre-requisit: Tenir instal·lat a *jamovi* el mòdul d'extensió anomenat Rj.¹³

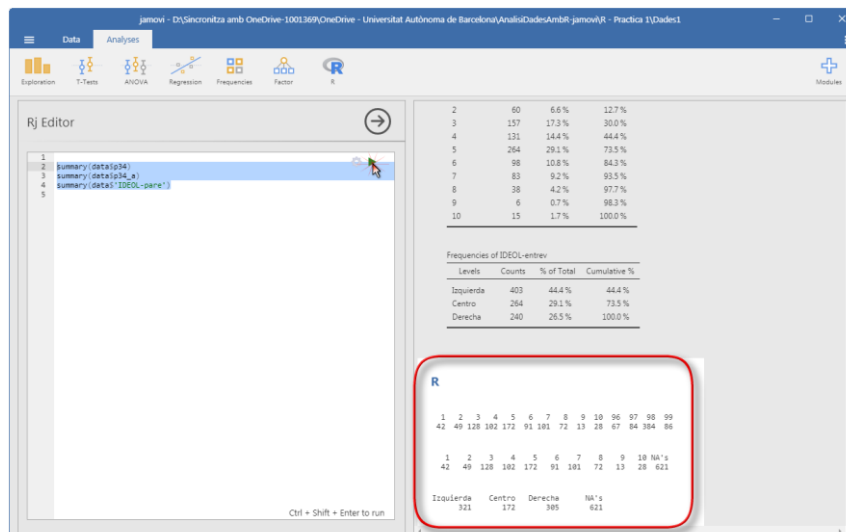
1. A l'apartat “Analyses”, clicar a la icona “R”. Al desplegable, seleccionar l'opció “Rj Editor”,
2. Apareix la “consola” d'ordres R, i per defecte, una ordre inactiva d'exemple: `# summary(data[1:3])`. Qualsevol línia que comenci per # es considera inactiva (està de color verd),
3. Escriu, poc a poc, les següents ordres a la consola, en tres línies distintes:

```
summary(data$p34)
summary(data$p34_a)
summary(data$'IDEOL-pare')
```

Com has pogut veure, *jamovi* t'ajuda proposant ordres i variables disponibles.
4. Aquestes simples ordres que has escrit en llenguatge R signifiquen, respectivament:
“mostra un resum descriptiu de la variable de la matriu de dades que es diu p34”,

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

- “mostra un resum descriptiu de la variable de la matriu de dades que es diu *p34_a*”,
“mostra un resum descriptiu de la variable de la matriu de dades que es diu *IDEOL-pare*”.
- 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.



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

- 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.

Pots comprovar com la primera transformació aplicada a *p34* ha “depurat” de valors perduts la variable (hi ha 621 perduts, o NA “Not Available”), mentre que la segona transformació aplicada a *p34_a* ha generat correctament la variable recodificada, ordinal, i sense perduts (*IDEOL-pare*).



EXERCICI

- Genera una nova variable *IDEOL-mare*. Utilitza les regles de transformació ja existents.
- Genera la nova variable “¿Cuando comenzaste a usar Internet? [De niño/a (antes de 10 años) | De joven (entre 10 y 25 años) | De mayor (después de 25 años)] a partir de les dades de la *p6*.

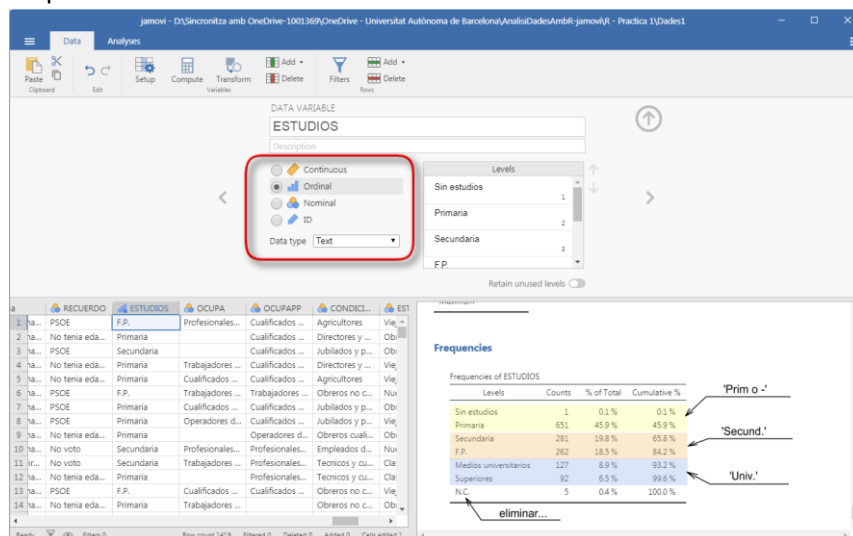
ATENCIÓ! Si feu servir regles de transformació creades prèviament, **compte** amb no modifica-les sense comprovar que no afecten a les “altres” variables a les que ja s’havien aplicat anteriorment. Es per això que a la finestra corresponent de la especificació de les regles o funcions de transformació hi ha un indicador de a quantes variables s’ha aplicat la regla: , a la part superior dreta.

Recodificació de variables qualitatives a qualitatives

En el següent exemple, a partir de les variable ESTUDIOS (amb 6 categories ordinals), es generarà una de nova que s'anomenarà "estudios"; amb només tres categories i prescindint dels N.C. (que es convertiran a perduts): 'Prim. o -' (Primaria o menys), 'Secund.' (que inclou FP), i 'Univ.' (Universitaris).

Segons es pot veure a Qüestionari.pdf, la variable original ESTUDIOS cal que sigui definida com a ORDINAL / TEXT i amb el valor perdut 'N.C.'. Com que les dades han estat IMPORTADES dins de *jamovi* d'es d'un arxiu de dades extern és possible que la seva definició "automàtica" al moment de realitzar-se la importació no sigui la correcta.

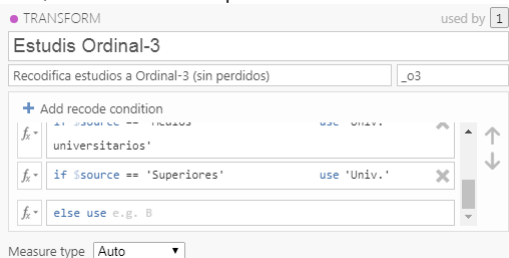
1. A *jamovi*, fes doble-clic al nom de la variable ESTUDIOS, que encapçala la columna de dades. S'obre el panell superior de definició de la variable.
2. Comprova que efectivament la variable ESTUDIOS estigui definida com de tipus ORDINAL, amb dades de tipus TEXT. Si no és el cas, **modifica la seva definició** fins que s'ajusti a aquests paràmetres.¹⁴



Imatge 67. Comprovació del tipus de variable, i criteris per a recodificar

3. Fes clic a la etiqueta "Data" del menú superior. Fes clic a la icona "Transform" del sub-menú de procediments amb dades. Es crea una nova columna a la dreta de la ESTUDIOS, per defecte anomenada "ESTUDIOS (2)" i basada en la variable d'origen.
4. Canvia el nom a "estudios". Pots incloure una descripció més expressiva.
5. A continuació cal crear la nova transformació requerida en aquest cas per fer la recodificació de variable qualitativa ordinal (ESTUDIOS, 6 intervals) a variable qualitativa ordinal (estudios, amb 3 intervals). Fes clic a el desplegable que actualment indica [None] per a "using transform", i llavors selecciona l'opció [Create New Transform].
6. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada "Transform 4", 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** per que sigui més significatiu, canvia "Transform 4" per "Estudios Ordinal-3". Si vols pots introduir a sota una descripció més detallada, com ara "Recodifica estudios a Ordinal-3 (sin perdidos)".
8. A continuació, a la dreta de la descripció cal **introduir el sufix** amb el que s'identificaran les noves variables que siguin transformades per aquesta regla: es pot utilitzar "_o3" (sense les cometes) per destacar que la variable transformada serà la versió ordinal amb 3 intervals.

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

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 -' (per recodificar Primaria o menys), 'Secund.' (que inclou FP), i 'Univ.' (tots dos nivells Universitaris).
 10. Fes clic a 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 'Sin estudios' llavors el valor a la variable transformada serà 'Prim. o -'*. En el llenguatge de programació:
IF \$source == 'Sin estudios' use 'Prim o -'
 12. Cal introduir **cada** condició en una **nova línia**. Fes clic a 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 'Secun.'
IF \$source == 'F.P.' use 'Secun.'
IF \$source == 'Medios universitarios' use 'Univ.'
IF \$source == 'Superiores' use 'Univ.'
- 
15. La condició final, que afecta a tots els valors de la variable original que no han estat encara canviats, es pot expressar programàticament com: `else use`
Això implicarà que tots els valors de la variable original que no hagin estat recodificats utilitzant els criteris previs (només queden els 'N.C.') quedaran en blanc, és a dir, seran considerats com a perduts per *jamovi*. Deixa-ho així, perquè és precisament el que és vol.
 16. A la part inferior de la finestra de condicions, deixa el tipus de mesura com a AUTO per a "copiar" el mateix de la variable original (ordinal).
 17. 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. Fes clic per tancar-la.
 18. Com pots veure ara al panell de transformació, la nova variable està basada en els valors originals de la ESTUDIO, i utilitza la transformació anomenada "Estudios Ordinal-3" que acabes de crear.
 19. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

Pots demanar un resum descriptiu simple de les dues variables ESTUDIOS i estudios, per comparar-les i comprovar si la recodificació s'ha generat amb èxit. També pots comprova si els valors 'N.C.' (a la variable original hi ha, p.e. a les files 584 i 586) s'han eliminat correctament de la variable recodificada.

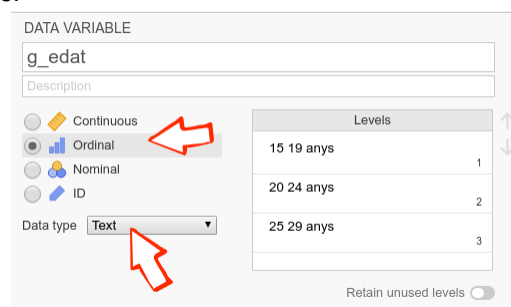
Recodificació de variables qualitatives a quantitatives

En el cas que vulguem 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 per que s'ajusti al format adient: INTEGER o DECIMAL.

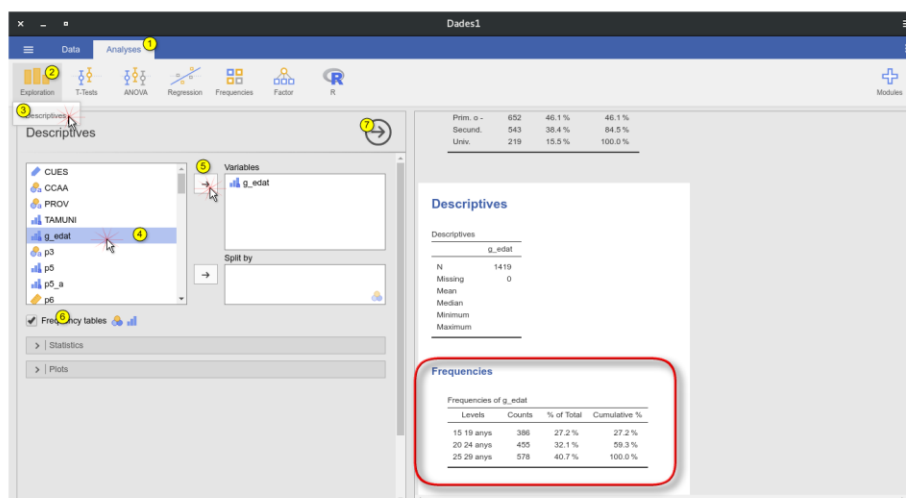
Per a transformar la variable *g_edat* (que és ordinal/text) a quantitativa cal assignar a cada valor el corresponent a la seva “marca de classe”. Així, la transformació es resumeix a passar de “15 19 anys” a 17,5; de “20 24 anys” a 22,5; i de “25 29 anys” a 27,5. Val a dir, a cadascuna de les tres etiquetes o nivells cal assignar el valor (que serà continuu/decimal) que correspon al punt mig de l'interval que representen. Per fer-ho, segueix els passos següents:

1. Double clicka el nom de la variable, que encapçala la columna de dades corresponent. Comprova que efectivament la variable “*g_edat*” estigui definida com de tipus ORDINAL, amb dades de tipus TEXT. Si no és el cas, **modifica la seva definició** fins que s'ajusti a aquests paràmetres.¹⁵



Imatge 68. Comprovació de definició de variable

2. Comprova si la variable que vols recodificar té valors perduts. Al menú “Analyses” fes clic a el sub-menú “Exploration” i tria l'opció “Descriptives”. S'obre el panell corresponent, on cal que seleccionis la variable, amb un clic a el seu nom al llistat de la part esquerra, i que la activis per a l'anàlisi fent clic a el botó quadrat FLETXA-DRETA de la part central superior. Cal que activis l'opció “[] Frequency tables”, ubicada a la part inferior del llistat de variables, per a visualitzar la taula de freqüències.



Imatge 69. Exemple de taula de freqüència de variable ordinal

3. Tot sembla correcte: cap valor perdut, hi ha dades per a tots els individus de la mostra. Pots tancar el panell de definició d'anàlisi fent clic a la FLETXA-ENCERCLADA de la part superior dreta (marcada amb un 7 a la il·lustració).
4. Selecciona la variable *g_edat*, amb clic a el nom que encapçala la columna. Fes clic a la etiqueta “Data” del menú superior. Fes clic a la icona “Transform” del sub-menú de

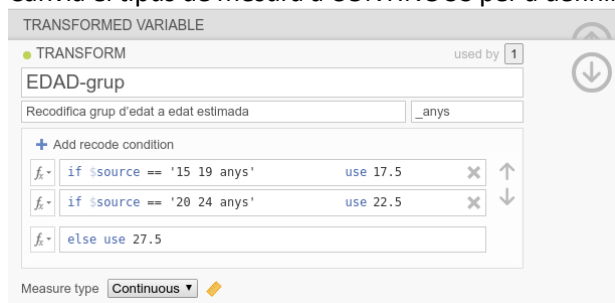
¹⁵ Pots consultar apartats anteriors d'aquesta mateixa pràctica, on s'ha vist com fer-ho.

- procediments amb dades. Es crea una nova columna a la dreta de la *g_edat*, per defecte anomenada “*g_edat (2)*” i basada en la variable d'origen.
5. No cal canvia el nom, però pots incloure una descripció més expressiva, com ara: “Edat estimada a partir de grups d'edat”.
 6. A continuació cal crear la nova transformació requerida en aquest cas per fer la recodificació de variable qualitativa ordinal (*g_edat*, 3 intervals) a variable quantitativa decimal (*g_edat_anys*, amb 1 decimal). Fes clic a el desplegable que actualment indica [None] per a “using transform”, i llavors selecciona l'opció [Create New Transform].
 7. S'obrirà la finestra de regles de transformació, amb una nova transformació per defecte anomenada “Transform 5”, amb una regla molt simple “els mateixos valors que la variable origen” (= \$source), i un tipus de variable automàticament dependent de la variable original.
 8. Primer **modifica el nom** per que sigui més significatiu, canvia “Transform 4” per “EDAD-grup”. Si vols pots introduir a sota una descripció més detallada, com ara “Recodifica grup d'edat a edat estimada”.¹⁶
 9. A continuació, a la dreta de la descripció cal **introduir el sufix** amb el que s'identificaran les noves variables que siguin transformades per aquesta regla: es pot utilitzar “_anys” (sense les cometes) per destacar que la variable transformada serà la versió numérica.
 10. 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*. Fes clic a 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 '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

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

Fixat en la diferencia entre la etiqueta de la variable original, entre cometes perquè és text, i el valor de la nova variable, sense cometes perquè és una xifra decimal.

12. Cal introduir cada condició en una nova línia, fent clic a el botó [+ Add recode condition]. Les següents condicions en llenguatge de programació segueixen la mateixa lògica:
IF \$source == '20 24 anys' use 22.5
13. La condició final, que afecta a 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
14. Canvia el tipus de mesura a CONTINUOUS per a definir la nova variable com a quantitativa.



15. 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.

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

16. Com pots veure ara al panell de transformació, la nova variable està basada en els valors originals de la ESTUDIO, i utilitza la transformació anomenada “Estudis Ordinal-3” que acabes de crear.
17. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.



EXERCICI

Pots demana un resum descriptiu simple de les dues variables *g_edat* i *g_edat_anys*, i comparar-les per a comprovar si la recodificació s'ha generat amb èxit. La mitjana d'edat hauria de ser 23,2 anys.

Recodificació de variables quantitatives a quantitatives

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

Càlculs amb variables quantitatives

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

Igualment, permet operar amb més d'una variable per generar noves columnes.

Per exemple, preguntades les persones entrevistades pel nombre de persones, contactes, amics, agregats tenen a xarxes socials, *p15* indica el nombre total declarat, mentre que *p15_a* indica quants d'aquells considera una “relació amistat consolidada”. Totes dues variables han estat definides com a CONTINUOUS i INTEGER, per tant són purament quantitatives. Totes dues tenen com a valors perduts 9997, 9998 i 9999.

Descriptives		
	p15	p15a
N	1171	1081
Missing	248	338
Mean	998	291
Median	200	20
Standard deviation	2618	1556
Minimum	1	1
Maximum	9999	9999

Per crear una nova variable, calculada a partir de *p15*, que detecti els casos amb valors perduts, i els elimini, segueix els següents passos:

1. Doble clic el nom de la variable, que encapçala la columna de dades corresponent. Comprova que efectivament la variable “p15” estigui definida com de tipus CONTINUOUS, amb dades de tipus INTEGER. Si no és el cas, **modifica la seva definició** fins que s'ajusti a aquests paràmetres.¹⁷
2. Selecciona la variable *p15*, amb clic el nom que encapçala la columna. Fes clic a la etiqueta “Data” del menú superior. Fes clic a la icona “Compute” del sub-menú de procediments amb

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

dades. 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 el nom a "p15_c" per denotar que la nova variable està calculada a partir de p15.
4. Fes clic a la finestra de fórmules, on cal escriure la funció que realitza el càlcul de la nova variable.

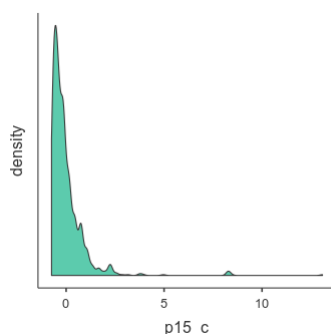
Imatge 70. Finestra de fórmules, pel càlcul de nova variable

5. Per començar eliminant els valors perduts 9997, 9998 i 9999, cal "filtrar la variable p15 de forma que només s'acceptin els valors menors que 9997". Aquest criteri s'escriu en llenguatge de programació comprensible per *jamovi* a la finestra de fórmules:
`FILTER(p15, p15 < 9997)`
6. Confirma la execució del càlcul prement la tecla [Intro]. Comprova com apareixen a la nova columna els valors filtrats, amb els valors perduts en blanc.
7. Com pots veure per la icona que acompanya al nom de la variable calculada, és de tipus NOMINAL. Això és incorrecte, perquè hauria de mantenir el mateix nivell de mesura que la variable original: CONTINUOUS/INTEGER.
Per canviar el nivell de mesura cal modificar la fórmula que s'acaba d'introduir.
8. Fes clic de nou a la finestra de fórmules i modifica el càlcul per afegir el següent criteri: "Fes que sigui un valor sencer (integer) tot això". Aquesta modificació s'escriu en llenguatge de programació comprensible per *jamovi* a la finestra de fórmules:
`INT( FILTER(p15, p15 < 9997) )`
Assegurat d'incloure correctament els parèntesis.
9. Comprova com ha canviat el nivell de mesura de la variable calculada, ara la icona indica que és quantitativa.

p15	p15_c
250	250
9998	
500	500
50	50
500	500
50	50

Imatge 71. Icones identificatives del tipus de variable

10. Actualment la mitjana és 249 (amb una desviació típica de 333). Pots comprovar-ho fent una descripció de la variable.
11. Si, a més, volguessis "normalitzar" la variable (per tal que la distribució de les dades s'ajusti a una mitjana de 0 i una desviació de 1) caldria modificar de nou la fórmula, incloent tot dins d'una altra funció entre parèntesis:
`Z( INT( FILTER(p15, p15 < 9997) ) )`
12. Ara la mitjana és 0 i la desviació és 1. Si es representa la funció de densitat es pot comprovar si s'ajusta o no a una típica "campana de Gauss". Pots comprovar com NO és així, hi ha un biaix clar cap a la dreta.



Imatge 72. Gràfica de densitat de variable normalitzada

13. 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 “sintaxi”: paraula-clau (objecte-al-que-s’aplica), p.e. $Z(p35)$.

Les diferents opcions de *jamovi* es poden consultar (i seleccionar per a utilitzar) fent clic el botó [$f_x =$] situat a la esquerra de la finestra de fórmules.

Operacions i càlculs entre diverses variables

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, o mitjana...) es pot generar una nova variable amb valors funcionalment dependents dels valors d’altres variables existents.

En un exemple del qüestionari utilitzat per la present pràctica, preguntades les entrevistades pel nombre de persones, contactes, amics, o agregats que tenen a xarxes socials, p15 indica el nombre total declarat mentre que p15_a indica quants d’aquells considera una “relació amistat consolidada”. Totes dues variables han estat definides com a CONTINUOUS i INTEGER, per tant són purament quantitatives. Totes dues tenen com a valors perduts 9997, 9998 i 9999.

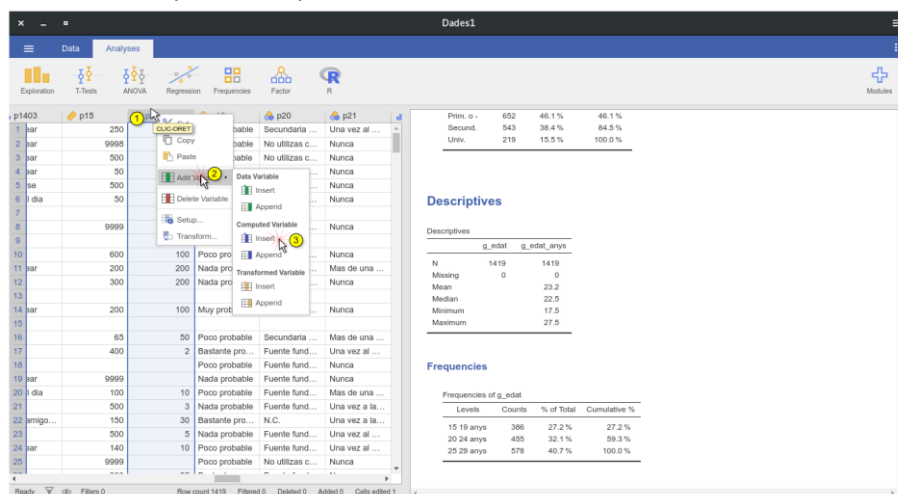
Si volem establir una relació entre la quantitat de contactes a xarxes socials i els “veritables amics” per a cada persona que ha respost a la enquesta es pot fer un simple càlcul amb les variables p15 i p15a. El quocient o raó $p15a/p15$ és un indicador del nivell de “contactes fiables” de cada persona enquestada; proper a 1 per casos amb màxima confiabilitat dels contactes a xarxes i proper a 0 per a mínima fiabilitat dels contactes a xarxes.

Per calcular la nova variable CONFIA-xarxes, a partir d’un càlcul amb p15 i p15a:

1. Comprova la existència de valors perduts, utilitza el procediment estadístic Data / Analyses / Exploration / Descriptives, selecciona les dues variables p15 i p15a, i activa ☐ Frequency tables. Per sortir del panell de “descriptives”, fes clic el botó FLETXA-DRETA de la part superior dreta.
2. Veuràs que només apareixen els estadístics descriptius, no les taules de freqüències. Recorda que *jamovi* només realitza aquestes taules per a variables de tipus NOMINAL o ORDINAL.

ATENCIÓ ara, que s'utilitza un procediment "informal" !

3. Cal "enredar" a *jamovi* per poder veure les freqüències de variables CONTINUES: Fes doble clic a el nom de p15 que encapçala la columna de dades, **canvia el tipus de variable** a ORDINAL. Fes el mateix amb la variable p15a. S'actualitza el panell de resultats a la dreta, amb les dues taules de freqüències.
4. Pots veure, a la part inferior de totes dues taules, com la variable p15 té valors perduts 9998 i 9999, mentre que la p15a té 9997, 9998 i 9999. Recorda que el valor perdut més petit és el **9997**. Aquesta informació serà fonamental més endavant.
5. Ja pots "retornar" les variables p15 i p15a al seu tipus correcte: CONTINUOUS i INTEGER. En el moment que ho facis desapareixeran les taules de freqüència de la finestra de resultats.
6. Clic al nom de la variable p15a, que encapçala la columna de dades, per seleccionar-la
7. Clic-dret al ratolí, per accedir al menú contextual. Selecciona l'opció "Add Variable". Selecciona l'opció "Computed Variable / Insert".

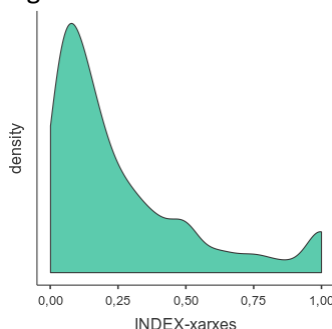


8. Es crea una nova columna a l'esquerra de la variable seleccionada p15a, per defecte anomenada amb una combinació de dues lletres.
9. Canvia el nom a "INDEX-xarxes". Pots incloure una descripció per recordar que la nova variable està calculada a partir de p15 i la p15a, p.e. "*Index de "fiabilitat de xarxes" (amics / contactes totals)*".
10. Fes clic a la finestra de fórmules, on cal escriure la funció que realitza el càlcul de la nova variable.
11. Cal afegir el següent criteri: "*Fes un càlcul dividint p15a per 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.
12. Comprova com ara la nova columna conté dades entre 0 i 1, però hi ha un problema: no apareixen els decimals. Això és degut a que les dues variables originals del càlcul estan definides com a CONTINUOUS/INTEGER. Només cal canviar el tipus de dades d'una de les variables originals per solucionar aquest problema.
13. Fes clic a el nom de qualsevol de les variables originals, que encapçala la columna de dades. Fes clic a el desplegable "Data type" que actualment indica "Integer" i selecciona "Decimal".
14. Solucionat, ara la nova variable conté les dades del índex calculat, amb decimals.

15. Fixat en la segona fila, **hi ha un problema**: La variable p15 té un valor perdut (que no ha estat eliminat o filtrat) i que s'ha tingut en compte pel càlcul, erròniament. Cal filtrar les variables "d'entrada" per què cap dels valors perduts de les dues variables originals sigui tingut en compte en fer el càlcul per a la nova variable.
16. Per eliminar els valors perduts 9997, 9998 i 9999, cal *"filtrar la variable p15 de forma que només s'acceptin els valors menors a 9997"* i cal *"filtrar la variable p15a de forma que només s'acceptin els valors menors a 9997"*. Aquests dos criteris s'escriuen, respectivament, en llenguatge de programació comprensible per *jamovi*:
`FILTER(p15, p15 < 9997)`
`FILTER(p15a, p15a < 9997)`
17. Per tant cal substituir la fórmula original, simple però errònia, per una que indiqui el següent criteri: *"Calcula el resultat dels valors de p15a per sota 9997, dividit pels valors de p15 per sota de 9997"*. A la finestra de fórmules, escriu la nova versió, que combina la funció de filtratge de variables amb la operació matemàtica entre variables:
`FILTER(p15a, p15a < 9997) / FILTER(p15, p15 < 9997)`
18. Executa el càlcul prement la tecla [Intro]. Apareixen a la nova columna els valors calculats.
19. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

Pots comprovar la construcció d'aquest índex, fes una descripció exploratòria a la que es demani una representació gràfica de la funció de densitat de la distribució. Hauria d'estar entre 0 i 1 i ajustar-se a una distribució com la següent:



Imatge 73. Diagrama de densitat d'indicador calculat



EXERCICI

- Realitza un altre exemple de càlcul utilitzant diverses variables: utilitza p13a01 i p13a02 (hores diàries i minuts diaris, respectivament, que l'individu està connectat a les xarxes socials) per tal de crear una nova variable (CONNECT-xarxes) que siguin els minuts totals que l'individu es passa connectat a les xarxes socials. Per passar d'hores a minuts, òbviament, cal multiplicar les hores per 60.
- Així, la "fórmula" o "funció" que cal utilitzar per fer la transformació seria:

$$\text{CONNECT-xarxes} = (\text{p13a01} * 60) + \text{p13a02}$$

- En el cas d'una persona enquestada que respon que passa 1 hora (p13a01 == 1) i 30 minuts (p13a02 == 30) diàriament connectat a xarxes socials, la transformació corresponent que genera el valor per a la nova variable és $(1 * 60) + 30 = 90$ minuts.
- Tingues en compte el problema fonamental del tractament dels valors perduts. Hi ha diferents formes de resoldre el problema: una possibilitat és crear noves variables "sense

perduts" i calcular la nova a partir d'aquestes variables, una altra és incorporar filtres a la fórmula del càlcul tal i com s'ha mostrat anteriorment, p.e. `FILTER(p13a01,p13a01<99)*60+FILTER(p13a02,p13a02<99)`

Transformacions complexes (1): Combinació de recodificacions i càlculs

Sovint ens pot interessar construir índexs a partir de transformacions de variables NOMINAL 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.

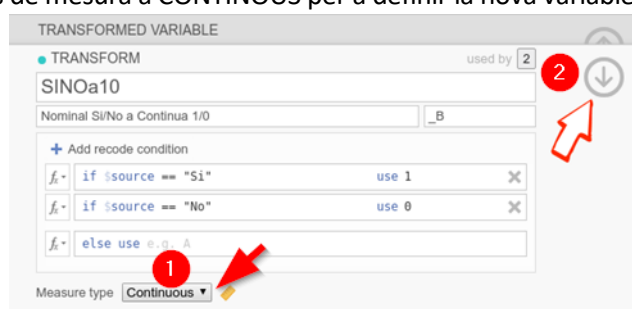
Si en comptes de "Si" o "No", tinguéssim valors numèrics 1 i 0, respectivament, seria molt senzill, en tant que sumariem directament les 14 variables.

1. Per això, en primer lloc, el que fem és TRANSFORMAR les variables inicials (de la p901 a la p914) en 14 noves variables amb uns o zeros, en funció de si en fa ús o no. Es tracta de transformar variables qualitatives (sí-no) a variables quantitatives (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 CALCULAR sumant aquestes variables i obtenir la variable resultat. La nova variable resultat serà quantitativa (tipus 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 dividim per 14 s'obtindrà 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 existent a les altres variables.

1. Selecciona la variable p901, amb doble clic el nom que encapçala la columna. Fes clic a la etiqueta "Data" del menú superior. Fes clic a la icona "Transform" del sub-menú 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. No cal canvia el nom, però pots incloure una descripció més expressiva, com ara: "Gestiones bancarias por Internet".
3. A continuació cal crear la nova transformació requerida en aquest cas per fer la recodificació de variable qualitativa nominal (p901, 2 categories) a variable quantitativa sencera (p901_c, 0 o 1). Fes clic 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** per que sigui més significatiu, canvia "Transform X" per "SINOa10". Si vols pots introduir a sota una descripció més detallada, com ara "Nominal Si/No a Continua 1/0".
6. A continuació, a la dreta de la descripció cal **introduir el sufix** amb el que s'identificaran les noves variables que siguin transformades per aquesta regla: es pot utilitzar "_B" (sense les cometes) per destacar que la variable transformada serà la versió numèrica Binària (0,1).

7. 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*. Fes clic a 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, fent clic a el botó [+ Add recode condition]. Les següents condicions en llenguatge de programació segueixen la mateixa lògica:
IF \$source == 'No' use 0
10. La condició final, que afecta a tots els valors de la variable original que no han estat encara canviats (els perduts), es pot expressar programàticament com: `else use .` Deixa en blanc la condició final. Això produirà coma resultat que els valors diferents de SI o NO quedin en blanc, i per tant són considerats perduts.
11. Canvia el tipus de mesura a CONTINUOUS per a definir la nova variable com a quantitativa.

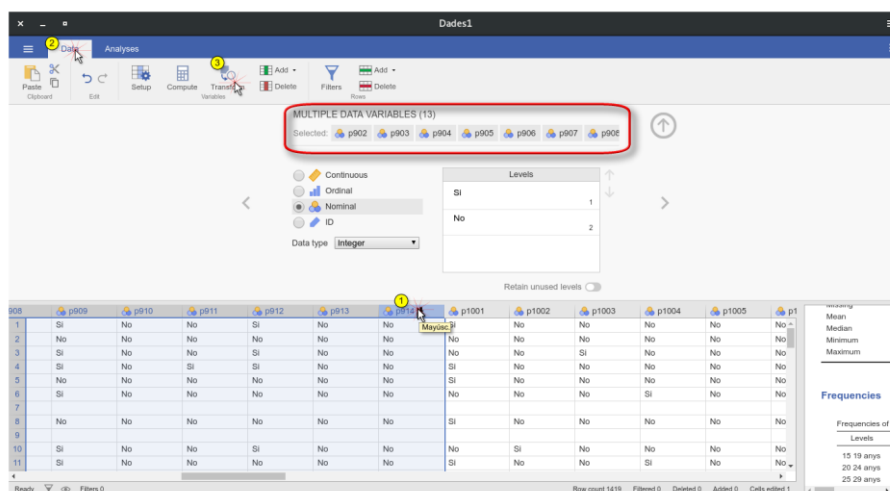


Imatge 74. Transformació de variable a numèrica binària (0,1)

12. 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.
13. Com pots veure ara 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.
14. 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, per cada una de elles, es generi 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:

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



Imatge 75. Selecció de conjunts de variables (per aplicar la mateixa transformació)

- Has fet una selecció múltiple de variables. Per a aplicar la mateixa transformació a totes elles, al menú principal superior fes clic a l'apartat "Data" i clic a l'opció "Transform". Al panell de transformacions múltiples, fes clic a 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. Pots tancar el panell de transformació fent clic a 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] totes amb el mateix tipus de variable i dades.

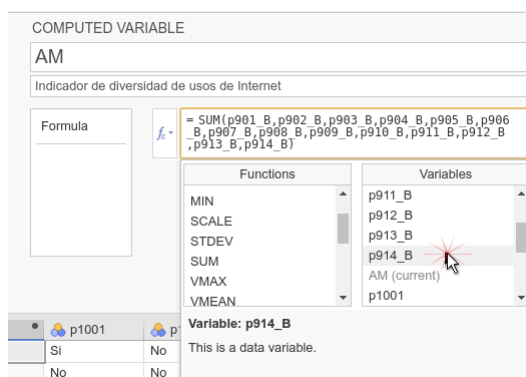
Per fer selecció múltiple també es pot combinar la tecla [Ctrl] amb clic, d'aquesta forma es pot seleccionar un conjunt format per variables no adjacents.

Finalment, tal i com s'ha expressat més amunt, caldrà de 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 *_B*, i (2) l'encapçalament de la columna té un petit punt (blau) que identifica la regla de transformació utilitzada.

Per a "sumar" els valors (1 i 0) de cada fila en una columna nova cal CALCULAR una nova variable (que es pot anomenar, p.e., "p9sum"). Cal seguir els següents passos:

- Vols crear una nova variable calculada, ubicada entre la p914_B i la p1001. Fes clic a el nom, que encapçala la columna, de la variable p1001.
- Clic a Menú / Data / Add / Computed Variable Insert. Apareix la nova columna de dades.
- Posa-li un nom curt, p.e., p9sum.
- Posa-li una descripció més descriptiva, p.e., "Indicador de diversidad de usos de Internet".
- En la ventana de fórmula cal incloure la operació algebraica que ha de ser realitzada amb els valors de les 14 variables quantitatives binaries. Es pot utilitzar la funció SUM() que està predeterminada en *jamovi* i aprofitar el sistema de construcció assistida de fórmules.
- Per a seleccionar la funció sumatori: fes clic a el botó central [$f_x =$]. Al desplegable, a la part esquerra "Functions", desplaça cap a baix fins a l'apartat "Statistical" on trobaràs la funció SUM. Fes clic per seleccionar-la; una finestra inferior mostra la seva funció.

7. Fes doble-clic sobre la funció SUM. El sistema *jamovi* la escriu la funció 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 =]. Desplaçat pel llistat dret de “Variables” fins que trobis la “primera”: p901_B. Fes-hi doble-clic per escriure-la a la finestra de fórmula.
9. Pica la tecla coma [,]. Torna a utilitzar el botó [f_x =] i el llistat de variables per a trobar la segona variable a sumar (p902_B). 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_B” que es vol sumar.



Imatge 76. Exemple de funció SUMATORI amb un conjunt de variables

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

Quan es fa un càlcul per generar noves variables, una fila tindrà valor “perdut” (en blanc) si qualsevol de les columnes originals que intervenen en la operació té 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 quasi tot (12 usos diferents). Sembla que la major part de persones enquestades utilitzen Internet per aproximadament 5 funcionalitats diferents.

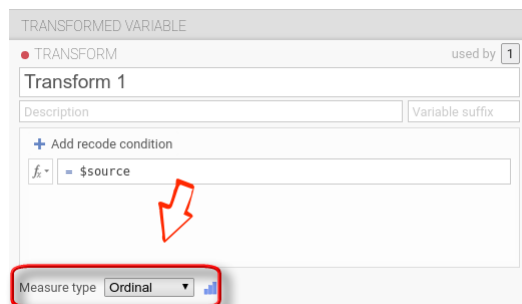
- Pots provar d'explorar descriptivament aquesta nova variable que acabes de construir?
- Pots veure la seva distribució de freqüències?

PERÒ, com s'aconsegueix fer taula de freqüències de variables quantitatives?

Una variable “sumada” mitjançant càlcul a partir d'altres (continues) és de tipus continu, i no hi ha forma de canviar el seu tipus. Per tant, no es podrà demanar una taula de freqüències.

Truc!

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 afegir una **nova** variable transformada que sigui una copia d'ella mateixa (= \$source) tot modificant el tipus de mesura a la part inferior de la pantalla de transformació.

***Transformacions complexes (2): Construcció de tipologies***

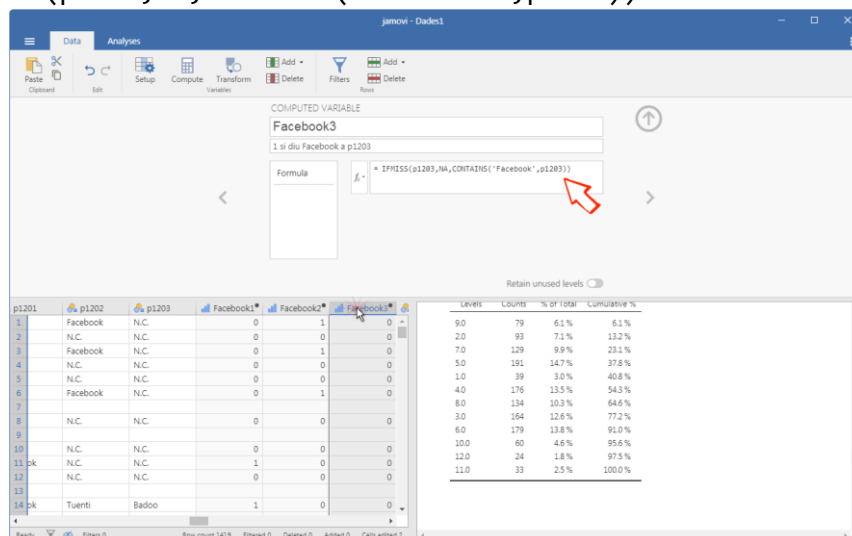
Una nova variable tipològica, o classificatòria, està construïda a partir d'informació continguda a més d'una variable original, de forma que cada categoria de la tipologia correspon a una determinada combinació de categories de les variables originals.

Dins del procés de creació de noves variables mitjançant TRANSFORMACIÓ, anteriorment has pogut veure com s'utilitza la funció condicional (IF): el valor de la nova variable transformada depèn del valor existent a la variable 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 a l'apartat anterior, sinó que implica **operacions lògiques o condicionals**. Aquesta funció IF és fonamental en la construcció de “tipologies”.

Exemple 1. 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 per separat i es vol crear una única variable que digui “*Si utilitza Facebook o no*”, per exemple. 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.

1. En forma de condició textual això es pot expressar com: “*Si el valor de la variable p1201 és ‘Facebook’ o el valor de la variable p1202 és ‘Facebook’ o el valor de la variable p1203 és ‘Facebook’, llavors el valor de la nova variable serà ‘Sí’, i en qualsevol altre cas el valor de la nova variable serà ‘No’*”.
2. En llenguatge de programació que jamovi pot entendre i executar, això és complicat de fer en un sol pas. Per això es procedirà combinant dos procediments en una seqüència d'operacions que acabarà produint el resultat:
(1er) Creant 3 noves variables quantitatives, amb valor 1 si el text “Facebook” apareix a la corresponent variable original, amb compte de deixar en blanc els casos “perduts”, i

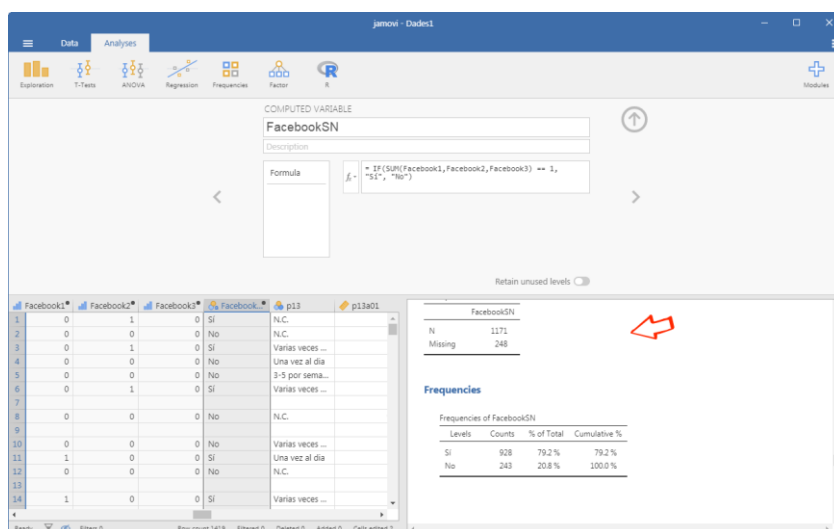
- (2on) Creant una variable final que assigni valor "Sí" si a la fila la suma de totes tres variables numèriques és igual a 1, i assigna valor "No" en cas contrari.
3. Selecciona les tres variables originals: clica el nom que encapçala la primera columna (p1201) i "arrossega" cap a la dreta fins arribar a la última (p1203). Han quedat destacades les tres columnes de dades.
 4. Clic a Menú / Data / Setup. S'ha obert el panell de definició. Comprova que les variables estan correctament definides: han de ser NOMINALS i amb dades tipus TEXT.
 5. Clic a Menú / Data / Compute. S'han creat tres noves variables a continuació de les originals i s'ha obert el panell de càlcul per a múltiples variables.
 6. A la finestra de fórmula escriu la funció següent i pica [Intro] al teclat:
CONTAINS(' Facebook ' , p1201)
 7. Pots comprovar a la matriu de dades que el perduts de les variables originals ara tenen valor 0. Això és incorrecte, caldrà utilitzar la funció IFMISS() perquè *jamovi* identifiqui els perduts i els deixi en blanc, i llavors executi la funció CONTAINS() només en el cas que no siguin valors perduts. Canvia, o modifica, la fórmula anterior fins que quedi així, i pica [Intro]:
IFMISS(p1201,NA,CONTAINS(' Facebook ' , p1201))



Imatge 77. Fórmula per a identificar si un text està contingut a una variable

8. Ara hi ha files en blanc; són els perduts, és correcte.
9. Clic al nom de la primera nova columna. Al panell de càlcul canvia el nom per 'Facebook1'. Si vols pots afegir una descripció més explicativa, p.e. "1 si diu Facebook a p1201".
10. Clic al nom de la segona nova columna. Al panell de càlcul canvia el nom per 'Facebook2'. Si vols pots afegir una descripció. Clic a la finestra de fórmula i canvia el nom de la variable utilitzada (canviar la darrera xifra, de 1 a 2). Pica [Intro].
11. Fes el mateix amb la tercera i última nova variable.
12. Ara clica el nom de la darrera variable creada "Facebook3". Clic al menú Data / Compute. S'han creat una nova variable a continuació i s'ha obert el panell de càlcul.
13. Anomena-la, p.e., "FacebookSN" (per "Facebook sí o no", o per "Facebook suma nombres"...))
14. El criteri que cal utilitzar ara s'expressa literalment com: "Si la suma dels valors de les variables Facebook1, Facebook2 i Facebook3 és 1 llavors assigna 'Sí' a la nova variable, en qualsevol altre cas assigna 'No'".
15. A la finestra de fórmula, escriu la funció següent, i pica [Intro] al teclat:
IF(SUM(Facebook1,Facebook2,Facebook3) == 1, "Sí", "No")
16. Recorda desar l'arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general [≡] / Save.

A la distribució d'aquesta nova variable (pots utilitzar *RjEditor* per a executar la ordre de *R summary(data\$FacebookSN)*, o pots fer *Analyses / Descriptives / Explore...* i activar la taula de freqüències) veiem que 928 persones fan servir *Facebook*, 243 no en fan servir, i hi ha 248 perduts. Aquests es corresponen a persones que no fan servir cap xarxa social.



Imatge 78. Resultat de càlcul de reconeixement de text a conjunts de variables

Exemple 2. Una altra estratègia de construcció de tipologies consisteix en treballar amb variables amb dades de tipus TEXT, que cal “combinar” en quant a categories per poder posteriorment fer una recodificació que generi la variable final. Així, per exemple, en creuar l’estat civil (p41) amb la convivència en parella (p42) s’obté la següent taula de distribució de freqüències creuades:¹⁸

Contingency Tables

p41	p42			Total
	Si	No	N.C.	
Soltero/a	219	1029	5	1253
Casado/a	149	3	1	153
Separado/a	1	7	0	8
Divorciado/a	0	4	0	4
Viudo/a	0	1	0	1
Total	369	1044	6	1419

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=Casado/a, en groc a la taula anterior) dels que viuen junts en cohabitació (en verd a la taula anterior)

Per fer-ho cal procedir en una seqüència de tres operacions combinades:

1. Crear dues noves variables, còpies de les originals però SENSE els valors perduts.
2. Crear una nova variable calculada, que COMBINI el text de categories de les dues originals. Es tracta d’ajuntar, o concatenar, les lletres de les categories nnominals/text. Per exemple,

¹⁸ Pots veure la generació d’aquest tipus de taules a l’apartat “Inferència amb una variable utilitzant *jamovi*” d’aquest tutorial.

en el cas de p41 i p42, els valors originals “Vidu/a” i “No” es combinaran en el valor final “Vidu/aNo”, i “Casado/a” i “Sí” es combinaran en “Casado/aSí”, etc.

3. Crear una nova variable transformada a partir de la variable combinada, que assigni els valors definitius a les categories resultants.

Amb aquest mètode genèric es pot generar qualsevol tipologia, mantenint el control detallat sobre cada decisió teòrica que justifiqui aquesta construcció.

1. Clic al nom de variable p41 i “arrossega” fins al nom de variable p42. Activar, si cal, el panell de definició de variables múltiples, amb clic a Data / Setup. Comprovar que totes dues variables originals són NOMINALS i amb dades tipus TEXT.
2. Amb les dues variables p41 i p42 seleccionades, clic a Data / Transform per a obrir el panell de regles de transformació. A “using transform” clic a [None] i al desplegable seleccionar la regla pre-existent “Eliminar valores N.C.” (que es va crear anteriorment i es re-utilitza ara).
3. Doble-Clic al nom de CADASCUNA les noves variables, per poder canviar el nom a p41_a i p42_a.
4. Clic al nom de p42_a, i clic a Data / Compute, per a crear una nova variable calculada. Canvia el nom a p4142. A la finestra de fórmula escriu la funció per a combinar text:
p41_a + p42_a
Comprova com s’han combinat els textos en una sola columna.
5. Clic a la nova variable p4142, i clic a Data / Transform, per a crear una nova variable recodificada. Canvia el nom de la variable final a “Cohabitacion”. A “using transform” clic a [None] i al desplegable seleccionar “Create New Transform”.
6. Canvia el nom a “Cohabit”. Afegeix totes les regles següents, clicant a [+] Add recode condition:
if \$source == “Casado/aSi” use “Matrimoni”
if \$source == “Soltero/aSi” use “Cohabita”
if \$source == “Separado/aSi” use “Cohabita”
if \$source == “Divorciado/aSi” use “Cohabita”
if \$source == “Viudo/aSi” use “Cohabita”
else use
7. Com pots veure, executant aquesta transformació la nova variable “Cohabitació” adopta els valors “Matrimoni” o “Cohabita” o perdut en funció de la combinació de les p41 i p42 originals.
8. Recordar desar l’arxiu *jamovi* freqüentment per no perdre el què vas fent, amb Menú general  / Save.

El resultat és la taula següent, com pots comprovar si fas Analyses / Exploration / Descriptives demanant taules de freqüències:

Descriptives

Cohabitacion	
N	369
Missing	1050

Frequencies of Cohabitacion

Levels	Counts	% of Total	Cumulative %
Matrimoni	149	40.4 %	40.4 %
Cohabita	220	59.6 %	100.0 %

Exemple 3. Fins ara has vist com s'han combinat variables numèriques (sumant, com al ex.1) o variables nominals (combinant text, com al ex.2). Amb aquest exemple veuràs una altra estratègia, utilitzada en el context de variables ORDINALS, que d'una banda tenen “etiquetes” (levels) que són textuals i per una altra tenen dades que són numèriques i, fins a cert punt, operables algebraicament.

Per exemple, volem construir una tipologia de “Homogamia” (*emparellament entre persones de característiques similars*) i una de les seves dimensions és la educativa: hi ha parelles homogàmiques en les que tots dos tenen el MATEIX nivell educatiu, i d'altres parelles amb “superioritat del pare” o “superioritat de la mare”. Cal treballar amb les variables “Nivell de estudis del pare” (p53) i “Nivell de estudis de la mare” (p54). Aquestes variables tenen molts valors textuals i bastant llargs, però poden permetre un conjunt reduït d'operacions que pot ser més pràctic utilitzar en determinades situacions.

p54	p55				N.S.	N.C.	Total
	Menos de primaria	Primaria	Secundaria	Universitarios			
Menos de primaria	79	22	4	1	1	0	107
Primaria	20	492	87	26	1	2	628
Secundaria	4	63	232	50	3	2	354
Universitarios	3	28	76	131	3	1	242
N.S.	2	9	7	6	37	0	61
N.C.	0	7	7	3	0	10	27
Total	108	621	413	217	45	15	1419

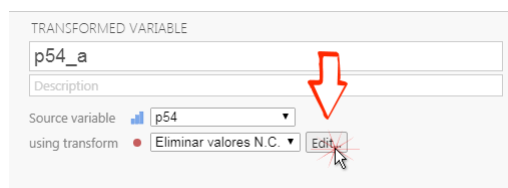
Prèviament, cal assegurar-se **eliminar els valors perduts** a les variables que s'utilitzaran:

1. Clic al nom de variable p54 i “arrossega” fins al nom de variable p55. Activar, si cal, el panell de definició de variables múltiples, amb clic a Data / Setup. Comprovar que totes dues variables originals són **ORDINALS** i amb dades tipus **INTEGER** (Han de ser INTEGER per poder fer els càlculs però sense perdre el seu ordenament).



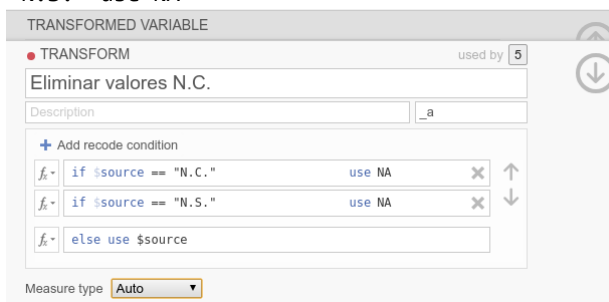
Imatge 79. Definició del tipus per a múltiples variables

2. Amb les dues variables p54 i p55 seleccionades, clic a Data / Transform per a obrir el panell de regles de transformació. A “using transform” clic a [None] i al desplegable seleccionar la regla pre-existent “Eliminar valores N.C.” (que es va crear anteriorment i es re-usa ara).
3. Originàriament la regla de transformació “Eliminar valores N.C.” realment elimina només els valors “N.C.”, però en el cas actual també hauria d'eliminar els “N.S.” com a perduts. Es pot modificar una regla de transformació en qualsevol moment, com faràs ara.
4. Al costat del desplegable que indica la regla aplicada hi ha un botó per a modificar-la



Imatge 80. Accés a finestra d'edició de regles de transformació

5. Clic al botó [Edit...]. S'obre el panell de transformació. Per a afegir una nova categoria que quedarà com a valor perdut, fes clic a [+] Add recode condition
6. Afegeix la categoria que addicionalment serà considerada com a perdut:
if \$source == "N.S." use NA



Imatge 81. Finestra de definició de regla de transformació

7. A la matriu de dades, baixa fins a la fila 197, i comprova si ha funcionat la transformació.
8. Pots veure com s'ha "perdut" la informació sobre què significa cada categoria de les noves variables. Això és degut a la transformació i al tipus de les dades originals (integer), però NO és un problema perquè aquestes variables ara seran sotmeses a operació algebraica, i per tant només interessa el número d'ordre, i no el text que etiqueta la categoria.

jamovi, igual que R, utilitza NA per fer referència als valors perduts (not available), tot i que també es pot deixar en blanc per a denotar el mateix.

En creuar ambdues variables després d'eliminar els possibles N.S. i N.C. s'obté la distribució:

	p55_a				
p54_a	1	2	3	4	Total
1	79	22	4	1	106
2	20	492	87	26	625
3	4	63	232	50	349
4	3	28	76	131	238
Total	106	605	399	208	1318

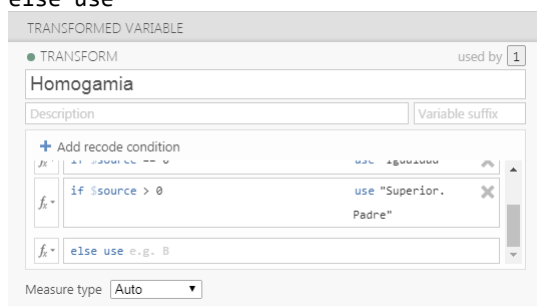
La nova variable "Homogamia" cal que compari el nivell d'estudis d'ambdós progenitors i que contingui tres valors: igualtat d'estudis ($P=M$, que es correspon a les cel·les en **groc**), el pare més que la mare ($P>M$, cel·les en **verd**) i el pare menys que la mare ($P<M$, cel·les en **blau**). Per construir-la tindrem en compte que es tracta d'una variable ordinal i que és funció del resultat de la diferencia numèrica entre les dues versions numèriques ($p54_a - p55_a$) de les variables originals:

- Si la diferencia és 0, llavors hi ha "Igualtat"

- Si la diferencia > 0 , llavors hi ha “Superioritat del pare”
- Si la diferencia < 0 , llavors hi ha “Superioritat de la mare”

Per a crear la nova variable tipològica a partir de les dues variables p54_a i p55_a:

1. Clic al nom de p55_a, i clic a Data / Compute, per a crear una nova variable calculada. Canviar el nom a “p5455_diff”. A la finestra de fórmula escriu la funció corresponent:
p54_a - p55_a
Pica la tecla [Intro] i comprova com s’ha operat amb les dades originals.
2. Clic a la nova variable p5455_diff, i clic a Data / Transform, per a crear una nova variable recodificada. Canvia el nom de la variable final a “Homogamia”. A “using transform” clic a [None] i al desplegable seleccionar “Create New Transform”.
3. Canvia el nom a “Homogamia”. Afegeix les regles, clicant a [+] Add recode condition:
if \$source < 0 use “Superior. Madre”
if \$source $== 0$ use “Igualdad”
if \$source > 0 use “Superior. Padre”
else use



4. Com pots veure, executant aquesta transformació la nova variable “Homogamia” adopta els seus valors en funció de la combinació de les p54 i p55 originals i segons la lògica tipològica indicada.
5. 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, per exemple, fent clic a Analyses / R / Rj Editor... i escrivint la simple ordre en llenguatge R: `summary(data$Homogamia)`. Recorda que per executar una ordre R cal picar una combinació de 3 tecles [Ctrl] + [Mayúsc.] + [Intro].

Descriptives

Homogamia	
N	1318
Missing	101

Frequencies of Homogamia

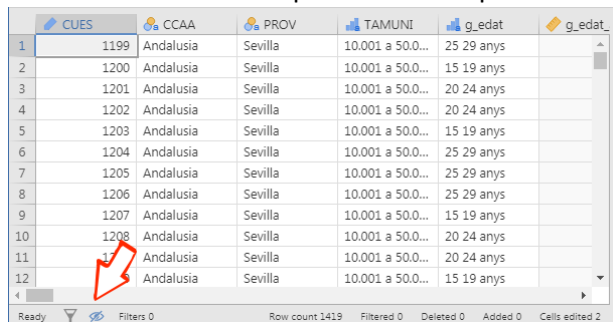
Levels	Counts	% of Total	Cumulative %
Superior. Madre	190	14.4 %	14.4 %
Igualdad	934	70.9 %	85.3 %
Superior. Padre	194	14.7 %	100.0 %

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 la edat (es a dir, només els casos que no es "desvien" de l'edat mitjana més de 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.

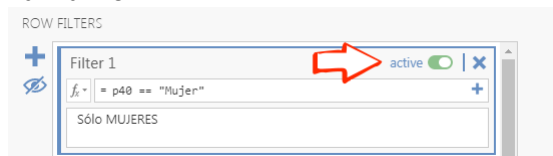


	CUES	CCAA	PROV	TAMUNI	g_edat	g_edat
1	1199	Andalusia	Sevilla	10.001 a 50.0...	25 29 anys	
2	1200	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	
3	1201	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
4	1202	Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
5	1203	Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	
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		Andalusia	Sevilla	10.001 a 50.0...	20 24 anys	
12		Andalusia	Sevilla	10.001 a 50.0...	15 19 anys	

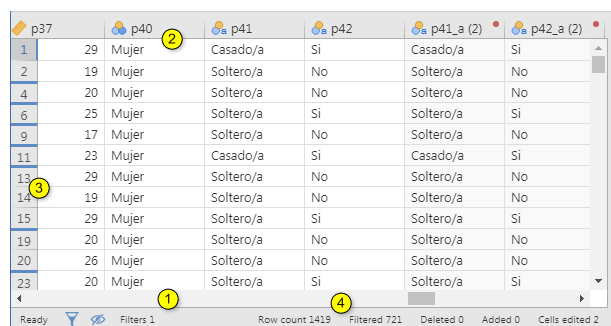
Imatge 82. Ubicació del control de filtratge de casos (selecció)

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

1. Clic a 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 == "Mujer"
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 *.omv



4. Qualsevol modificació sobre la fórmula d'un filtre triga un temps en fer-se efectiva, perquè *jamovi* re-calculta totes les transformacions i anàlisis de l'arxiu per ajustar-los a la sub-mostra 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.

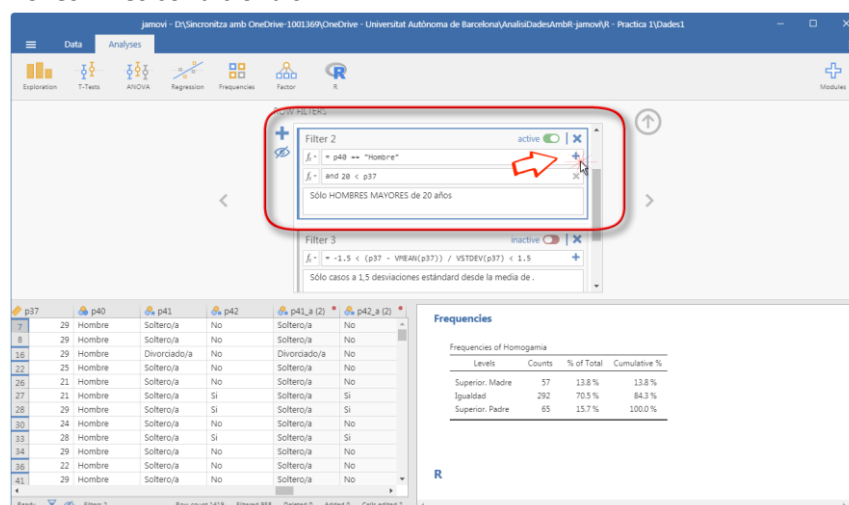


	p37	p40	p41	p42	p41_a (2)	p42_a (2)
1	29	Mujer	Casado/a	Si	Casado/a	Si
2	19	Mujer	Soltero/a	No	Soltero/a	No
4	20	Mujer	Soltero/a	No	Soltero/a	No
6	25	Mujer	Soltero/a	Si	Soltero/a	Si
9	17	Mujer	Soltero/a	No	Soltero/a	No
11	23	Mujer	Casado/a	Si	Casado/a	Si
13	29	Mujer	Soltero/a	No	Soltero/a	No
14	19	Mujer	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 83. Exemple de filtre activat

Desactiva tots els altres filtres abans de crear un de nou. Pot haver conflictes entre filtres tals que *jamovi* entra en un “bucle” que no es pot resoldre i bloqueja tot el sistema.¹⁹

Es pot aplicar una condició de filtratge complexa, o combinada, utilitzant el botó lateral dret [+] per introduir noves línies condicionals.



Imatge 84. 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 estadístics) tinguin en compte només als **homes majors de 20 anys**:

1. Clic a la icona de filtratge. S'obre el panell de filtres.
2. Escriu la primera condició que han de satisfer els casos que quedaran seleccionats:
 $p40 == \text{"Hombre"}$
3. Clic al botó dret blau [+] per afegir condicions.
4. Escriu la segona condició que han de satisfer els casos que quedaran seleccionats:
 $20 < p37$
5. Òbviament, cal que les condicions siguin compatibles, per evitar situacions de “bloqueig”.
6. 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 minuts...) en fer-se efectiva, perquè *jamovi* re-calcu la totes les transformacions i anàlisis de l'arxiu per ajustar-los a la sub-mostra seleccionada.

Tots els filtres, actius o no, queden desats en l'arxiu *.omv juntament amb totes les dades, transformacions i resultats de forma que es poden recuperar més endavant i en diferents sistemes.



EXERCICI

- Crea un filtre amb el següent criteri: $-1.5 < (p37 - VMEAN(p37)) / VSTDEV(p37) < 1.5$.
- Activa'l i comprova com afecta als valors extrems (mínim i màxim) de les distribucions.

¹⁹ Similarment al “robloqueo” que pateixen els robots als relats d'Isaac Asimov (1950).

- 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

Errors en variables individuals

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.

Per avaluar possibles errors uni-variable cal estudiar les freqüències de totes les variables, tant de les quantitatives (CONTINUOUS) com les qualitatives (NOMINAL o ORDINAL). La simple observació d’aquestes freqüències permet detectar valors impossibles o altament improbables.

1. Clic a Analyses / R / Rj Editor... Escriu la ordre corresponent de R:
`summary(data)`
2. Clic al 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 la freqüència de la variable p1001. Per error, s’han entrat els codis “SI” i “si”, en comptes del correcte “Sí”.

1. Clic a Analyses / R / Rj Editor... Escriu la ordre corresponent de R:
`table(data$p1001)`
2. Clic al triangle verd a la part superior dreta del panell consola R, per executar-la.

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 casos que només continguin aquests dos casos.

1. Clic a la icona de filtratge. S’obre el panell de filtres. Si cal desactivar-los tots.
2. Escriu la primera condició que han de satisfer els casos que quedaran seleccionats:
`p1001 == "SI"`
3. Clic al botó dret blau [+] per afegir condicions.
4. Escriu la segona condició que han de satisfer els casos que quedaran seleccionats:
`p1001 == "si"`
5. Els casos buscats ja estan identificats i a la vista.

L’acció requerida depèn de si es disposa o no del suport original de les dades (qüestionari): mitjançant l’identificador de qüestionari (variable CUES) es pot consultar el valor real al qüestionari en paper i corregir la matriu de dades.

Per detectar un error en una variable **quantitativa** hem de mirar més atentament els valors un cop hem fet les freqüències i comprovar si hi ha valors impossibles.

1. És el que succeeix en la variable p6 (edat que tenia quan comença a utilitzar Internet), que no hi pot haver cap valor superior a 29 (edat màxima dels entrevistats).

Errors en creuaments de variables

Per avaluar possibles errors que afecten a 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 a 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).
- En 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 que ha contestat que no fa servir mai les xarxes socials ha respost aquesta pregunta, mentre que si creues la p11 amb la p1403 sí que hi ha una incongruència, algú 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 R executades des de la consola.

1. Clic a Analyses / R / Rj Editor... Escriu les ordres corresponent de R:
`table(data$p1401,data$p11)`
`table(data$p1403,data$p11)`
2. Clic al triangle verd a la part superior dreta del panell consola R, per executar-les.

Imatge 85. Exemple de control d'incongruència a les dades, mitjançant taules creuades

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 *P5_Dades2.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 respostes "N.S." i "N.C.". Reconverteix aquestes respostes a valors perduts. Enganxa en un processador de textos la freqüència d'aquesta variable abans de la conversió i després.

Fes el mateix amb el total de contactes que tenen a les xarxes socials (p15). En aquest cas enganxa en el processador informació sobre la mitjana d'amics abans d'eliminar els valors perduts i després. Inclou també el valor més petit i el valor més elevat que pren la variable.

(b) Observa com és la variable p13. Volem crear una nova variable que tingui només dues categories, aquells que fan servir les xarxes socials cada dia, i els que les fan servir amb una freqüència menor. Crea aquesta variable i comprova fent freqüències de la nova variable i de la vella que el resultat és correcte. Enganxa ambdós resultats en un processador de textos.

(c) 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?
- L'10% dels que es connecten més estona quantes hores es connecta?
- Quant val la mediana?
- I el coeficient de variació de la variable?

(d) A la pregunta 10 del qüestionari 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 **quantas activitats** ha afectat Internet. Mostra la taula de freqüències de la nova variable i interpreta-la.

(e) 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. Enganxa en un processador de textos una freqüència de la nova variable.

(f) A partir de la informació disponible a les variables p41 i p42, crea una variable que, a més de les categories "Matrimoni" i "Cohabitant", inclogui una categoria dels que "No conviuen en parella" (1.044 persones). Enganxa en un processador de textos la freqüència d'aquesta nova variable.

(g) 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. Comprova fent

frequències de la nova variable que el resultat és correcte. Tingues en compte que no hi hagi valors perduts (9998, 9999).

(h) A partir de les variables TAMUNI i PROV, filtra la base de dades perquè que contingui només els residents a les ciutats de Madrid i de Barcelona. En aquesta nova matriu de dades demana freqüències de la comunitat autònoma i comprova que tots els valors són 0 excepte els corresponents a Catalunya i Madrid. Comprova, a més, que hi ha 42 residents a Barcelona ciutat i 90 a Madrid ciutat. Enganxa aquesta taula en un processador de textos.

(i) Identifica en quins qüestionaris es troben el "SI" i el "si" de la variable p1001. Canvia, en la base de dades original, aquests dos valors per el codi "Si". Copia la freqüència de la variable corregida en un processador de textos.

Comprova que en la p6 hi ha un codi 39 i un de 49 i canvia'ls pel codi 99 (no contesta) copia la freqüència de la variable corregida en un processador de textos.

(j) Identifica quin és el qüestionari que ens dona una incongruència en creuar la p11 amb la p1403. Canvia el valor de la p1403 deixant-lo en blanc (NA). Copia la freqüència de la variable corregida en un processador de textos.

(k) 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.

Abans de fer-ho recorda d'eliminar els valors perduts d'ambdues variables i un cop fet fes una taula de freqüències de la nova variable.

(l) 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.

Abans de fer-ho recorda d'eliminar els valors perduts d'ambdues variables considerant perduts tots els valors superiors a 9 000 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.

(m) 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 respostes N.S. i N.C. Converteix aquestes respostes a valors perduts. Enganxa en un processador de textos la freqüència d'aquesta variable abans de la conversió i després.

Fes el mateix amb el total de contactes que tenen a les xarxes socials (p15). En aquest cas enganxa en el processador informació sobre la mitjana d'amics abans d'eliminar els valors perduts i després. Inclou també el valor més petit i el valor més elevat que pren la variable.

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).

Con en d'altres apartats anteriors, al Tutorial s'utilitza l'arxiu de dades originals "*P4_Dades1.rdata*". 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 2011.²⁰ De totes les variables de l'estudi només se n'han conservat algunes, les que podeu consultar en el document "*Qüestionari.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: "*P4_Dades2.rdata*", i "*Qüestionari.pdf*" corresponents a l'enquesta del CIS.

1. Descarrega aquests dos arxius a una carpeta local del teu PC.
2. Executa l'aplicació *jamovi* al teu ordinador.
3. Per a obrir el fitxer *P4_Dades1.rdata* amb *jamovi*, fes clic el menú d'arxius (dalt, esquerra).
4. Selecciona l'opció Open / This PC. Al panell central, fes clic a [Browse] (dalt, dreta)
5. S'obre una finestra per seleccionar l'arxiu de dades a obrir. Tria l'arxiu "*P4_Dades1.rdata*" que has descarregat al teu PC local.
6. A la part inferior dreta de la pantalla principal apareix una notificació del procés d'obertura d'arxiu i càrrega de dades.
7. Una vegada completat el procés, *jamovi* disposa de la matriu de dades, amb 1419 individus entrevistats i 70 variables.

Recorda que aquest arxiu original NO ESTÀ DEPURAT: pot contenir valors perduts que has de tractar i errors de dades que cal detectar i corregir. En cas que ja hagi DEPURAT aquestes dades en apartats anteriors del present Tutorial, i si prèviament has desat en format OMV la versió depurada de l'arxiu "*Dades2*" ara pots aprofitar-lo obrint directament a *jamovi*.

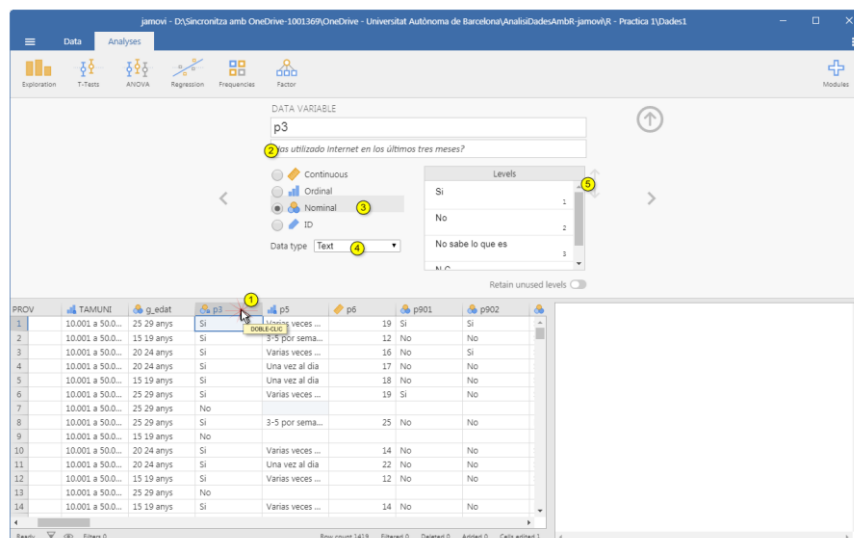
Construcció d'un interval de confiança per a una proporció (π).

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 probablement la proporció corresponent per a tota la població (π) –de la que s'ha extret 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.

²⁰ 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.

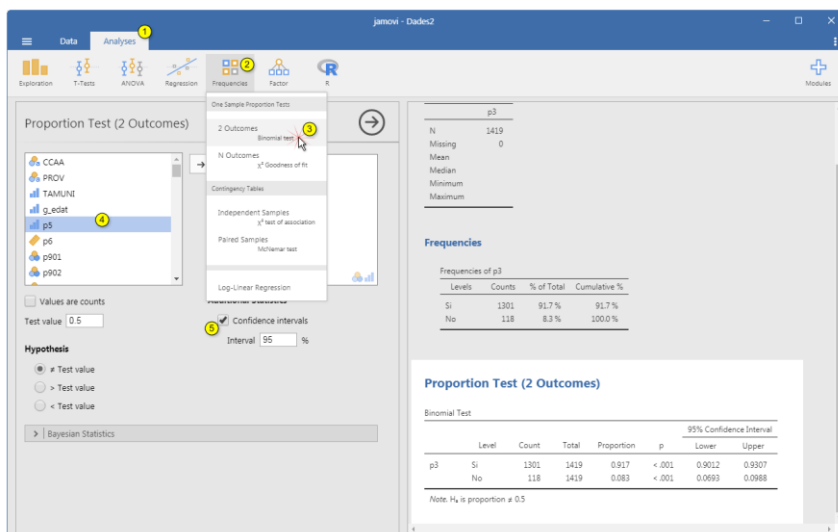
El procediment estadístic per a obtenir intervals de confiança per a proporcions amb el sistema de menús de *jamovi* és un test de comparació d'una proporció observada, anomenat “Test binomial per a 2 resultats” (2 Outcomes Binomial Test).

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.



Imatge 86. Definició correcta de la variable nominal p3

2. Fes ara clic a la etiqueta “Analyses” del menú superior de *jamovi*.
3. Selecciona l'opció “Frequencies”, cap a la dreta del sub-menú 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 amb un clic 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 2 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. Fes clic a 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 "tants per 100" o "percentatges".

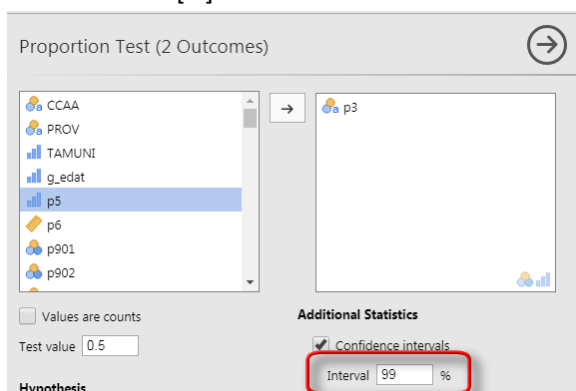
Binomial Test

						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. 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". Si cal, transforma la variable original per poder aconseguir-ho.

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 la 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 ("¿Podrías decirme cuántos contactos, amigos/as, personas agregadas, tienes, en total, en tus redes sociales?") cal fer:

1. Comprova que la variable està definida (Setup) com de tipus CONTINUOUS i dades 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 amb una transformació de la variable.
3. Selecciona a menú *Analyses / T-Tests / One Sample 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. The 'One Sample T-Test' option is selected. The variable 'p15_sp' is added to the 'Dependent Variables' box. In the 'Additional Statistics' section, the following options are checked: 'Mean difference', 'Confidence interval' (set to 95%), and 'Descriptives'. The 'Results' panel on the right displays the 'One Sample T-Test' output table.

One Sample T-Test					95% Confidence Interval	
	Statistic	df	p	Mean difference	Lower	Upper
p15_sp	Student's t	24.6	1080	< .001	249	

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

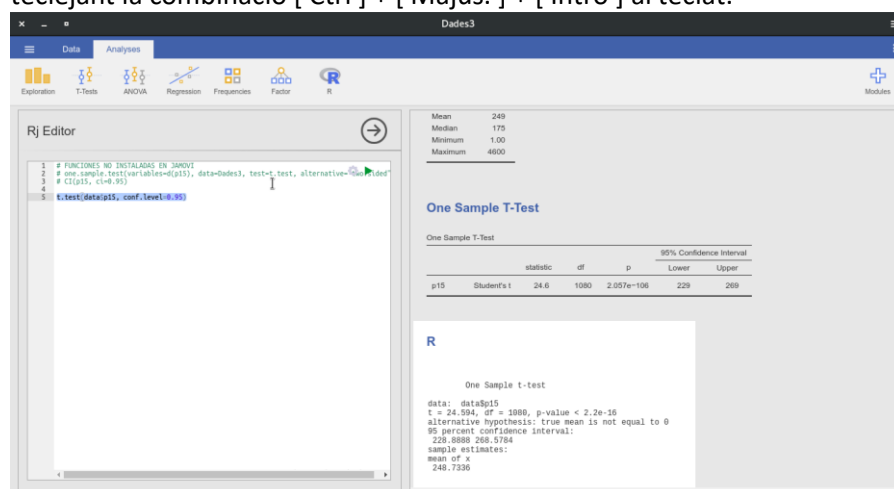
Per tant, es pot afirmar, amb un nivell d'error del 5% o menys (que és el mateix que dir 95% nivell de confiança), que la mitjana de contactes a xarxes socials pel conjunt de la població estarà entre 229 i 269.

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, si la variable és continua cal definir-la "temporalment" com a ORDINAL per poder demanar taula de freqüències.

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. 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. Executa la instrucció R, amb clics el triangle verd de la part superior dreta del panel, o be teclejant la combinació [Ctrl] + [Majus.] + [Intro] al teclat.



Imatge 90. Codi R per interval de confiança d'una mitjana (amb n.c. 99%).

Com pots comprovar, per canviar el nivell de confiança, per exemple al 99% caldrà escriure i executar (en vermell el canvi introduït): `t.test(data$p15, conf.level=0.99)`

De manera que ara el resultat serà:

One Sample t-test


```
data: data$p15_sp
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.

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

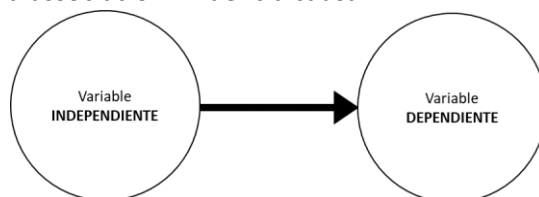
(1) A partir de la pregunta 13 del qüestionari (variable p13) es demana que facis la següent estimació puntual: quin percentatge dels que utilitzen les xarxes socials ho fa com a mínim una vegada al dia? Un cop la tinguis, construeix-ne (i interpreta) un interval de confiança utilitzant un nivell de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

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.



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”.



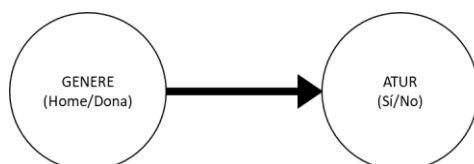
En aquest cas cal utilitzar les eines estadístiques corresponents a l'anàlisi bi-variable (objecte de la resta del Tutorial) o multi-variable.

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
	Anàlisi de regressió. La regressió lineal
	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 91. 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 *Qüestionari.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: “Qüestionari.pdf” i “Dades2.rdata”.
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 i clic a [Browse] (dalt, dreta).
6. S’obre una finestra per seleccionar l’arxiu de dades a obrir. Tria l’arxiu “Dades2.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* ("Dades2.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").

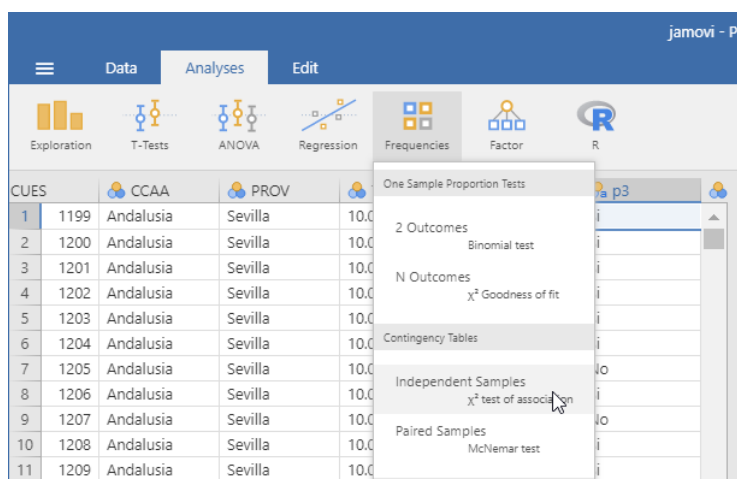
Compte! Fixa't que els termes "independent" i "dependent" (o aparellades), referit a les mostres o sub-grups d'una població que es volen comparar, té un significat TOTALMENT DIFERENT a quan la mateixa terminologia es refereix al paper que juga una variable en la relació bivariable.

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

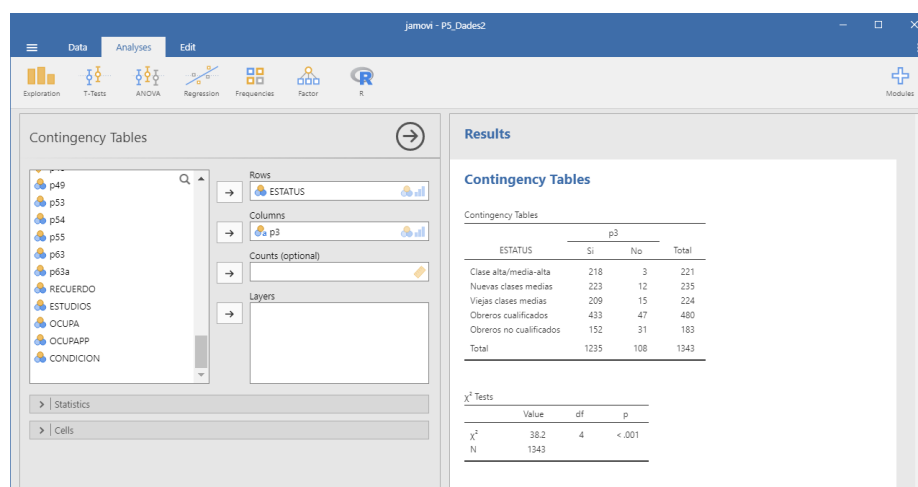
has après anteriorment al Tutorial: pots ajudar-te de l'arxiu "Questionari.pdf" per a aquesta tasca, i per a introduir les descripcions per a cada variable.

2. Selecciona el procediment de Taules per a mostres Independents, a menú "Analyses" / "Frequències" / "Contingency Tables" / "Independent Samples – χ^2 test of association".



Imatge 92. Menú per generar taules de contingència amb mostres independents.

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



Imatge 93. 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

Contingency Tables

ESTATUS	p3		Total
	Si	No	
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.

The screenshot shows the jamovi interface with the 'Contingency Tables' analysis selected. The 'Rows' variable is 'ESTATUS' and the 'Columns' variable is 'p3'. The 'Percentages' section is set to 'Row', indicated by a red arrow. The 'Results' panel displays a contingency table and chi-square test results.

ESTATUS		p3		Total
		Sí	No	
Clase alta/media-alta	% within row	98.6 %	1.4 %	100.0 %
Nuevas clases medias	% within row	94.9 %	5.1 %	100.0 %
Viejas clases medias	% within row	93.3 %	6.7 %	100.0 %
Obreros cualificados	% within row	90.2 %	9.8 %	100.0 %
Obreros no cualificados	% 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 94. 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.

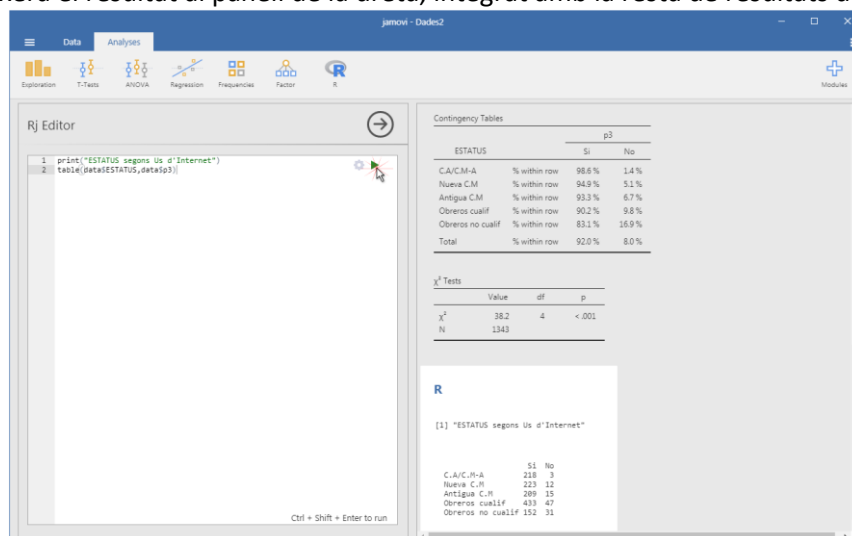
Per a obtenir el resultat d'una **taula de contingència directament executant codi R** amb l'extensió (Rj) de *jamovi* cal:

1. Tenir instal·lat el mòdul o extensió de *jamovi* anomenat Rj Editor.²¹
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)
```
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.

²¹ 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*.

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



Imatge 95. Taula de contingència generada amb codi R

- 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.

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

- Molta atenció a com la taula definitiva es va construir, 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. Escriu i executa, una a una, les següents ordres:
TITOL

TPerc
TP2dec

6. 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.
7. 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.

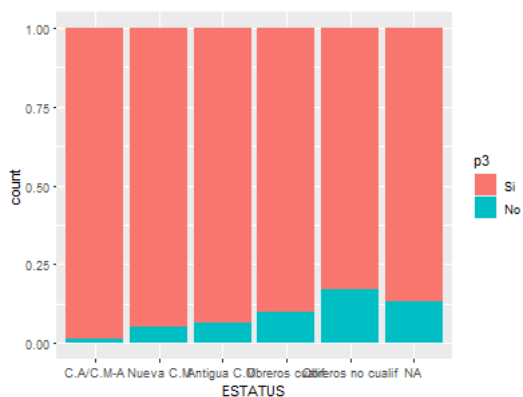
Per poder generar amb *jamovi* aquest tipus de gràfics, a dia d'avui (desembre de 2019) cal utilitzar la consola Rj Editor; no hi ha cap menú ni existeix cap mòdul habilitat per fer-ho. Cal tenir en compte que *jamovi* és un projecte en desenvolupament i que incorpora la potència de R. D'aquí a un temps 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:²²

VERSIÓ SIMPLE

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

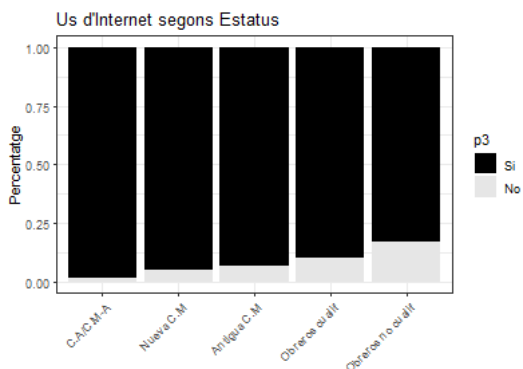
²² Cal executar la instrucció que "activa" la part de R que realitza els gràfics: **library(ggplot2)**, només una vegada abans d'executar el primer gràfic. Per a la resta de gràfics ja no cal executar-la una altra vegada, tot i que aquí la incloguem a cada gràfic.



Imatge 96. 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))
```



Imatge 97. 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.

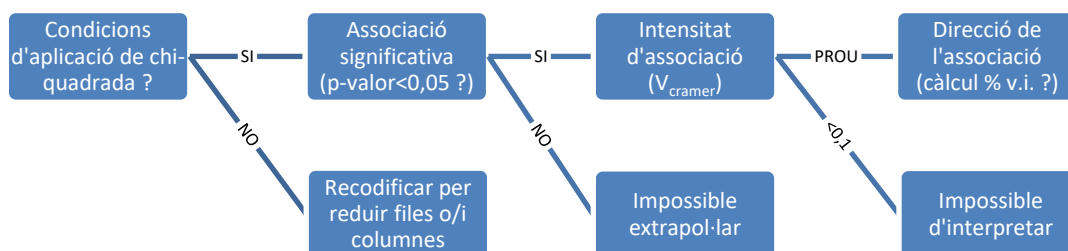
L'objectiu de l'estadística inferencial, expressada aquí a partir del test de la khi quadrat (χ^2), és avaluar fins a quin punt aquestes diferències –que s'han trobat, recordem-ho, en una mostra–, són extrapolables al conjunt de la població. La conclusió pot ser:

- 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.
- 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.
- 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 98. 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 esperàriem 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 **frequències 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 frequències esperades ("Expected"):

1. Fes clic a la taula de frequències 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
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).

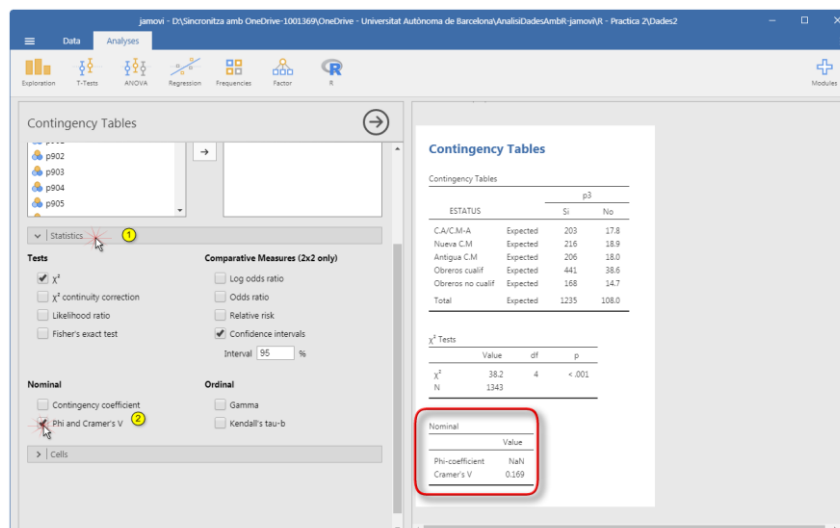
Un cop establert que hi ha relació significativa entre ambdues variables se sol calcular una mesura per avaluar la intensitat d'aquesta relació. Es habitual fer-ho a partir de l'indicador d'associació bilateral **V de Cramer**, que varia entre 0 i 1: Un valor de 0 equivaldria a dir que no hi ha cap relació (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 99. 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 de residuals** d'una taula de contingència a la qual s'ha identificat una variable com a dependent i una altra com a independent implica:

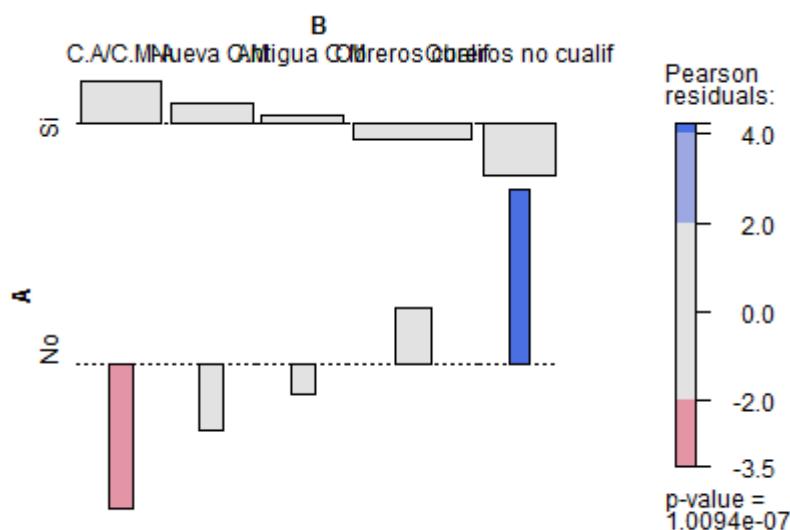
- Demanar a R un gràfic d'associació de la variable dependent en funció de la variable independent, i a continuació,
- Detectar visualment (per colors extrems: vermell i blau) les desviacions de les freqüències esperades (*residuals*) més elevades, per a identificar **QUINES categories** contribueixen 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”.²³
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)
```
4. Aquestes simples ordres que has escrit en llenguatge R signifiquen:
“Activa a R la possibilitat de fer gràfics i càlculs per analitzar variables categòriques”, i “Mostra un gràfic d'associació amb la variables p3 com a dependent i la variable ESTATUS com a independent o causa, acoloreix els valors residuals segons la seva desviació a la freqüència esperada”.
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.

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



Imatge 100. Representació gràfica de residus 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**.²⁴ Els residuals 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 (les línies horitzontals).

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 diferència i, per tant, a la dependència entre les variables.

En aquest cas, pots veure com en la cel·la estatus Alt/Mitjà-Alt i NO ús d'Internet, hi ha un residual negatiu i molt significatiu (vermell). Això vol dir que en aquesta cel·la s'hi acumulen menys casos dels esperats, de manera que els casos realment observats d'estatus Alt/Mitjà-Alt tenen un **major** ús d'Internet que el conjunt. Correlativament, a la part superior del gràfic es pot veure que hi ha més casos dels esperats (hi ha més gent d'estatus Alt/Mitjà-Alt que fa ús d'Internet)

En l'extrem contrari es dona una situació simètrica, és a dir una major acumulació d'individus d'estatus Obrer no qualificat que NO utilitzen Internet (blau). Per altra banda, es pot observar com existeix un patró de "creixement" lineal que permet concloure que el nivell d'estatus socioeconòmic 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.

²⁴ La ordre R "assoc" només suporta el càlcul dels residus de Pearson (no estandarditzats, ni tipificats).

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.	Secund	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 podrem 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.²⁵
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.

²⁵ Les transformacions de variables i recodificació es van veure a la pàgina núm 60 d'aquest tutorial.

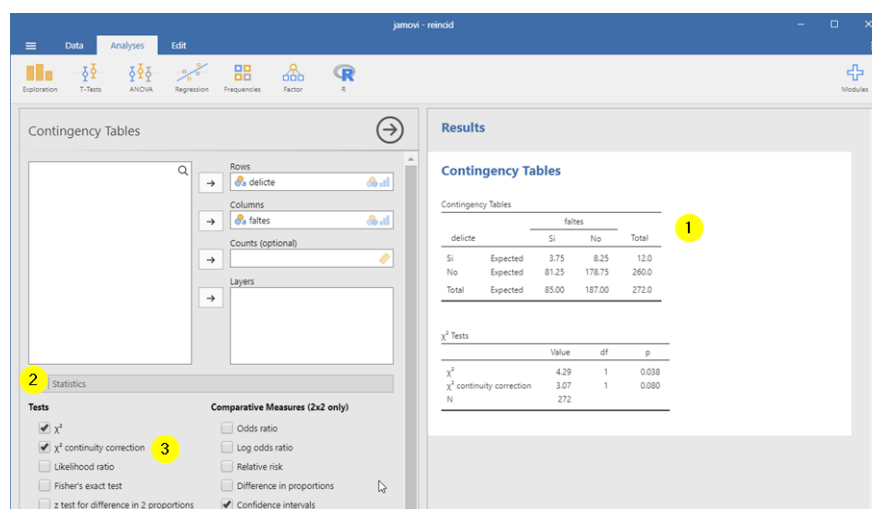
Anem a veure-ho amb un exemple extret d'un fitxer provinent del món de la criminologia, que només conté dues variables, el fitxer *reincid.rdata*. En aquest fitxer trobaràs les variables *faltes* i *delicte*. La primera fa referència ha haver tingut faltes de comportament durant la condemna (Si/No), mentre que la segona ens diu si ha reincidit o no, dos anys després de sortir de la presó.

1. Comprova els tipus de variables i de dades: han de ser NOMINAL/TEXT. Si no és així, modifica la seva definició ("Setup").
2. Selecciona el procediment de Taules per a mostres Independents, a menú "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	
		Si	No
delicte	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.



Imatge 101. 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 desactivis 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 formula 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.²⁶ 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*).

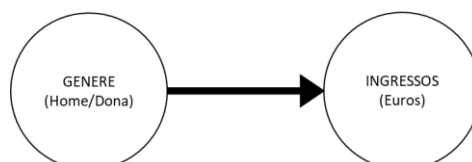
(8) Tal i com mostra la pregunta 21 del qüestionari un dels usos d'Internet és el d'obtenir informació sobre l'actualitat política. Digues si hi ha alguna relació entre aquesta variable i la ideologia política (*p33*), arribant a totes les conclusions possibles. Fes el test corresponent amb un 99% de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

(9) Des de la *Conselleria 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.

²⁶ 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’extrapolar-la al conjunt de la població (aspecte inferencial de l’anàlisi).



Imatge 102. 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.²⁷

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:

²⁷ 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.²⁸

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.

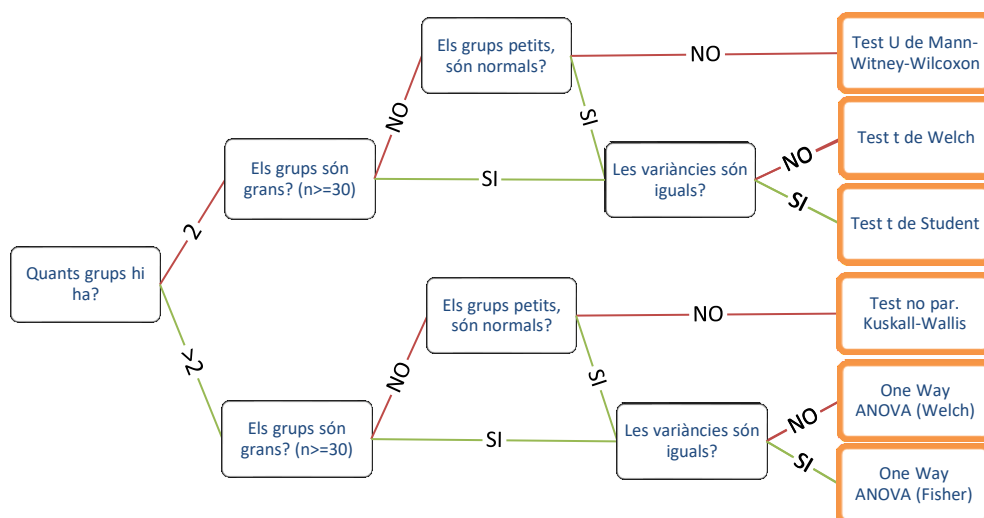
Tests no paramètrics. Tenen l'inconvenient que en la seva aplicació es perd molta informació, en tant que no treballen amb els valors quantitatius sinó, habitualment, amb el rang, amb la posició que ocupen les observacions. En aquest sentit només els farem servir quan, per qüestions de mostra o normalitat,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

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



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 "Dades3.omv" que és pràcticament el mateix que en la pràctica anterior, i que es basa en el qüestionari "Qüestionari.pdf", que necessitareu per anar seguint la practica.

Per a disposar de la matriu de dades corresponent:

1. Localitza a l'espai Moodle de l'assignatura l'apartat corresponent a aquest apartat dedicat a COMPARACIÓ DE MITJANES.
2. Descarrega al teu PC els arxius necessaris: "Qüestionari.pdf" i "Dades3.omv".
3. Executa l'aplicació *jamovi* al teu ordinador.
4. Fes clic 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).

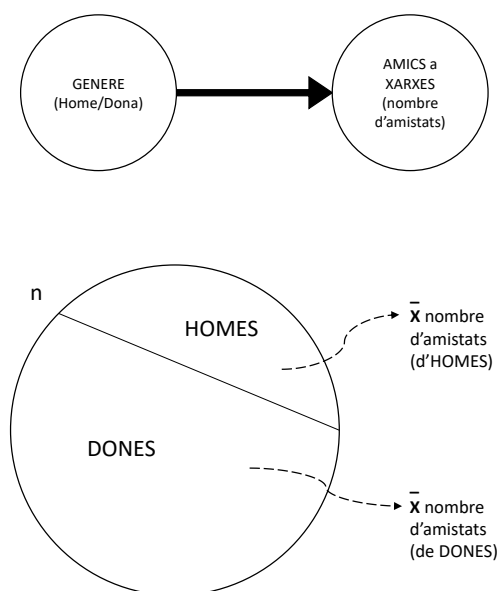
Com veuràs més endavant, caldrà repassa les variables de l'arxiu "Dades2.rdata" utilitzades a aquest apartat, comprovant que estan correctament definides a *jamovi*.

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").

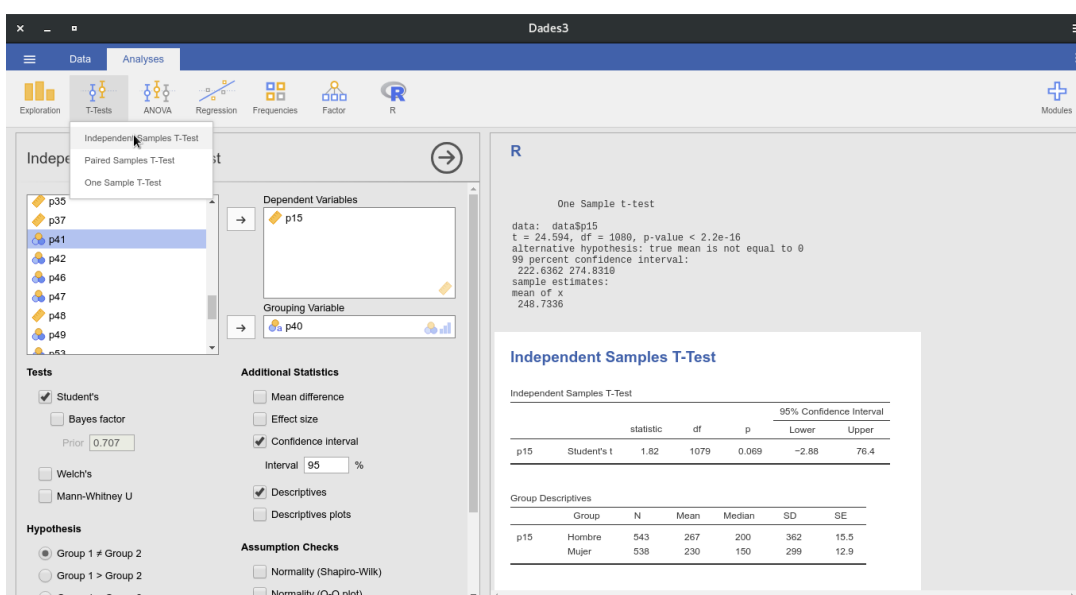
L'estadística descriptiva en la comparació de mitjanes.

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).



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.

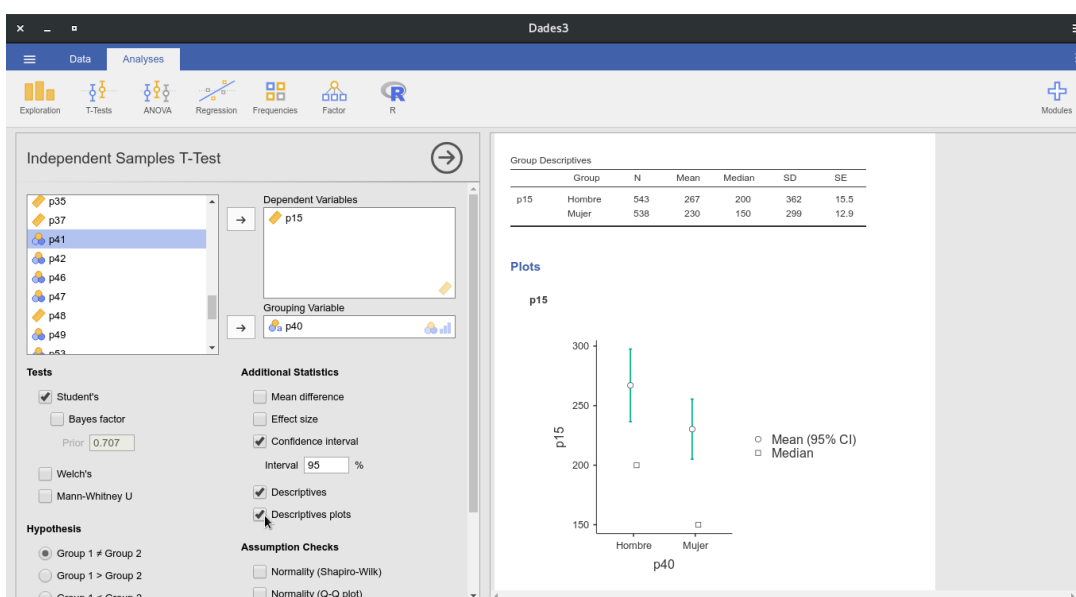


Imatge 103. Anàlisi paramètrica de comparació de dues mitjanes de mostres independents.

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

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**.



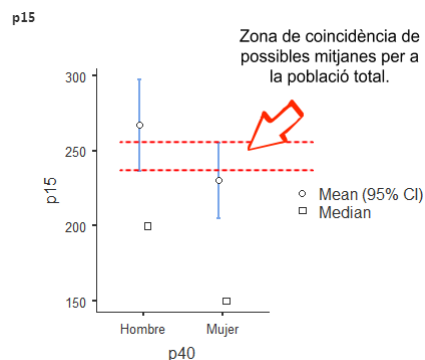
Imatge 104. Gràfic de BARRES D'ERROR per a comparació de dues mitjanes de mostres independents.

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.



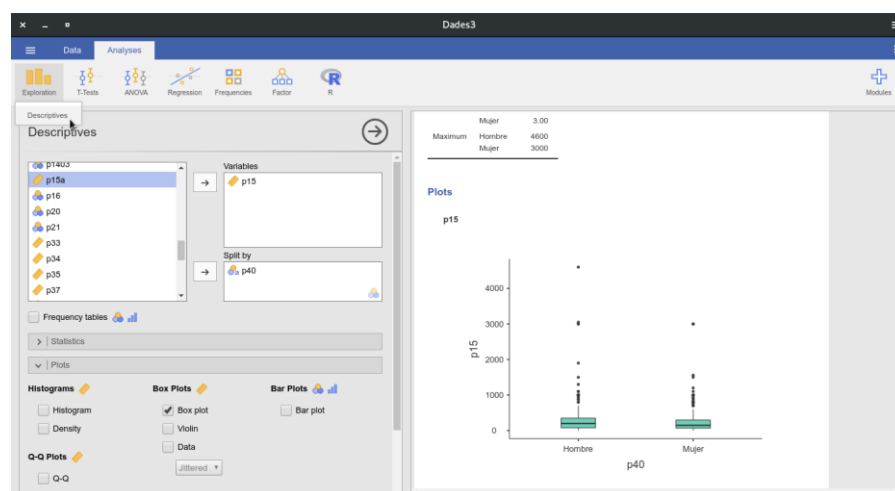
Imatge 105. Solapament d'interval·ls de confiança.

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” és contrària al resultat que s’ha obtingut a la mostra.

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 106. 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 faríem 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

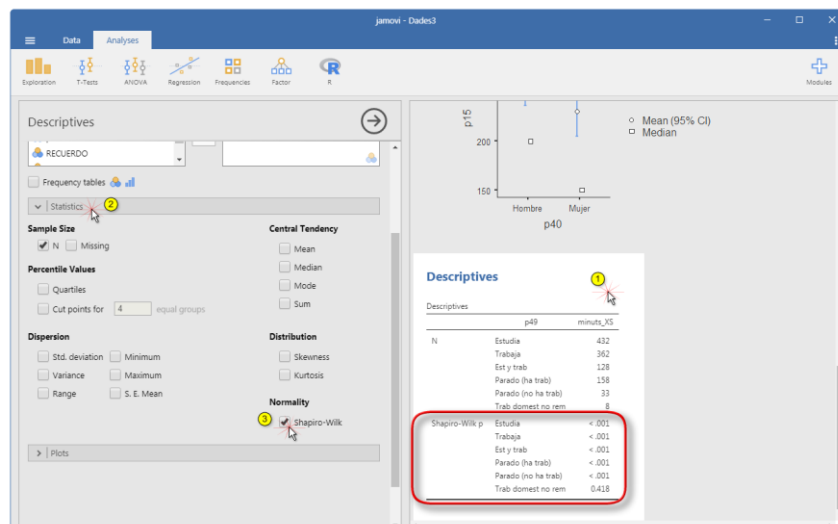
Observa que hi ha un grup molt poc nombrós, que són les persones que manifesten que la seva situació laboral és la de treball domèstic no remunerat ("*Trab domest no rem*"), dels quals només disposem de 8 observacions vàlides per a la variable minuts_XS.²⁹

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

²⁹ En realitat hi ha 26 persones en aquesta situació, però 18 tenen valors perduts per a la variable "minuts_XS".

3. Troba, a la part inferior dreta, l'aparta **Normality**, i activa l'opció [] Shapiro-Wilk.



Imatge 107. 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.

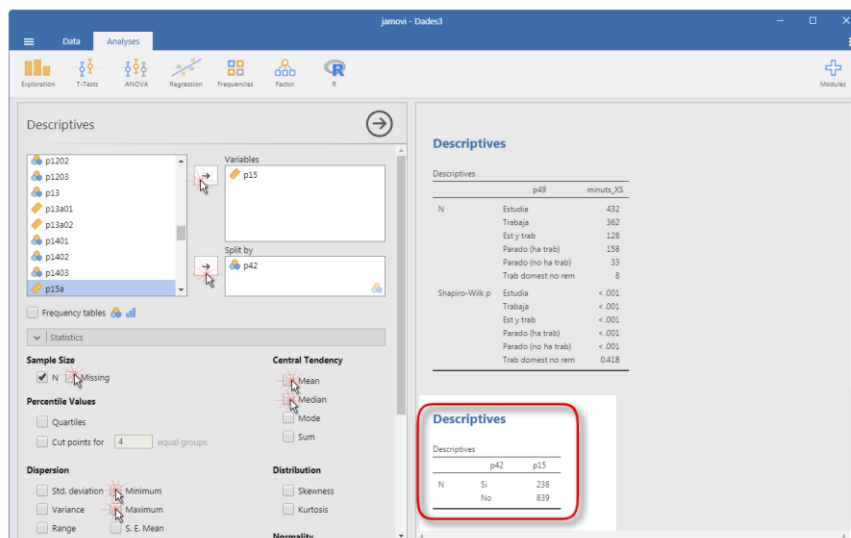
Per a comparar la mitjana d'una variable resposta (el nombre d'amics a xarxes, p15) entre dos grups independents (conviuen o no en parella, p42) es pot fer servir la prova anomenada "T d'Student per a mostres independents". En aquest cas el que volem contrastar s'expressa en un test com:

- $H_0: \mu_{si} = \mu_{no}$ (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 108. Estimació de la mida dels crups 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.

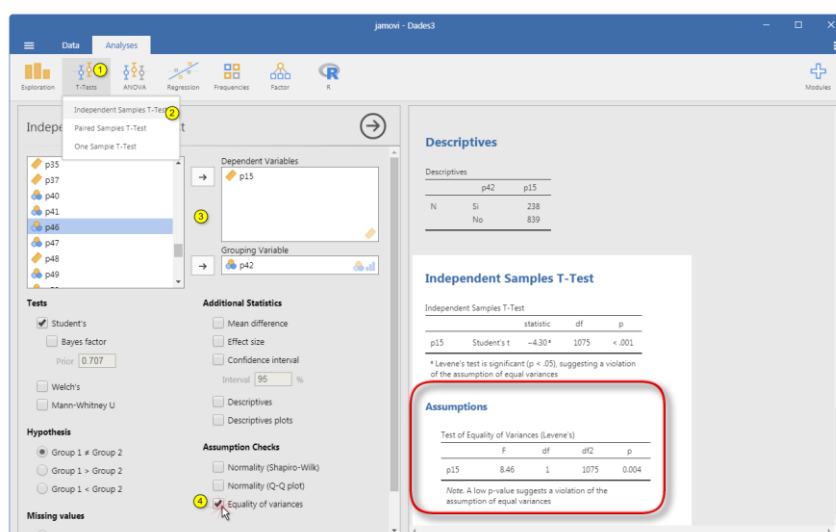
Tal i com s'esmenta en la introducció, hi ha dues maneres diferents de fer aquesta comparació paramètrica: (1) quan les variàncies són iguals i (2) quan no ho són; de manera que en primer lloc haurem de comprovar si es dona la condició d'aquesta igualtat entre variàncies (homoscedasticitat). Per fer-ho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



Imatge 109. 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.³⁰

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.

³⁰ Per saber quin test cal fer en cada cas, consulta l'arbre de decisió Quin test cal fer?.

5. Es convenient activar també l'opció "Descriptives" a l'apartat **Additional Statistics**.

Independent Samples T-Test

		statistic	df	p
p15	Student's t	-4.30 ^a	1075	< .001
	Welch's t	-4.94	481	< .001

^a 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

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-valor $\leq 0,05$ ens quedem amb la hipòtesi alternativa, mentre si el p-valor $> 0,05$ ens quedem amb la nul·la. A més, la hipòtesi nul·la sempre ha d'estar formulada en termes d'igualtat (normalitat, homoscedasticitat, igualtat de mitjanes) mentre que l'alternativa en termes de diferència (no normal, heteroscedasticitat, diferència de mitjanes...).

Comprova a l'apartat **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 variància³¹ (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:

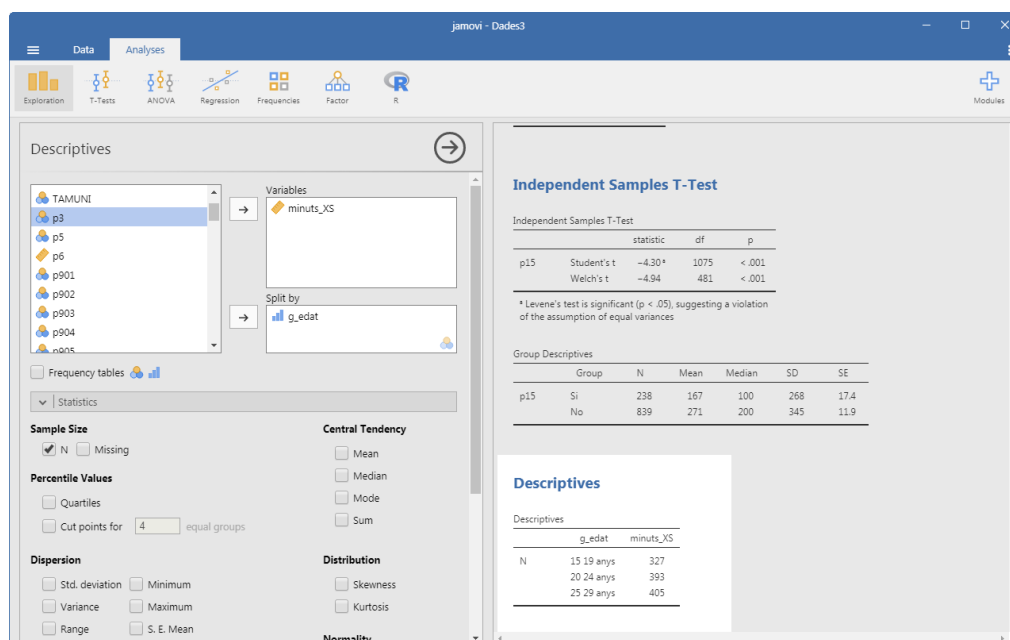
- $H_0: \mu_{15-19} = \mu_{20-24} = \mu_{25-29}$ (la mitjana de temps dedicat a les xarxes socials és la mateixa en els tres grups d'edat)
- $H_1: \mu_i \neq \mu_j$, per algun i, j (la mitjana de temps dedicat a les xarxes socials no és la mateixa en els tres grups d'edat, alguna mitjana difereix de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

³¹ La denominació habitual ANOVA es, òbviament, un acrònim anglès de *analysis of variance*.

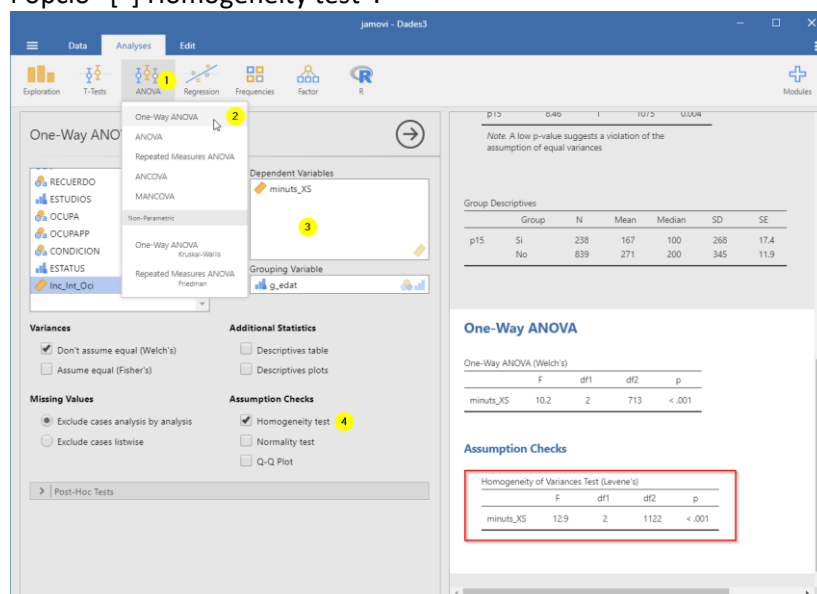


Imatge 110. Estimació de la mida dels crups o mostres per comparar DIVERSES mitjanes.

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 111. 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.

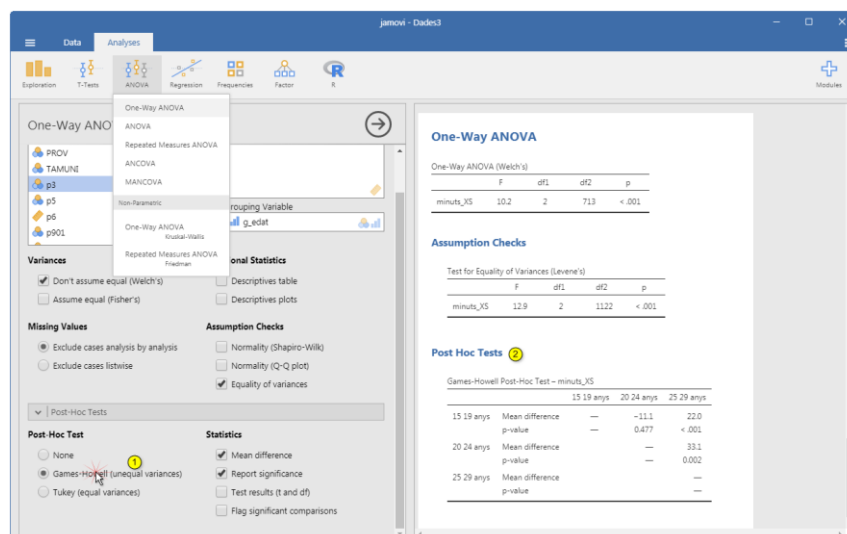
Dir que hi ha relació entre edat i temps és equivalent a dir que el temps dedicat a les xarxes socials és diferent en funció de l'edat. Ara bé, aquesta anàlisi no ens diu entre quins grups es donen les diferències. Per veure-ho hem de comparar cada possible parella de grups, el que s'anomena comparacions 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.



Imatge 112. 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 conclouem 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.

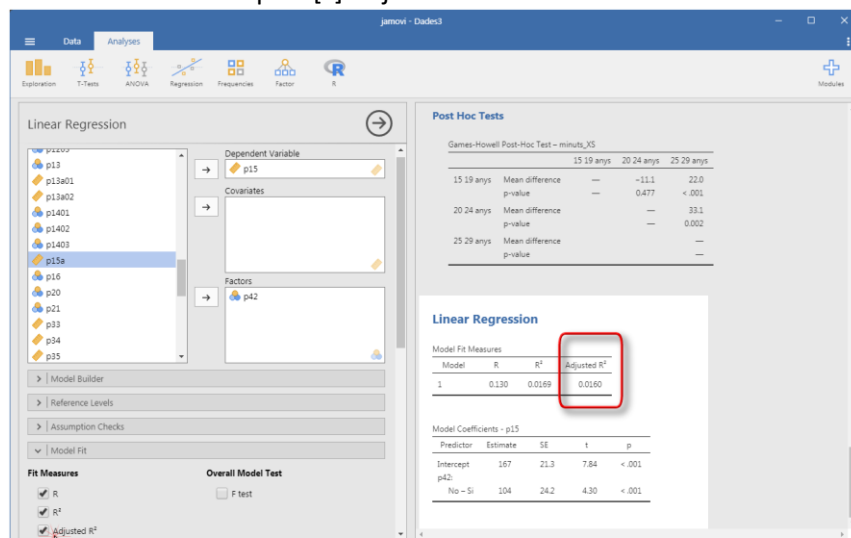
Mentre que en les taules de contingència s'avalua la intensitat de la relació a partir de la *V de Cramer*, quan comparem mitjanes ho fem mitjançant l'estadístic R^2 , anomenat "coeficient de determinació". La diferència més important respecte de la *V de Cramer* (que varia entre 0 i 1) és que R^2 es 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 R².



Imatge 113. Intensitat de la relació (R²) 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).³² 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:

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

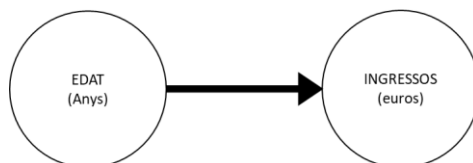
# de usos	Recompte
6	179
7	129
8	134
9	79
10	60
11	33
12	24

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.

³² Per construir-lo basat en la primera pràctica.

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 114. 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: “Qüestionari.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 solem utilitzar tres estratègies:

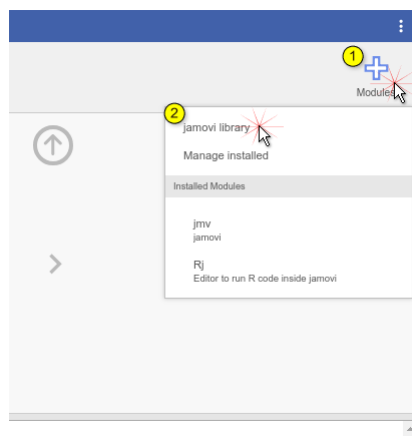
1. Dibuixar gràficament el “núvol de punts” que representa la relació, utilitzant l’anomenat diagrama de dispersió o diagrama XY.
2. Avaluar el nivell o intensitat de la relació entre les dues variables a partir de l’anomenat coeficient de correlació de Pearson (r).
3. Estimar els coeficients que determinen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

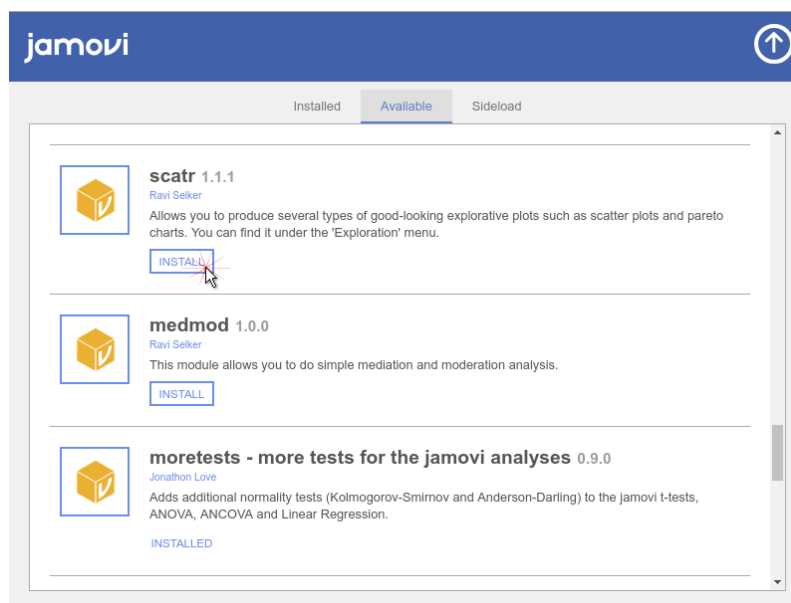
Per a generar un diagrama de dispersió, o diagrama XY (“*scatter plot*”) a *jamovi*, si és que no s’ha fet prèviament, cal instal·lar un mòdul addicional anomenat “*scatr*”.

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 115. 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**]



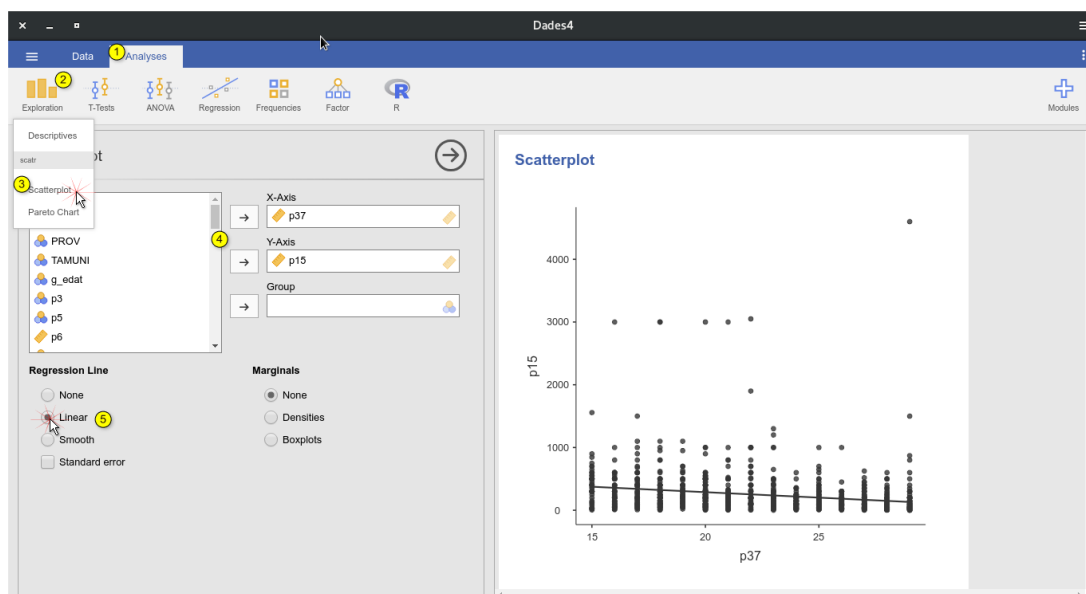
Imatge 116. Instal·lació de mòdul per a generar diagrames XY

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:

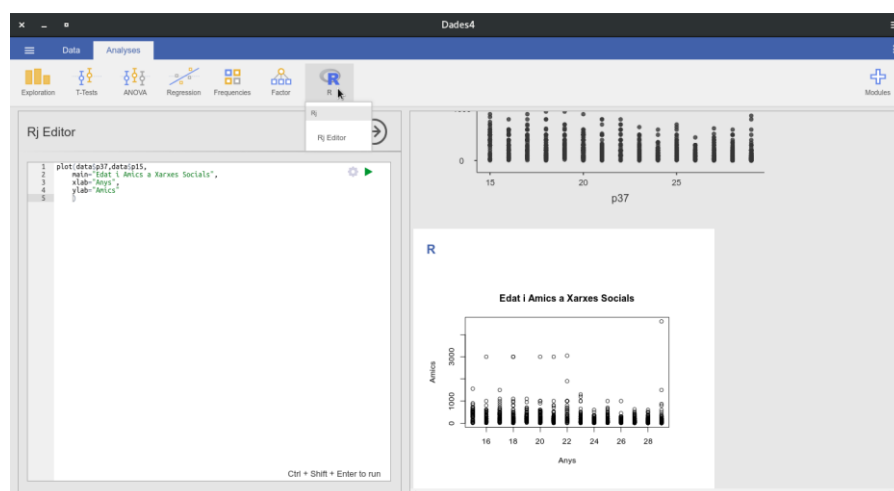


Imatge 117. Diagrama de dispersió o XY amb jamovi.

Diagrama de dispersió amb codi R

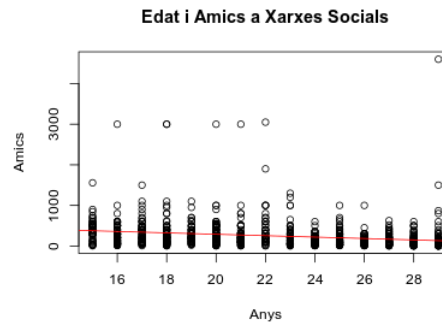
Si cal modificar el resultat gràfic, per exemple incloent-hi títols, etc..., o controlar la escala dels eixos, és preferible utilitzar codi R per executar a la consola de *jamovi*. A més, per generar el diagrama de dispersió directament amb codi R no cal tenir instal·lat el mòdul d'extensió "scatr" (però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



Imatge 118. Diagrama de dispersió o XY amb codi R

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



Imatge 119. Diagrama XY i recta amb codi R

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

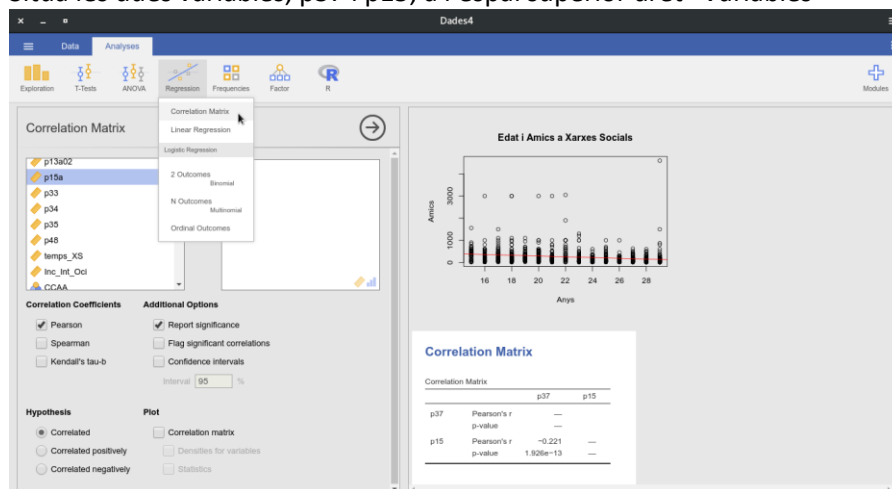
Aquest gràfic de dispersió ens mostra que hi ha una certa correlació inversa (o negativa) entre l'edat i el nombre d'amics a les xarxes socials, de manera que les persones més joves són les que tenen més amics, i que aquests van descendint a mesura que augmenta l'edat. A més, observem una forta dispersió de les dades, de manera que la correlació no serà massa intensa.

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 120. 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 .

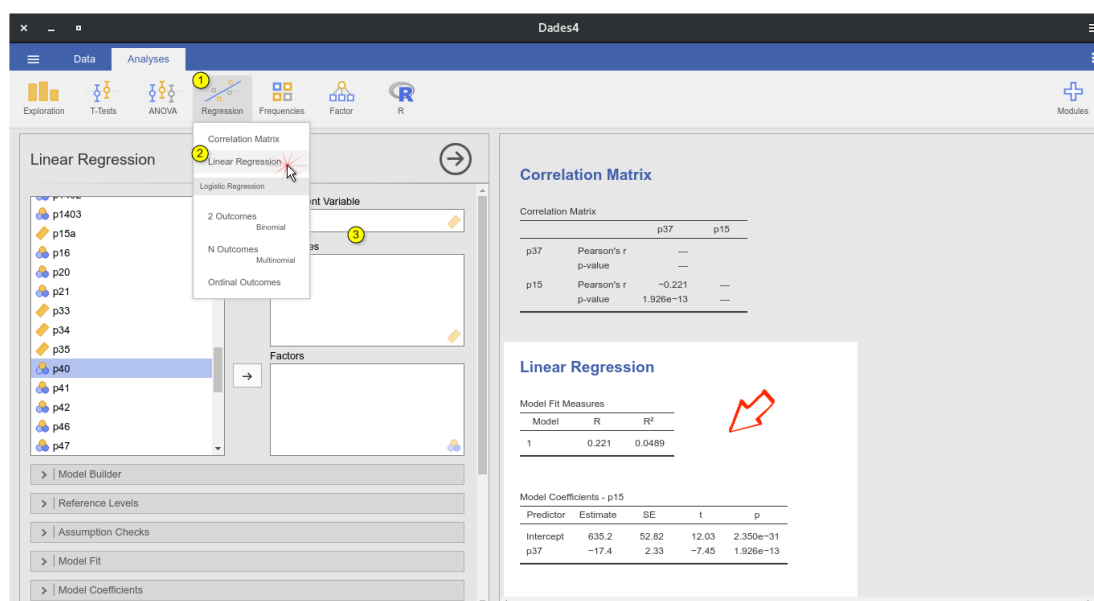
A més, sol ser habitual que ens interessi quina és la funció de la recta de regressió (la recta que *jamovi* ha dibuixat en el gràfic de dispersió). Aquesta funció ens permetrà predir, per a cada valor de la variable independent, el valor que pren, en mitjana, la variable dependent.

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 121. 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:

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 $\text{edat}=0$, 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).

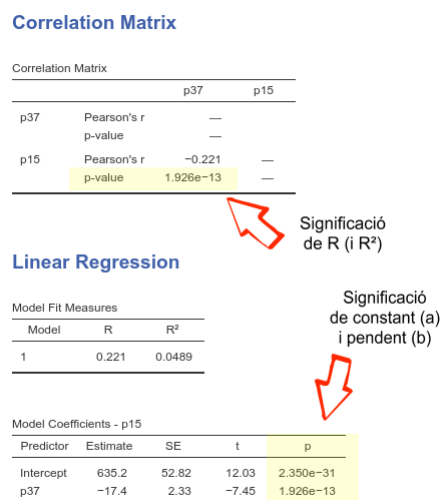
Inferència estadística en els models de regressió

Pots comprovar com en els anteriors apartats hem estimat quatre paràmetres: coeficient de correlació (R), coeficient de determinació (R^2), 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 cap 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 , R^2 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 122. 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 0,000 ³³; 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:

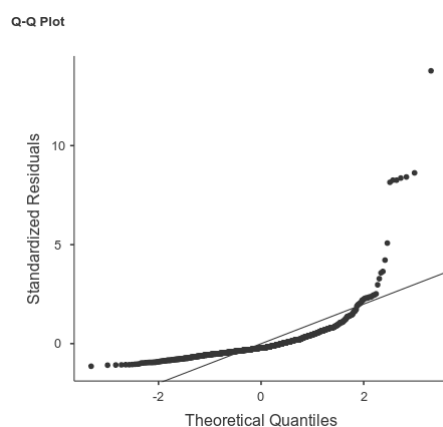
1. El model està correctament especificat.

[illegible]

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 residus. Per fer això:

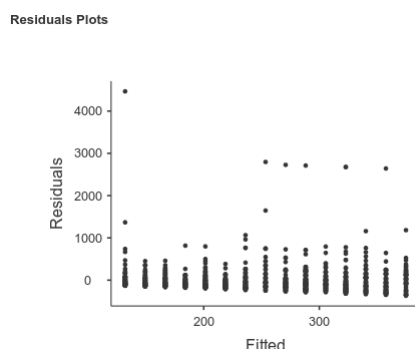
1. Fes clic a la taula de resultats de la Regressió Lineal corresponent. S'obrirà el panell de definició del procediment.
2. Desplega l'apartat "> Assumption Checks" i activa les opcions ☐ Q-Q plot of residuals, i ☐ Residual plots.



Imatge 123. Anàlisi de residus 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 "extremes" (*outliers*), situats més enllà de 2 desviacions típiques del punt central.

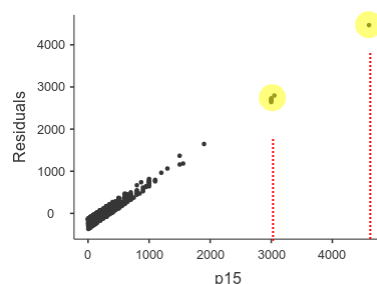


Imatge 124. Anàlisi de residus al model RLS. Gràfic de Residuals.

El gràfic "Residuals plots" representa els residus (diferència entre el valor real i el valor ajustat/predit per la recta) respecte dels valors ajustats. Les condicions del model ens diuen que (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 125. 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:

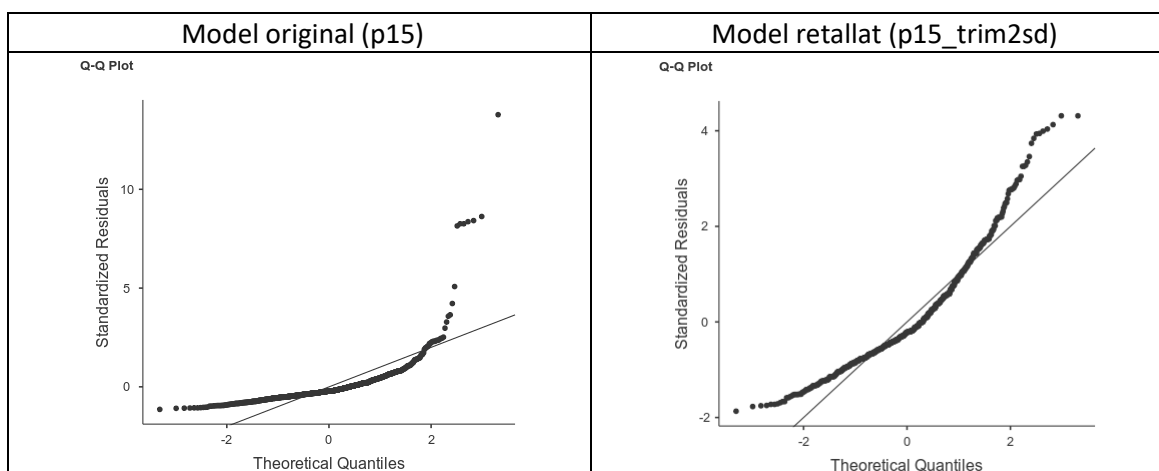
```
if $source > 2 use NA
else use p15
```

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



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

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".
=> **Microsoft-Excel**
=> **LibreOffice-Calc** <<https://es.libreoffice.org/descarga/libreoffice/>>
4. Analitza respostes a qüestionaris: Realitza les operacions de transformació i anàlisi de la matriu de dades pertinents per a justificat 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.

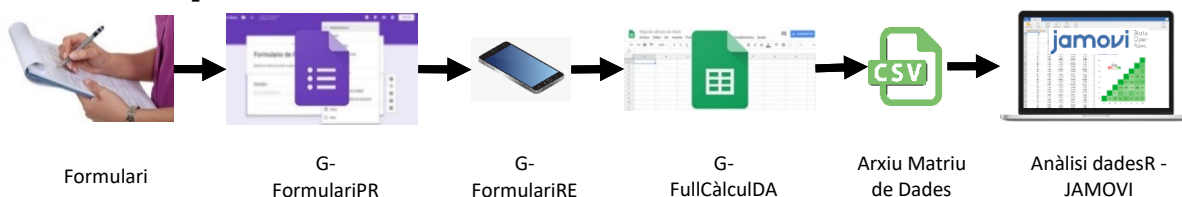
Prerequisits

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)

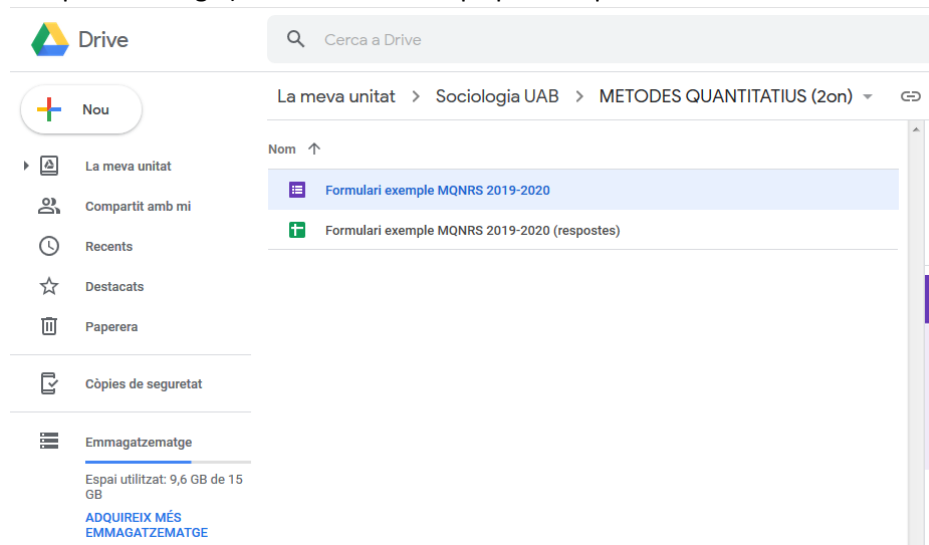
- A. Primer, cal construir el formulari (amb textos, preguntes i respostes). Es convenient utilitzar un PC d'escriptori.
- B. Segon, cal accedir al formulari en mode "visualitza", per poder aplicar-ho a les persones seleccionades. Es convenient utilitzar un dispositiu portàtil, telefonia mòbil o tauleta 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

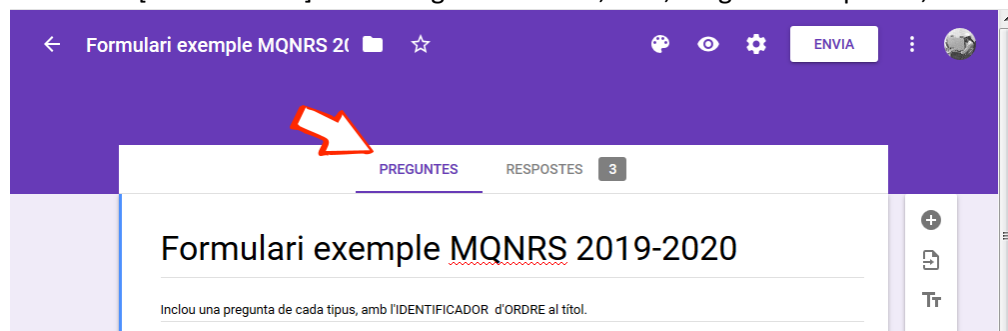
1. Navegar a Google Drive <https://www.google.com/intl/es_ALL/drive/>
Es recomana crear una carpeta especial pel projecte del qüestionari. Aquesta carpeta es pot COMPARTIR amb la resta de membres del grup (cal que també disposin d'un compte de Google). Qui l'ha creat és qui pot compartir-la.



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?,...)

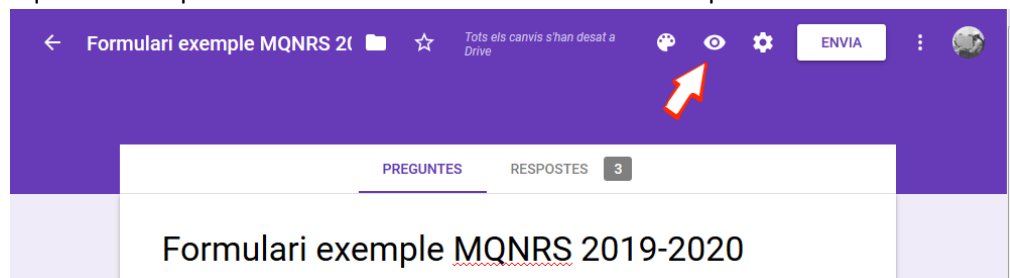
5. Per a TOTES les preguntes:

- 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.

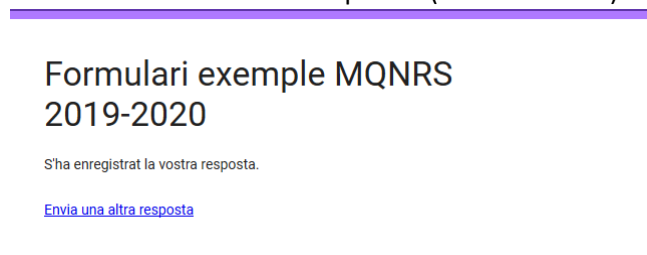
6. NO CAL, però si t'interessa separar les preguntes en blocs diferents, pots clicar la última pregunta d'un bloc i clicar a la icona lateral { Quadrat-Línia-Blanca-Mig } per afegir una nova "secció" del qüestionari.

(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.

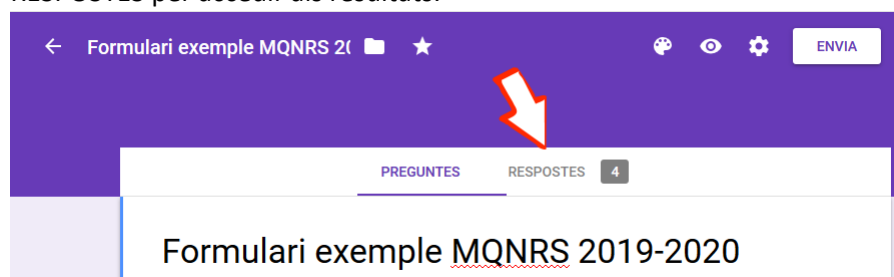


2. Una altra opció és aconseguir el vincle URL associat al formulari de respostes. Qualsevol persona que disposi d'aquesta adreça URL pot respondre al qüestionari, per la qual cosa és important mantenir **estrictament CONTROLADA** la seva difusió.
 - Per aconseguir la URL del formulari: A G-Formularis, 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.



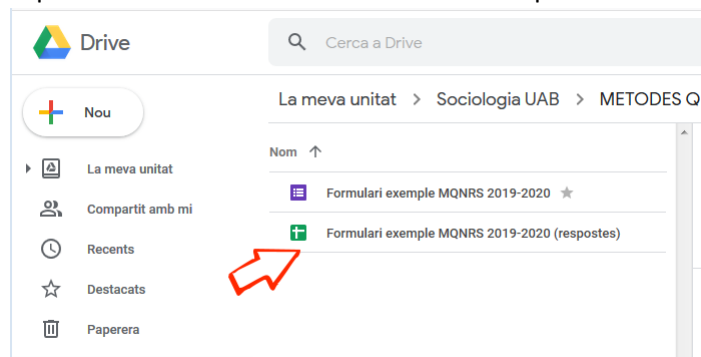
(C) Gestió de les Respostes

1. Cada G-Formulari disposa de 2 pestanyes: "PREGUNTES" i "RESPOSTES". Fes clic a RESPONSTES 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 Edita 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	dfbxdgfnxfgd	Cotxe	Sí	No
3	08/10/2019 12:10:24	Sí		Bus	Sí	Sí
4	08/10/2019 12:17:00	Sí	ghmncvncv	Cotxe	Sí	No
5	08/10/2019 13:10:59	No	xdfbxvcb	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	dfhdfghdfhcfgh. aoi aishj	Bus	Sí	Sí
9	15/10/2019 12:47:24	No sap / No contesta	hdfgjdjdfhjd	Tren	No	Sí

(D) Exportació de les Respostes

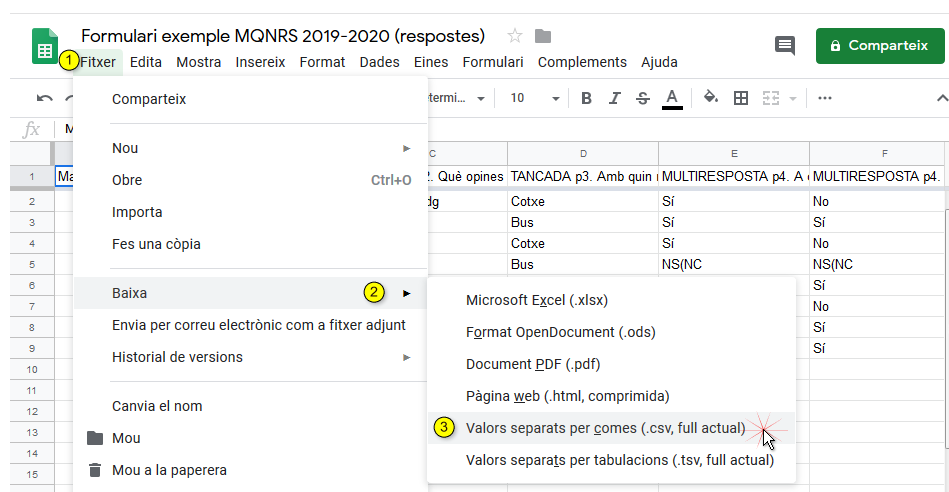
Un fitxer de tipus CSV conté les dades corresponents a les files i columnes d'un full de càlcul, organitzades per "registres". Cada fila és un registre, on cada columna està separada per comes. La primera fila inclou els encapçalaments de cada columna, igualment separats per comes.

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

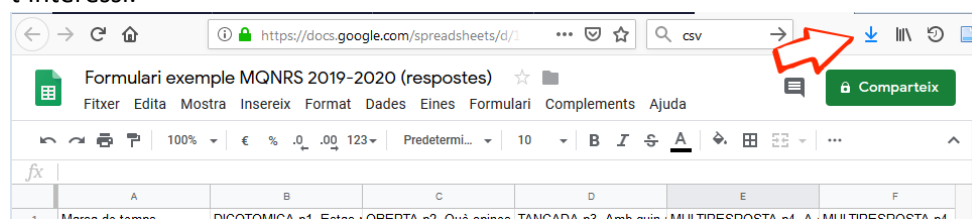
```
Data,Grup,Prova,NIU,Aprovat?,"Curs, si s'escou"
12/02/2019,A,1234567,Parcial,No,1
12/03/2019,B,2222222,Parcial,Si,1
24/10/2019,A,7654321,Final,Si,2
```

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.



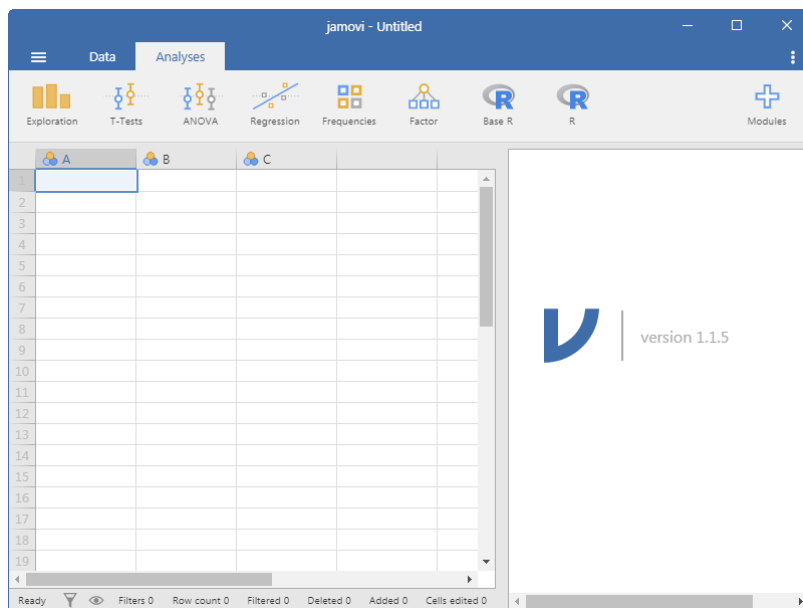
- Ja tens les dades de totes les respostes a un arxiu compatible de tipus CSV.

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*

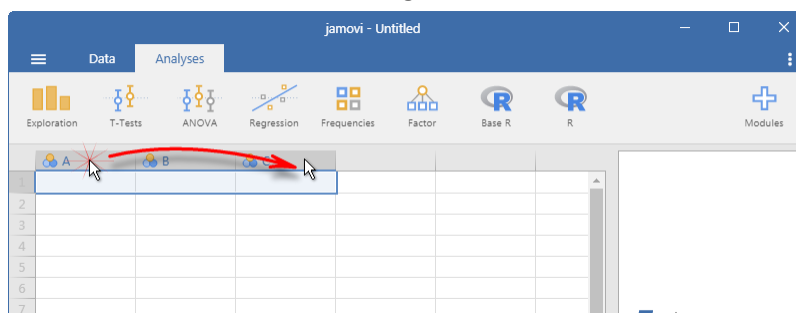
PREREQUISITS

- 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)

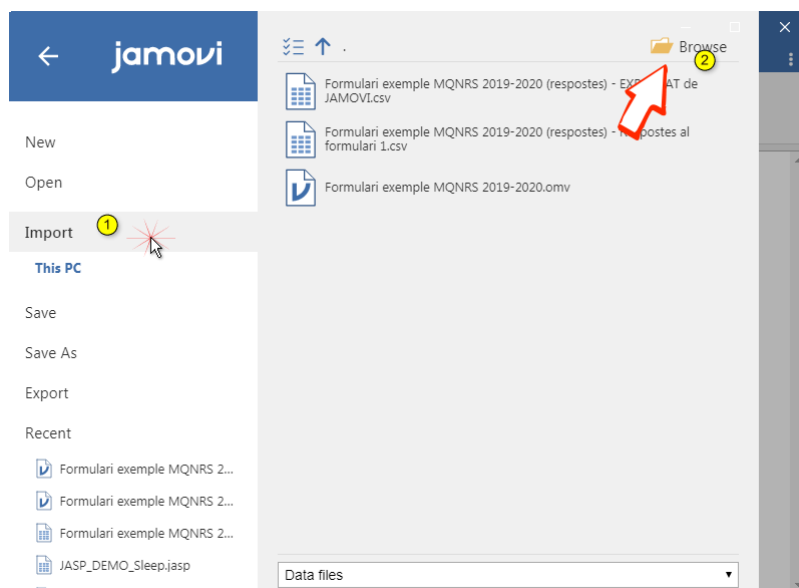


L'aplicació de gestió i anàlisi de dades JAMOVI permet importar arxius de dades tipus CSV. No pot importar "Descripcions", ni tampoc la "ordinalitat" d'algunes variables (O). Aquests són dos aspectes que CALDRÀ AJUSTAR després de fer la importació de les dades.

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.

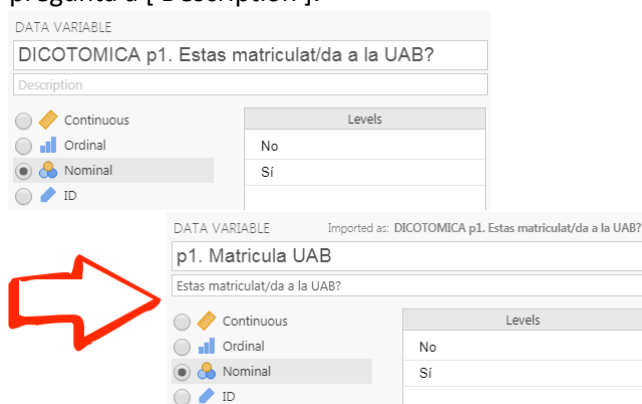


2. Fes clic al menú Data. Fes clic a l'opció "Delete" o icona { Columna-Vermell }  Delete. Confirma per a eliminar les 3 columnes. Ara la matriu està buida, preparada per a importar les teves dades CSV.
3. Fes clic a la icona de menú general { Tres-Línies-Horitzontals } , a la dreta del menú Data.
4.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

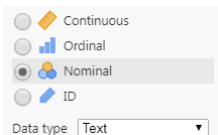


5. **CAL MODIFICAR la definició de les variables**, fes clic a icona { Setup }

- JAMOMI 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].

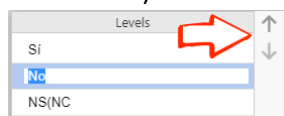


- La variable "marca de temps", cal canviar-la a tipus ID.
- Per passar a la definició de la següent variable fes clic a la icona lateral dreta { > } >, o directament un clic a la columna corresponent. La icona { Fletxa-Amunt-Encerclada }  serveix per tancar la zona de definició de variables (columnes).
- Per a variables de tipus DICOTÓMIC: Comprova que són Nominal i Text.

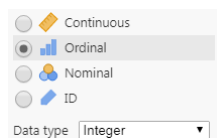


- Per a variables de tipus OBERTA: Comprova que són Nominal i Text.
- Per a variables de tipus TANCADA: Comprova que són Nominal i Text.
- Per a variables de tipus MULTIRESPONSA: Comprova que són Nominal i Text. Re-numera la ID de les respostes per distingir-les, tot i ser resposta múltiple a la mateixa 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

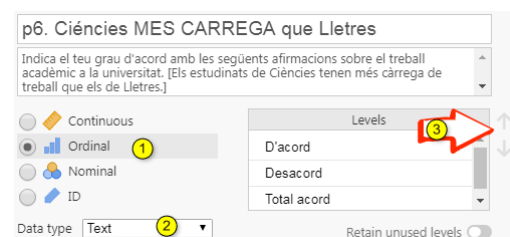
lateral dret).



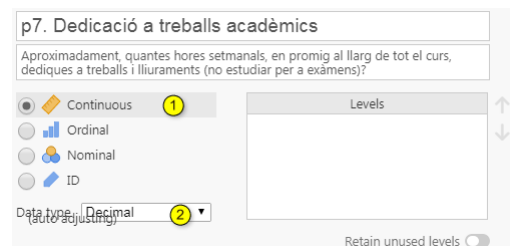
- Per a variables de tipus ESCALA NUMERICA: Comprova que són Ordinal i Integer.



- Per a variables de tipus ESCALA ORDINAL: Comprova que són Ordinal i Text. Renombra 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).



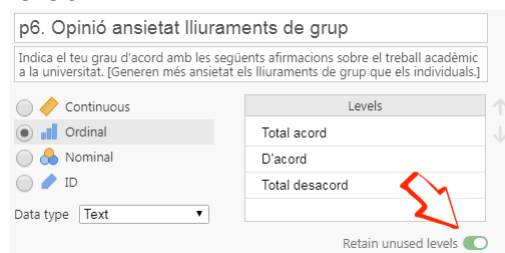
- Per a variables de tipus CONTINUA: Comprova que són Continuous i Decimal (o Integer si no tenen decimals). ATENCIÓ, com que JAMOMI és una aplicació en anglès utilitza el PUNT com a separador decimal (i no la coma).



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):

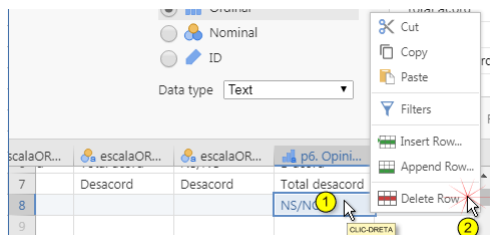
- Fes clic a la columna de la variable que vols modificar. Double-clic a l'encapçalament de columna si està "tancada" la zona de modificació de la variable.
- A sota del llistat de "Nivells" (Levels) de la variable, activa l'opció "Retain unused levels".



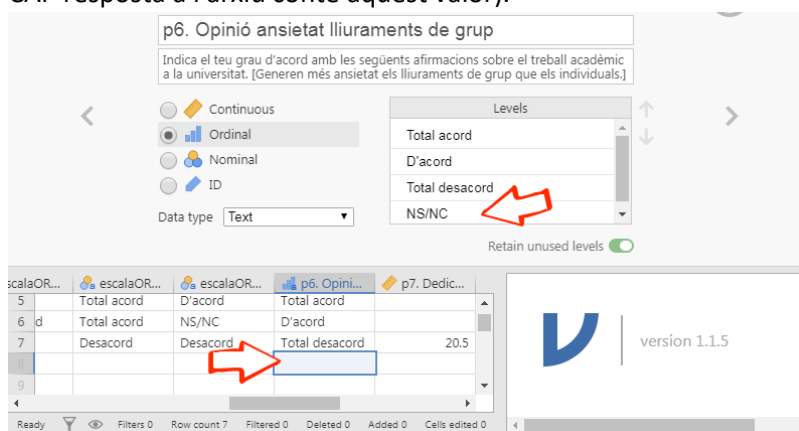
- Fes clic a una fila de la variable que estigui per sota de la última dada importada des de l'arxiu CSV. Escriu el text pel Nivell que vulguis afegir. Pica la tecla [Intro] per confirmar.

	Desacord	Desacord	Total desacord	20.5
7				
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. És 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.