Mark A Goss-Sampson

Anàlisi estadística amb JASP: una guia per a estudiants

PID_00265006

Traducció revisada per:

Julio Meneses

Traducció revisada per Julio Meneses, professor agregat dels Estudis de Psicologia i Ciències de l'Educació de la Universitat Oberta de Catalunya (UOC).
Av. Tibidabo, 39-43, 08035 Barcelona
Realització editorial: FUOC

Continguts

PREFACI

L'acrònim JASP prové de l'expressió anglesa **Jeffrey's Amazing Statistics Program**, com a reconeixement al pioner de la inferència bayesiana Sir Harold Jeffrey. Es tracta d'un paquet estadístic de codi obert multiplataforma, desenvolupat i actualitzat ininterrompudament (en la seva versió 0.9.2 de desembre de 2018) per un grup d'investigadors de la Universitat d'Amsterdam. El seu objectiu era desenvolupar un programa lliure i de codi obert que inclogués tant els estàndards com les tècniques estadístiques més avançades, posant especial èmfasi a aconseguir una interfície d'usuari simple i intuïtiva.

En contrast amb molts altres paquets d'estadística, JASP facilita una interfície simple d'arrossegar i deixar anar, menús de fàcil accés, anàlisi intuïtiva amb computació en temps real i visualització de tots els resultats. Totes les taules i els gràfics estan presentats en format APA i poden ser copiats directament i/o independentment. Les taules també poden ser exportades de JASP a format LaTeX.

JASP es pot descarregar des del lloc web https://jasp-stats.org/ i està disponible per a Windows, Mac OS X i Linux. També es pot descarregar una versió preinstal·lada per a Windows que funcionarà directament des d'una unitat USB o un disc dur extern, sense necessitat d'instal·lar-lo localment. L'instal·lador Wix per a Windows permet triar una ruta per a la instal·lació d'JASP –no obstant això, aquesta opció pot estar bloquejada en algunes institucions a causa de normes administratives locals.

El programa també inclou una llibreria de dades amb una col·lecció inicial amb més de 50 conjunts de dades procedents del llibre d'Andy Field*, Discovering Statistics using IBM SPSS statistics*,[1] i de *The Introduction to the Practice of Statistics*,[2] de Moore, McCabe i Craig.

Des de maig de 2018, JASP també pot executar-se directament des del navegador via rollApp ™ sense necessitat d'instal·lar res a l'ordinador (https://www.rollapp.com/app/jasp). No obstant això, podria no tractar-se de la versió més recent de JASP.

És important parar esment en les actualitzacions regulars de JASP, i als vídeos i els posts d'ajuda del seu blog!!

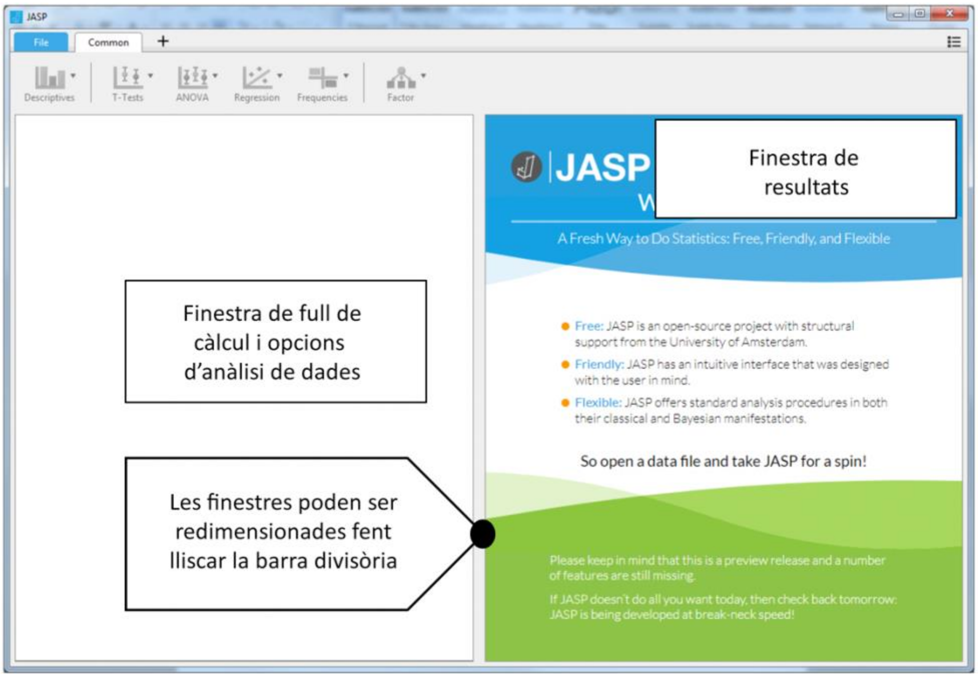
Aquest document és un recull de capítols independents que cobreixen les anàlisis estadístiques més habituals (basades en el model freqüentista) utilitzades pels estudiants de ciències biològiques. Els conjunts de dades utilitzats en aquest document estan disponibles per a la seva descàrrega a http://bit.ly/2wlbMvf.

Dr. Mark Goss-Sampson
Centre per a la Ciència i la Medicina en l'Esport
Universitat de Greenwich
2018

[1] A Field. (2017) Discovering Statistics Using IBM SPSS Statistics (5th Ed.) SAGE Publications.
[2] D Moore, G McCabe, B Craig. (2011) Introduction to the Practice of Statistics (7th Ed.) W H Freeman.

ÚS DE LA INTERFÍCIE DE JASP

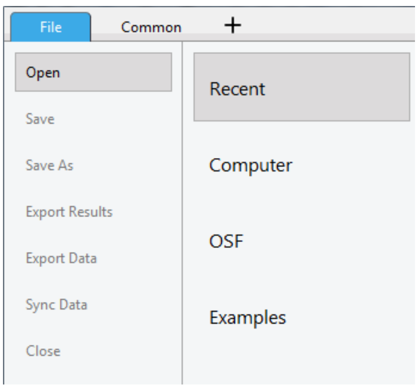
Obriu JASP:

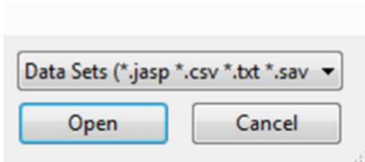


JASP té el seu propi format **.jasp** però accepta una gran varietat de formats de conjunts de dades, com:

- **.csv** (*comma separated values*, valors separats per comes), normalment guardats en Excel
- **.txt** (text pla) també pot ser guardat en Excel
- .**sav** (arxiu de dades IBM SPSS)
- .**ods** (*open document spreadsheet*, full de càlcul de codi obert)

Fent clic a la pestanya «File» o a «So open a data file and take JASP for a spin» de la pantalla d'inici es poden obrir els arxius recents, buscar entre les carpetes de l'equip i accedir a l'Open Science Framework (OSF), o a un ampli ventall d'exemples inclosos en JASP.



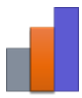


Tots els arxius han d’incloure una etiqueta de capçalera a la primera fila. Un cop carregat, el conjunt de dades apareix a la finestra esquerra:

	Game	Country code	Number of England Injuries
1	1	France	7
2	2	Tonga	4
3	3	New Zealand	2
4	4	France	5
5	5	Tonga	1
6	6	Wales	2
7	7	Wales	5
8	8	New Zealand	4
9	9	Wales	4
10	10	Tonga	3
11	11	Wales	5
12	12	Wales	3
13	13	France	6

En conjunts de dades grans, la icona de la mà permet desplaçar-s’hi fàcilment.

En importar, JASP prova d’assignar de manera automàtica les dades als diferents tipus de variables:

Nominal  **Ordinal**  **Contínua** 

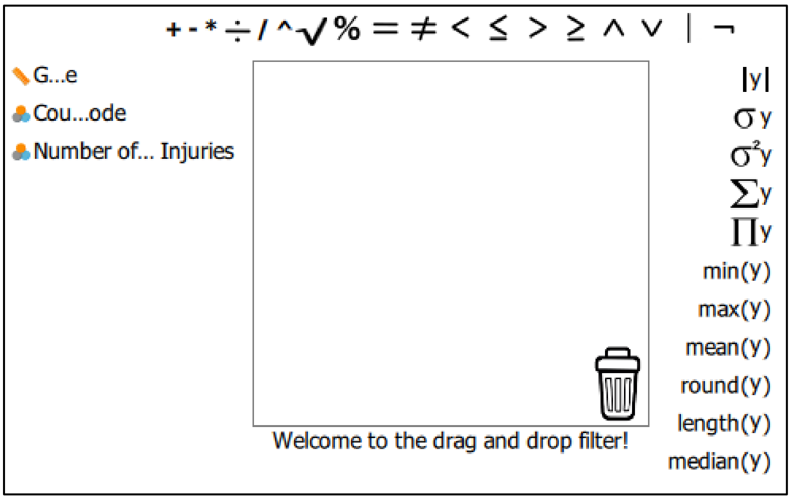
Si JASP ha identificat incorrectament el tipus de dada, només s’ha de fer clic sobre la icona adient en el títol de columna per canviar-la al format correcte.

Si s’han codificat les dades, es pot clicar sobre el nom de variable per obrir la finestra següent que permet etiquetar cada codi. Aquestes etiquetes reemplacen els codis en la visualització del full de càlcul. Si es guarda aquest document com a arxiu **.jasp**, aquests codis, així com totes les anàlisis i les notes, es guardaran automàticament. Això permet que l’anàlisi de dades sigui totalment reproduïble.

Filter	Value	Label
✓	1	Tonga
✓	2	New Zealand
✓	3	France
✓	4	Wales

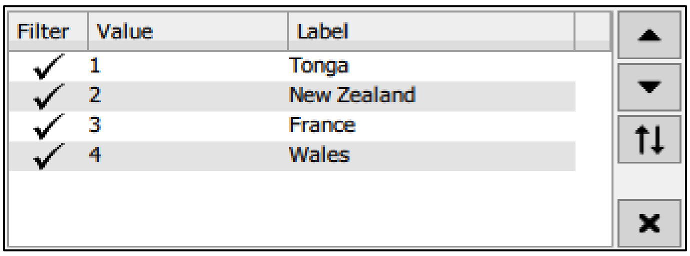
En aquesta finestra també es pot dur a terme un filtrat simple de dades; per exemple, si es desmarca l’etiqueta «Wales», no es farà servir en les anàlisis subsegüents.

 Clicant en aquesta icona de la finestra del full de càlcul s'obre un conjunt d'opcions de filtrat de dades molt més complet:



L'ús d'aquesta opció no es descriu en aquest document. Per a informació detallada sobre l'ús de filtres més complexos, consulteu el següent enllaç: https://jasp-stats.org/2018/06/27/how-to-filter-your-data-in-jasp/

Per defecte, JASP grafica les dades segons el valor (p. ex., 1-4). L'ordre es pot canviar seleccionant l'etiqueta i movent-la amunt o avall utilitzant els cursors pertinents:

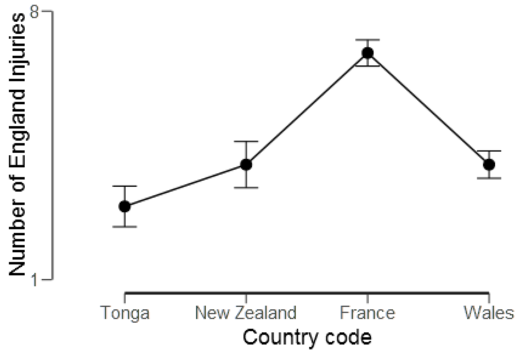


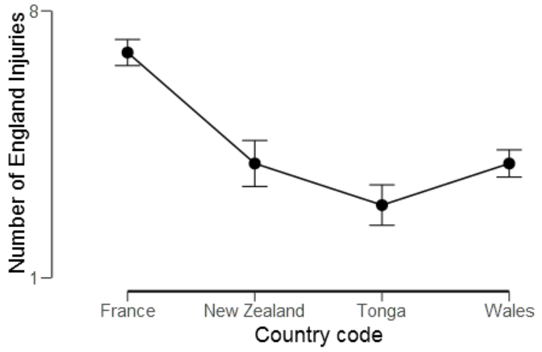
Moure amunt

Moure avall

Invertir l'ordre

Tancar

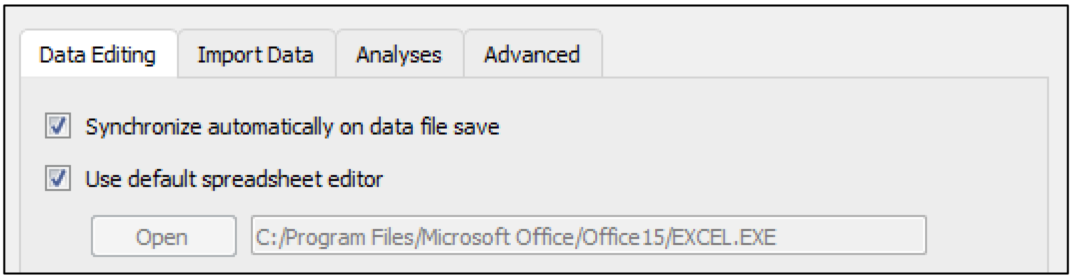




Filter	Value	Label
✓	1	Tonga
✓	2	New Zealand
✓	3	France
✓	4	Wales

Filter	Value	Label
✓	3	France
✓	2	New Zealand
✓	1	Tonga
✓	4	Wales

Si s’han d’editar les dades al full de càlcul, només cal fer doble clic sobre la cel·la i la dada s’obrirà en el full de càlcul original, p. ex., en Excel. Es pot canviar l’opció de l’editor de fulls de càlcul que s’utilitza clicant sobre la icona a la cantonada superior dreta de la finestra de JASP i seleccionant «**Preferences**».



En aquesta finestra es pot canviar l’opció del full de càlcul a SPSS, ODS, etc. Tornarem sobre les preferències més endavant.

Un cop editades les dades i guardat el full de càlcul original, JASP s’actualitzarà automàticament per reflectir els canvis que s’hagin realitzat, sempre que no s’hagi modificat el nom de l’arxiu.

MENÚ D'ANÀLISI DE JASP

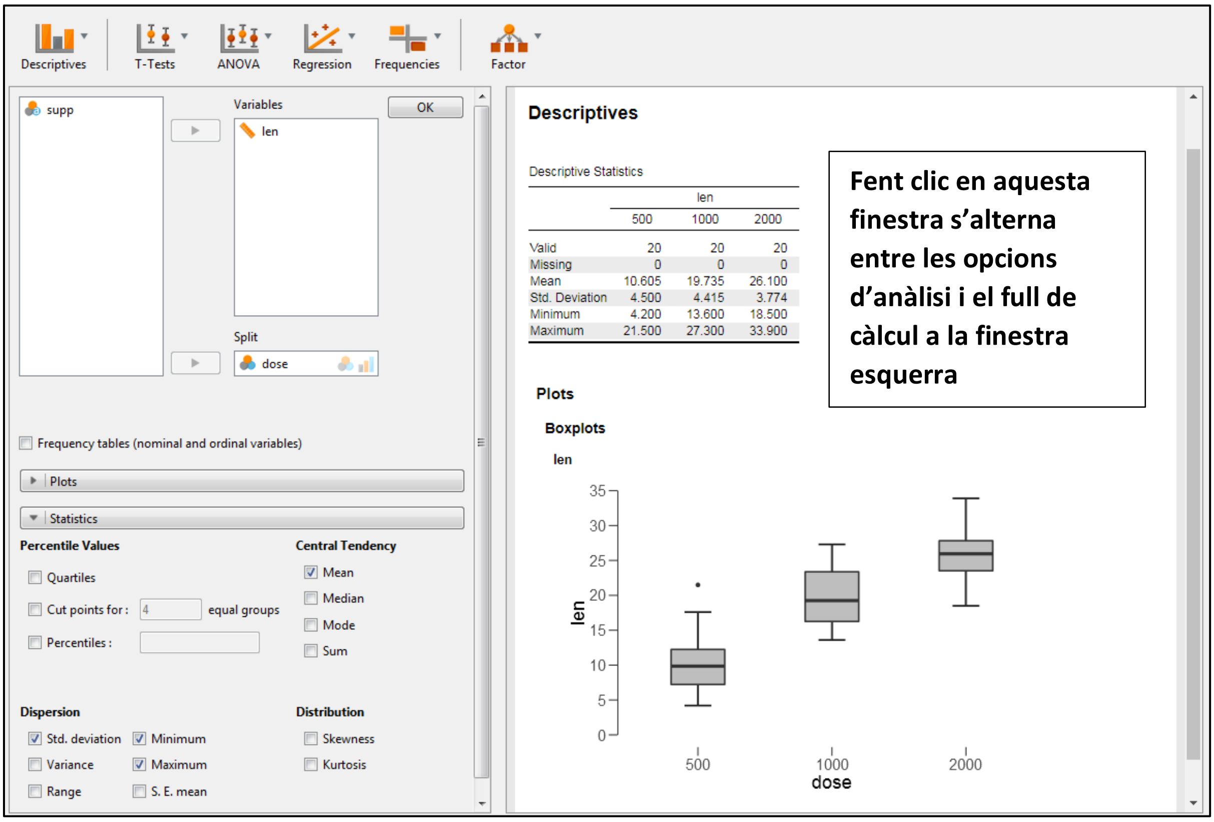
Es pot accedir a les opcions d'anàlisi més **comunes** des de la barra d'eines principal. Actualment (v0.9.0.1), ofereix les següents proves basades en el model freqüentista (estadística més habitual) i les alternatives bayesianes següents:

Descriptives • Estadística descriptiva • Anàlisi de fiabilitat*	**Regressió** • Correlació • Regressió lineal • Regressió logística
Proves *t* • Per a dues mostres independents • Per a dues mostres aparellades • Per a una mostra única	**Freqüències** • Test binomial • Test multinomial • Taules de contingència • Regressió log-lineal*
ANOVA • Mesures independents • Mesures repetides • ANCOVA*	**Anàlisi Factorial** • Anàlisi de components principals (ACP, PCA en anglès)* • Anàlisi factorial exploratòria (AFE, EFA en anglès)*

* No es tracta en el present document

Clicant sobre la icona + del menú superior es pot accedir a les opcions avançades, incloent anàlisi de xarxes, metaanàlisi, models d'equacions estructurals i estadística bayesiana.

Després de seleccionar l'anàlisi requerida, totes les opcions estadístiques possibles apareixen a la finestra esquerra i els resultats es mostren a la finestra dreta.



Si se situa el cursor a sobre de «Results», apareix la icona ▼ i, clicant, es pot accedir a diverses opcions que inclouen:

- **Remove all.** Elimina totes les anàlisis de la finestra de resultats.
- **Remove.** Elimina les anàlisis seleccionades.
- **Collapse.** Oculta el resultat.
- **Add notes.** Afegeix notes a cada resultat.
- **Copy**. Copiar.
- **Copy special (LaTeX code).** Copiat especial (codi LaTeX).
- **Save image as.** Guardar imatge com a.

L'opció «Add notes» permet afegir fàcilment anotacions als resultats i exportar-los a un arxiu HTML seleccionant «File» → «Export results».

ANOVA - Number of England Injuries

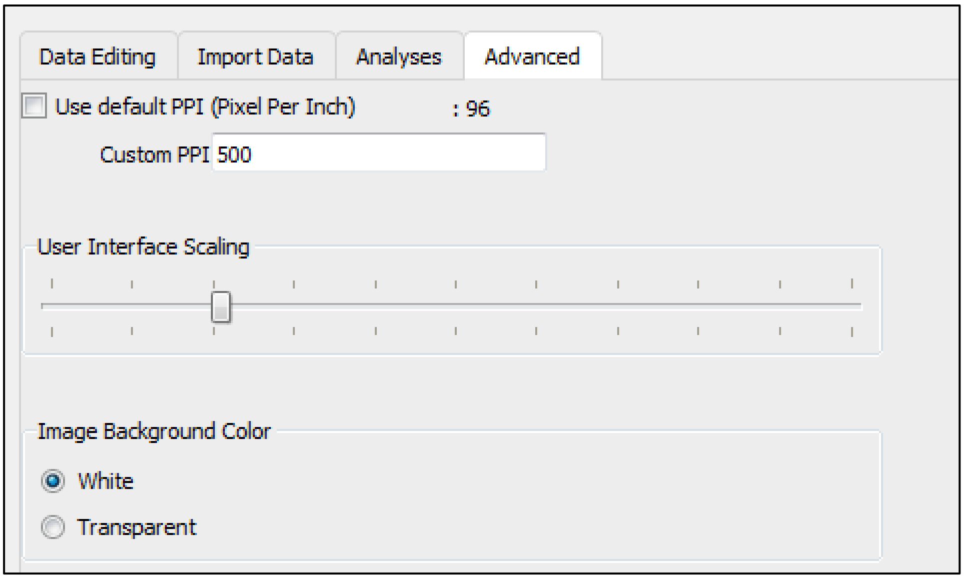
Cases	Sum of Squares	df	Mean Square	F	p
Country code	97.09	3	32.364	13.23	< .001
Residual	97.82	40	2.445		

Note. Type III Sum of Squares

One way ANOVA of injuries received by England rugby players against Tonga, New Zealand, France and Wales

Es pot canviar la mida de totes les taules i els gràfics fent servir ctrl+ (augmentar) ctrl- (reduir) ctrl= (tornar a la mida per defecte). Els gràfics també poden ser redimensionats arrossegant la cantonada inferior dreta del gràfic.

Com s'ha esmentat anteriorment, totes les taules i figures compleixen amb l'estàndard APA i es poden copiar directament en qualsevol altre document. Des de la v0.9.2, totes les imatges poden ser copiades o guardades amb fons blanc o transparent. Això es pot seleccionar a «Preferences» → «Advanced»:

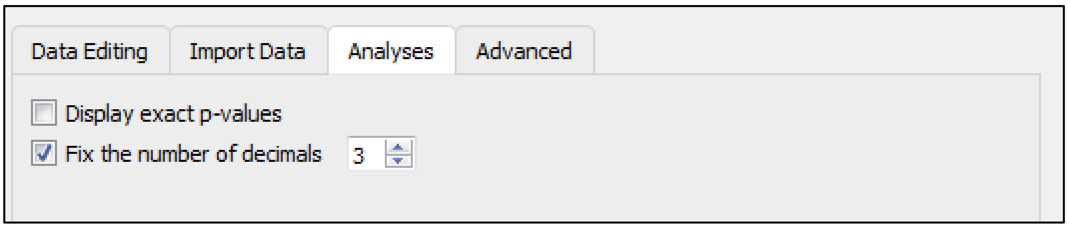


En la mateixa finestra també es pot canviar la mida de la font de la interfície del full de càlcul mitjançant el controlador d'escala d'interfície d'usuari («User Interface Scaling»).

	Trial 1	Trial 2	Trial 3
1	12	12	12
2	36	26	16
3	27	27	27

	Trial 1	Trial 2	Trial 3
1	12	12	12
2	36	26	16
3	27	27	27

Un últim consell en relació amb les preferències («Preferences»): perquè les taules estiguin menys saturades es pot ajustar el nombre de decimals que es mostren, així com mostrar els valors p exactes; per exemple, de $p < 0,001$ a $p < 0,00084$.



Hi ha molts més recursos sobre l'ús de JASP al lloc web https://jasp-stats.org/

ESTADÍSTICA DESCRIPTIVA

És molt difícil per al lector visualitzar o fer inferències a partir d'una presentació de les dades brutes. L'estadística descriptiva i els gràfics relacionats són una manera concisa de descriure i resumir les dades, però no proven cap hipòtesi. Hi ha diferents tipus d'estadístics que es poden utilitzar per descriure les dades:

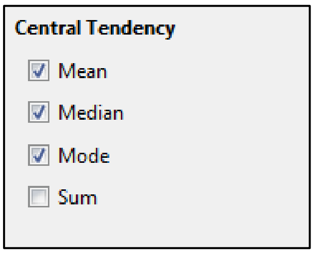
- Mesures de tendència central.
- Mesures de dispersió.
- Percentils.
- Mesures de distribució.
- Gràfics descriptius.

Per estudiar aquestes mesures, carregueu **Descriptive data.csv** a JASP. Aneu a «Descriptives» → «Descriptive statistics» i traslladeu les dades variables a la caixa «Variables» de la dreta.

TENDÈNCIA CENTRAL

Pot ser definida com la tendència de les variables a agrupar-se al voltant d'un valor central. Les tres formes de descriure aquest valor central són la mitjana, la mediana o la moda. Si es considera el total de la població, s'utilitza el terme mitjana, mediana o moda poblacionals. Si s'analitza una mostra / subconjunt de població, s'utilitza el terme mitjana, mediana o moda mostrals. Les mesures de tendència central es mouen cap a un valor constant quan la mida de la mostra és suficient per ser representativa de la població.

En les opcions estadístiques, cal assegurar-se que tot està desmarcat excepte la mitjana, la mediana i la moda.



Descriptive Statistics

	Variable
Valid	810
Missing	0
Mean	17.71
Median	17.90
Mode	20.00

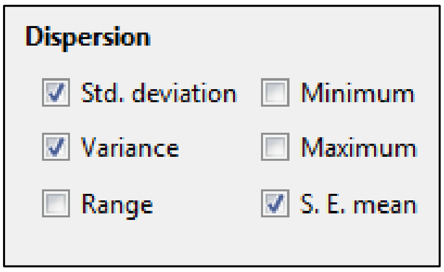
La ***mitjana*****, M o x̄** (17,71), és igual a la suma de tots els valors dividida pel nombre de valors de la taula. És a dir, la mitjana dels valors. S'utilitza per descriure dades contínues. Proporciona un model estadístic simple del centre de la distribució dels valors i és una estimació teòrica del "valor típic". No obstant això, pot quedar fortament influenciada per valors "extrems".

La ***mediana*****, Mdn** (17,9) és el valor central en un conjunt de dades que ha estat ordenat del valor més petit al més gran, i és la mesura tradicional utilitzada per a dades contínues ordinals o contínues no paramètriques. És menys sensible als valors atípics i a les distribucions asimètriques.

La ***moda*** (20,0) és el valor més freqüent en el conjunt de dades i normalment la barra més alta en un histograma d'una distribució.

DISPERSIÓ

En les opcions estadístiques, assegureu-vos que tot està desmarcat menys la desviació estàndard («Std. Deviation»), la variància («Variance») i l'error estàndard de la mitjana («S. E. mean»).



Descriptive Statistics

	Variable
Valid	810
Missing	0
Std. Error of Mean	0.24
Std. Deviation	6.94
Variance	48.10

La ***desviació estàndard (Standard deviation), S o SD*** (6,94) s'utilitza per quantificar el grau de dispersió de les dades respecte a la mitjana. Una desviació estàndard baixa indica que els valors estan a prop de la mitjana, mentre que una desviació estàndard alta indica que el rang de dispersió dels valors és més ampli.

La ***variància (Variance)*** (S^2 = 48,1) és una altra estimació de fins a quin punt les dades se separen de la mitjana. També és el quadrat de la desviació estàndard.

L'error estàndard de la mitjana (The standard error of the mean), SE (0,24) és una mesura que expressa fins a quin punt s'espera que la mitjana obtinguda a partir d'una mostra difereixi de la mitjana real de la població. A mesura que augmenta la mida de la mostra, l'SE disminueix en comparació amb la S i la vertadera mitjana de la població és coneguda amb major especificitat.

Els **intervals de confiança (CI),** encara que no es mostrin en els resultats de l'estadística descriptiva, s'utilitzen en molts altres tests estadístics. Quan es pren una mostra de la població per obtenir una estimació de la mitjana, els intervals de confiança representen un rang de valors dins del qual s'està n% segur que s'inclou la vertadera mitjana. Un CI del 95% és, per tant, un rang de valors del qual un pot estar un 95% segur que conté la vertadera mitjana de la població. Això **no** és el mateix que un rang que contingui el 95% de **tots** els valors.

Per exemple, en una distribució normal, s'espera que el 95% de les dades tingui una SD de ± 1,96 respecte a la mitjana, i el 99% una SD de ± 2,576.

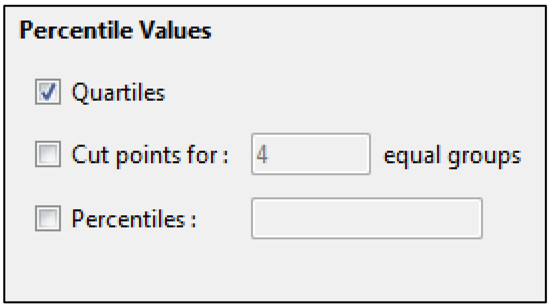
95% CI = M ± 1,96 * l'error estàndard de la mitjana.

Basant-nos en les dades a les quals ens hem referit fins ara, M = 17,71; SE = 0,24; això serà 17,71 ± (1,96 * 0,24) o 17,71 ± 0,47.

Per tant, el CI del 95% per a aquest conjunt de dades és 17,24-18,18 i suggereix que la mitjana real es troba dins d'aquest rang en un 95% de les ocasions.

QUARTILS

En les opcions estadístiques, assegureu-vos que tot està desmarcat excepte els quartils.



Descriptive Statistics

	Variable
Valid	810
Missing	0
25th percentile	13.05
50th percentile	17.90
75th percentile	22.30

Els quartils són els punts en els quals els conjunts de dades es divideixen en 4 parts iguals, a partir dels valors de les medianes un cop ordenades les dades. Per exemple, per a aquest conjunt de dades:

1	1	2	2	3	3	4	4	4	4	5	5	5	6	7	8	8	9	10	10	10
				25%						50%					75%					

El valor de la mediana que divideix les dades pel 50% = percentil 50 = 5.

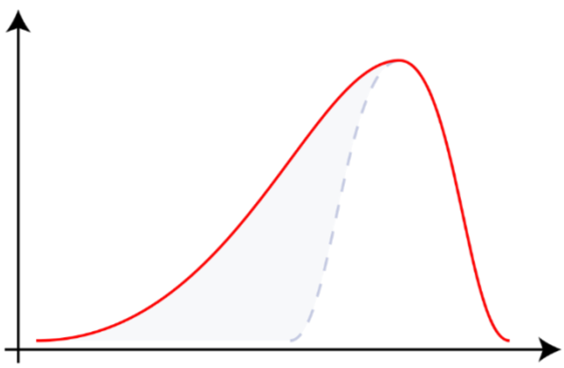
El valor de la mediana del costat esquerre = percentil 25 = 3.

El valor de la mediana del costat dret = percentil 75 = 8.

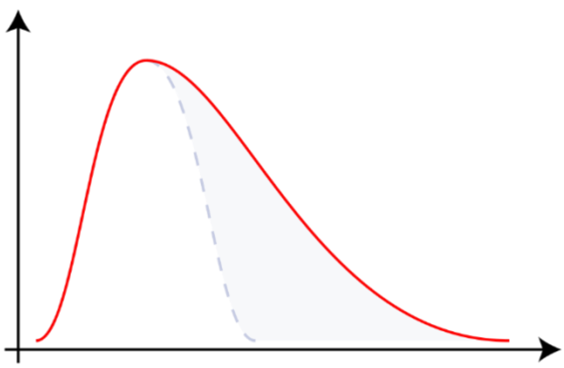
A partir d'això, es pot calcular el rang interquartil (IQR), és a dir, la diferència entre els percentils 75 i 25, és a dir, 5. Aquests valors s'utilitzen per construir, més endavant, els gràfics de caixa descriptius.

DISTRIBUCIÓ

L'asimetria descriu el desplaçament de la distribució respecte a una distribució normal. Una asimetria negativa mostra que la moda es mou cap a la dreta donant com a resultat una cua esquerra dominant. Una asimetria positiva mostra que la moda es desplaça cap a l'esquerra resultant en una cua dreta dominant.

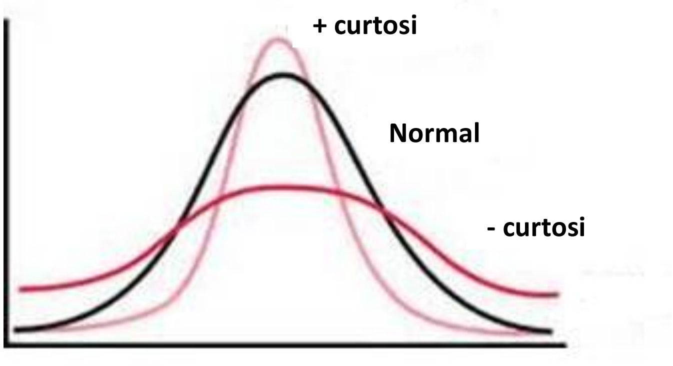


Asimetria negativa

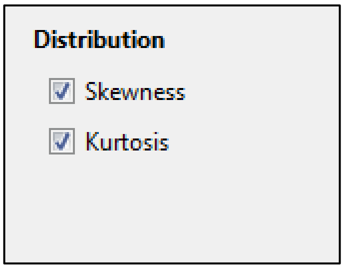


Asimetria positiva

La curtosi descriu com de pronunciades o suaus són les cues. Una curtosi positiva dona com a resultat un “vèrtex” de la distribució més agut, amb cues més pronunciades (llargues); en canvi, una curtosi negativa mostra una distribució molt més uniforme o aplanada, amb cues suaus (curtes).



En les opcions estadístiques, assegureu-vos que tot està desmarcat excepte l’asimetria (*skewness*) i la curtosi (*kurtosis*).



Descriptive Statistics

	Variable
Valid	810
Missing	0
Skewness	−0.004
Std. Error of Skewness	0.086
Kurtosis	−0.410
Std. Error of Kurtosis	0.172

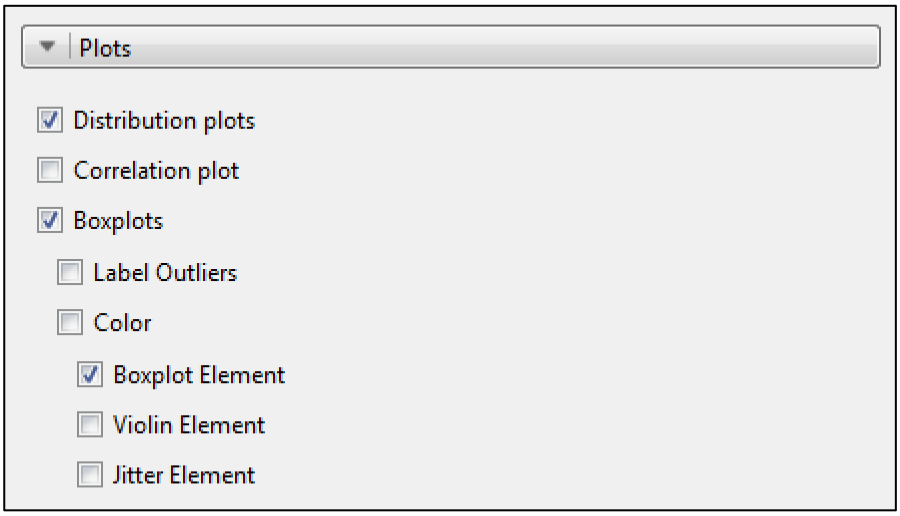
Podem fer servir els resultats descriptius per a calcular les asimetries i les curtosis. Per a una distribució normal, ambdós valors haurien de ser propers a zero (vegeu “Exploració de la integritat de les dades en JASP” per a més detalls).

GRÀFICS DESCRIPTIUS EN JASP

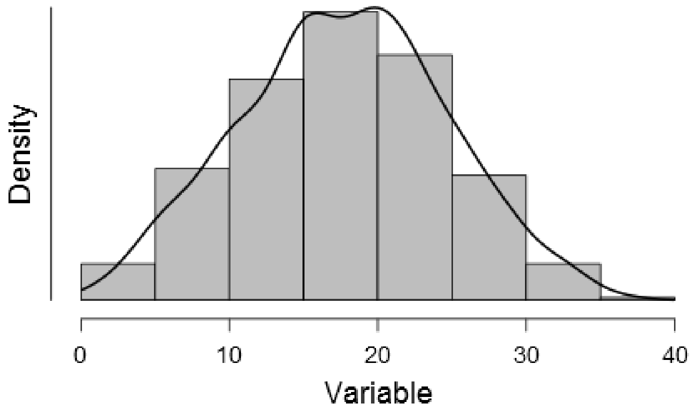
Actualment, JASP produeix tres tipus de gràfics descriptius:

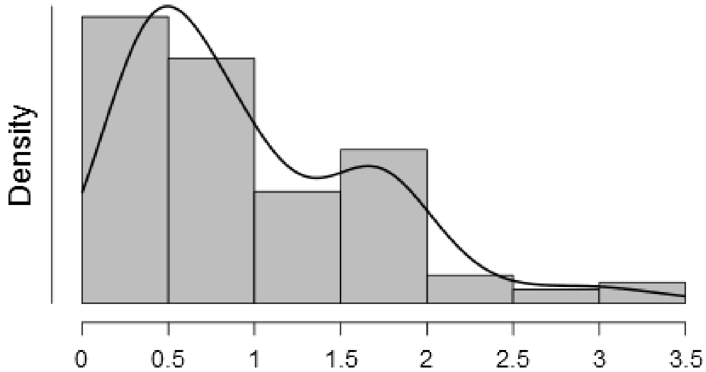
- Gràfics de distribució («Distribution plots»).
- Gràfics de correlació («Correlation plot»).
- Gràfics de caixa, amb 3 opcions («Boxplots»):
    - Element gràfic de caixa («Boxplot Element»).
    - Element violí («Violin Element»).
    - Element *jitter* («Jitter Element»).

De nou, utilitzant **Descriptive data.csv,** un cop introduïdes les variables a la caixa «Variables», aneu a les opcions estadístiques i a sota de «Plots», seleccioneu «Distribution plots» i «Boxplots» – «Boxplot Element».



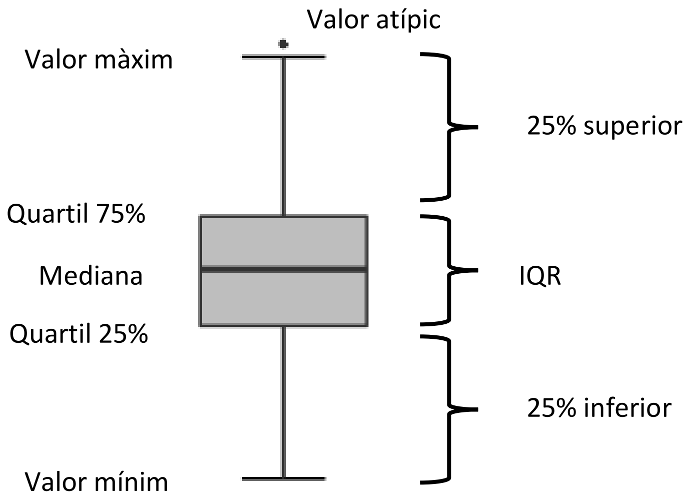
El gràfic de distribució («Distribution plots») està basat en una divisió de les dades en intervals de freqüència, que se superposa a la corba de distribució. Com s'ha dit anteriorment, la barra més alta és la moda (el valor més freqüent en el conjunt de dades). En aquest cas, la corba sembla gairebé simètrica, el que suggereix que les dades es distribueixen d'una manera aproximadament normal. El segon gràfic de distribució és d'un altre conjunt de dades, que mostren una asimetria positiva.



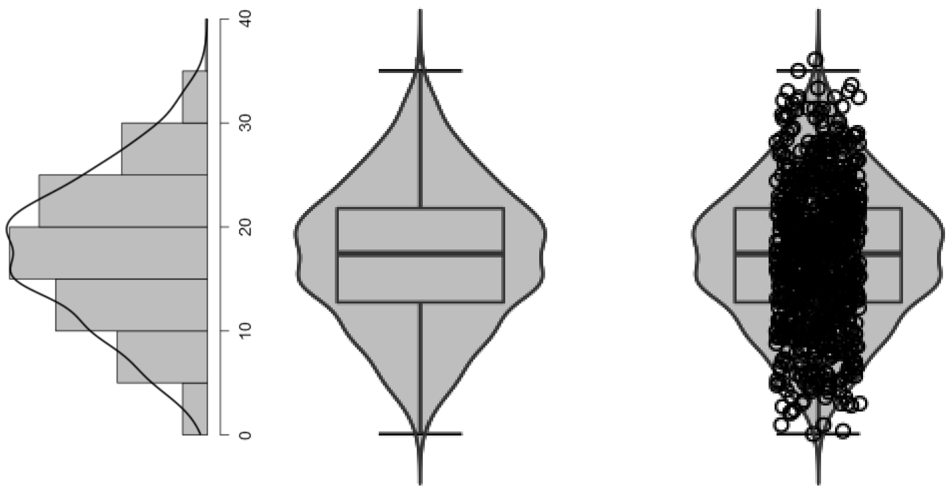


Els gràfics de caixa mostren els estadístics descrits anteriorment en un gràfic:

- Mediana.
- Quartils del 25% i el 75%.
- Rang interquartil (IQR), és a dir, valors del quartil de 75%-25%.
- Valors màxims i mínims representats una vegada exclosos els valors atípics.
- Si es demana, també es mostren els valors atípics.



Torneu a les opcions estadístiques. A «Descriptive plots», marqueu «Boxplot Element» i «Violin Element», i vegeu com ha canviat el gràfic. Després d'això, seleccioneu els elements «Boxplot Element», «Violin Element» i «Jitter Element». El gràfic de violí ha adoptat la corba suavitzada del gràfic de distribució, girant-la 90 ° i superposant-la en el gràfic de caixa. El gràfic *jitter* ha agregat, a més, tots els punts de les dades.



Gràfic de caixa + gràfic de violí

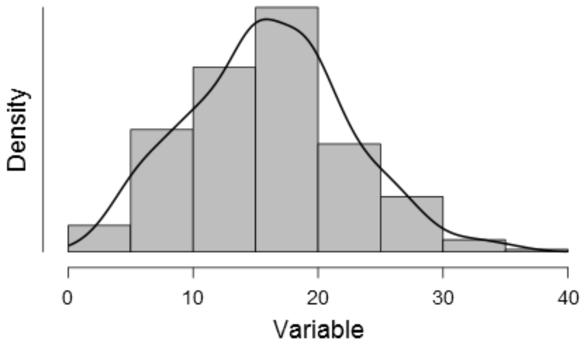
Gràfic de caixa + gràfic de violí + *jitter*

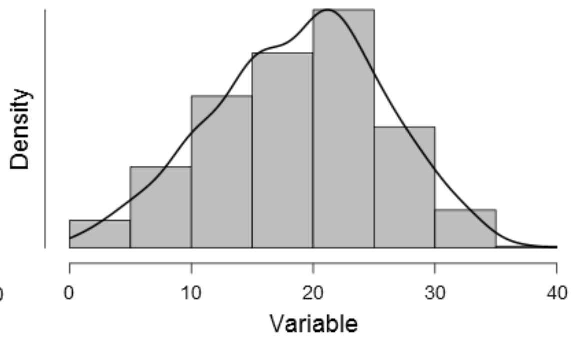
DIVISIÓ DELS ARXIUS DE DADES

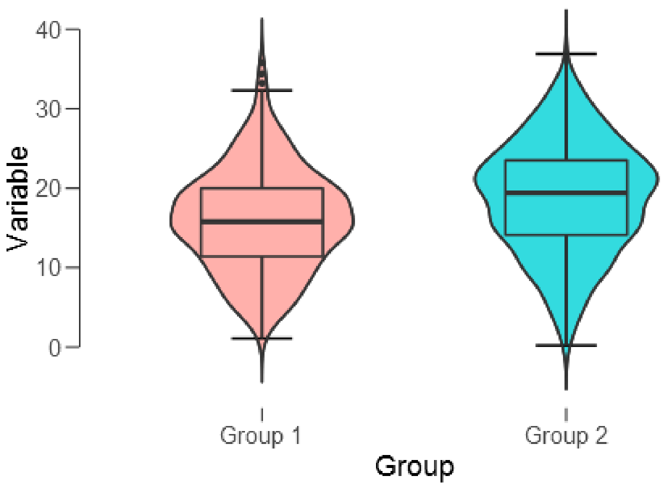
Si hi ha una variable d'agrupació (categòrica o ordinal), es poden elaborar gràfics i estadístics descriptius per a cada grup. Utilitzant **Descriptive data.csv** amb les variables a la caixa «Variables», afegiu una variable d'agrupació a la caixa «Split». El resultat es mostrarà com segueix:

Descriptive Statistics ▼

	Variable	
	Group 1	Group 2
Valid	315	495
Missing	0	0
Mean	16.021	18.787
Median	15.800	19.400
Mode	20.000	20.200
Std. Deviation	6.424	7.040
Variance	41.269	49.556
Skewness	0.200	−0.176
Std. Error of Skewness	0.137	0.110
Kurtosis	−0.101	−0.397
Std. Error of Kurtosis	0.274	0.219
Minimum	1.100	0.200
Maximum	35.800	36.900







EXPLORACIÓ DE LA INTEGRITAT DE LES DADES

Les dades obtingudes a partir d'una mostra s'utilitzen per estimar els paràmetres de la població, tenint en compte que un paràmetre és una característica mesurable d'una població, com la mitjana, la desviació estàndard, l'error estàndard o els intervals de confiança, etc.

Quina és la diferència entre un estadístic i un paràmetre? Suposem que realitzem una enquesta sobre la qualitat del bar estudiantil a un grup d'estudiants seleccionats aleatòriament, i que el 75% dels mateixos es mostra satisfet. Això és un **estadístic** mostral ja que només s'enquestaria una mostra de la població. Es calcularia el que la població probablement faria sobre la base de la mostra. Si es preguntés a **tots** els estudiants de la universitat i un 90% es declarés satisfet s'obtindria un **paràmetre**, ja que s'hauria enquestat el total de la població universitària.

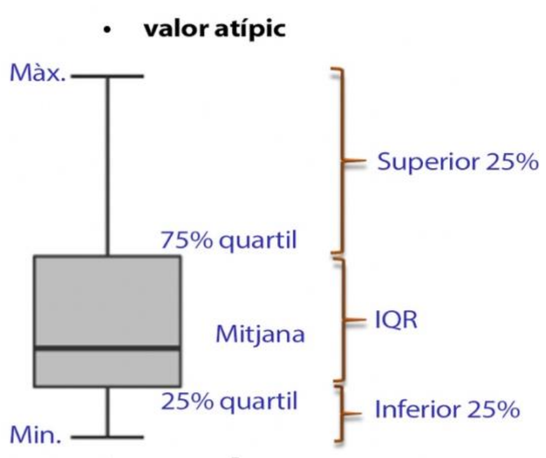
El biaix pot ser definit com la tendència d'un procés de mesura a sobreestimar –o subestimar– el valor d'un paràmetre d'una població. Hi ha molts tipus de biaix que poden aparèixer en el disseny de la investigació i la recollida de dades, entre ells:

- Biaix en la selecció de participants –alguns són més propensos que altres a ser seleccionats per a l'estudi.
- Biaix en l'exclusió de participants –per l'exclusió sistemàtica de certs individus.
- Biaix analític –a causa de la manera com s'avaluen els resultats a l'estudi.

No obstant això, el biaix estadístic pot afectar: a) l'estimació dels paràmetres; b) els errors estàndard i els intervals de confiança; o c) els tests estadístics i els valors de *p*. Llavors, com podem comprovar si hi ha biaix?

SÓN CORRECTES LES VOSTRES DADES?

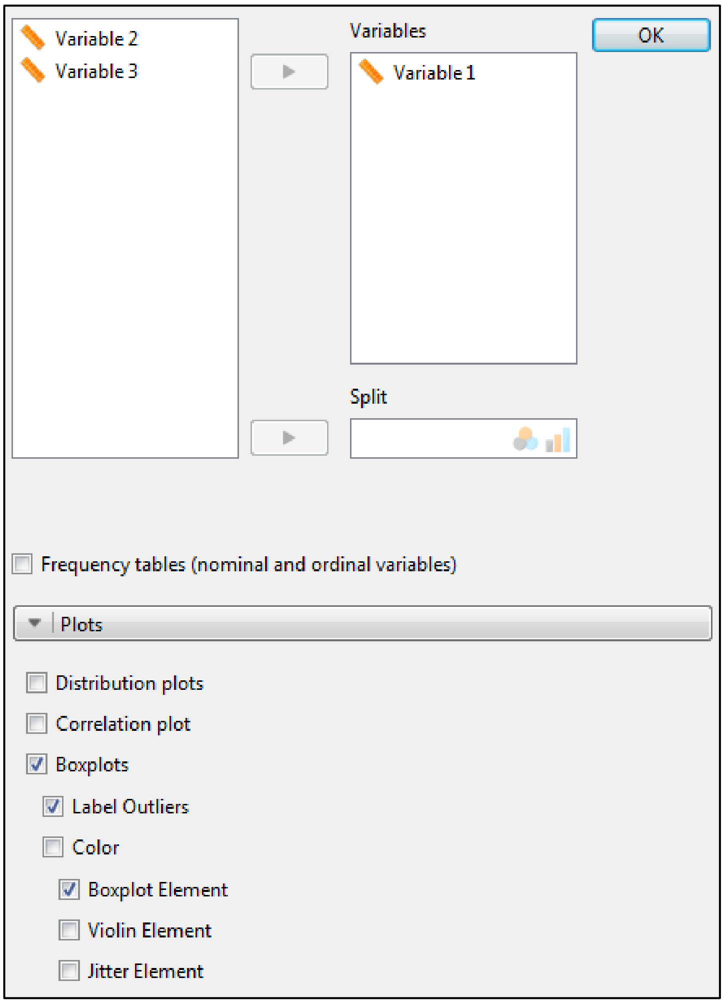
Els valors atípics són punts de les dades que es troben anormalment lluny d'altres punts. Un valor atípic pot ser degut a diferents motius, com errors en la introducció de dades o errors analítics comesos en el moment de la recollida de les dades. Els gràfics de caixa permeten visualitzar fàcilment aquests punts de dades en què els valors atípics són fora del límit del quartil superior (75% + 1,5 * IQR) o inferior (25% - 1,5 * IQR).



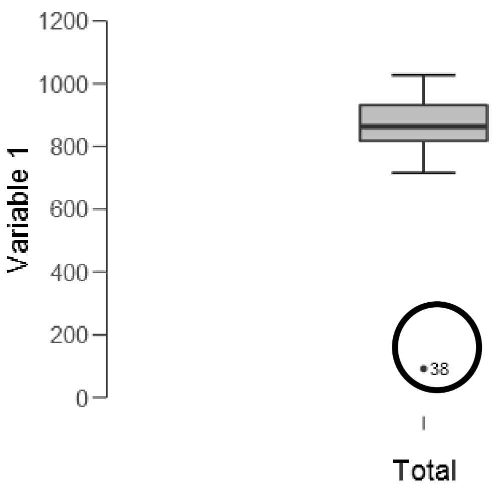
El gràfic de caixa mostra:

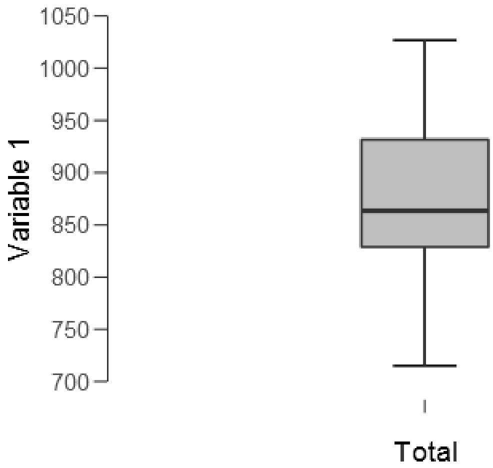
- Mediana.
- Quartils del 25% i 75%.
- IQR –Rang interquartil.
- Valors màxims i mínims representats una vegada exclosos els valors atípics.
- Si es demana, també es mostren els valors atípics.

Carregueu **Exploring Data.csv** a JASP. A «Descriptives» → «Descriptive statistics», afegiu la variable 1 a la caixa «Variables». En gràfics («Plots»), seleccioneu gràfics de caixa («Boxplots»), etiquetar valors atípics («Label Outliers») i element gràfic de caixa («Boxplot Element»).



El gràfic de caixa resultant que es mostra a l'esquerra es veu molt comprimit i es pot observar un valor atípic evident a la fila 38 del conjunt de dades. Això es pot deure a un error en la introducció de les dades, en introduir 91,7 en comptes de 917. El gràfic de caixa de la dreta mostra les dades "netes".





Com es manegi un valor atípic dependrà de la seva causa. La majoria de les proves paramètriques són molt sensibles als valors atípics, mentre que les no paramètriques generalment no ho són.

Corregir-lo? – Comprovem les dades originals per assegurar que no es tracti d'un error d'introducció de les dades; si és així, el corregim i executem l'anàlisi de nou.

Mantenir-lo? – Fins i tot en conjunts de dades amb distribució normal es poden esperar dades atípiques per a mostres grans i no s'han de descartar automàticament si es dona el cas.

Eliminar-lo? – És una pràctica controvertida en conjunts de dades petits en què no es pot assumir una distribució normal. Es poden excloure els valors atípics deguts a un error de lectura en l'instrument, però primer s'han de verificar.

Reemplaçar-lo? – També coneguda com a *winsorització*. Aquesta tècnica reemplaça els valors atípics pels valors màxims i/o mínims rellevants, trobats després d'excloure el valor atípic.

Qualsevol mètode que s'utilitzi ha d'estar justificat per la metodologia estadística adoptada i les anàlisis subsegüents.

FEM MOLTES SUPOSICIONS SOBRE LES NOSTRES DADES

Quan utilitzem proves paramètriques, partim d'una sèrie de suposicions sobre les nostres dades i si es violen aquests supòsits es produirà un biaix, en particular:

- Normalitat.
- Homogeneïtat de la variància o homoscedasticitat.

Moltes proves estadístiques són en realitat un conjunt de proves "òmnibus", algunes de les quals verifiquen aquests supòsits.

PROVA DEL SUPÒSIT DE NORMALITAT

La normalitat no significa necessàriament que les dades estiguin normalment distribuïdes *per se*, sinó si el conjunt de dades pot estar ben modelat per una distribució normal. La normalitat pot explorar-se per diferents vies:

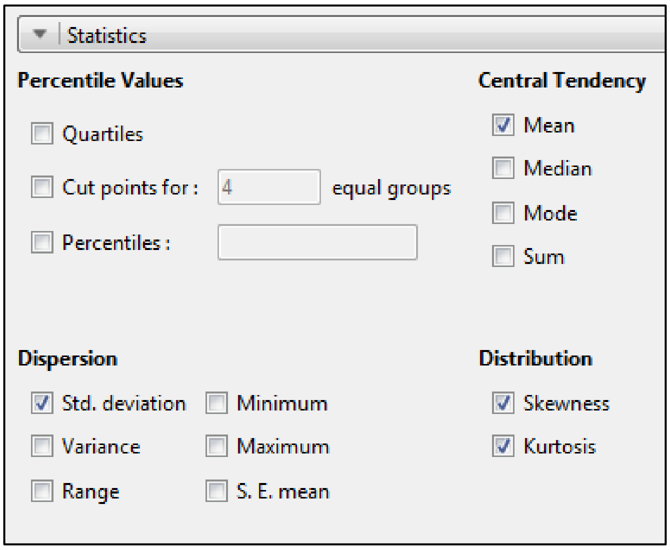
- Numèricament.
- Visualment / gràficament.
- Estadísticament.

Numèricament, podem utilitzar els resultats descriptius per a calcular l'asimetria i la curtosi. En una distribució normal, ambdós valors haurien de ser propers a zero. Per a determinar la significació de l'asimetria o la curtosi, calculem les puntuacions z (*z-scores*) dividint-les pels seus errors estàndard respectius:

$$\text{Asimetria}_z = \frac{\text{asimetria}}{\text{error estàndard de l'asimetria}} \qquad \text{Curtosi}_z = \frac{\text{curtosi}}{\text{error estàndard de la curtosi}}$$

Significació de la puntuació z: $p < 0{,}05$ si $z > 1{,}96$ $p < 0{,}01$ si $z > 2{,}58$ $p < 0{,}001$ si $z > 3{,}29$

Utilitzant **Exploring data.csv**, aneu a «Descriptives» → «Descriptive statistics» i mogueu la variable 3 a la caixa «Variables»; en el menú desplegable de «Statistics», seleccioneu «Mean», «Std. Deviation», «Skewness» i «Kurtosis» tal com es mostra a continuació en la corresponent taula de resultats.



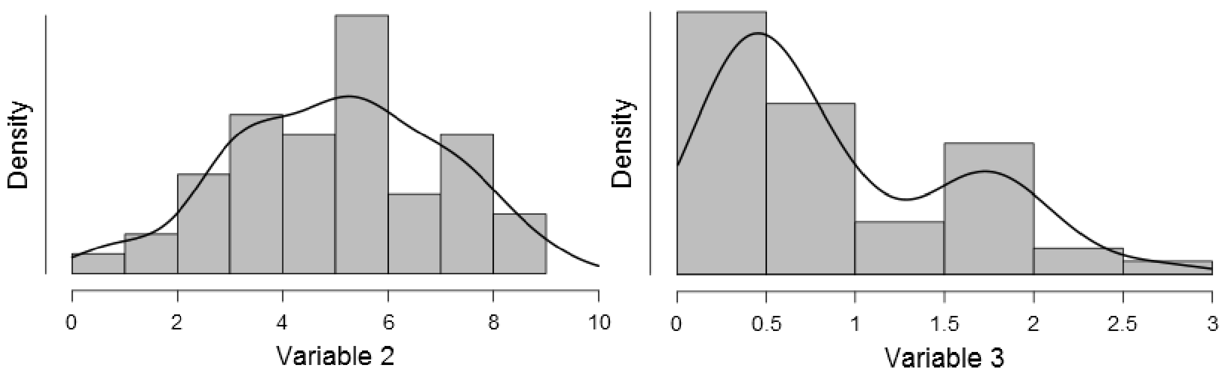
	Variable 3
Valid	50
Missing	0
Mean	0,893
Std. Deviation	0,673
Skewness	0,839
Std. Error of Skewness	0,337
Kurtosis	-0,407
Std. Error of Kurtosis	0,662

Es pot veure que l'asimetria i la curtosi no són properes a 0. L'asimetria positiva suggereix que les dades estan més distribuïdes cap a l'esquerra (veure els gràfics a continuació), mentre que la curtosi negativa suggereix una distribució plana. En calcular les seves puntuacions Z, es pot veure que les dades són asimètriques (p <0,05).

$$\text{Asimetria}_{Z} = \frac{0{,}839}{0{,}337} = 2{,}49 \qquad \text{Curtosi}_{Z} = \frac{-0{,}407}{0{,}662} = 0{,}614$$

[Noteu, com a advertència, que l'asimetria i la curtosi es mostren significatives en grans conjunts de dades, encara que la distribució sigui normal.]

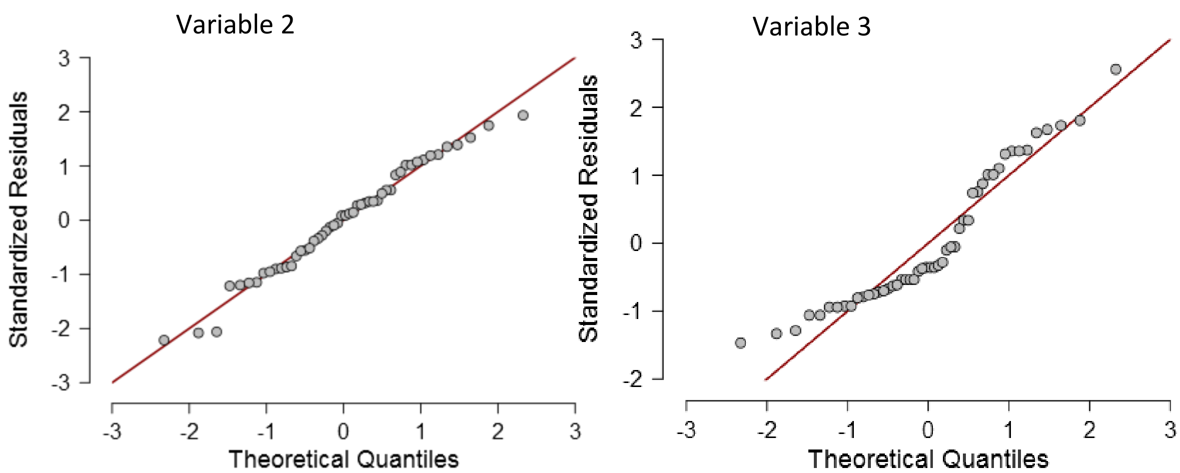
Ara, afegiu la variable 2 a la caixa «Variables» i, a «Plots», seleccioneu «Distribution plots». Això proporcionarà els dos gràfics següents:



Resulta fàcil veure que la variable 2 té una distribució simètrica. La variable 3 presenta una asimetria cap a l'esquerra, com confirma la puntuació Z de l'asimetria.

Una altra manera de comprovar gràficament la normalitat és mitjançant un gràfic Q-Q. Aquest procediment forma part de la comprovació dels supòsits de la **regressió** i l'**ANOVA**. Els gràfics Q-Q mostren els quantils de les dades reals enfront dels esperats per a una distribució normal.

Si les dades es distribueixen normalment, tots els punts estaran a prop de la línia diagonal de referència. Si els punts "cauen" per sobre o per sota de la línia, hi ha un problema amb la curtosi. Si els punts serpentegen al voltant de la línia, llavors el problema és l'asimetria. Tot seguit, es mostren els gràfics Q-Q per a les variables 2 i 3. Compareu-los amb els gràfics de distribució i les puntuacions Z d'asimetria / curtosi anteriors.



La prova de Shapiro-Wilk és una forma estadística utilitzada per JASP per verificar el supòsit de normalitat. S'utilitza en les proves t per a dues mostres independents (distribució dels dos grups) i aparellades (distribució de diferències entre parells). El test proporciona un valor de W, on els valors petits indiquen que la mostra no està distribuïda normalment (la hipòtesi nul·la que la població està distribuïda normalment si els seus valors estan per sota d'un cert llindar pot, per tant, ser rebutjada). La taula següent és un exemple de la taula de resultats de Shapiro-Wilk que no mostra cap desviació significativa de la normalitat en els 2 grups.

Test de normalitat (Shapiro-Wilk)

		W	p
Variable 2	Control	0,971	0,691
	Test	0,961	0,408

Nota. Els resultats significatius suggereixen una desviació de la normalitat.

La limitació més important és que la prova pot estar esbiaixada per la mida de la mostra. Com més gran sigui la mostra, més alta serà la probabilitat d'obtenir un resultat estadísticament significatiu.

Provant el supòsit de normalitat. Nota d'advertència!

Perquè la majoria dels tests paramètrics siguin fiables, un dels supòsits és que les dades es distribueixen de manera **aproximadament** normal. Una distribució normal assoleix el seu punt màxim al mig i és simètrica respecte a la mitjana. No obstant això, les dades no han d'estar distribuïdes de manera perfectament normal perquè els tests siguin fiables.

Llavors, calia estendre's tant sobre els tests de normalitat?

El teorema del límit central estableix que, a mesura que la mida de la mostra augmenta –és a dir, > 30 punts de dades– la distribució de les mitjanes mostrals s'aproxima a una distribució normal. Per tant, com més punts de dades es tinguin, més normal semblarà la distribució i més s'acostarà la mitjana de la mostra a la mitjana de la població.

Els conjunts de dades grans poden donar com a resultat proves significatives de normalitat; és a dir, mostrar Shapiro-Wilk o puntuacions Z d'asimetria i curtosi significatives quan els gràfics de distribució semblen bastant normals. I, al contrari, els conjunts de dades petits reduiran la potència estadística per detectar la no normalitat.

No obstant això, les dades que definitivament no compleixen amb el supòsit de normalitat oferiran resultats deficients en certs tipus de tests (en concret, aquells que assumeixen que s'ha de complir amb aquest supòsit). Fins a quin punt han d'ajustar-se les seves dades a una distribució normal? Per prendre una decisió en relació amb aquest supòsit, és millor observar les dades.

QUÈ FAIG SI LES MEVES DADES NO ES DISTRIBUEIXEN NORMALMENT?

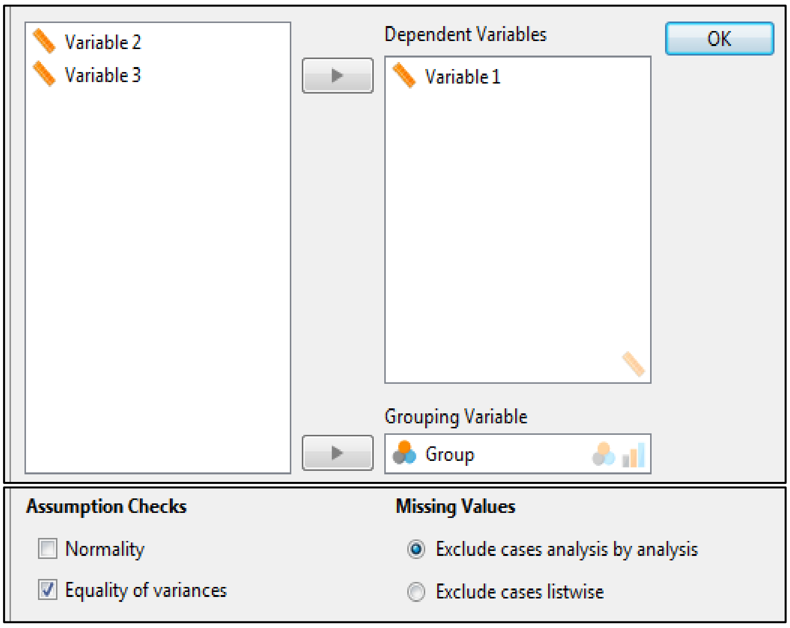
S'han de transformar les dades i realitzar novament comprovacions de normalitat per a les dades transformades. Les transformacions comunes inclouen calcular el logaritme o l'arrel quadrada de les dades.

És millor utilitzar tests no paramètrics, atès que es tracta de proves de distribució lliure i es poden emprar en comptes del seu equivalent paramètric.

PROVES D'HOMOGENEÏTAT DE LA VARIÀNCIA

El test de Levene s'utilitza freqüentment per provar la hipòtesi nul·la que les variàncies en els diferents grups són iguals. El resultat del test (F) es reporta com a valor de p; si no és significatiu, es pot assumir que la hipòtesi nul·la ha de ser mantinguda (que les variàncies són iguals); si el valor de p és significatiu, llavors la implicació és que les variàncies són desiguals. El test de Levene s'inclou en la **prova t independent i l'ANOVA**, en JASP, com a part de la comprovació dels supòsits.

Utilitzant **Exploring data.csv**, aneu a «T-Tests» → «Independent samples t-test», traslladeu la variable 1 a la caixa «Variables», la variable Group a la caixa «Grouping Variable» i marqueu «Assumption Checks» → «Equality of variances».

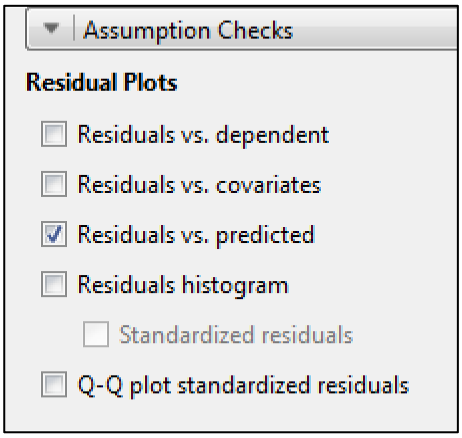


Test d'igualtat de variàncies (test de Levene)

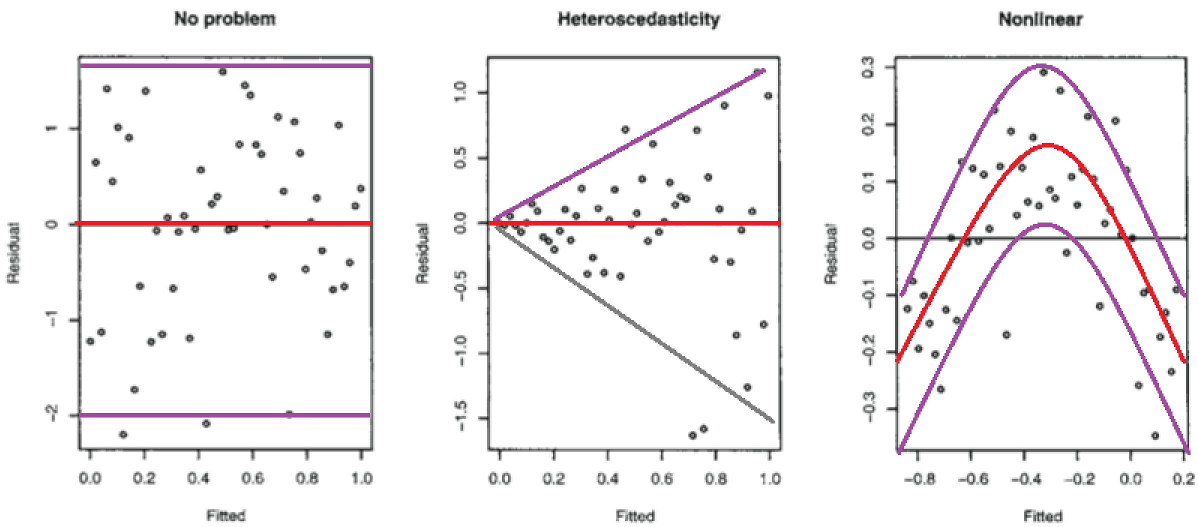
	F	df	p
Variable 1	0,218	1	0,643

En aquest cas, no hi ha diferències significatives en la variància entre els dos grups: F (1) = 0,218, p = 0,643.

El supòsit d'homoscedasticitat (igualtat de variància) és important en els models **de regressió lineal**, com ho és la linealitat. Aquesta prova assumeix que la variància de les dades al voltant de la línia de regressió és la mateixa per a tots els punts de dades de les variables predictores. L'heteroscedasticitat (la violació de l'homoscedasticitat) es presenta quan la variància difereix en els valors d'una variable independent. Això es pot avaluar visualment en una regressió lineal representant els residus obtinguts en relació amb els residus predits pel model.

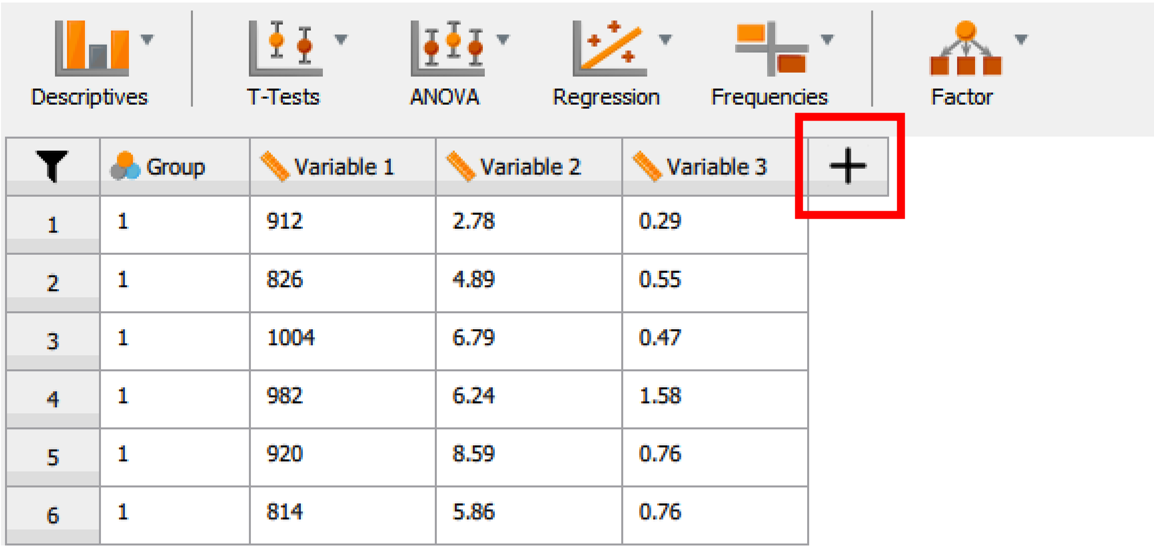


Si no es violen l'homoscedasticitat i la linealitat, no hi hauria d'haver una relació entre allò que el model prediu i els seus errors, com mostra el gràfic de l'esquerra. Qualsevol tipus de canalització (gràfic del mig) suggereix que s'ha violat l'homoscedasticitat i qualsevol corba (gràfic de la dreta) suggereix que no s'ha complert amb el supòsit de linealitat.



TRANSFORMACIÓ DE LES DADES

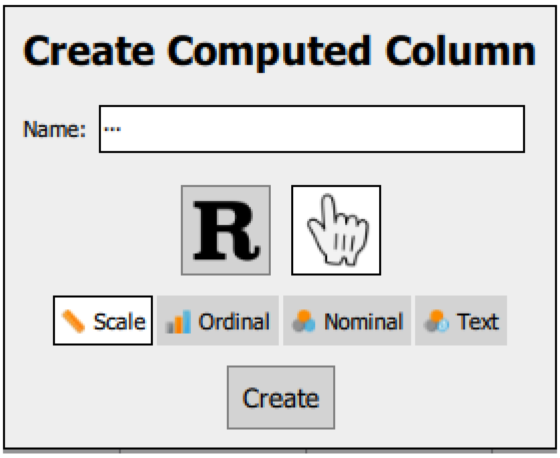
La capacitat per calcular noves variables o transformar dades va ser introduïda en la versió 0.9.1. En alguns casos, pot ser útil calcular les diferències entre mesures repetides o, perquè un conjunt de dades estigui distribuït d'una manera més normal, aplicar una transformació logarítmica, per exemple. Quan un conjunt de dades estigui carregat, hi haurà un signe més (**+**) al final de les columnes.



	Group	Variable 1	Variable 2	Variable 3
1	1	912	2.78	0.29
2	1	826	4.89	0.55
3	1	1004	6.79	0.47
4	1	982	6.24	1.58
5	1	920	8.59	0.76
6	1	814	5.86	0.76

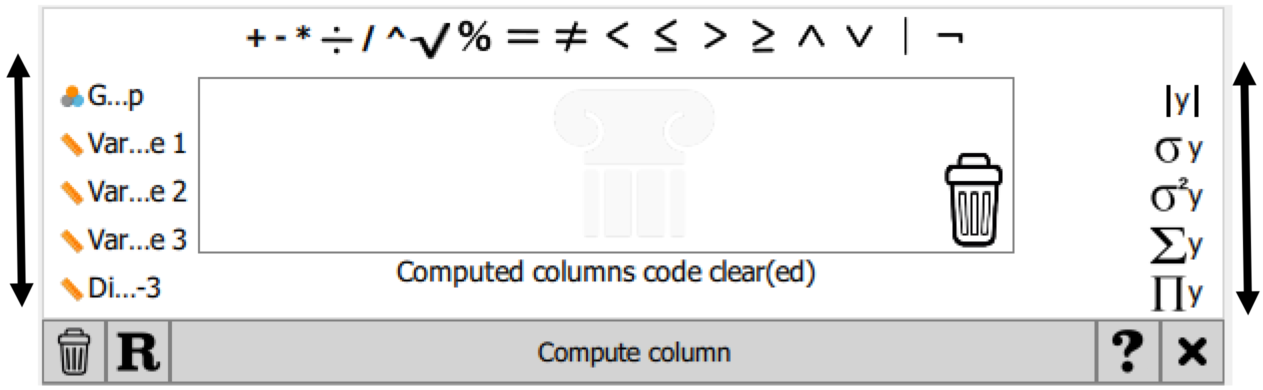
Fent clic a + s'obre un petit quadre de diàleg en el qual es pot:

- Introduir el nom d'una nova variable o de la variable transformada.
- Seleccionar si s'introdueix el codi R directament o es fan servir els comandaments integrats en JASP.
- Seleccionar quin tipus de dada es requereix.

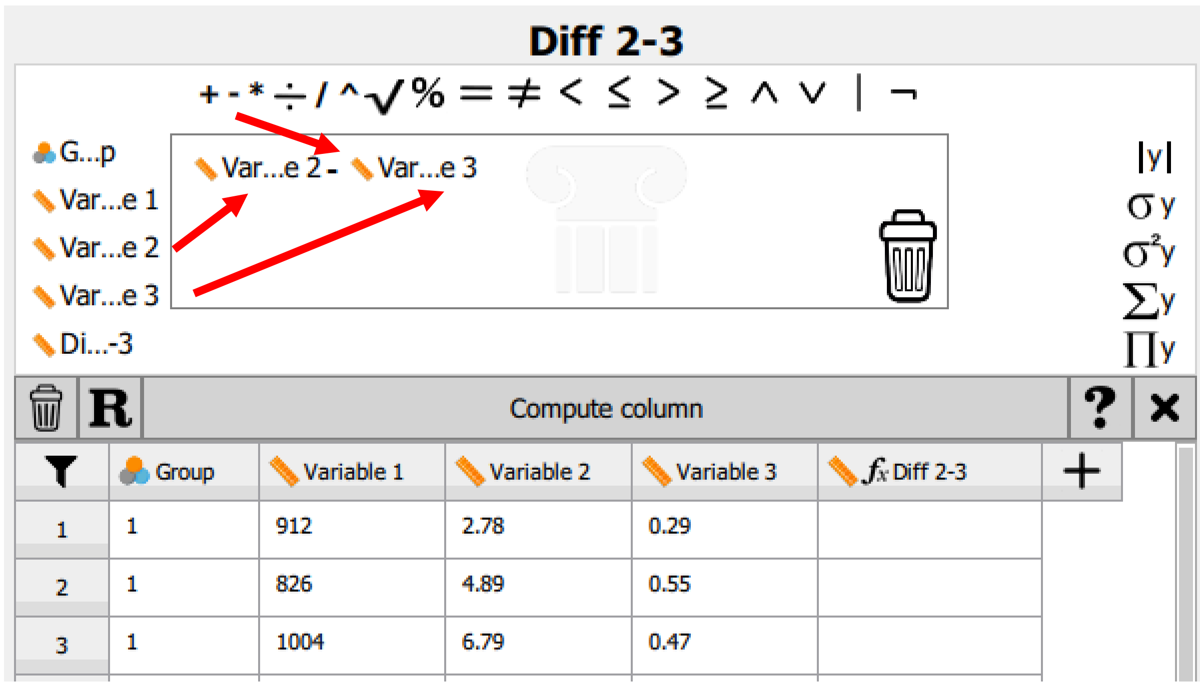


Un cop anomenada la nova variable i triades les altres opcions, cliqueu «Create».

Si es tria l'opció manual en comptes del codi R, s'obriran totes les opcions integrades per crear i transformar. Tot i no ser gaire intuïtiu, es pot navegar per les opcions que hi ha a mà esquerra i a mà dreta per trobar més variables i altres operadors, respectivament.

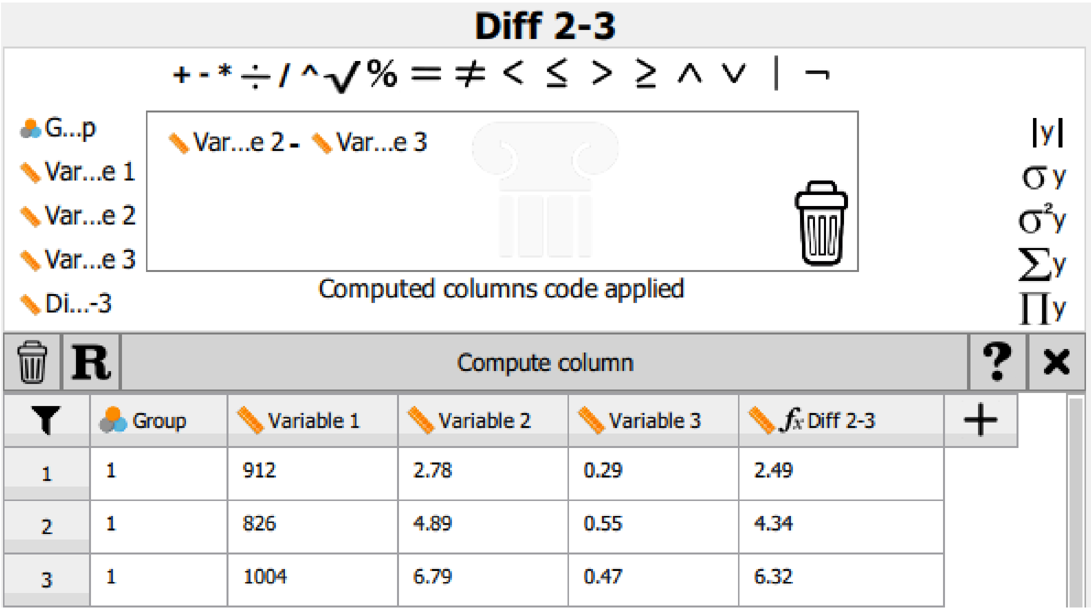


Per exemple, volem crear una columna de dades que mostri la diferència entre la variable 2 i la variable 3. Un cop introduït el nom de la columna al quadre de diàleg «Create computed column», aquest apareixerà a la finestra del full de càlcul. Ara caldrà definir les operacions matemàtiques. En aquest cas, arrossegueu la variable 2 fins a la caixa d'equacions, feu el mateix amb el signe "menys" i finalment arrossegueu la variable 3.



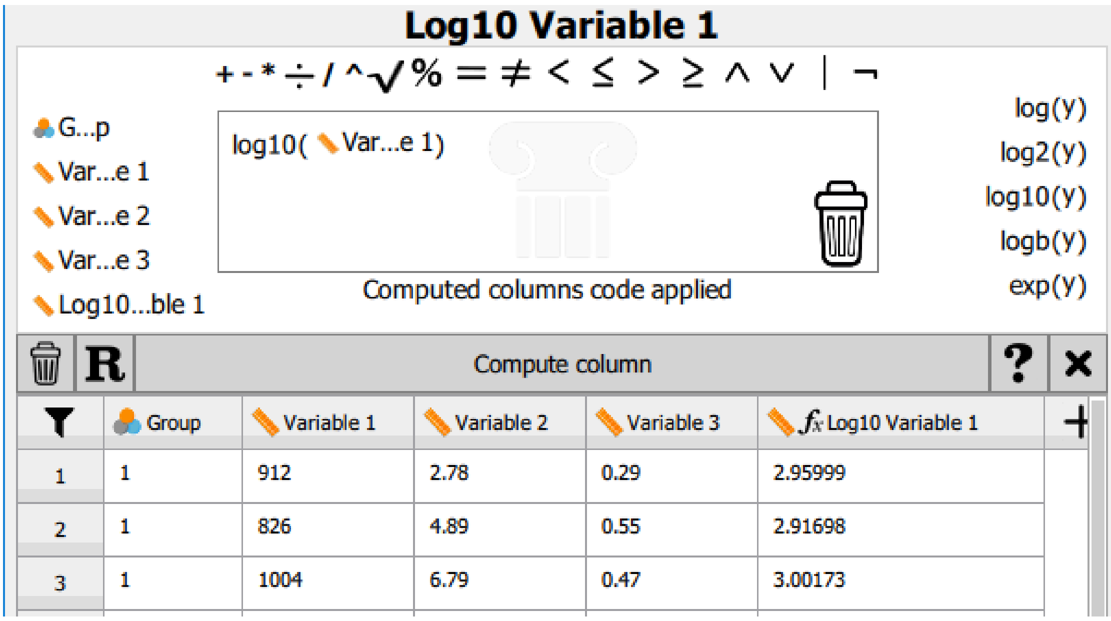
Si heu comès algun error, per exemple, si heu fet servir una variable o un operador erronis, elimineu-lo arrossegant l'ítem a la paperera que es troba a la cantonada inferior dreta.

Quan estigueu d'acord amb l'equació / operació, cliqueu a «Compute column» i la dada quedarà incorporada.

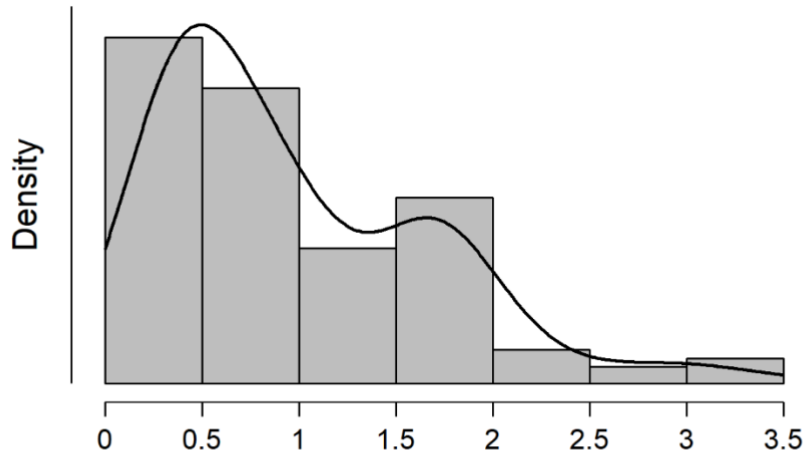


Si es decideix no conservar les dades derivades, es pot eliminar la columna clicant a l'altra icona de la paperera situada al costat de **R**.

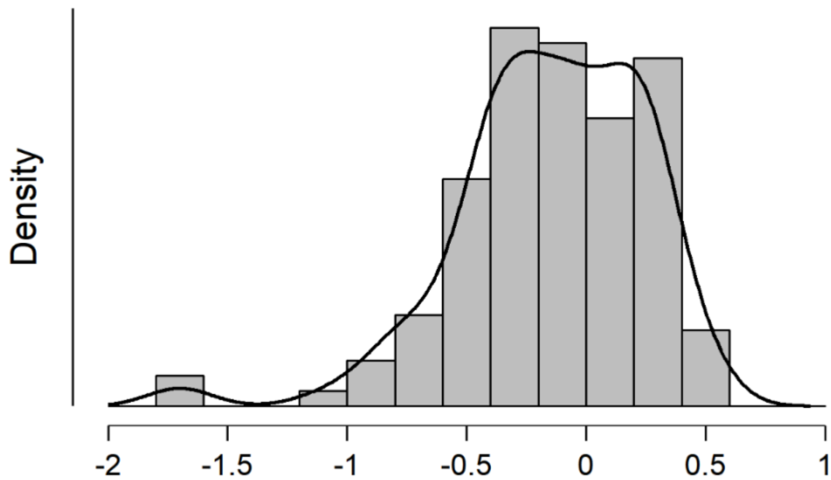
Un altre exemple seria realitzar una transformació logarítmica de les dades. En el cas següent, la variable 1 ha estat transformada desplaçant-se pels operadors de l'esquerra i seleccionant l'opció «log10(y)». Reemplaceu la *y* amb la variable que vulgueu transformar i després cliqueu a «Compute column». En acabar, feu clic a **X** per tancar el diàleg.



Els dos gràfics següents mostren les dades sense transformar i les transformades amb log10. Les dades clarament asimètriques han estat transformades en un perfil amb una distribució més normal.



Sense transformar



Transformats amb log10

La funció «Export» també exportarà totes les noves variables que hagin estat creades.

PROVA T PER A UNA MOSTRA ÚNICA

La investigació es porta a terme, normalment, amb mostres obtingudes d'una població, però, com de prop està la mostra de reflectir el conjunt de la població? La prova t paramètrica per a una mostra única determina si la mitjana de la mostra és estadísticament diferent de la mitjana coneguda o hipotètica de la població.

La hipòtesi nul·la (H_o) que es posa a prova és que la mitjana de la mostra és igual a la mitjana de la població.

SUPÒSITS

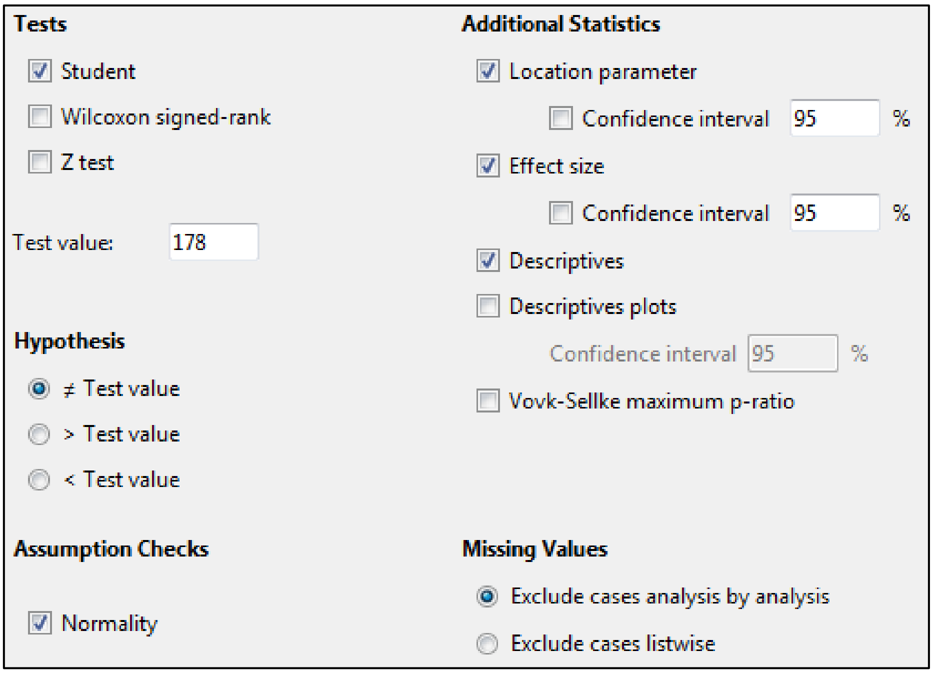
Es requereixen tres supòsits per a obtenir un resultat vàlid en la prova t per a una mostra única:

- La variable de la prova s'ha de mesurar en una escala **contínua**.
- Les dades de la variable de la prova han de ser **independents**, és a dir, sense relació entre cap dels punts de dades.
- Les dades han de seguir una **distribució aproximadament normal**.
- No hi ha d'haver **valors atípics** significatius.

EXECUTANT LA PROVA T PER A UNA MOSTRA ÚNICA

Obriu **one sample t-test.csv.** Aquest arxiu conté dues columnes de dades que representen l'alçada (cm) i les masses corporals (kg) d'una mostra d'homes utilitzada en un estudi. El 2017, les mitjanes de la població adulta masculina al Regne Unit eren **178 cm** d'alçada i **83,6 kg** de massa corporal.

Aneu a «T-Tests» → «One sample t-test» i afegiu, en primera instància, l'alçada a la caixa d'anàlisi de la dreta. Després d'això, seleccioneu les opcions següents i afegiu **178** com a valor de prova:



ENTENENT EL RESULTAT

El resultat hauria de contenir tres taules.

Test of Normality (Shapiro-Wilk)

	W	p
height	0.969	0.507

Note. Significant results suggest a deviation from normality.

La comprovació del supòsit de normalitat (Shapiro-Wilk) no és significativa, la qual cosa suggereix que les altures estan distribuïdes normalment; per tant, aquest supòsit no ha estat violat. Si l'anàlisi mostrés una diferència significativa, hauria de repetir-se utilitzant l'equivalent no paramètric, **la prova dels rangs amb signe de Wilcoxon** (*Wilcoxon's signed rank test*), provada sobre la mediana d'alçada de la població.

One Sample T-Test ▼

	t	df	p	Mean Difference	Cohen's d
height	−0.382	22	0.706	−0.391	−0.080

Note. Student's t-test.
Note. For the Student t-test, location parameter is given by mean difference *d*.
Note. For the Student t-test, effect size is given by Cohen's *d*.
Note. For all tests, the alternative hypothesis specifies that the population mean is different from 178.

Aquesta taula mostra que no hi ha diferències significatives entre les mitjanes: p = 0,706.

Descriptives

	N	Mean	SD	SE
height	23.000	177.609	4.915	1.025

Les dades descriptives mostren que l'alçada mitjana de la mostra era de 177,6 cm comparada amb la mitjana de 178 cm dels homes britànics.

Repetiu el procediment reemplaçant alçada per massa i canviant el valor de prova a 83,6.

Test of Normality (Shapiro-Wilk)

	W	p
mass	0.941	0.185

Note. Significant results suggest a deviation from normality.

La comprovació del supòsit de normalitat (Shapiro-Wilk) no és significativa, la qual cosa suggereix que les masses estan distribuïdes normalment.

One Sample T-Test

	t	df	p	Mean Difference	Cohen's d
mass	−7.159	22	< .001	−10.487	−1.493

Note. Student's t-test.
Note. For the Student t-test, location parameter is given by mean difference *d*.
Note. For the Student t-test, effect size is given by Cohen's *d*.
Note. For all tests, the alternative hypothesis specifies that the population mean is different from 83.4.

Aquesta taula mostra una diferència significativa entre la mitjana de la mostra (72,9 kg) i la massa corporal de la població (83,6 kg): p < 0,001.

Descriptives

	N	Mean	SD	SE
mass	23.000	72.913	7.025	1.465

REPORTANT ELS RESULTATS

Una prova t per a una mostra única no va exhibir diferències significatives en l'alçada en comparació amb la mitjana de la població: t (22) = -0,382, p = 0,706. No obstant això, els participants eren significativament més prims (menor massa corporal) que la mitjana de la població masculina del Regne Unit: t (22) = -7,159, p < 0,001.

TEST BINOMIAL

El test binomial és una versió no paramètrica de la prova t per a una mostra única destinat a utilitzar-se amb conjunts de dades categòriques dicotòmiques (és a dir, sí / no). Aquesta prova serveix per determinar si la freqüència de la mostra és estadísticament diferent de la freqüència poblacional coneguda o hipotètica.

La hipòtesi nul·la (H_0) que es posa a prova és que la freqüència de la mostra és igual a la freqüència poblacional esperada.

SUPÒSITS

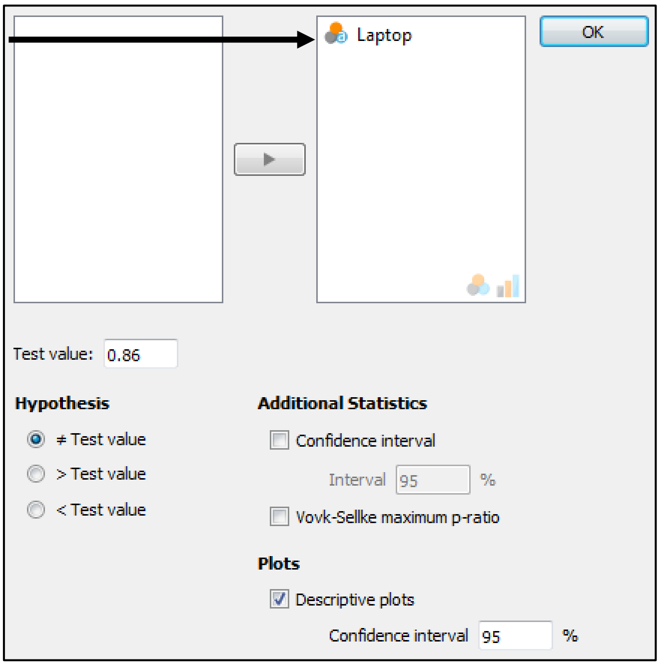
Es requereixen tres supòsits perquè un test binomial ofereixi un resultat vàlid:

- La variable del test ha de tenir una escala dicotòmica (com sí/no, masculí/femení, etc.).
- Les respostes de la mostra han de ser independents.
- La mida de la mostra és més petita, però segueix sent representativa de la població.

EXECUTANT EL TEST BINOMIAL

Obriu **binomial.csv**. Aquest arxiu conté una columna de dades que mostra el nombre d'estudiants que utilitzen o bé un portàtil Windows, o bé un MacBook a la universitat. El gener de 2018, comparant aquests dos sistemes operatius, la quota de mercat de Windows al Regne Unit era del 86%, i la de Mac IOS del 14%.[3]

Aneu a «Frequencies» → «Binomial test». Traslladeu la variable Laptop a la finestra de dades i indiqueu el valor de prova en 0,86 (86%). Seleccioneu, també, «Descriptive plots».



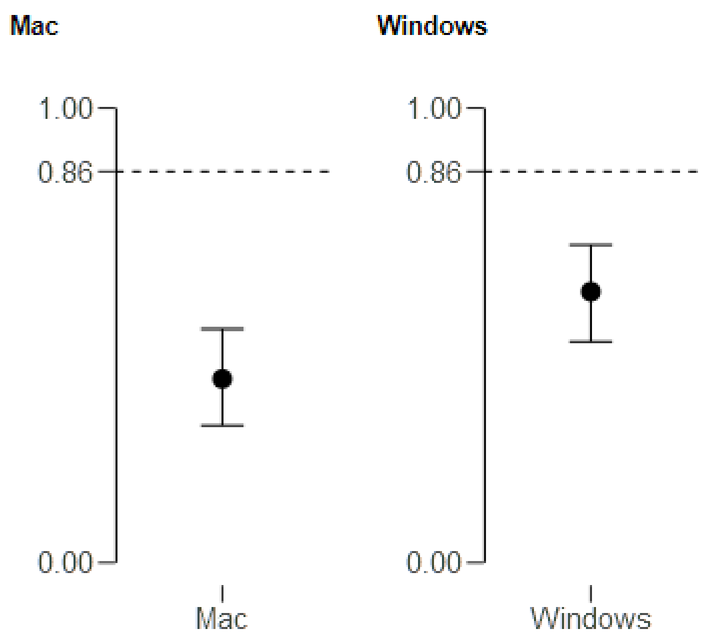
[3] https://www.statista.com/statistics/268237/global-market-share-held-by-operating-systems-since-2009/

La taula i el gràfic següents mostren que les freqüències d'ambdós portàtils són significativament inferiors al 86%. En particular, aquests estudiants estan utilitzant portàtils Windows d'una manera significativament inferior al que s'esperava, comparat amb la quota de mercat al Regne Unit.

Binomial Test

	Level	Counts	Total	Proportion	p
Laptop	Mac	36	89	0.404	< .001
	Windows	53	89	0.596	< .001

Note. Proportions tested against value: 0.86.

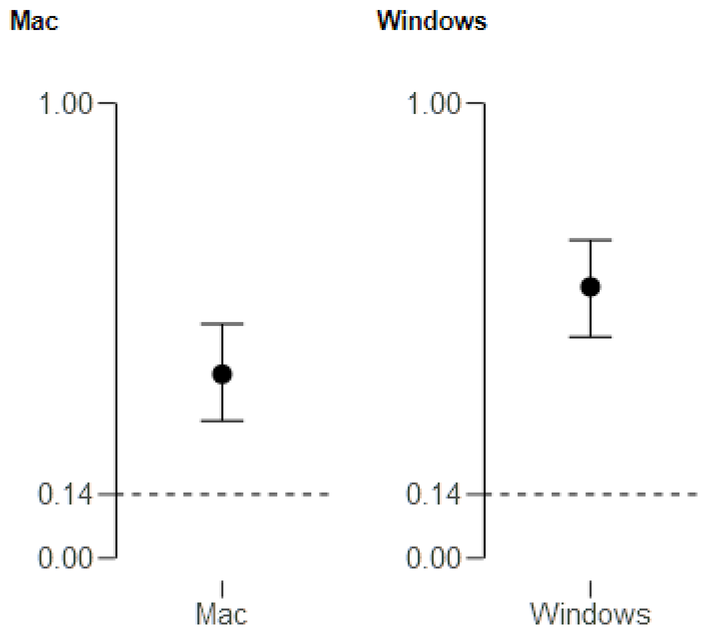


Succeeix el mateix amb els usuaris de MacBook? Torneu a la finestra d'opcions i canvieu el valor de prova per 0,14 (14%). Aquest cop, la freqüència és significativament superior al 14%. Això mostra que els estudiants fan servir MacBooks d'una manera significativament superior al que s'esperava, comparat amb la quota de mercat al Regne Unit.

Binomial Test

	Level	Counts	Total	Proportion	p
Laptop	Mac	36	89	0.404	< .001
	Windows	53	89	0.596	< .001

Note. Proportions tested against value: 0.14.



REPORTANT ELS RESULTATS

La proporció reportada d'usuaris britànics de Windows i MacBook va ser, respectivament, del 86% i del 14%. En una cohort d'estudiants universitaris (N = 90), un test binomial va revelar que la proporció d'estudiants usuaris de portàtils Windows era significativament inferior (59,6%, $p < 0,001$) i els que utilitzaven MacBooks ho feien de manera significativament superior (40,4%, $p < 0,001$) al que s'esperava.

TEST MULTINOMIAL

El test multinomial és una extensió del test binomial, destinat a utilitzar-se amb conjunts de dades categòriques que continguin tres o més factors. Aquesta prova serveix per determinar si la freqüència de la mostra és o no és estadísticament diferent d'una freqüència poblacional hipotètica (test multinomial) o coneguda (test de "bondat d'ajustament" chi quadrat).

La hipòtesi nul·la (H_0) que es posa a prova és que la freqüència de la mostra és igual a la freqüència poblacional esperada.

SUPÒSITS

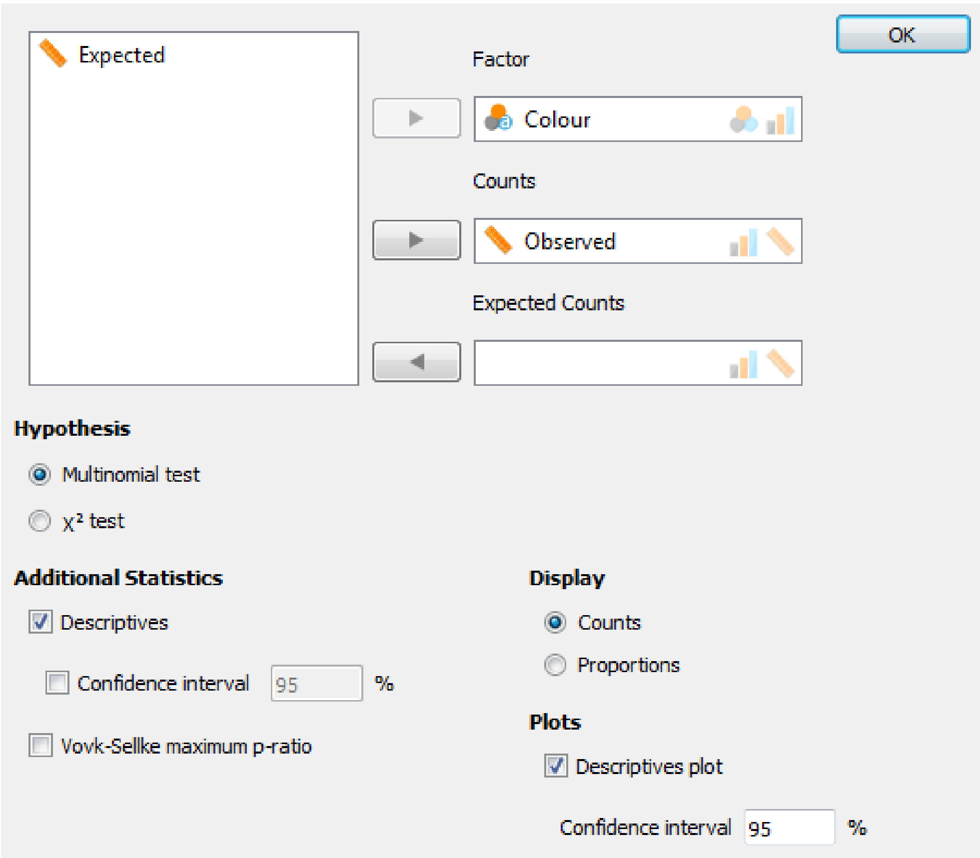
Es requereixen tres supòsits perquè un test multinomial proporcioni un resultat vàlid:

- La variable del test ha de tenir una escala categòrica amb 3 o més factors.
- Les respostes de la mostra han de ser independents.
- La mida de la mostra és més petita, però segueix sent representativa de la població.

EXECUTANT EL TEST MULTINOMIAL

Obriu **multinomial.csv.** Aquest arxiu conté tres columnes de dades que mostren el nombre de M&M de diferents colors repartits en cinc bosses. Sense cap coneixement previ, es podria suposar que els M&M de diferents colors es distribueixen uniformement.

Aneu a «Frequencies» → «Multinomial test». Traslladeu el color de l'M&M a «Factor» i el nombre observat de M&M a «Counts». Seleccioneu «Descriptives» i «Descriptives plot».



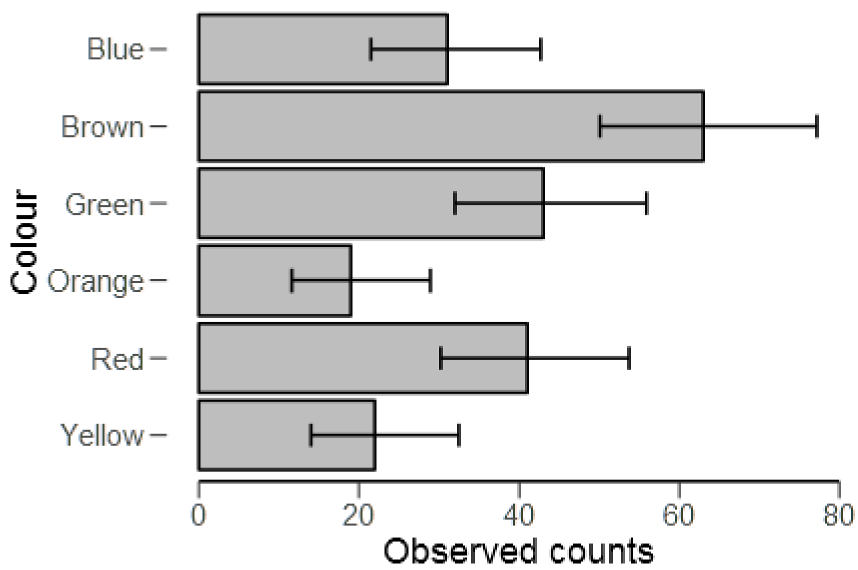
Com es pot veure a la taula de descriptives, el test assumeix una mateixa expectativa per a les proporcions de M&M de colors (36 de cada color). Els resultats del test multinomial mostren que la distribució observada és significativament diferent ($p < 0{,}001$) a una distribució equitativa.

Multinomial Test

	χ^2	df	p
Multinomial	35.932	5	< .001

Descriptives

Colour	Observed	Expected: Multinomial
Blue	31	36
Brown	63	36
Green	43	36
Orange	19	36
Red	41	36
Yellow	22	36



TEST DE “BONDAT D’AJUSTAMENT” CHI QUADRAT

No obstant això, investigacions addicionals mostren que els fabricants produeixen M&M de colors en diferents proporcions:

Color	Blau	Marró	Verd	Taronja	Vermell	Groc
Proporció	24	13	16	20	13	14

Ara, aquests valors poden ser utilitzats com a recomptes estimats, per tant moveu la variable Expected a la caixa «Expected Counts». Això executa automàticament el test de “bondat d’ajustament” χ^2 tot deixant en gris les opcions d’hipòtesis.

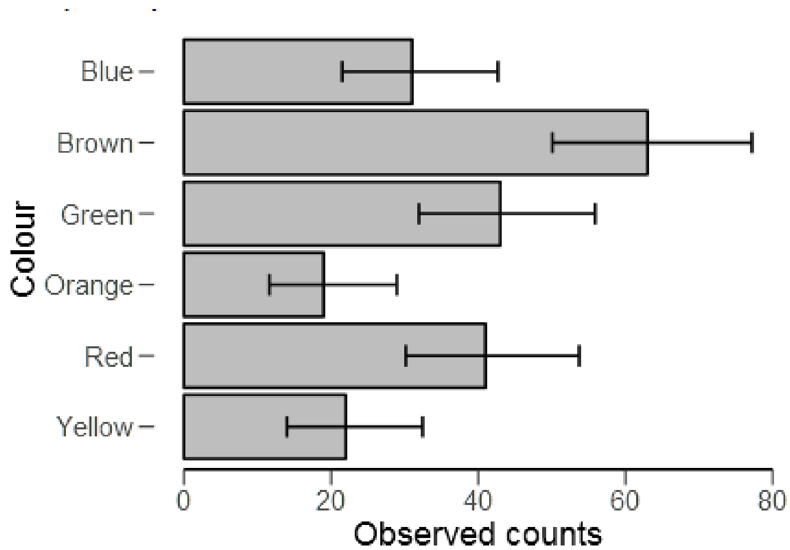
Com es pot veure a la taula de descriptives, JASP ha calculat els números esperats dels M&M de diferents colors sobre la base de la ràtio de producció reportada pels fabricants. Els resultats del test mostren que les proporcions observades per als M&M de diferents colors són significativament diferents (χ^2 = 74,5, p < 0,001) de les proporcions declarades pel fabricant.

Multinomial Test

	χ^2	df	p
Expected	74.535	5	< .001

Descriptives

Colour	Observed	Expected: Expected
Blue	31	52
Brown	63	28
Green	43	35
Orange	19	43
Red	41	28
Yellow	22	30



TEST MULTINOMIAL I DE "BONDAT D'AJUSTAMENT" X^2

JASP també proporciona una altra opció mitjançant la qual ambdues proves es poden executar alhora. Torneu a la finestra d'opcions i afegiu la variable Colour a la caixa «Factor» i Observed a la caixa «Counts»; elimineu Expected de la caixa «Expected Counts» si la variable encara es troba aquí. A «Hypothesis», marqueu el test χ^2. Això obrirà una petita finestra de full de càlcul que mostrarà el color i H_o (a) amb un 1 a cada cel·la. Això implica que les proporcions de cada color són les mateixes (test multinomial).

En aquesta finestra, afegiu una altra columna que s'etiquetarà automàticament com a H_o (b). Ara es poden introduir les proporcions estimades per a cada color.

	H_0 (a)	H_0 (b)
Brown	1	13
Green	1	16
Orange	1	20
Red	1	13
Yellow	1	14

Add column

Delete column

Reset

Ara, una vegada executada l'anàlisi, es mostren els resultats de les proves per a les dues hipòtesis. H_o (a) comprova la hipòtesi nul·la que les proporcions de cada color estan distribuïdes uniformement, mentre que H_o (b) comprova la hipòtesi nul·la que les proporcions són les mateixes que les esperades. Com es pot observar, ambdues hipòtesis són rebutjades. En concret, l'evidència indica que els colors dels M&M no coincideixen amb les proporcions publicades pels fabricants.

Multinomial Test

	χ^2	df	p
H_0 (a)	35.932	5	< .001
H_0 (b)	74.535	5	< .001

Descriptives

		Expected	
Colour	Observed	H_0 (a)	H_0 (b)
Blue	31	36	52
Brown	63	36	28
Green	43	36	35
Orange	19	36	43
Red	41	36	28
Yellow	22	36	30

COMPARACIÓ DE DOS GRUPS INDEPENDENTS

PROVA T PER A DUES MOSTRES INDEPENDENTS

La prova t paramètrica per a dues mostres independents, també coneguda com a prova t de Student (*Student's t-test*), s'utilitza per determinar si hi ha diferència estadística entre les mitjanes de dos grups independents. La prova requereix una variable dependent contínua (p. ex., massa corporal) i una variable independent que contingui dos grups (p. ex., homes i dones).

Amb aquesta prova s'obté una puntuació t (*t-score*) que és el quocient de les diferències entre els dos grups i les diferències dins dels dos grups:

$$t = \frac{\boldsymbol{mitjana\ grup\ 1 - mitjana\ grup\ 2}}{\boldsymbol{error\ estàndard\ de\ les\ mitjanes}}$$

$$t = \frac{(x_1 - x_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

X = mitjana

S = desviació estàndard

n = nombre de punts de dades

Una puntuació t alta indica que hi ha una gran diferència entre els grups. Com més baixa sigui la puntuació t, més gran serà la similitud entre els grups. Una puntuació t de 5 indica que els grups són cinc vegades més diferents entre ells del que ho són a dins de cadascun d'ells.

La hipòtesi nul·la (H_o) que es posa a prova és que les mitjanes poblacionals dels dos grups no relacionats són iguals.

SUPÒSITS DE LA PROVA T PARAMÈTRICA PER A DUES MOSTRES INDEPENDENTS

Independència del grup:

Ambdós grups han de ser independents entre si. Cada participant només proporcionarà un punt de dades per a un sol grup. Per exemple, el participant 1 només pot estar en u grup, masculí o femení, però no en tots dos. Les mesures repetides s'avaluen amb la **prova t per a dues mostres aparellades** (*paired t-test*).

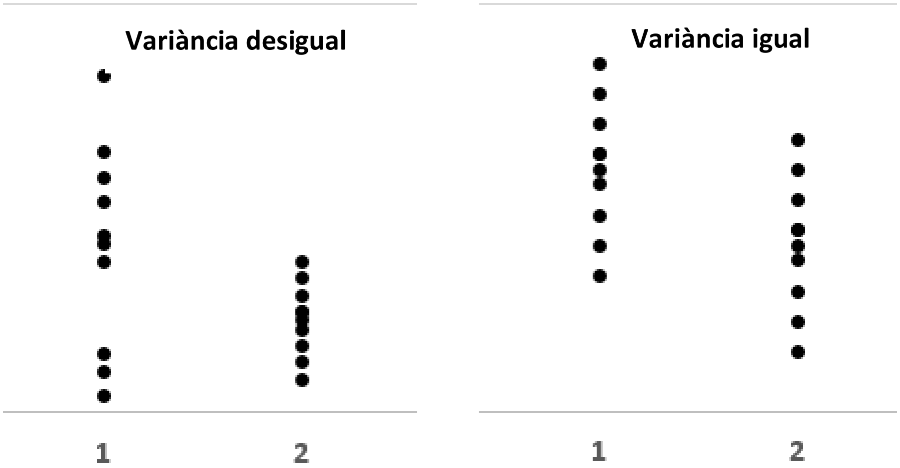
Normalitat de la variable dependent:

La variable dependent també s'ha de mesurar en una escala contínua i ha de tenir una distribució aproximadament normal, sense valors atípics significatius. Això es pot comprovar mitjançant el test Shapiro-Wilk. La prova t és força robusta, de manera que es poden acceptar petites desviacions de la normalitat. Amb tot, això no és així en el cas de grups amb mides molt diferents. Com a regla general, la ràtio entre les mides de grup ha de ser < 1,5 (p. ex., grup A = 12 participants i grup B = > 8 participants).

Si la normalitat ha estat violada, podeu provar de transformar les dades (p. ex., transformacions logarítmiques o arrel quadrada) o, si les mides de grup són molt diferents, utilitzar el test **U de Mann-Whitney**, l'equivalent no paramètric que no requereix el supòsit de normalitat (veure més endavant).

Homogeneïtat de la variància:

Les variàncies de la variable dependent han de ser iguals en cada grup. Això es pot comprovar amb el test d'igualtat de variàncies de Levene.

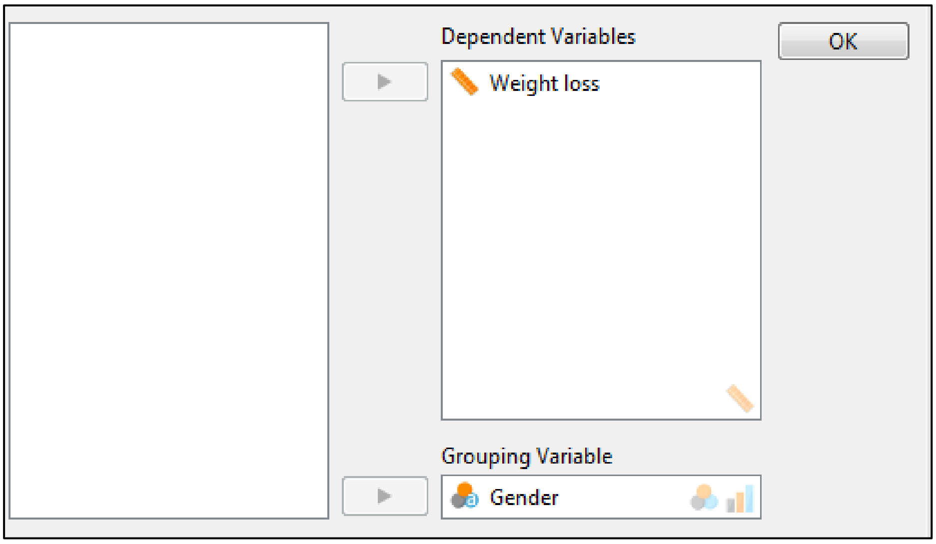


Si el test de Levene és estadísticament significatiu, indicant que les variàncies dels grups són desiguals, es pot corregir aquesta violació utilitzant una prova t ajustada segons el mètode de **Welch**.

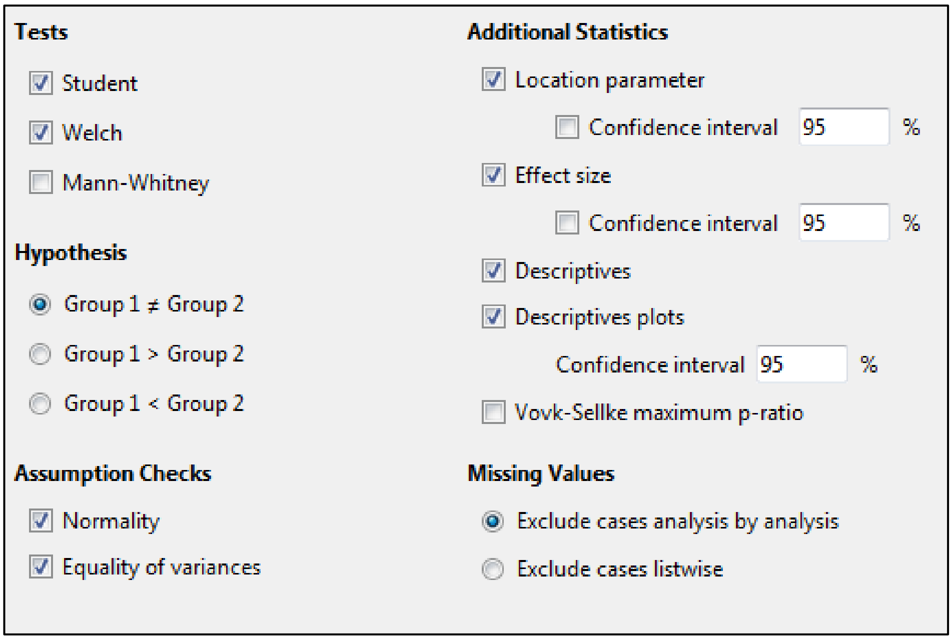
EXECUTANT LA PROVA T PER A DUES MOSTRES INDEPENDENTS

Obriu **Independent t-test.csv**. Aquest arxiu conté la pèrdua de pes amb una dieta autocontrolada de 10 setmanes entre homes i dones. És una bona pràctica comprovar la distribució i els gràfics de caixa a «Descriptives», per verificar visualment la distribució i els valors atípics.

Aneu a «T-tests» → «Independent samples t-test», i introduïu la pèrdua de pes a la caixa «Dependent Variables» i el gènere (variable independent) a la caixa «Grouping Variable».



A la finestra d'anàlisi, seleccioneu les opcions següents:



ENTENENT ELS RESULTATS

El resultat ha de contenir quatre taules i un gràfic. En primer lloc, cal comprovar que no es violen els supòsits paramètrics requerits.

Test of Normality (Shapiro-Wilk)

		W	p
Weight loss	Females	0.968	0.282
	Males	0.971	0.310

Note. Significant results suggest a deviation from normality.

El test Shapiro-Wilk mostra que els dos grups tenen dades distribuïdes normalment, per la qual cosa no es viola el supòsit de normalitat. Si un o ambdós fossin significatius, caldria considerar l'ús del test equivalent no paramètric de **Mann-Whitney**.

Test of Equality of Variances (Levene's)

	F	df	p
Weight loss	2.278	1	0.135

La prova de Levene mostra que no hi ha diferència en la variància, per tant, no es viola el supòsit d'homogeneïtat de la variància. Si la prova de Levene fos significativa, s'hauria de reportar la prova t amb la correcció de Welch, els graus de llibertat i els valors de p.

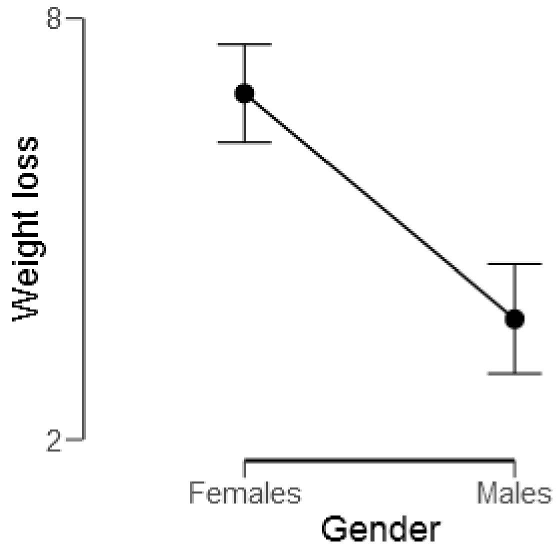
Independent Samples T-Test

	Test	Statistic	df	p	Mean Difference	SE Difference	Cohen's d
Weight loss	Student	6.160	85.000	< .001	3.209	0.521	1.322
	Welch	6.191	84.544	< .001	3.209	0.518	1.325

Aquesta taula mostra el càlcul de les dues proves t (**Student i Welch**). Hem de recordar que l'estadístic t s'obté dividint la diferència de mitjanes per l'error estàndard de la diferència. Ambdós mostren que hi ha una diferència estadística significativa entre els dos grups ($p < 0,001$), i la d de Cohen suggereix que es tracta d'un efecte important.

Group Descriptives

	Group	N	Mean	SD	SE
Weight loss	Females	42	6.929	2.242	0.346
	Males	45	3.720	2.588	0.386



A partir de les dades descriptives, es pot veure que les dones van tenir una pèrdua de pes més gran que els homes.

REPORTANT ELS RESULTATS

Una prova t per a dues mostres independents va mostrar que les dones han perdut significativament més pes després de 10 setmanes de dieta que els homes: **$t(85) = 6,16$, $p < 0,001$**. La d de Cohen (1,322) suggereix que es tracta d'un efecte important.

PROVA U DE MANN-WITNEY

Si es dona el cas que les dades no estan normalment distribuïdes (resultat significatiu del test de Shapiro-Wilk) o si la distribució és ordinal, la prova no paramètrica per a dues mostres independents equivalent és la **prova U de Mann-Whitney**.

Obriu **Mann-Whitney pain.csv**. Aquest arxiu conté puntuacions de dolor subjectiu (0-10) amb i sense tractament amb gel. **Nota**: comproveu que el tractament sigui categòric i que la puntuació del dolor sigui ordinal. Aneu a «T-test» → «Independent t-test» i afegiu la puntuació del dolor a la caixa «Dependent Variables», utilitzant el tractament com a variable d'agrupació.

En les opcions d'anàlisi, seleccioneu només:

- ✓ Mann-Whitney.
- ✓ Paràmetre de localització (*Location parameter*).
- ✓ Mida de l'efecte (*Effect size*).

No hi ha cap motiu per comprovar els supòsits, atès que Mann-Whitney no assumeix el supòsit de normalitat ni el d'homogeneïtat de la variància requerits per les proves paramètriques.

ENTENENT EL RESULTAT

Aquesta vegada només s'obté una taula:

Independent Samples T-Test

	W	p	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Pain score	207.000	< .001	3.000	0.840

Note. Mann-Whitney U test.

El test estadístic U de Mann-Whitney (JASP la reporta com a *W*, ja que es tracta d'una adaptació del test dels rangs amb signe de Wilcoxon) és altament significatiu: **U = 207, p < 0,001**.

El paràmetre de localització, l'estimació Hodges-Lehmann, és la diferència **mediana** entre els dos grups. La correlació de rang biserial (*Rank-Biserial Correlation*, r_B) pot ser considerada com a mida de l'efecte i interpretada de la mateixa manera que la r de Pearson, de manera que 0,84 és una mida de l'efecte important.

Per a dades no paramètriques, s'han de reportar valors **medians** com a estadística descriptiva i utilitzar gràfics de caixa en comptes de gràfics de línies i intervals de confiança, barres SD / ES. Aneu a «Descriptive statistics», introduïu la puntuació de dolor a la caixa «Variables» i el tractament a la caixa «Split».

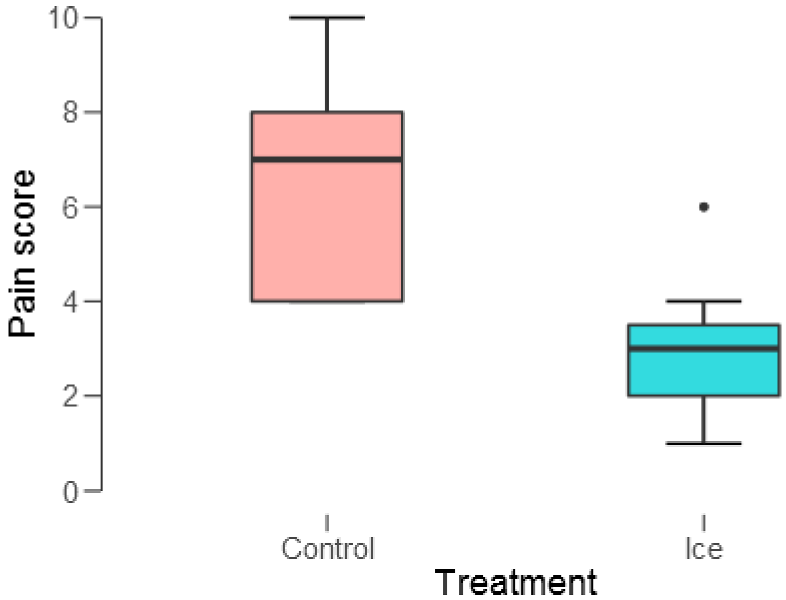
Descriptive Statistics

	Pain score	
	Control	Ice
Valid	15	15
Missing	0	0
Median	7.000	3.000
Minimum	4.000	1.000
Maximum	10.000	6.000

Plots

Boxplots

Pain score



REPORTANT ELS RESULTATS

El test de Mann-Whitney va mostrar que el tractament amb gel redueix significativament les puntuacions de dolor (Mdn = 3), en comparació amb el grup de control (Mdn = 7), U = 207, p < 0,001.

COMPARACIÓ DE DOS GRUPS RELACIONATS

PROVA T PER A DUES MOSTRES APARELLADES

Com succeeix amb la prova t per a dues mostres independents, JASP ofereix ambdues opcions: la paramètrica i la no paramètrica. La prova t paramètrica per a dues mostres aparellades (també coneguda com a prova t per a mostres dependents o prova t per a mesures repetides) compara les mitjanes entre dos grups relacionats en la mateixa variable contínua dependent. Per exemple, observant la pèrdua de pes abans i després de les 10 setmanes de dieta.

$$\text{Estadístic } t \text{ aparellat} = \frac{\text{mitjana de les diferències entre les parelles dels grups}}{\text{errors estàndard de les diferències de les mitjanes}}$$

Amb la prova t per a dues mostres aparellades, la hipòtesi nul·la (H_0) que es posa a prova és que la diferència entre les parelles dels dos grups és zero.

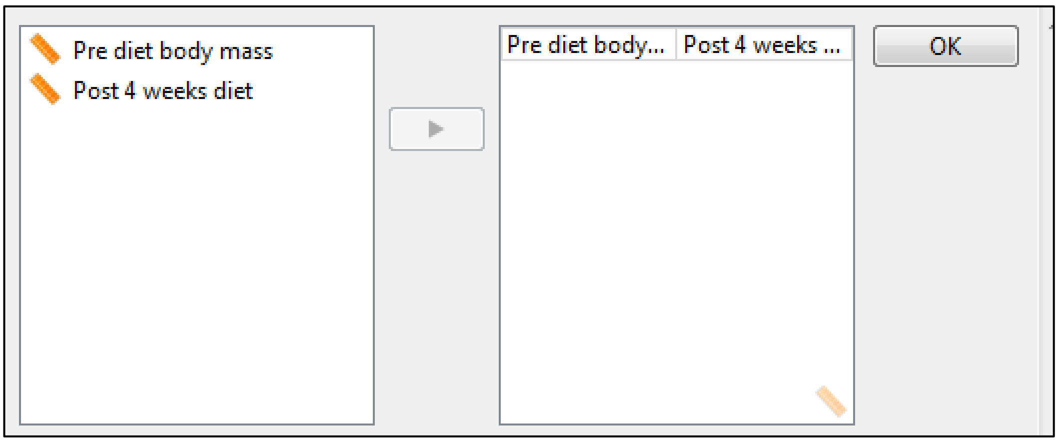
SUPÒSITS DE LA PROVA T PARAMÈTRICA PER A DUES MOSTRES APARELLADES

Perquè la prova t paramètrica proporcioni un resultat vàlid, es requereixen quatre supòsits:

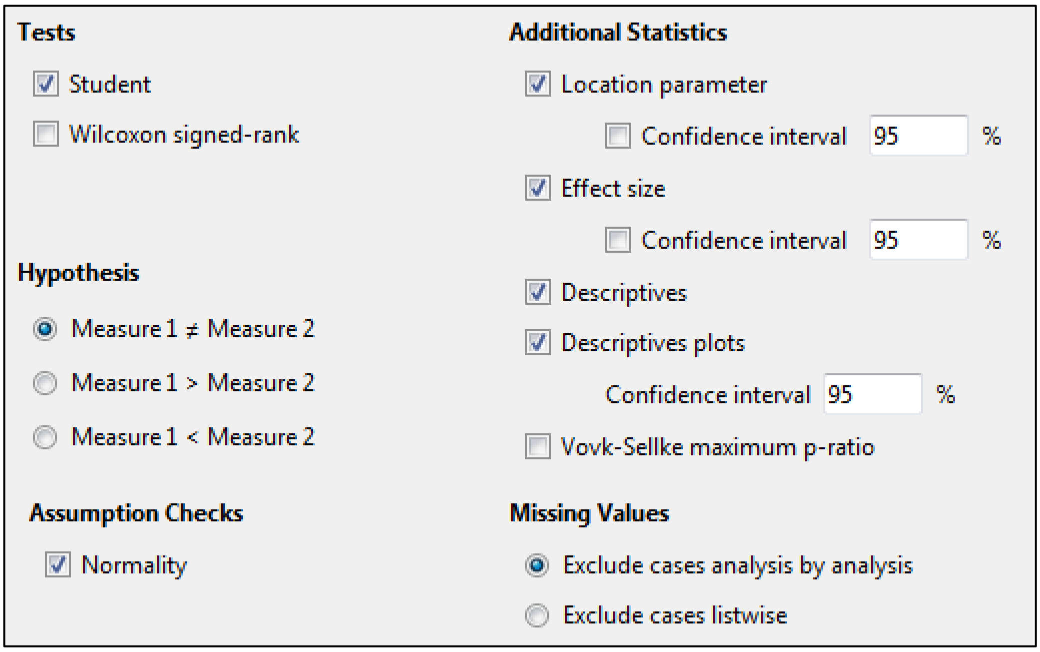
- La **variable dependent** ha de ser mesurada en una escala contínua.
- La **variable independent** ha de comptar amb 2 grups categòrics relacionats / aparellats, és a dir, que cada participant apareix en ambdós grups.
- Les diferències entre les parelles han d'estar aproximadament **distribuïdes normalment**.
- No hi ha d'haver **valors atípics** significatius en les diferències entre els 2 grups.

EXECUTANT LA PROVA T PER A MOSTRES APARELLADES

Obriu **Paired t-test.csv** a JASP. Aquest arxiu conté dues columnes de dades aparellades: massa corporal anterior a la dieta i després de 4 setmanes fent dieta. Aneu a «T-test» → «Paired samples t-test». Feu clic sobre ambdues variables tot mantenint la tecla Ctrl pressionada i afegiu-les a la caixa d'anàlisi de la dreta.



En les opcions d'anàlisi, marqueu el següent:



ENTENENT EL RESULTAT

El resultat ha d'incloure tres taules i un gràfic.

Test of Normality (Shapiro-Wilk)

			W	p
Pre diet body mass	-	Post 4 weeks diet	0.975	0.124

Note. Significant results suggest a deviation from normality.

La comprovació del supòsit de normalitat (Shapiro-Wilk) no és significativa, suggerint que les diferències aparellades estan distribuïdes normalment, de manera que es compleix el supòsit. Si mostrés una diferència significativa, s'hauria de repetir l'anàlisi utilitzant l'equivalent no paramètric, la **prova de rangs amb signe de Wilcoxon.**

Paired Samples T-Test

			t	df	p	Mean Difference	SE Difference	Cohen's d
Pre diet body mass	-	Post 4 weeks diet	13.039	77	< .001	3.782	0.290	1.476

Note. Student's t-test.

Això mostra que hi ha una diferència significativa de massa corporal entre les condicions prèvies i les posteriors a la dieta, amb una diferència mitjana (paràmetre de localització) de 3,783 kg. La d de Cohen estableix que es tracta d'un efecte important.

El gràfic i l'estadística descriptiva mostren que hi va haver una reducció de massa corporal després de seguir la dieta durant 4 setmanes.

Descriptives

	N	Mean	SD	SE
Pre diet body mass	78	72.526	8.723	0.988
Post 4 weeks diet	78	68.744	9.009	1.020

REPORTANT ELS RESULTATS

Els participants van perdre, de mitjana, 3,78 kg (SE: 0,29 kg) de massa corporal seguint un programa de dieta de 4 setmanes. La prova t per a mostres aparellades va mostrar que aquesta disminució és significativa (**t(77) = 13,039, p < 0,001**). La d de Cohen suggereix que es tracta d'un efecte important.

EXECUTANT LA PROVA NO PARAMÈTRICA PER A MOSTRES APARELLADES

PROVA DE RANGS AMB SIGNE DE WILCOXON

Si s'observa que les dades no estan normalment distribuïdes (resultat significatiu del test Shapiro-Wilk) o si la distribució és ordinal, la prova no paramètrica equivalent és la prova de rangs amb signe de Wilcoxon. Obriu **Wilcoxon's rank.csv**. Aquest arxiu conté dues columnes: una amb les puntuacions d'ansietat abans del tractament i una altra amb les puntuacions després d'un tractament amb hipnoteràpia (de 0 a 50). En mostrar-se el conjunt de dades, assegureu-vos que ambdues variables estan assignades com a variables ordinals.

Aneu a «T-test» → «Paired samples t-test» i seguiu les instruccions explicades anteriorment, però aquesta vegada seleccioneu, únicament, les opcions següents:

- ✓ Rang amb signe de Wilcoxon (*Wilcoxon signed rank*).
- ✓ Paràmetre de localització (*Location parameter*).
- ✓ Mida de l'efecte (*Effect size*).

El resultat es mostrarà en una única taula:

Paired Samples T-Test

			W	p	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Pre-anxiety	-	Post-anxiety	322.000	< .001	8.000	0.480

Note. Wilcoxon signed-rank test.

L'estadístic W de Wilcoxon és altament significatiu, $p < 0{,}001$.

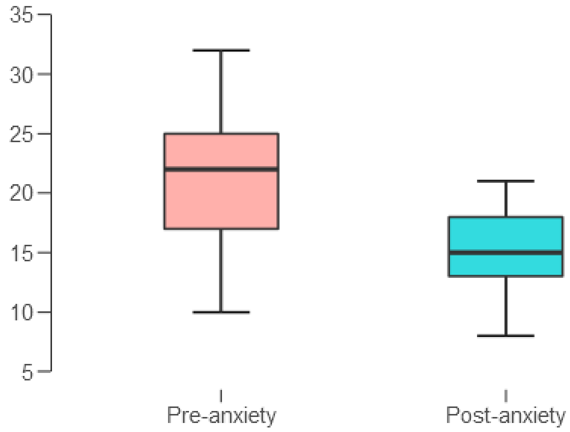
El paràmetre de localització, l'estimació Hodges-Lehmann, és la diferència mediana entre els dos grups. La correlació de rang biserial (*Rank-Biserial Correlation*, r_B) pot ser considerada com una mida de l'efecte i s'interpreta com la r de Pearson, de manera que 0,48 és una mida de l'efecte entre mitjana i gran.

Mida de l'efecte	Irrellevant	Petit	Mitjà	Gran
Rang biserial (r_B)	< 0,1	0,1	0,3	0,5

Per a dades no paramètriques, s'han de reportar els valors medians com a estadística descriptiva i utilitzar gràfics de caixa en comptes de gràfics de línia i intervals de confiança, barres SD / ES.

Descriptive Statistics

	Pre-anxiety	Post-anxiety
Valid	29	29
Missing	0	0
Median	22.0	15.0
Minimum	10.0	8.0
Maximum	32.0	21.0



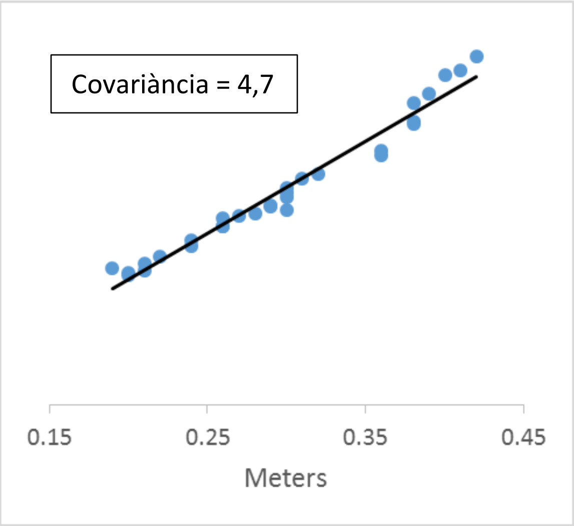
REPORTANT ELS RESULTATS

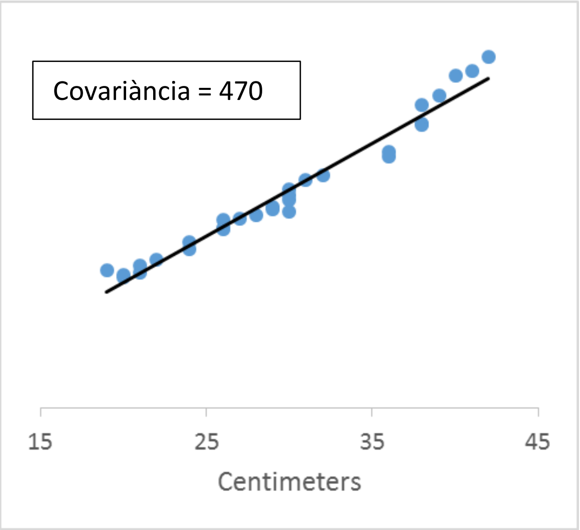
La prova de rangs amb signe de Wilcoxon va mostrar que la hipnoteràpia redueix significativament les puntuacions d'ansietat (Mdn = 15), en comparació amb les puntuacions d'ansietat anteriors al tractament (Mdn = 22), W = 322, p < 0,001.

ANÀLISI DE CORRELACIÓ

La correlació és una tècnica estadística que es pot utilitzar per determinar si hi ha parells de variables relacionats i amb quina força ho estan. La correlació només és apropiada per a dades quantificables que tinguin significat, com ara dades contínues o ordinals. No es pot utilitzar amb dades purament categòriques; per a aquestes, el més adient és l'anàlisi de taula de contingència (vegeu Anàlisi chi quadrat en JASP).

En essència, diferents variables covarien? És a dir, es donen canvis en una variable que tinguin el seu reflex en canvis similars en una altra variable? Si una variable es desvia de la seva mitjana, l'altra variable es desvia de la seva mitjana en la mateixa direcció o en l'oposada? Això es pot avaluar mesurant la covariància, encara que no es un mètode estandarditzat. Per exemple, es pot mesurar la covariància de dues variables mesurades en metres. No obstant això, si transformem els valors a centímetres obtenim la mateixa relació, encara que amb un valor de la covariància completament diferent.





Per superar aquesta situació, s'utilitza una covariància estandarditzada, coneguda com a **coeficient de correlació de Pearson** (*Pearson's correlation coefficient*, o r). Adopta un valor en l'interval entre -1,0 i +1,0. Com més a prop està r de +1 o -1, més estretament relacionades entre si estan les dues variables. Si r és proper a 0, no hi ha relació. Si r és (+), quan els valors d'una variable són més alts, els de l'altra també ho són. Si r és negatiu (-), quan els valors d'una variable són més alts, els de l'altra són més baixos (anomenada de vegades correlació "inversa").

No s'ha de confondre el coeficient de correlació (r) amb R^2 –coeficient de determinació (*coefficient of determination*)–, ni amb R –coeficient de correlació múltiple (*multiple correlation coefficient*), tal com s'utilitza en la regressió.

El supòsit principal en aquesta anàlisi és que les dades tenen una distribució normal i són lineals. Aquesta anàlisi no funcionarà bé amb relacions curvilínies.

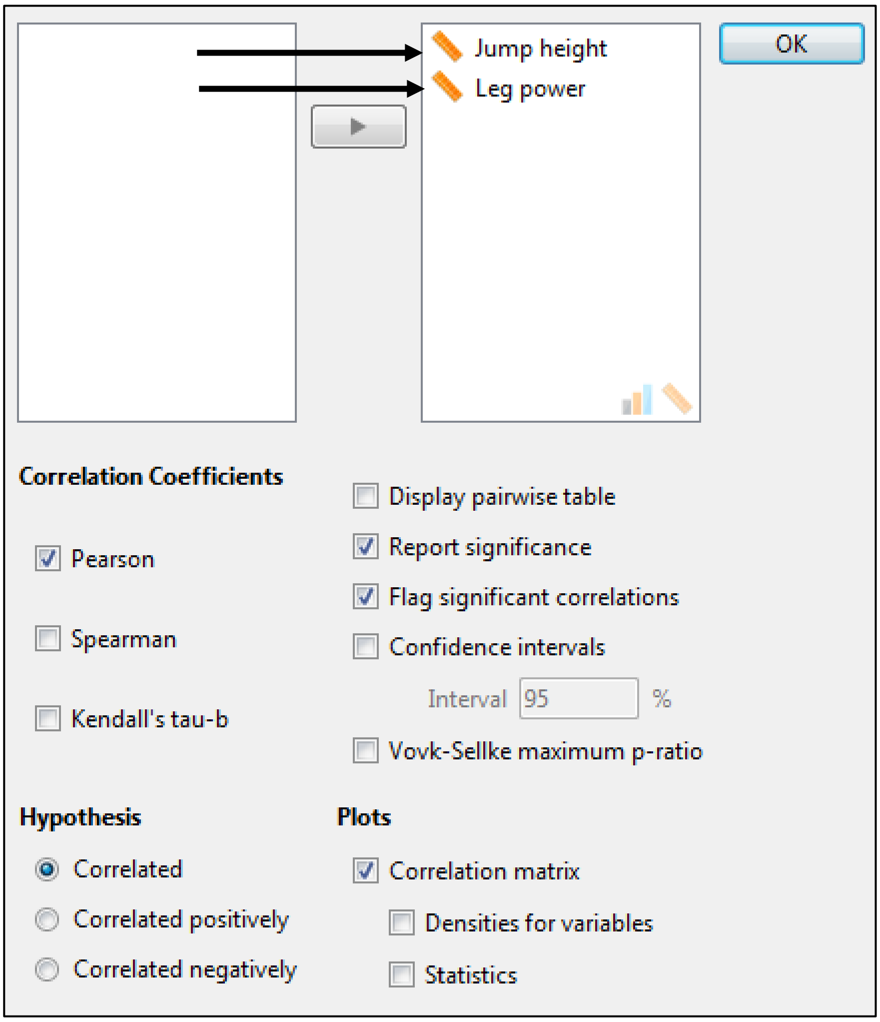
EXECUTANT LA CORRELACIÓ

L'anàlisi posa a prova la hipòtesi nul·la (H_0) que no hi ha relació entre dues variables.

De les dades d'exemple, obriu **Jump height correlation.csv.** Aquest arxiu conté 2 columnes de dades, Jump height (m) i Leg power (W). En primer lloc, aneu a «Descriptive statistics» i comproveu els gràfics de caixa per si hi hagués valors atípics.

Per a executar l'anàlisi de correlació, aneu a «Regression» → «Correlation matrix». Traslladeu les 2 variables a la caixa d'anàlisi de la dreta. Marqueu:

- ✓ Pearson.
- ✓ Reportar significació («Report significance»).
- ✓ Marcar correlacions significatives («Flag significant correlations»).
- ✓ Matriu de correlació («Correlation matrix») (a «Plots»).



ENTENENT EL RESULTAT

La primera taula mostra la matriu de correlació amb els valors de la r de Pearson i les seves p. S'observa una correlació altament significativa ($p < 0,001$), amb un valor de r proper a 1 ($r = 0,984$), que ens permet rebutjar la hipòtesi nul·la.

Pearson Correlations

		Jump height	Leg power
Jump height	Pearson's r	—	
	p-value	—	
Leg power	Pearson's r	0.984***	—
	p-value	< .001	—

* p < .05, ** p < .01, *** p < .001

Per a correlacions simples com aquesta, resulta més senzill observar la taula de valors per parelles. Retrocediu fins a l'anàlisi i seleccioneu l'opció «Display pairwise table». Això substitueix la matriu de correlació en els resultats i pot facilitar la seva lectura.

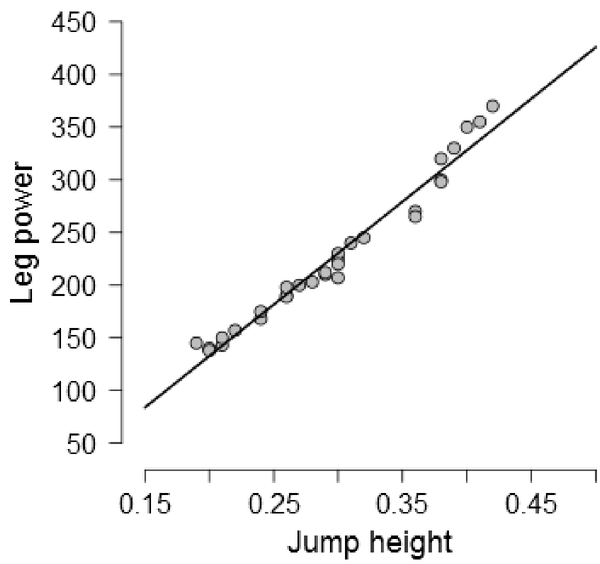
Pearson Correlations

			Pearson's r	p
Jump height	-	Leg power	0.984***	< .001

* p < .05, ** p < .01, *** p < .001

En realitat, el valor r de Pearson mostra una mida de l'efecte on < 0,1 és irrellevant, de 0,1 a 0,3 és un efecte petit, de 0,3 a 0,5 és un efecte moderat i > 0,5 és un efecte gran.

El gràfic permet visualitzar d'una manera simple aquesta forta correlació positiva ($r = 0,984$, $p < 0,001$).



ANANT UN PAS MÉS ENLLÀ

Si es pren el coeficient de correlació r i s'eleva al quadrat, s'obté el coeficient de determinació (R^2). És un procés de mesura estadística de la proporció de la variància d'una variable que s'explica per l'altra variable. O:

R^2 = Variància explicada / Variància total.

R^2 produeix sempre un valor entre 0 i 100% en el qual:

- un 0% indica que el model no explica res sobre la variabilitat de les dades al voltant de la seva mitjana, i
- un 100% indica que el model explica tota la variabilitat de les dades al voltant de la seva mitjana.

En l'exemple anterior, r = 0,984, de manera que R^2 = 0,968. Això suggereix que l'altura de salt representa un 96,8% de la variància en la potència de cama.

REPORTANT ELS RESULTATS

La correlació de Pearson va mostrar una correlació significativa entre l'altura de salt i la potència de cama (r = 0,984, p < 0,001), representant l'altura de salt un 96,8% de la variància en la potència de cama.

EXECUTANT LA CORRELACIÓ NO PARAMÈTRICA: LA TAU DE KENDALL I LA RHO DE SPEARMAN

Si les dades són ordinals o si són dades contínues que han violat els supòsits requerits per a l'ús de l'estadística paramètrica (normalitat i/o variància), haurieu d'utilitzar alternatives no paramètriques al coeficient de correlació de Pearson.

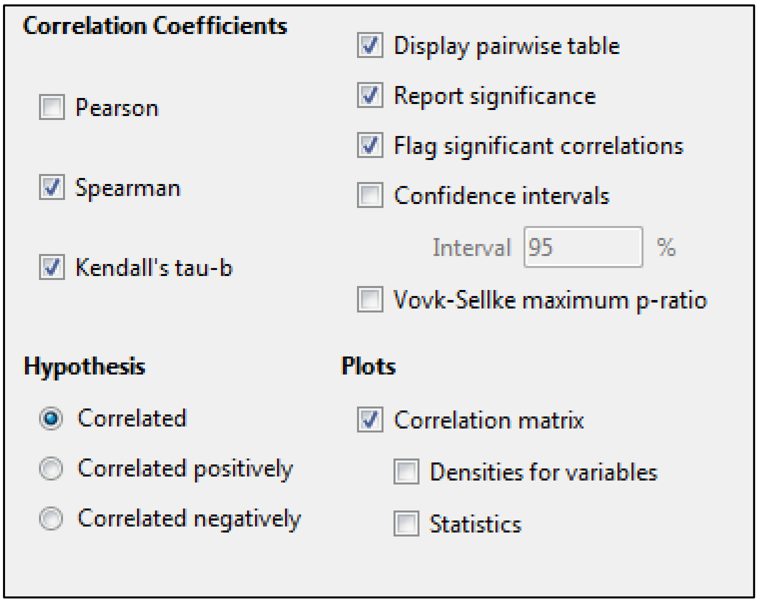
Les alternatives són els coeficients de correlació de Spearman (rho) o Kendall (tau). Ambdós estan basats en dades de classificació (ordenats de major a menor), i no estan afectats per la presència de valors atípics o violacions de la variància / normalitat.

La rho de Spearman s'utilitza habitualment amb dades d'escala ordinal i la tau de Kendall s'utilitza en mostres petites o quan hi ha molts valors amb la mateixa puntuació (empats). En la majoria dels casos, la tau de Kendall i el coeficient de correlació de Spearman són molt similars i, per tant, condueixen invariablement a les mateixes inferències.

Les mides de l'efecte són les mateixes que la r de Pearson. La principal diferència és que es pot fer servir rho^2 com una aproximació no paramètrica al coeficient de determinació, cosa que no succeeix en el cas de la tau de Kendall.

De les dades d'exemple, obriu **Non-parametric correlation.csv**. Aquest arxiu conté 2 columnes de dades: una amb puntuacions de creativitat i una altra amb les posicions en la competició d'"El mentider més gran del món" (*World's biggest liar*, gràcies a Andy Field).

Executeu l'anàlisi com en el cas anterior, però aquesta vegada utilitzant els coeficients de Sperman i tau-b de Kendall en comptes del de Pearson.

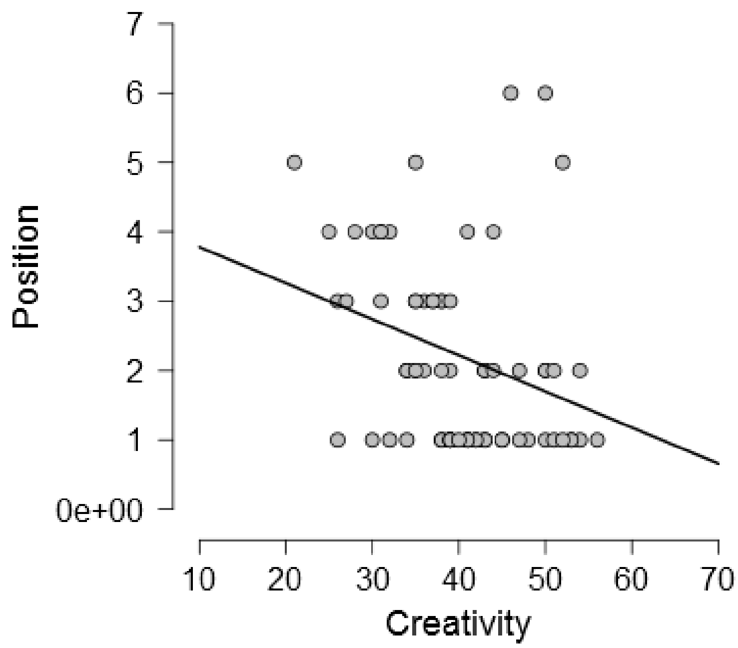


Correlation Table

			Spearman		Kendall	
			rho	p	tau B	p
Creativity	-	Position	−0.373**	0.002	−0.300**	0.001

* p < .05, ** p < .01, *** p < .001

Com es pot veure, hi ha una correlació significativa entre les puntuacions de creativitat i la posició final a la competició *World's biggest liar*: com més gran és la puntuació, millor és la posició final en la competició. No obstant això, la mida d'efecte és moderada.



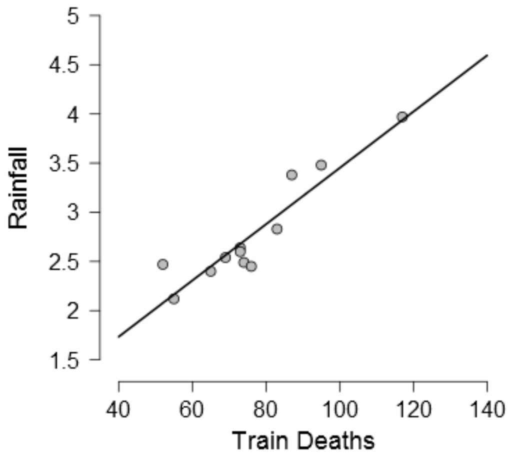
NOTA D’ADVERTÈNCIA

En realitat, la correlació només ofereix informació sobre la fortalesa de l’associació. No informa sobre la direcció, és a dir, sobre quina variable fa que l’altra canviï. Per això, no pot ser utilitzada per afirmar que una cosa és causa d’una altra. Sovint, una correlació significativa no vol dir absolutament res i és purament casual, especialment si es correlacionen milers de variables. Això es pot veure en correlacions estranyes com les següents:

El número de vianants morts en un atropellament de tren correlaciona amb la pluja a Missouri

Pearson Correlations

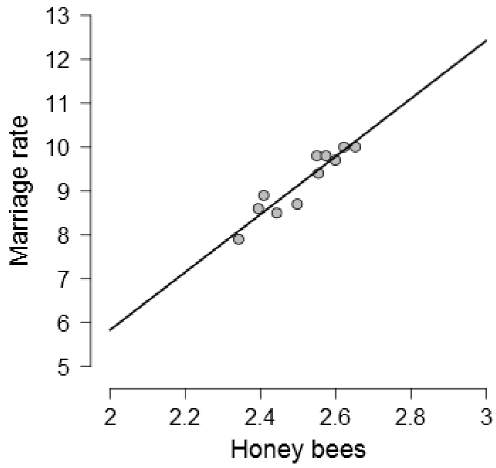
			Pearson's r	p
Train Deaths	-	Rainfall	0.928	< .001



El nombre de colònies d’abelles productores de mel (per 1.000) correlaciona fortament amb la taxa de matrimonis a Carolina del Sud (per 1.000 matrimonis)

Pearson Correlations

			Pearson's r	p
Honey bees	-	Marriage rate	0.938	< .001



REGRESSIÓ

Mentre que les proves de correlació es fan servir per a les associacions entre variables, la regressió és el pas següent utilitzat habitualment per a les anàlisis predictives, és a dir, per predir una variable de resultat dependent a partir d'una (regressió simple) o més (regressió múltiple) variables predictives independents.

La regressió resulta en un model hipotètic de relació entre la variable resultat i una o més variables predictives. El model utilitzat és lineal, definit per la fórmula:

$$y = c + b*x + \varepsilon$$

- y = puntuació de la variable de resultat dependent estimada
- c = constant
- b = coeficient de regressió
- x = puntuació de la variable independent predictiva
- ε = component d'error aleatori (basat en els residus)

La regressió lineal proporciona tant la constant com el o els coeficients de regressió.

La regressió lineal assumeix els següents supòsits:

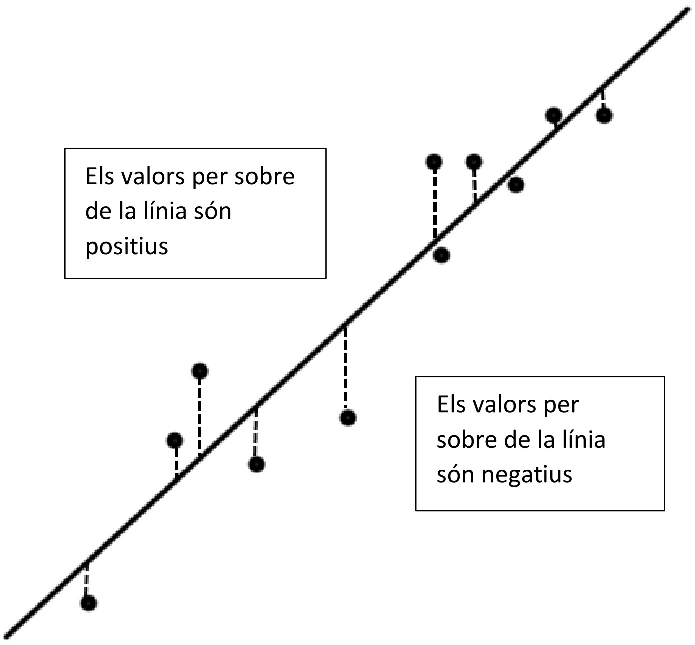
1. **Relació lineal**: és important revisar els valors atípics, ja que la regressió lineal és sensible als seus efectes.
2. **Independència** de les variables.
3. **Normalitat multivariant**: requereix que totes les variables estiguin distribuïdes normalment.
4. **Homoscedasticitat**: homogeneïtat de la variància dels residus.
5. **Multicol·linealitat / autocorrelació mínima**: quan les variables independents / els residus estan molt correlacionats entre si.

Respecte a les mides de les mostres, hi ha molta literatura sobre diferents regles generals que van des dels 10-15 punts de dades per predictor inclòs en el model (és a dir, 4 variables predictives requeriran entre 40 i 60 punts de dades) a 50 punts + (8*nombre de predictors). Així, 4 variables requeririen 82 punts de dades (50 + 8 * 4 = 50 + 32 = 82). En qualsevol cas, com més gran sigui la mida de la mostra, millor serà el model.

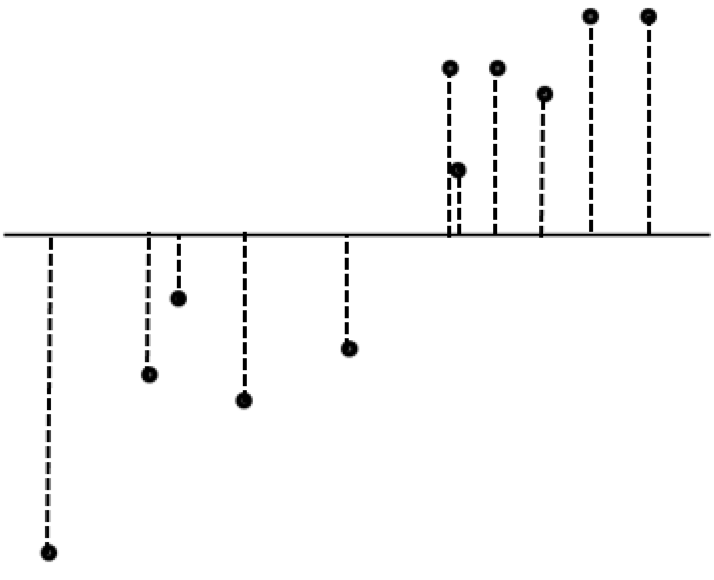
SUMES DE QUADRATS (Avorrit, però bàsic per a l'avaluació del model de regressió)

La majoria de les anàlisis de regressió produiran el millor model possible, però aquest model, com de bo és en realitat i quant d'error es comet amb ell?

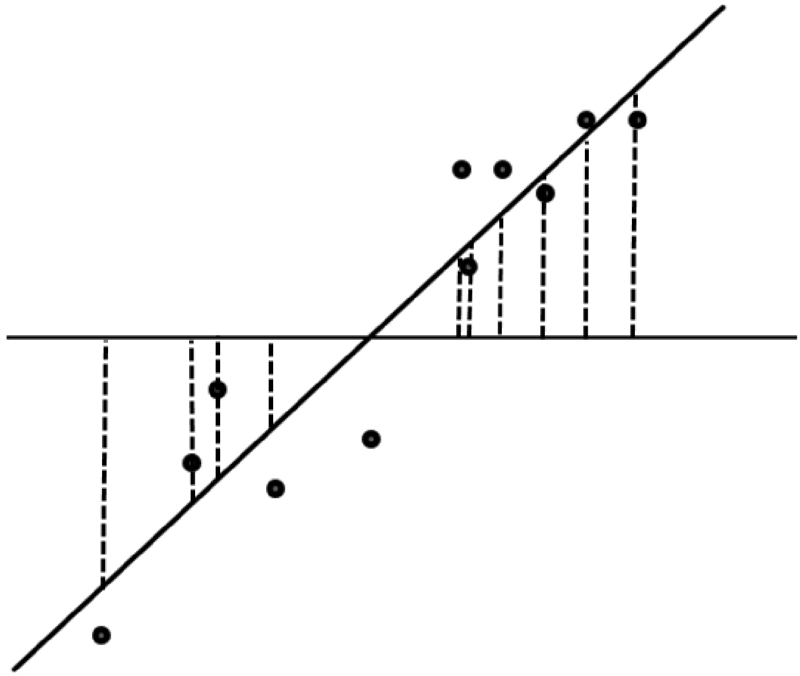
Això es pot determinar comprovant la "bondat d'ajustament" basada en les sumes de quadrats. Es tracta d'una mesura per determinar com de prop estan els punts de dades reals de la línia de regressió modelada.



La diferència vertical entre els punts de dades i la línia de regressió predita es coneixen amb el nom de **residus**. Aquests valors s'eleven al quadrat per a eliminar els números negatius i després se sumen per obtenir **SS_R** (SQ_R, suma de quadrats dels residus, en català). Aquest és, efectivament, l'error del model o "**bondat d'ajustament**"; de manera que com més petit sigui el valor, menor serà l'error en el model.



Es pot calcular la diferència vertical entre els punts de dades i la mitjana de la variable resultat. Aquests valors s'eleven al quadrat per a eliminar els números negatius i després se sumen per obtenir la suma **total** de quadrats **SS_T** (SQ_T, suma de quadrats total en català). Això mostra com de bo és el valor mitjà com a model de les puntuacions de la variable resultat.



Ara, podem determinar la diferència vertical entre la mitjana de la variable resultat i la línia de regressió predita. De nou, aquests valors s'eleven al quadrat per a eliminar els números negatius i després se sumen per a obtenir la suma de quadrats del model **SS_M** (SQ_M, suma de quadrats del model en català). Això indica com de bo és el model comparat amb l'ús únicament de la mitjana de la variable resultat.

Per tant, com més gran sigui **SQ_M** millor serà el model per predir el resultat comparat amb el valor mitjà per si sol. Si ve acompanyat d'un petit **SQ_R** el model també tindrà un error petit.

R^2 és similar al coeficient de determinació en la correlació, en tant que mostra fins a quin punt la variació en la variable resultat pot ser predita per la(es) variable(s) predictiva(es).

$$R^2 = \frac{SQ_M}{SQ_R}$$

En la regressió, el model s'avalua mitjançant l'estadístic F que es basa en la millora de la predicció del model (**SQ_M**) i l'error (**SQ_R**). Com més gran sigui el valor de F, millor serà el model.

$$F = \frac{\text{Mitjana } SQ_M}{\text{Mitjana } SQ_R}$$

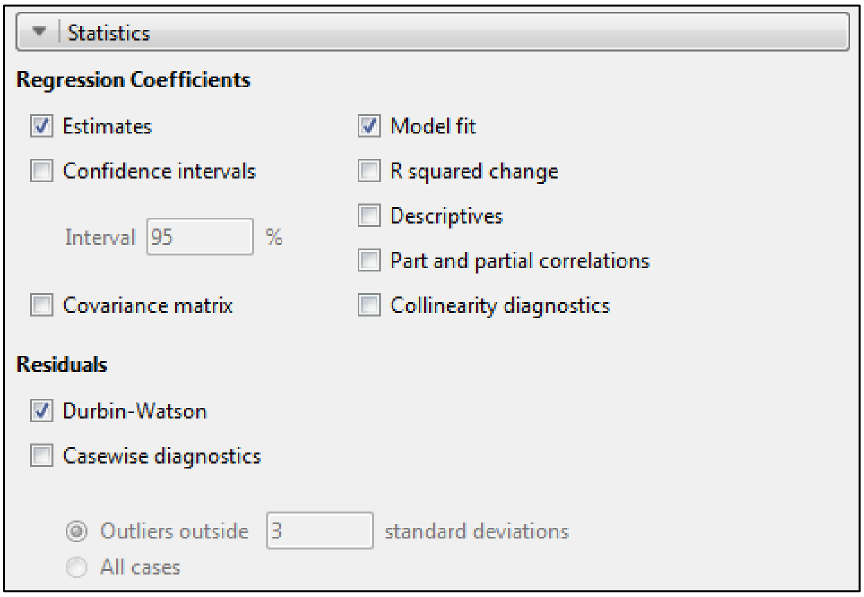
REGRESSIÓ SIMPLE

La regressió posa a prova la hipòtesi nul·la (H_0) que la(es) variable(s) predictiva(es) no podrà(an) predir significativament la variable dependent (resultat).

Obriu **Rugby kick regression.csv**. Aquest arxiu conté dades sobre xuts en el rugbi, incloent la distància recorreguda, la força i la flexibilitat de la cama dreta / esquerra, i la força de cama bilateral.

Primer, aneu a «Descriptives» → «Descriptive statistics» i comproveu els gràfics de caixa per si hi hagués valors atípics. En aquest cas no n'hi hauria d'haver cap, però la comprovació és una bona pràctica.

Per a aquesta regressió simple, aneu a «Regression» → «Linear regression» i introduïu la distància a «Dependent variable» (outcome), i R_Strength a la caixa «Covariates» (Predictor). Marqueu les següents opcions a «Statistics»:



ENTENENT EL RESULTAT

Ara obtindreu el resultats següents:

Model Summary

Model	R	R²	Adjusted R²	RMSE	Durbin-Watson
1	0.784	0.614	0.579	55.285	1.524

Aquí es pot veure que la correlació (R) entre les dues variables és alta (0,784). El valor R^2 de 0,614 ens diu que la força de la cama dreta representa el 61,4% de la variància en la distància de xut. Durbin-Watson comprova les correlacions entre els residus, cosa que podria invalidar el test. Hauria d'estar per sobre d'1 i per sota de 3, idealment a prop de 2.

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	53589.863	1	53589.863	17.533	0.002
	Residual	33621.061	11	3056.460		
	Total	87210.923	12			

La taula ANOVA mostra totes les sumes dels quadrats abans esmentats. “Regression” és el model i “Residual” l’error. L’estadístic F és significatiu p = 0,002. Això ens diu que el model és un predictor de la distància de xut significativament millor que la distància mitjana.
Informeu de la següent manera: **F (1, 11) = 17,53, p < 0,001.**

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	(Intercept)	57.105	103.588		0.551	0.592
	R_Strength	6.425	1.534	0.784	4.187	0.002

Aquesta taula proporciona els coeficients no estandarditzats (“Unstandardized”) que poden introduir-se en l’equació lineal.

y = c + b*x

y = puntuació estimada de la variable dependent resultat.

c = constant (“**(Intercept)**”).

b = coeficient de regressió (“**R_Strength**”).

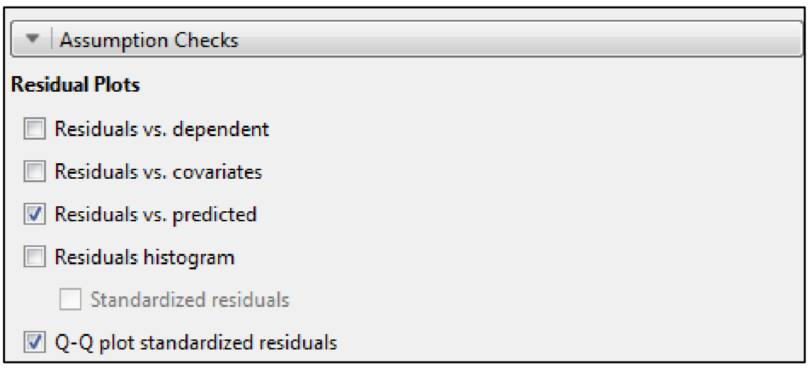
x = puntuació en la variable predictiva independent.

Per exemple, per a una força de cama de 60 kg, la distància de xut es pot predir amb la fórmula següent:

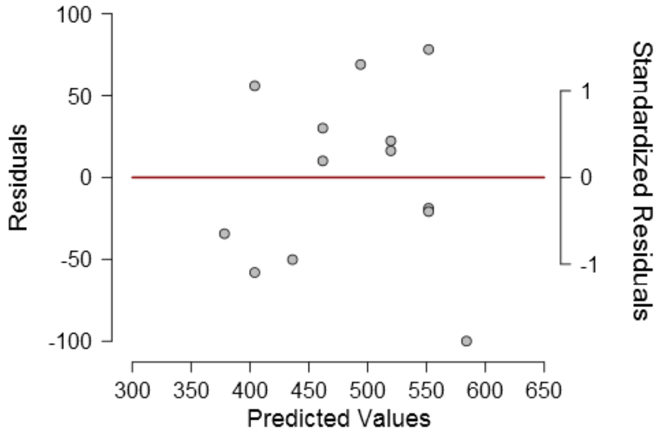
Distància = 57,105 + (6,452 * 60) = 454,6 m

COMPROVACIONS ADDICIONALS

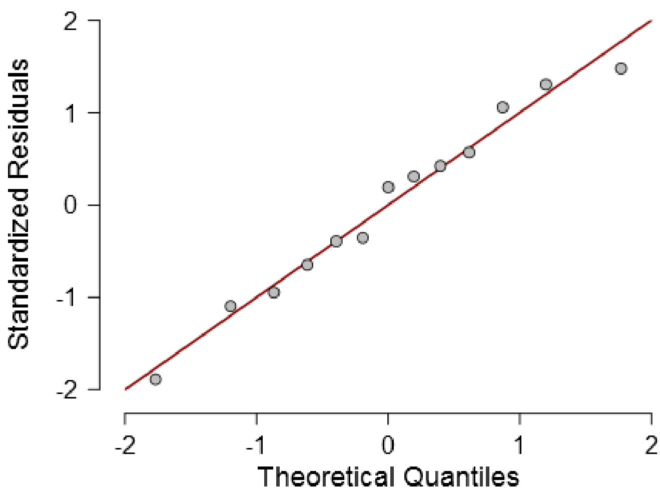
A «Assumption Checks», seleccioneu les dues opcions següents:



A partir d'això s'obtindran dos gràfics:



Aquest gràfic mostra una distribució aleatòria equilibrada dels residus al voltant de la línia de base, suggerint que el supòsit d'homoscedasticitat no ha estat violat. Vegeu "Exploració de la integritat de les dades en JASP" per a més detalls.



El gràfic Q-Q mostra que els residus estandarditzats coincideixen amb la diagonal, suggerint que els dos supòsits de normalitat i la linealitat no han estat violats.

REPORTANT ELS RESULTATS

La regressió lineal mostra que la força de la cama dreta pot predir significativament la distància de xut $F(1,11) = 17,53$, $p < 0,001$ emprant l'equació de regressió següent:

Distància = 57,105 + (6,452 * força de la cama dreta)

REGRESSIÓ MÚLTIPLE

El model utilitzat segueix sent lineal, definit per la fórmula:

$$y = c + b*x + \varepsilon$$

- y = puntuació estimada de la variable dependent resultat.
- c = constant.
- b = coeficient de regressió.
- x = puntuació en la variable predictiva independent.
- ε = component d'error aleatori (basat en els residus).

No obstant això, ara tenim més d'1 coeficient de regressió per a la puntuació en cada variable predictiva. És a dir:

$$y = c + b_1*x_1 + b_2*x_2 + b_3*x_3 \ldots b_n*x_n$$

Mètodes d'entrada de dades

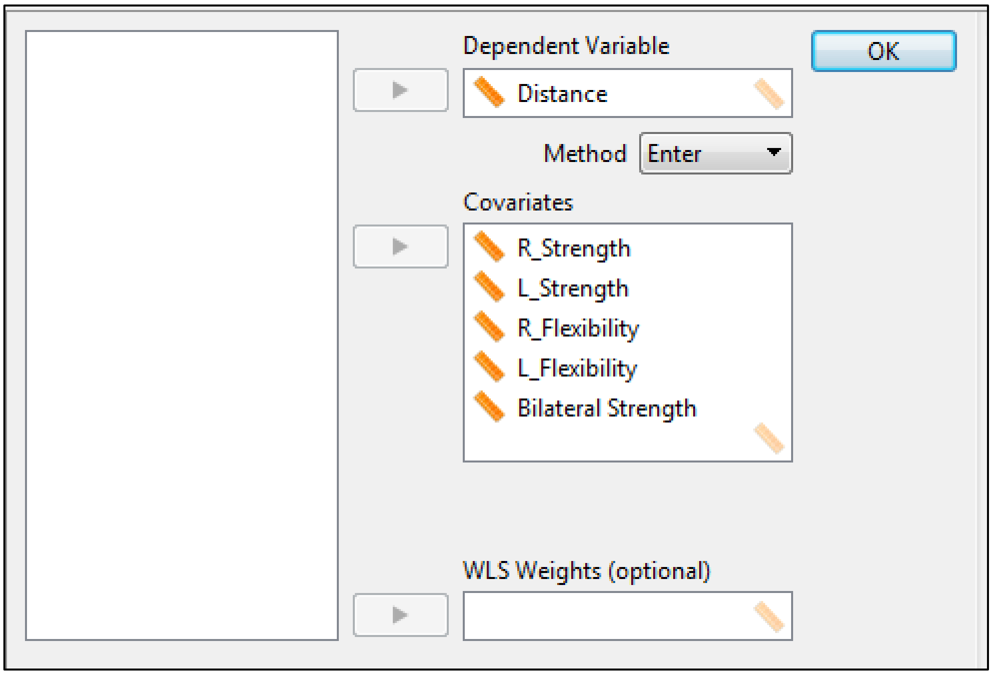
Si les variables predictives no estan correlacionades, el seu ordre d'entrada no té importància per al model. En la majoria dels casos, les variables predictives estan d'alguna manera correlacionades i, per això, l'ordre en què s'introdueixin pot tenir conseqüències. Els diferents mètodes disponibles han estat objecte d'un gran debat.

a) Entrada **forçada** («**Enter**»): aquest és el **mètode per defecte** en el qual es força l'entrada de les variables predictores en l'ordre en què apareixen a la caixa de covariables («Covariates»). Es considera el millor mètode.
b) Entrada **per blocs** (*hierarchical entry*): l'investigador, normalment basat en coneixements i estudis previs, decideix en primer lloc l'ordre en què s'introdueixen les variables predictores, en funció de la seva importància en la predicció de la variable resultat. En passos posteriors s'introdueixen predictores addicionals.
c) Entrada **per passos cap enrere** («**Backward**»): totes les variables predictores s'introdueixen inicialment en el model i es calcula la contribució de cadascuna d'elles. S'eliminen les predictores amb un nivell de contribució inferior al nivell establert ($p < 0{,}1$). Es repeteix el procés fins que totes les variables predictores que es conserven en el model són estadísticament significatives.
d) Entrada **per passos cap endavant** («**Forward**»): s'introdueix, en primer lloc, la variable predictora amb la correlació simple més alta respecte a la variable resultat. Les predictores subsegüents es trien en funció de la mida de la seva correlació semiparcial respecte a la variable resultat. Aquest procés es repeteix fins que han quedat incloses totes les predictores que contribueixen amb una variació única significativa al model.
e) Entrada **per passos** («**Stepwise**»): similar al mètode d'entrada cap endavant («Forward»), excepte que cada vegada que s'afegeix una variable predictora al model, es realitza un test per eliminar la predictora menys útil. El model es revisa constantment per comprovar si les predictores redundants poden ser eliminades.

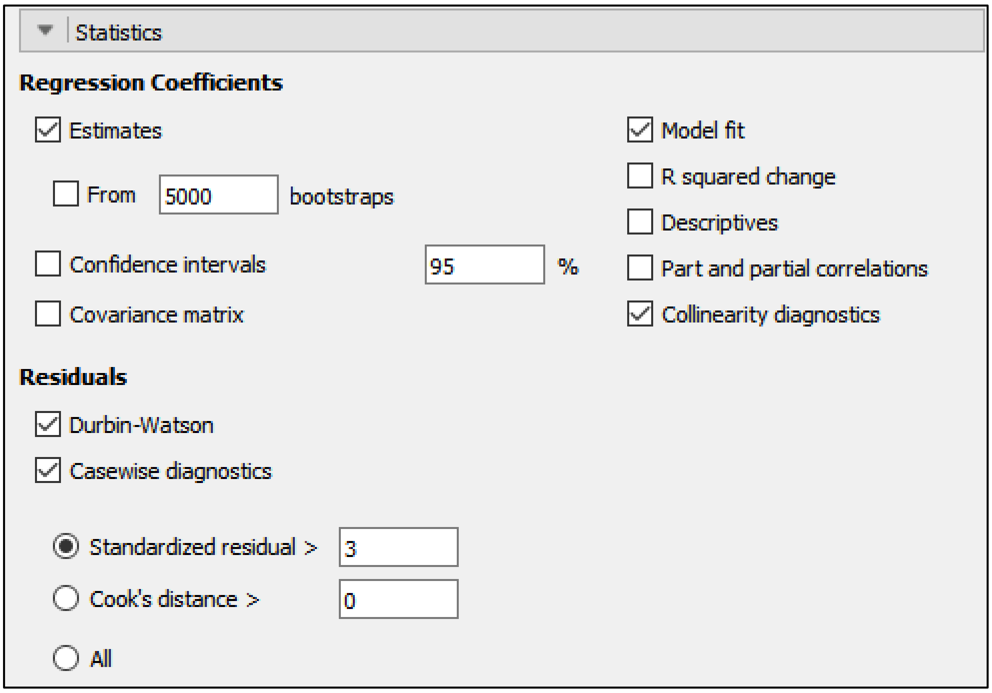
S'han descrit molts inconvenients relacionats amb l'ús de mètodes d'entrada per passes. No obstant això, el mètode **cap enrere** pot ser útil per a explorar variables predictores no utilitzades prèviament o per afinar el model per tal de seleccionar les millors entre les disponibles.

EXECUTANT LA REGRESSIÓ MÚLTIPLE

Obriu **Rugby kick regression.csv**, que també hem utilitzat per a la regressió simple. Aneu a «Regression» → «Linear regression», introduïu la distància a la caixa «Dependent Variable» (resultat) i la resta de variables a la caixa «Covariates» (predictores).



A «Method» deixeu el mètode d'entrada forçada («**Enter**») que apareix per defecte. Marqueu les següents opcions a «Statistics options»: «Estimates», «Model fit», «Collinearity diagnostics» i marqueu «Durbin-Watson».



ENTENENT EL RESULTAT

Ara, obtindreu el següents resultats:

Model Summary

Model	R	R²	Adjusted R²	RMSE	Durbin-Watson
1	0.902	0.814	0.681	48.132	1.328

L'R^2 ajustat (usat per a múltiples predictores) mostra que es pot predir un 68,1% de la variància de la variable resultat. Durbin-Watson comprova que les correlacions entre els residus es troben entre 1 i 3, com es requereix.

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	70994.078	5	14198.816	6.129	0.017
	Residual	16216.845	7	2316.692		
	Total	87210.923	12			

La taula ANOVA mostra que l'estadístic F és significatiu p = 0,017, suggerint que el model prediu significativament millor la distància de xut que la distància mitjana.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
1	(Intercept)	−92.367	218.389		−0.423	0.685		
	R_Strength	1.747	3.321	0.213	0.526	0.615	0.162	6.180
	L_Strength	0.703	3.590	0.086	0.196	0.850	0.138	7.231
	R_Flexibility	4.078	4.759	0.373	0.857	0.420	0.140	7.125
	L_Flexibility	−1.339	2.447	−0.135	−0.547	0.601	0.438	2.281
	Bilateral Strength	1.665	0.946	0.423	1.759	0.122	0.458	2.181

Aquesta taula mostra un model i la constant ("(Intercept)"), i els coeficients de regressió ("Unstandardized") per a tots els predictors forçats en el model. Tot i que la taula d'ANOVA mostri que el model és significatiu, cap dels coeficients de regressió predictius ho és!

Els estadístics de colinealitat, tolerància i VIF (sigles en anglès de *Variance Inflation Factor*, o factor d'inflació de la variància) comproven el supòsit de multicolinealitat. Com a regla general, si el VIF > 10 i la tolerància < 0,1, el supòsit ha estat àmpliament violat. Si la **mitjana** dels valors del VIF > 1 i la tolerància < 0,2, el model podria estar esbiaixat. En aquest cas, la mitjana del VIF és bastant gran (al voltant de 5).

La taula de diagnòstic per casos ("Casewise Diagnostics") és buida! Són bones notícies. Aquesta taula mostra els casos (files) amb residus que es trobin a 3 o més desviacions estàndard respecte a la mitjana. Aquests casos amb els errors més grans podrien ser valors atípics. La presència de massa valors atípics tindrà un impacte sobre el model i haurien de tractar-se de la manera habitual (vegeu "Exploració de la integritat de les dades").

Casewise Diagnostics

Case Number	Std. Residual	Distance	Predicted Value	Residual	Cook's Distance
-	-	-	-	-	-

Com a comparació, torneu a executar les anàlisis, però aquest cop escollint «Backward» (cap enrere) com a mètode d'entrada.

Els resultats són els següents:

Model Summary

Model	R	R²	Adjusted R²	RMSE	Durbin-Watson
1	0.902	0.814	0.681	48.132	
2	0.902	0.813	0.720	45.146	
3	0.897	0.805	0.740	43.505	
4	0.884	0.782	0.738	43.618	1.676

JASP ara ha calculat 4 models de regressió potencials. Es pot veure que cada model consecutiu incrementa l'R^2 ajustat, on el model 4 explica el 73,5% de la variable resultat. La puntuació Durbin-Watson també és més alta que amb el mètode d'entrada forçada (*Enter*).

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	70994.078	5	14198.816	6.129	0.017
	Residual	16216.845	7	2316.692		
	Total	87210.923	12			
2	Regression	70905.329	4	17726.332	8.697	0.005
	Residual	16305.594	8	2038.199		
	Total	87210.923	12			
3	Regression	70176.855	3	23392.285	12.359	0.002
	Residual	17034.068	9	1892.674		
	Total	87210.923	12			
4	Regression	68185.712	2	34092.856	17.920	< .001
	Residual	19025.211	10	1902.521		
	Total	87210.923	12			

La taula ANOVA indica que cada model successiu és millor, tal com mostra l'augment del valor de la F i la millora del valor de p.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	−92.367	218.389		−0.423	0.685		
	R_Strength	1.747	3.321	0.213	0.526	0.615	0.162	6.180
	L_Strength	0.703	3.590	0.086	0.196	0.850	0.138	7.231
	R_Flexibility	4.078	4.759	0.373	0.857	0.420	0.140	7.125
	L_Flexibility	−1.339	2.447	−0.135	−0.547	0.601	0.438	2.281
	Bilateral Strength	1.665	0.946	0.423	1.759	0.122	0.458	2.181
2	(Intercept)	−110.347	185.840		−0.594	0.569		
	R_Strength	2.218	2.148	0.271	1.033	0.332	0.340	2.938
	R_Flexibility	4.501	3.978	0.411	1.131	0.291	0.177	5.658
	L_Flexibility	−1.370	2.291	−0.138	−0.598	0.566	0.440	2.272
	Bilateral Strength	1.605	0.840	0.408	1.910	0.092	0.512	1.954
3	(Intercept)	−116.892	178.772		−0.654	0.530		
	R_Strength	2.710	1.911	0.331	1.418	0.190	0.399	2.505
	R_Flexibility	2.886	2.814	0.264	1.026	0.332	0.328	3.048
	Bilateral Strength	1.642	0.807	0.418	2.033	0.073	0.515	1.944
4	(Intercept)	46.251	81.820		0.565	0.584		
	R_Strength	3.914	1.512	0.478	2.588	0.027	0.641	1.561
	Bilateral Strength	2.009	0.725	0.511	2.770	0.020	0.641	1.561

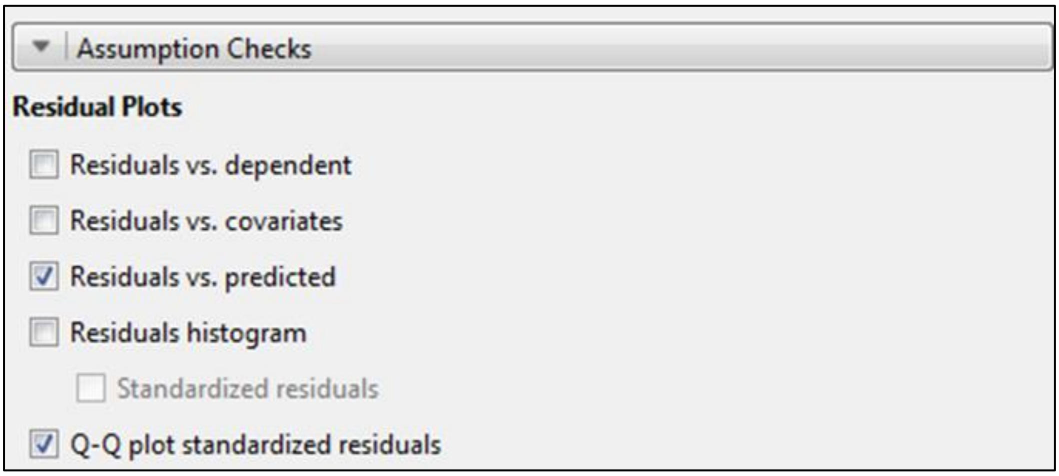
El model 1 és el mateix que el mètode d'entrada forçada (*Enter*) utilitzat en primer lloc. La taula mostra que a mesura que s'eliminen de manera seqüencial les predictores amb una contribució significativament menor, acabem obtenint un model amb dos coeficients de regressió predictius significatius: la força de la cama dreta (R_Strength) i la força de cama bilateral (Bilateral Strength). Tant la tolerància com el VIF són acceptables.

Ara podem reportar que l'entrada de les variables predictores per passos cap enrere (*Barckward*) resulta en un model altament significatiu: **F (2, 10) = 17,92, p < 0,001**, i una equació de regressió com la que segueix:

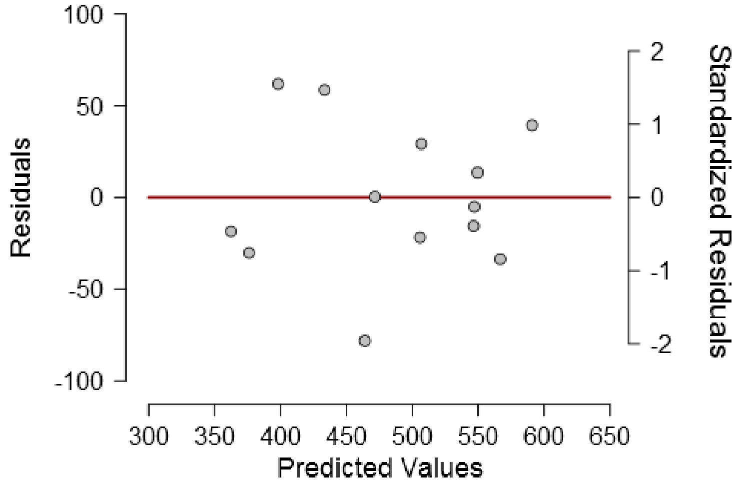
Distància = 46,251 + (3,914 * R_Strength) + (2,009 * Bilateral Strength)

COMPROVACIÓ DE SUPÒSITS ADDICIONALS

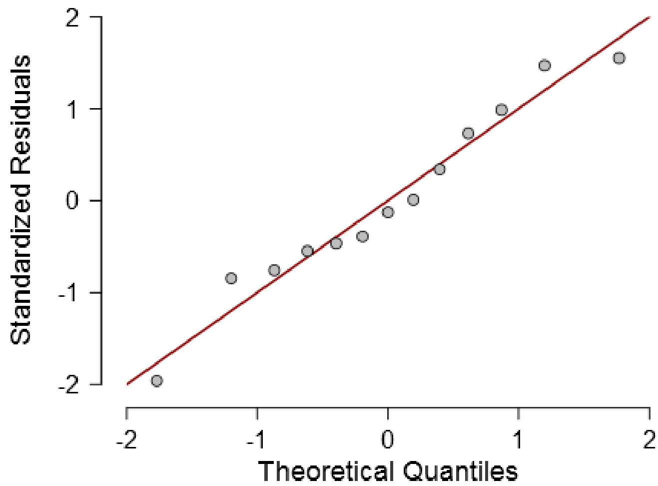
Com en l'exemple de regressió lineal simple, seleccioneu les opcions següents.



Residuals vs. Predicted



Q-Q Plot Standardized Residuals ▼



La distribució equilibrada dels residus al voltant de la línia base suggereix que el supòsit d'homoscedasticitat no ha estat violat.

El gràfic Q-Q mostra que els residus estandarditzats s'ajusten al llarg de la diagonal, la qual cosa suggereix que ambdós supòsits de normalitat i linealitat tampoc no han estat violats.

REPORTANT ELS RESULTATS

La regressió lineal múltiple basada en el mètode d'entrada per passos cap enrere mostra que la força de cama dreta i la força bilateral poden predir significativament la distància de xut **F (2,10) = 17,92, p < 0,001** utilitzant l'equació de regressió:

Distància = 57,105 + (3,914 * R_Strength) + (2,009 * Bilateral Strength)

EN RESUM

R^2 proporciona informació sobre quanta variància pot ser explicada utilitzant les variables predictores introduïes en el model.

L'estadístic F proporciona informació sobre com de bo és el model.

El valor dels coeficients no estandarditzats proporciona una constant que reflecteix la força de la relació entre cadascuna de les variables predictores i la variable resultat.

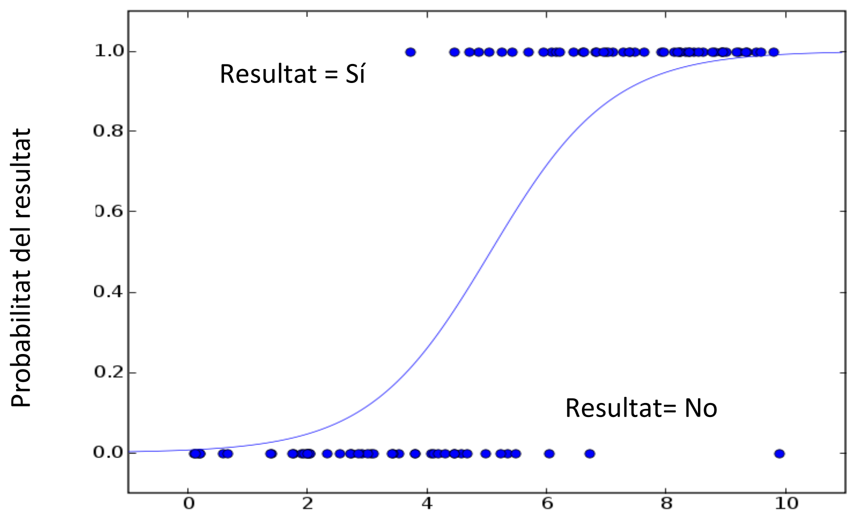
La violació dels supòsits pot ser comprovada utilitzant el valor de Durbin-Watson, els valors de tolerància / VIF i els gràfics de residus *vs*. predits i Q-Q.

REGRESSIÓ LOGÍSTICA

En la regressió lineal simple i múltiple, la variable resultat i les variables predictores eren contínues. Però què passaria si la variable resultat fos una mesura binària / categòrica? Podeu, per exemple, predir una variable resultat de sí o no, a partir d'altres variables contínues o categòriques? La resposta és que sí, si s'utilitza una regressió logística binària. Aquest mètode es fa servir per predir la probabilitat d'una variable resultat binària de sí o no.

La hipòtesi nul·la que es posa a prova és que no hi ha relació entre les variables resultat i les predictores.

Com es pot veure en el gràfic següent, una línia de regressió lineal entre les respostes de sí i no tindria poc sentit com a model predictiu. En el seu lloc, s'ajusta una corba de regressió logística sigmoide amb un mínim en 0 i un màxim en 1. Es pot veure que alguns valors de la variable predictora se superposen entre el sí i el no. Per exemple, un valor de 5 tindria una probabilitat del 50% de resultar en un sí o un no. Per tant, es calcula un llindar per determinar si el valor en una variable predictora es classificarà com un sí o com un no en la variable resultat.



SUPÒSITS DE LA REGRESSIÓ LOGÍSTICA BINÀRIA

- La variable dependent ha de ser binària, és a dir, sí o no, home o dona, bo o dolent.
- Una o més (variables predictives) independents que poden ser variables categòriques o contínues.
- Una relació lineal entre les variables independents contínues i la transformació *logit* (logaritme natural de la probabilitat que la variable resultat equivalgui a una de les categories) de la variable dependent.

MÈTRIQUES DE LA REGRESSIÓ LOGÍSTICA

AIC (per les sigles en anglès d'*Akaike Information Criteria,* o Criteri d'Informació Akaike) i **BIC** (per *Bayesian Information Criteria*, o Criteri d'Informació Bayesià) són mesures d'ajustament per al model; el millor model té els valors AIC i BIC més baixos.

En JASP es calculen tres valors **pseudo R²**: McFadden, Nagelkerke i Tjur. Aquests són anàlegs a l'R^2 en la regressió lineal i tots proporcionen valors diferents. El que constitueix un bon valor de R^2 varia, però són útils quan es comparen diferents models amb les mateixes dades. Es considera que el millor model és el que posseeixi un valor en l'estadístic R^2 més alt.

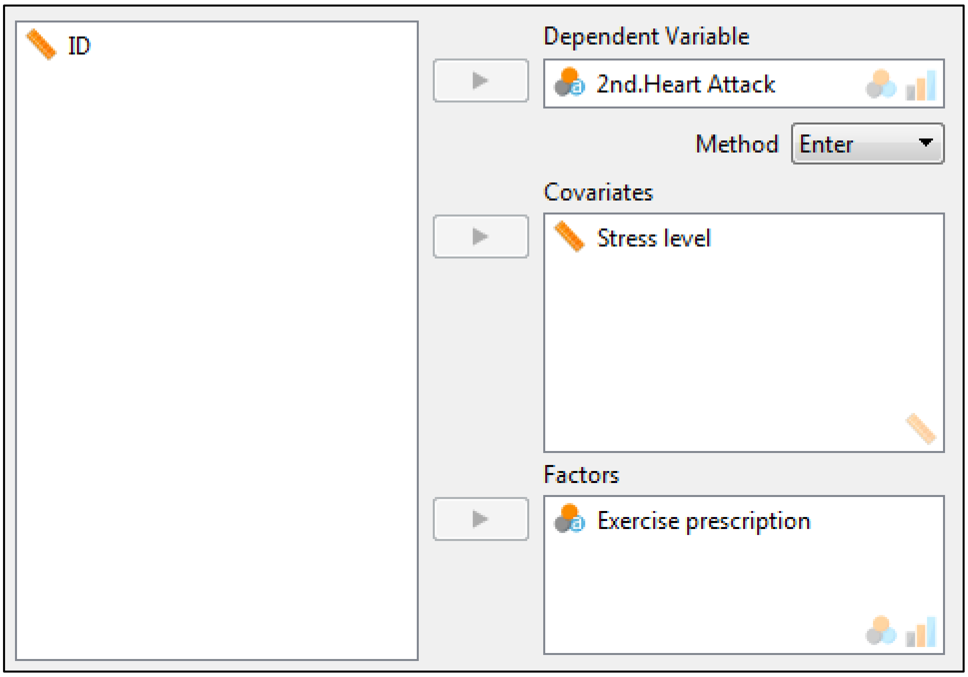
La **matriu de confusió** (***confusion matrix***) és una taula que mostra els resultats reals *vs*. els predits, i pot ser utilitzada per determinar la precisió del model. A partir d'ella, es poden derivar la sensibilitat i l'especificitat.

La **sensibilitat** (***sensitivity***) és el percentatge de casos en què el resultat observat va ser predit correctament pel model (és a dir, vertaders positius).

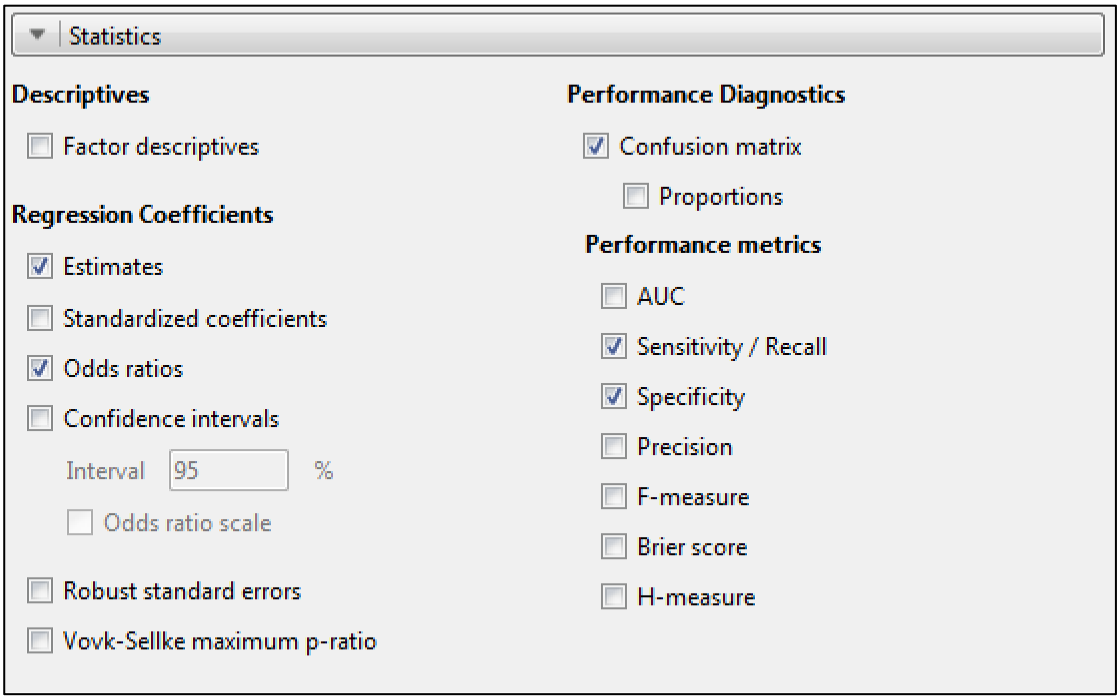
L'**especificitat** (***specificity***) és el percentatge d'observacions que també es van predir correctament com aquelles que no tenien els resultats observats (és a dir, vertaders negatius).

EXECUTANT LA REGRESSIÓ LOGÍSTICA

Obriu **Heart attack.csv** a JASP. Aquest arxiu conté 4 columnes de dades: la identificació de pacient (Patient ID), si van patir un segon atac de cor (sí / no), si se'ls va prescriure exercici (sí / no) i els seus nivells d'estrès (valor alt = estrès alt).
Poseu la variable de resultat (2nd.Heart.Attack) a la caixa «Dependent Variable», afegiu els nivells d'estrès a «Covariates» i la prescripció d'exercici a la caixa «Factors». Deixeu el mètode d'entrada com a forçada («Enter»).



A «Statistics», marqueu «Estimates», «Odds ratios», «Confusion matrix», «Sensitivity» i «Specificity».



ENTENENT EL RESULTAT

El resultat inicial consisteix en 4 taules.

El resum del model mostra que H1 (amb les puntuacions d'AIC i BIC més baixes) suggereix una relació significativa ($X^2(37) = 21{,}257$, $p < 0{,}001$) entre la variable resultat (Segon atac de cor) i les variables predictores (prescripció d'exercici i nivells d'estrès).

Model summary

Model	Deviance	AIC	BIC	df	X^2	p	McFadden R^2	Nagelkerke R^2	Tjur R^2
H_0	55.452	57.452	59.141	39					
H_1	34.195	40.195	45.261	37	21.257	< .001	0.383	0.550	0.126

L'R^2 de McFadden = 0,383. Sovint s'accepta que un valor en un rang entre 0,2 i 0,4 indica un bon ajustament del model.

Coefficients

	Estimate	Standard Error	Odds Ratio	z	p
(Intercept)	−4.368	2.550	0.013	−1.713	0.087
Stress level	0.089	0.041	1.093	2.159	0.031
Exercise prescription (Yes)	−2.043	0.890	0.130	−2.295	0.022

Note. 2nd.Heart Attack level 'Yes' coded as class 1.

Tant el nivell d'estrès com la prescripció d'exercici són variables predictores significatives (p = 0,031 i 0,022, respectivament). Els valors més importants en la taula de coeficients són les raons de probabilitats ("Odds Ratio"). Per a una variable predictora contínua, una raó de probabilitats més gran que 1 suggereix una relació positiva mentre que < 1 implica una relació negativa. Això suggereix que alts nivells d'estrès estan significativament relacionats amb una major probabilitat de tenir un segon atac de cor. La raó de probabilitats de 0,13 es pot interpretar com l'existència de només un 13% de probabilitats de patir un segon atac cardíac quan es realitza exercici físic.*

Confusion matrix

	Predicted	
Observed	No	Yes
No	15.000	5.000
Yes	5.000	15.000

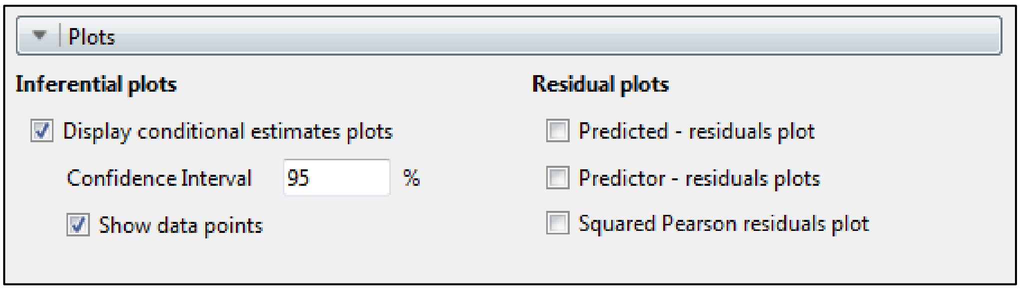
Performance metrics

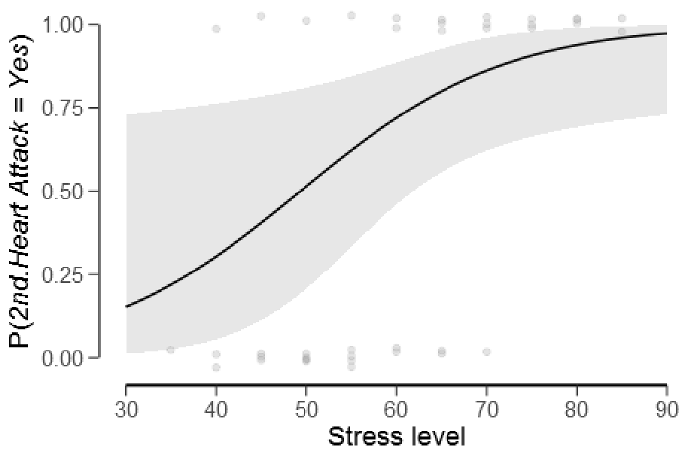
	Value
Sensitivity	0.750
Specificity	0.750

La matriu de confusió mostra que el model ha predit correctament 15 casos de vertaders negatius i altres 15 més de vertaders positius, mentre que ha comés un error en 5 casos de falsos negatius i uns altres 5 de falsos positius. Aquests resultats es reflecteixen en les mètriques de rendiment ("Performance metrics"), on la sensibilitat (% dels casos amb el resultat correctament predit) i l'especificitat (% de casos correctament predits com aquells que no tenen resultat; és a dir, vertaders negatius) són ambdues del 75%.

GRÀFICS

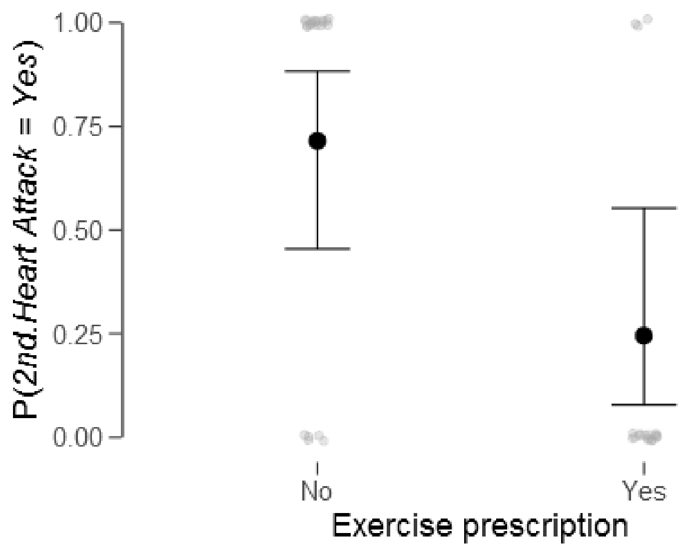
Aquests resultats es poden visualitzar fàcilment a través dels gràfics inferencials («Inferential plots»).





* O el que és el mateix, l'exercici físic està associat amb una reducció del 87% (0,13 - 1 = -0,87) de les probabilitats de patir un segon atac de cor en comparació amb aquells que no ho van fer. (Nota del revisor.)

Si s'incrementa el nivell d'estrès, augmenta la probabilitat de patir un segon atac de cor.



Si no es realitza exercici, augmenta la probabilitat de patir un segon atac de cor, mentre que es redueix si ha prescrit.

REPORTANT ELS RESULTATS

Es va realitzar una regressió logística per determinar els efectes de l'estrès i la prescripció d'exercici físic sobre la probabilitat que els participants patissin un segon atac de cor. El model de regressió logística va ser estadísticament significatiu, χ^2 (37) = 21,257, $p < 0{,}001$. El model va classificar correctament el 75,0% dels casos. L'increment de l'estrès es va associar amb un augment de la probabilitat de patir un segon atac cardíac, mentre que una reducció de l'estrès es va associar amb una disminució d'aquesta probabilitat. La prescripció d'un programa d'exercici va reduir al 13% la probabilitat d'un segon atac de cor.

COMPARACIÓ DE MÉS DE DOS GRUPS INDEPENDENTS

ANOVA

Mentre que les proves t comparen les mitjanes de dos grups / condicions, l'anàlisi de variància (**ANOVA,** *ANalysis Of VAriance*) d'un sol factor (o unifactorial) compara les mitjanes de 3 o més grups / condicions. En JASP hi ha disponibles els dos tipus d'ANOVA, de mesures independents (o mostres independents) i de mesures repetides (o mostres relacionades). L'ANOVA ha estat descrita com una "prova òmnibus" (global) que proporciona un estadístic F que compara si la variància explicada és significativament més gran que la variància no explicada. **La hipòtesi nul·la que es posa a prova és que no hi ha diferència significativa entre les mitjanes de tots els grups**. Si la hipòtesi nul·la és rebutjada, l'ANOVA simplement afirma que hi ha una diferència significativa entre els grups, però no on es troben aquestes diferències. Per determinar en quins grups es troben les diferències, a continuació s'han de dur a terme proves post hoc (del llatí *post hoc*, que significa "després d'això").

Per què no realitzar simplement múltiples comparacions entre parells? Si hi ha 4 grups (A, B, C, D), per exemple, i les diferències fossin comparades utilitzant múltiples proves t:

- A *vs*. B   P < 0,05   95% sense error Tipus I
- A *vs*. C   P < 0,05   95% sense error Tipus I
- A *vs*. D   P < 0,05   95% sense error Tipus I
- B *vs*. C   P < 0,05   95% sense error Tipus I
- B *vs*. D   P < 0,05   95% sense error Tipus I
- C *vs*. D   P < 0,05   95% sense error Tipus I

Assumint que cada prova fos independent, la probabilitat global seria doncs:

$$0,95 * 0,95 * 0,95 * 0,95 * 0,95 * 0,95 = 0,735$$

Això es coneix com a *familywise error* o error de Tipus I acumulatiu, i en aquest cas resulta en només un 73,5% de probabilitat que no cometem un error de Tipus I, de manera que la hipòtesi nul·la podria ser rebutjada quan en realitat és vertadera. Això s'evita amb les proves post hoc, que realitzen comparacions múltiples per parelles amb criteris d'acceptació més estrictes i permeten així prevenir aquest tipus d'error.

SUPÒSITS

L'ANOVA de mesures independents té els mateixos supòsits que la majoria dels altres tests paramètrics.

- La variable independent ha de ser categòrica i la variable dependent ha de ser contínua.
- Els grups han de ser independents entre si.
- La variable dependent ha de tenir una distribució aproximadament normal.
- No hi hauria d'haver valors atípics significatius.
- Hi hauria d'haver homogeneïtat de variància entre els grups; d'una altra manera, el valor de p per a l'estadístic F podria no ser fiable.

Normalment, els 2 primers supòsits es controlen amb un disseny d'investigació adequat.

Si els tres últims supòsits són violats, llavors s'hauria de considerar la prova de Kruskal-Wallis, el seu equivalent no paramètric.

PROVES POST HOC

JASP proporciona 4 alternatives per portar a terme amb la prova d'ANOVA de mesures independents:

a) ***Bonferroni*** – pot ser molt conservador, però ofereix garanties de control de l'error de Tipus I a risc de reduir la potència estadística.
b) ***Holm*** – el test Holm-Bonferroni, un mètode Bonferroni seqüencial menys conservador que el test Bonferroni original.
c) ***Tukey*** – un dels tests més freqüentment utilitzats que proporciona un error de Tipus I controlat per a grups amb la mateixa mida de mostra i la mateixa variància.
d) ***Scheffe*** – controla el nivell global de confiança si els grups tenen diferents mides de mostra.

JASP també proporciona 4 tipus:

a) ***Standar*** – tal com es descriuen a dalt les quatre alternatives.
b) ***Games-Howell*** – s'utilitza quan no tenim seguretat sobre la igualtat de les variàncies dels grups.
c) **Dunnett's** – es fa servir quan volem comparar tots els grups amb un de sol, és a dir, el grup de control.
d) ***Dunn*** – un test post hoc no paramètric que s'utilitza per posar a prova petits subgrups de parells.

MIDA DE L'EFECTE

JASP proporciona 3 càlculs alternatius de la mida de l'efecte per utilitzar amb la prova ANOVA de mesures independents:

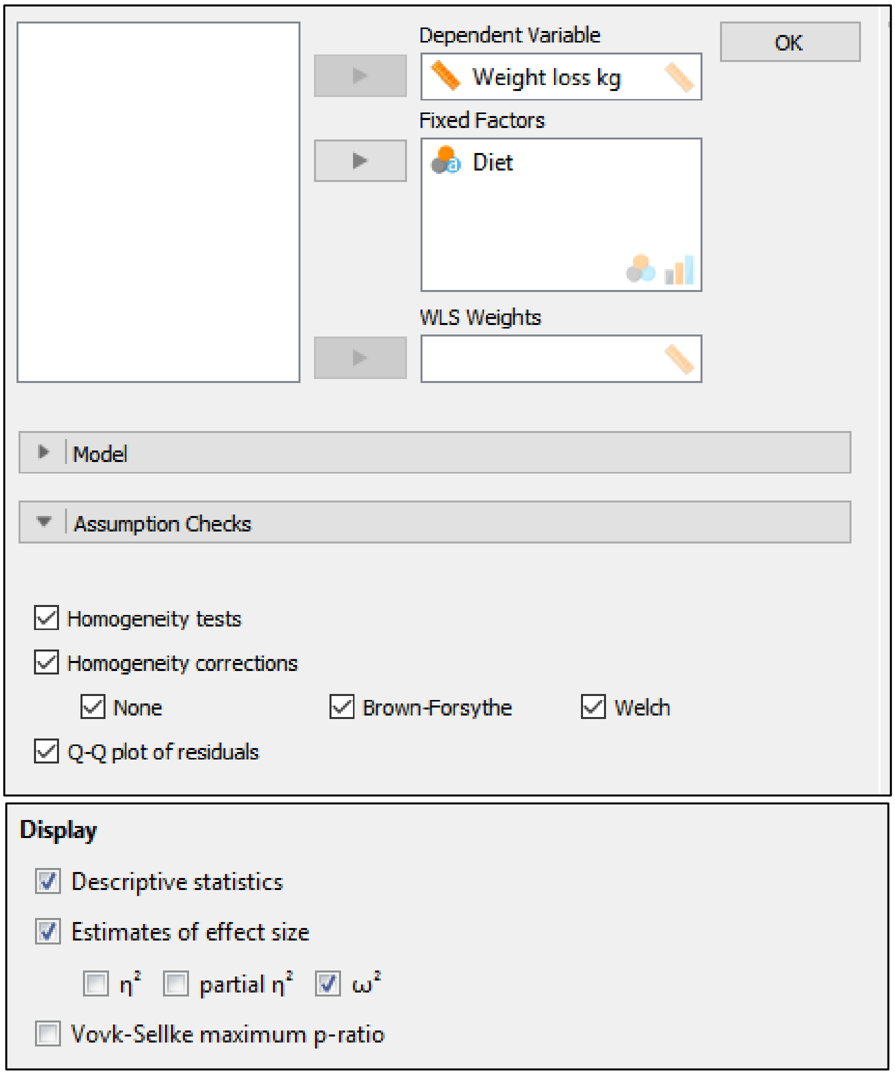
a) ***Eta quadrat (η^2)*** – precís per a estimar la variància explicada a la mostra però sobreestima la variància en la població. Dificulta la comparació de l'efecte de la mateixa variable en diferents estudis.
b) ***Eta quadrat parcial (η_p^2)*** – resol el problema de la sobreestimació de la variància en la població permetent la comparació de l'efecte de la mateixa variable en diferents estudis.
c) ***Omega quadrat (ω^2)*** – normalment, el biaix estadístic esdevé molt petit a mesura que s'incrementa la mida de mostra, però quan tenim mostres petites ($n < 30$) ω^2 proporciona una mesura no esbiaixada de la mida de l'efecte.

Test	Mida	Irrellevant	Petit	Mitjà	Gran
ANOVA	**Eta**	**< 0,1**	**0,1**	**0,25**	**0,37**
	Eta parcial	**< 0,01**	**0,01**	**0,06**	**0,14**
	Omega quadrat	**< 0,01**	**0,01**	**0,06**	**0,14**

EXECUTANT L'ANOVA DE MESURES INDEPENDENTS

Carregueu **Independent ANOVA diet.csv**. Aquest arxiu conté una columna A amb 3 dietes utilitzades (A, B i C) i una altra columna amb la quantitat total de pes perdut després de 8 setmanes seguint una de les 3 dietes diferents. És una bona pràctica comprovar l'estadística descriptiva i els gràfics de caixa per si hi hagués valors atípics extrems.

Aneu a «ANOVA» → «ANOVA», introduïu la pèrdua de pes a la caixa «Dependent Variable» i les agrupacions segons la dieta a la caixa de factors fixos («Fixed Factors»). En primera instància, seleccioneu les comprovacions de supòsits («Assumption Checks») i, en les opcions addicionals, marqueu «Descriptive statistics» i ω^2 com a mida de l'efecte:



Això donarà un resultat en 3 taules i un gràfic Q-Q.

ENTENENT EL RESULTAT

ANOVA - Weight loss kg

Cases	Homogeneity Correction	Sum of Squares	df	Mean Square	F	p
Diet	None	92.37	2.000	46.184	10.83	< .001
Diet	Brown-Forsythe	92.37	2.000	46.184	10.83	< .001
Diet	Welch	92.37	2.000	46.184	11.45	< .001
Residual	None	294.37	69.000	4.266		
Residual	Brown-Forsythe	294.37	64.352	4.574		
Residual	Welch	294.37	44.987	6.544		

Note. Type III Sum of Squares

La taula d'ANOVA principal mostra que l'estadístic F és significatiu ($p < 0{,}001$) i que hi ha una mida de l'efecte gran. Per tant, hi ha una diferència significativa entre les mitjanes dels 3 grups de dietes.

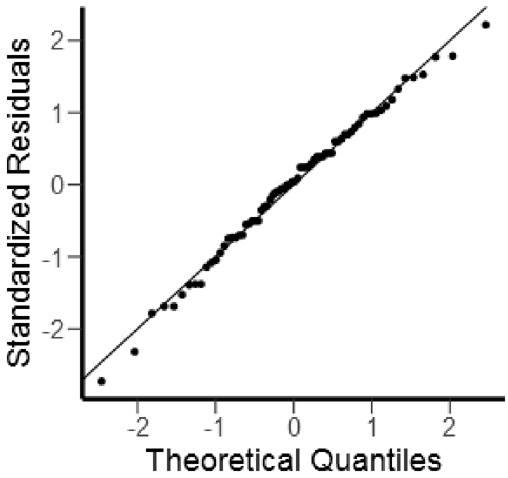
COMPROVACIÓ DELS SUPÒSITS

Abans de donar per bons aquests resultats, s'ha de comprovar que no es violen els supòsits requerits per la prova d'ANOVA.

Test for Equality of Variances (Levene's)

F	df1	df2	p
1.298	2.000	69.000	0.280

El test de Levene mostra que l'homogeneïtat de la variància no és significativa. No obstant això, si la prova de Levene mostra una diferència significativa en la variància, s'hauria de reportar la correcció Brown-Forsythe o la de Welch.



El gràfic Q-Q mostra que les dades semblen tenir una distribució normal i que són lineals.

Descriptives - Weight loss kg

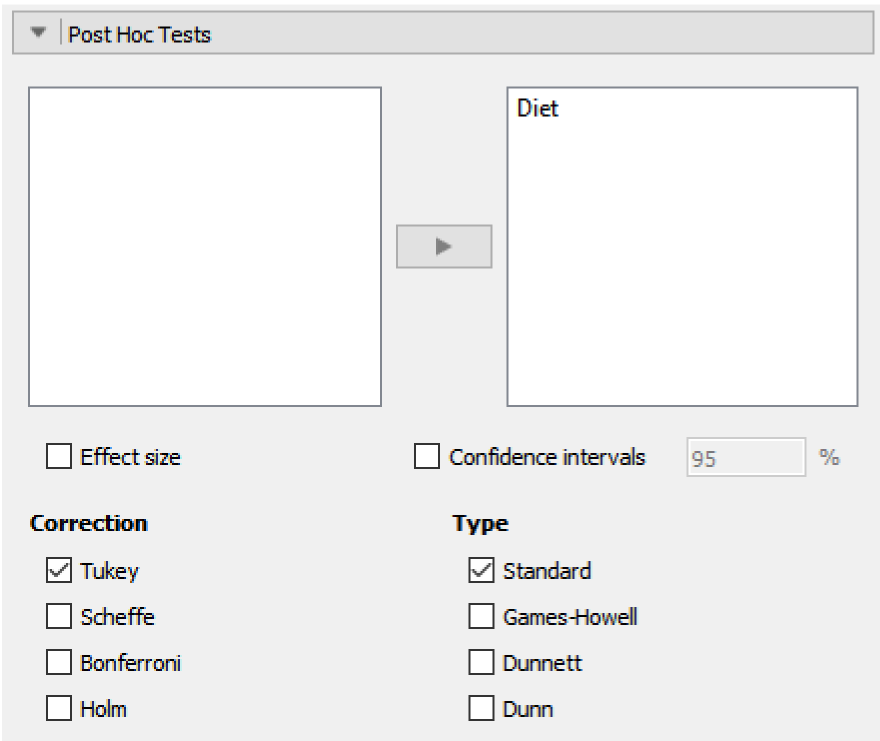
Diet	Mean	SD	N
Diet A	3.008	1.668	24.000
Diet B	3.413	2.361	24.000
Diet C	5.588	2.108	24.000

L'estadística descriptiva suggereix que la dieta 3 aconsegueix la major pèrdua de pes després de 8 setmanes.

Si l'ANOVA no reporta una diferència significativa, no heu de prosseguir amb l'anàlisi.

PROVES POST HOC

Si l'ANOVA és significativa, ara es pot dur a terme l'anàlisi post hoc. A «Post Hoc Tests», afegiu Diet a la caixa d'anàlisi de la dreta, seleccioneu «Effect size» i, en aquest cas, utilitzeu «Tukey» per a la correcció post hoc.



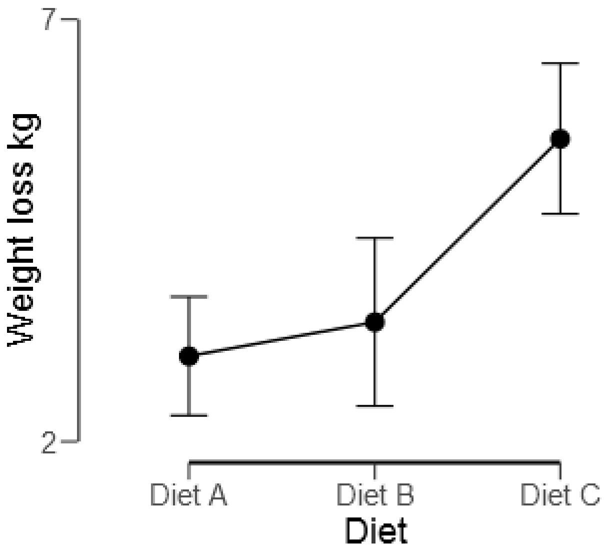
Afegiu també, en «Descriptive plots», el factor Diet a l'eix horitzontal i trieu «Display error bars».

Post Hoc Comparisons - Diet

		Mean Difference	SE	t	Cohen's d	p_{tukey}
Diet A	Diet B	−0.404	0.596	−0.678	−0.198	0.777
	Diet C	−2.579	0.596	−4.326	−1.357	< .001
Diet B	Diet C	−2.175	0.596	−3.648	−0.972	0.001

Note. Cohen's d does not correct for multiple comparisons.

L'anàlisi post hoc mostra que no hi ha diferència significativa en la pèrdua de pes entre les dietes A i B. Tanmateix, és significativament superior en la dieta C comparada amb la dieta A ($p < 0,001$) i la dieta B ($p = 0,001$). La d de Cohen mostra que aquestes diferències es corresponen amb una mida de l'efecte gran.



REPORTANT ELS RESULTATS

L'ANOVA unifactorial va mostrar un efecte significatiu del tipus de dieta sobre la pèrdua de pes després de 8 setmanes **($F (2, 69) = 46,184$, $p < 0,001$, $\omega^2 = 0,214$)**.

L'anàlisi post hoc mitjançant la correcció de Tukey va revelar que la dieta C va aconseguir una pèrdua de pes significativament superior que la dieta A ($p < 0,001$) o la dieta B ($p = 0,001$). No hi va haver diferències significatives de pèrdua de pes entre les dietes A i B ($p = 0,777$).

KRUSKAL-WALLIS: L'ANOVA NO PARAMÈTRIC

Si les dades no compleixen amb els supòsits paramètrics o són de naturalesa nominal, la prova *H* de Kruskal-Wallis és un equivalent no paramètric a l'ANOVA per a mesures o mostres independents. Es pot utilitzar per comparar dos o més grups independents amb una mida de mostra igual o diferent. Com les proves de Mann-Whitney i Wilcoxon, és un test basat en rangs.

Com l'ANOVA, la prova *H* de Kruskal-Wallis (també coneguda com a ANOVA unifactorial per rangs) és una prova global que no especifica quins grups de la variable independent són significativament diferents entre si. Per poder fer-ho, JASP proporciona l'opció d'executar la prova post hoc de Dunn. Aquesta prova de comparacions múltiples pot ser molt conservadora, especialment quan es realitzen un gran nombre de comparacions.

Carregueu el conjunt de dades **Kruskal-Wallis ANOVA.csv** a JASP. Aquest conjunt de dades conté puntuacions de dolor subjectiu en participants que no reben tractament (control), que reben crioteràpia o que reben una combinació de crioteràpia amb compressió per tractar el dolor muscular d'aparició tardana després de l'exercici.

EXECUTANT LA PROVA DE KRUSKAL-WALLIS

Aneu a «ANOVA» → «ANOVA». A la finestra d'anàlisi, afegiu la puntuació de dolor a la caixa de variable dependent («Dependent Variable») i el tractament a la caixa de factors fixos («Fixed Factors»). Comproveu que la puntuació de dolor està assignada com a variable ordinal. Això executarà, automàticament, l'ANOVA de mesures independents convencional. A «Assumption Checks», seleccioneu «Homogeneity tests» i «Q-Q plot of residuals».

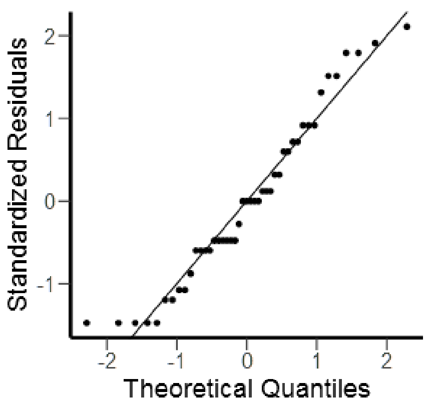
ANOVA - Pain Score

Cases	Sum of Squares	df	Mean Square	F	p
Treatment	98.844	2.000	49.422	16.457	< .001
Residual	126.133	42.000	3.003		

Note. Type III Sum of Squares

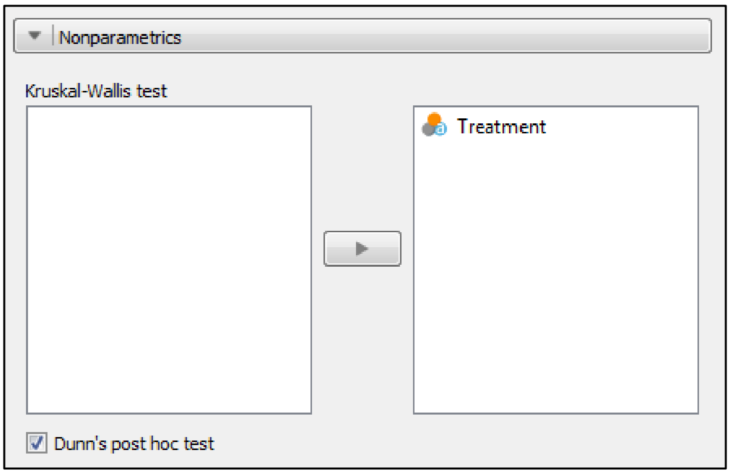
Test for Equality of Variances (Levene's)

F	df1	df2	p
3.832	2.000	42.000	0.030



Encara que l'ANOVA mostra un resultat significatiu, les dades no compleixen amb els supòsits d'homogeneïtat de la variància tal com es pot observar pel test significatiu de Levene, i només mostren linealitat al centre del gràfic Q-Q que es corba en els extrems, indicant la presència de valors extrems. Això, afegit al fet que la variable dependent està basada en puntuacions de dolor subjectiu, suggereix l'ús d'una alternativa no paramètrica.

Torneu a les opcions estadístiques i obriu l'opció «Nonparametrics», situada al final. Per obtenir el test de Kruskal-Wallis, moveu la variable Treatment a la caixa de la dreta i seleccioneu «Dunn's post hoc test».



ENTENENT EL RESULTAT

El resultat mostra dues taules. La prova de Kruskal-Wallis mostra que hi ha una diferència significativa entre les tres modalitats de tractament.

Kruskal-Wallis Test

Factor	Statistic	df	p
Treatment	19.693	2	< .001

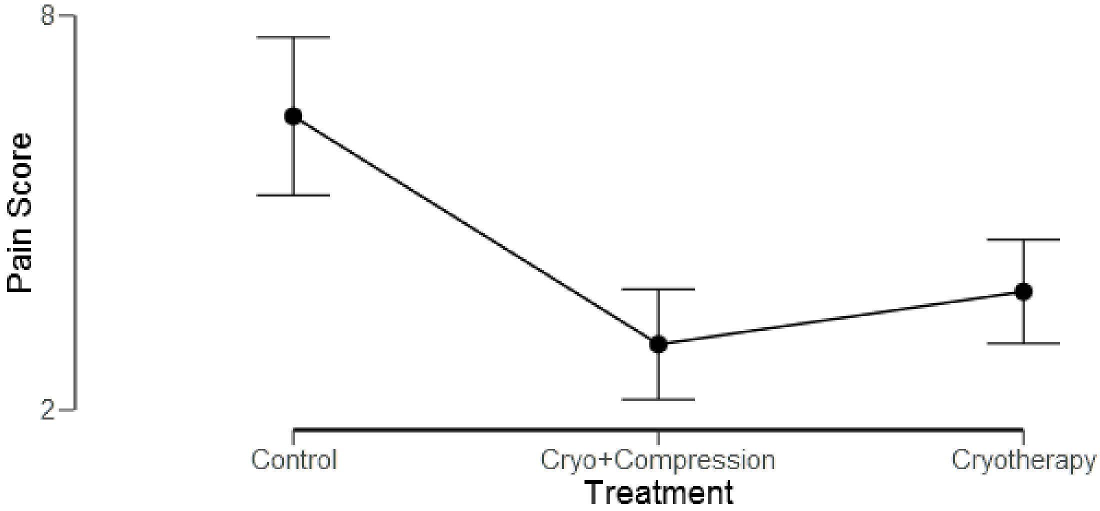
Dunn's Post Hoc Comparisons - Treatment

		z	W_i	W_j	p	p_{bonf}	p_{holm}
Control	Cryo+Compression	4.317	34.600	14.200	< .001	< .001	< .001
	Cryotherapy	3.048	34.600	20.200	0.001	0.003	0.002
Cryo+Compression	Cryotherapy	−1.270	14.200	20.200	0.102	0.306	0.102

El test post hoc de Dunn facilita el seu propi valor de p, així com els de Bonferroni i la correcció de Holm. Com es pot veure, ambdues condicions de tractament són significativament diferents respecte al grup control, però no entre si.

REPORTANT ELS RESULTATS

Les puntuacions de dolor van estar afectades significativament per la modalitat de tractament **H (2) = 19,693, p < 0,001**. Les comparacions dos a dos van mostrar que, tant la crioteràpia com la crioteràpia amb compressió, redueixen significativament les puntuacions de dolor (p = 0,001 i p < 0,001, respectivament) en comparació amb el grup de control. No hi va haver diferències significatives entre la crioteràpia i la crioteràpia amb compressió (p = 0,102).



COMPARACIÓ DE MÉS DE DOS GRUPS RELACIONATS

ANOVA MR

L'ANOVA d'un sol factor de mesures repetides (**ANOVA MR**) s'utilitza per avaluar si hi ha diferències en les mesures entre 3 o més grups (on els participants són els mateixos en cada grup) que hagin estat tractats en diverses ocasions o sota diferents condicions. Com a disseny de la investigació, per exemple, els mateixos participants podrien ser tractats prenent una mesura de resultat en 1, 2 i 3 setmanes o que la mesura de resultat fos presa sota les condicions 1, 2 i 3.

La hipòtesi nul·la que es posa a prova és que no hi ha diferència significativa entre les mitjanes de les diferències entre tots els grups.

La variable independent hauria de ser categòrica i la variable dependent ha de ser una mesura contínua. En aquesta anàlisi les categories de la variable independent són **nivells** designats, és a dir, són els grups relacionats. Per tant, en el cas que una variable resultat fos mesurada en 1, 2 i 3 setmanes, els 3 nivells serien setmana 1, setmana 2 i setmana 3.

L'**estadístic F** es calcula dividint els quadrats mitjans de la variable (variància explicada pel model) pels quadrats mitjans del seu error (variància no explicada). Com més gran sigui l'estadístic F, més probable serà que la variable independent hagi tingut un efecte significatiu sobre la variable dependent.

SUPÒSITS

L'ANOVA MR té els mateixos supòsits que la majoria dels tests paramètrics.

- La variable dependent hauria de tenir una distribució aproximadament normal.
- No hi hauria d'haver valors atípics significatius.
- Esfericitat, que té a veure amb la igualtat de les variàncies de les diferències entre els nivells del factor de mesures repetides.

Si els supòsits han estat violats, llavors s'hauria de considerar el **test de Friedman**, el seu equivalent no paramètric descrit més endavant en aquesta mateixa secció.

ESFERICITAT

Si un estudi té 3 nivells (A, B i C), l'esfericitat assumeix el següent:

$$\text{Variància (A-B)} \approx \text{Variància (A-C)} \approx \text{Variància (B-C)}$$

L'ANOVA MR comprova el supòsit d'esfericitat utilitzant el test d'esfericitat de Mauchly (pronunciat com "Mockley"). Aquest test posa a prova **la hipòtesi nul·la que les variàncies de les diferències són iguals**. En molts casos, les mesures repetides violen el supòsit d'esfericitat, el que pot conduir a un error de Tipus I. Si aquest és el cas, es poden aplicar correccions a l'estadístic F.

JASP ofereix dos mètodes de correcció de l'estadístic F: les correccions èpsilon (ε) **de Greenhouse-Geisser** i de **Huynh-Feldt**. Com a recomanació general, si els valors ε són < 0,75 s'ha d'utilitzar la correcció de Greenhouse-Geisser i, si són > 0,75, la correcció de Huynh-Feldt.

PROVES POST HOC

Tot i que l'anàlisi post hoc és limitada en el cas de l'ANOVA MR, JASP proporciona dues alternatives:

a) ***Bonferroni*** – pot ser molt conservador, però garanteix el control de l'error de tipus I a risc de reduir la potència estadística.
b) ***Holm*** – el test Holm-Bonferroni és un mètode seqüencial Bonferroni menys conservador que el test Bonferroni original.

Si se sol·liciten les correccions post hoc de Tukey o Scheffe, JASP reportarà un error NaN (per l'anglès *not a number*, o "no és un nombre").

MIDA DE L'EFECTE

JASP proporciona les mateixes alternatives al càlcul de la mida de l'efecte que les utilitzades en la prova d'ANOVA de mesures independents:

a) ***Eta quadrat (η^2)*** – precís per estimar la variància explicada en la mostra però sobreestima la variància en la població. Dificulta la comparació de l'efecte d'una mateixa variable en diferents estudis.
b) ***Eta quadrat parcial (η_p^2)*** – resol el problema de la sobreestimació de la variància en la població permetent la comparació de l'efecte de la mateixa variable en diferents estudis. Aquesta és la manera més habitual de reportar la mida de l'efecte en l'ANOVA de mesures repetides.
c) ***Omega quadrat (ω^2)*** – normalment, el biaix estadístic esdevé molt petit a mesura que s'incrementa la mida de mostra, però quan tenim mostres petites (n < 30) ω^2 proporciona una mesura no esbiaixada de la mida de l'efecte.

Nivells de mida de l'efecte:

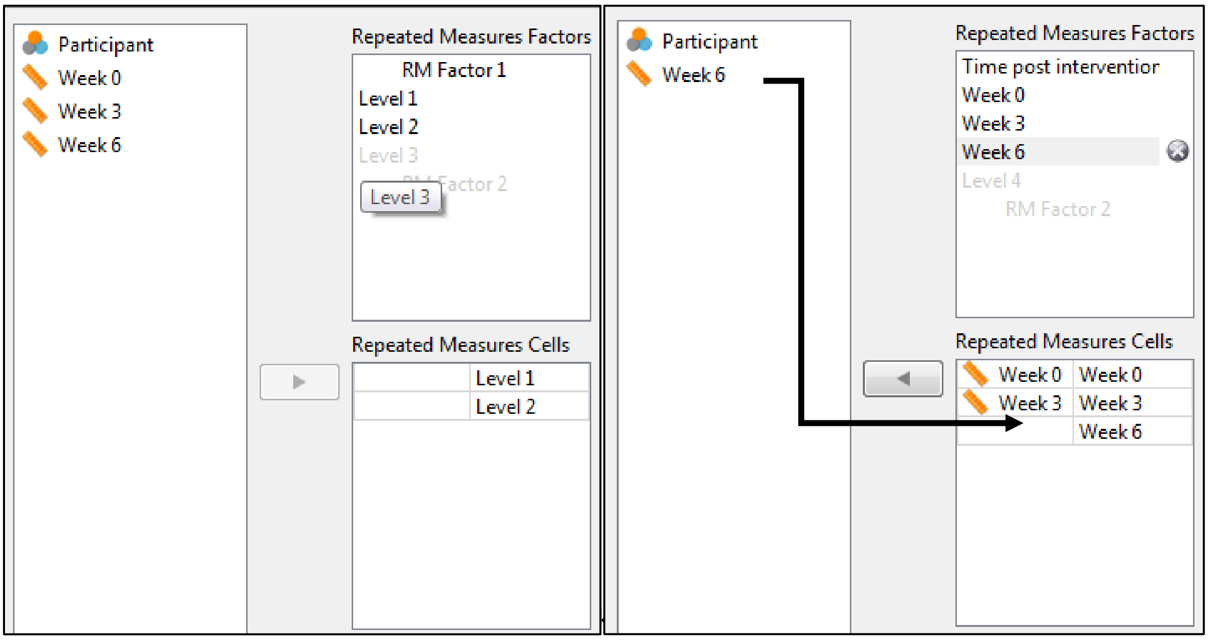
Test	Mida	Irrellevant	Petit	Mitjà	Gran
ANOVA	**Eta**	**< 0,1**	**0,1**	**0,25**	**0,37**
	Eta parcial	**< 0,01**	**0,01**	**0,06**	**0,14**
	Omega quadrat	**< 0,01**	**0,01**	**0,06**	**0,14**

EXECUTANT L'ANOVA DE MESURES REPETIDES

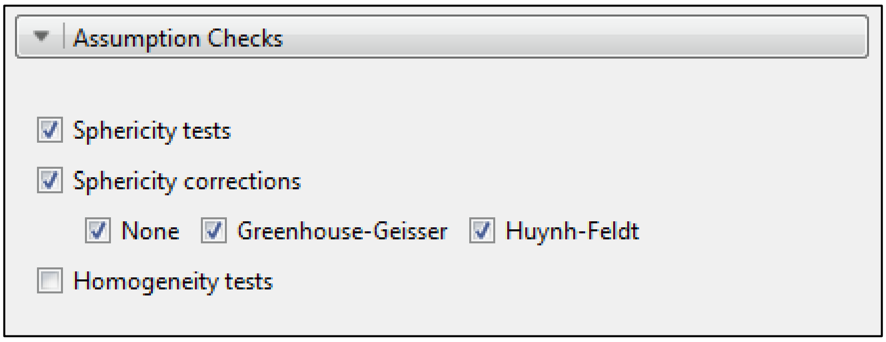
Carregueu **Repeated ANOVA cholesterol.csv.** Aquest arxiu conté una columna amb les ID dels participants i unes altres 3 columnes més, una per a cada mesura repetida del colesterol en sang després d'una intervenció. És una bona pràctica revisar l'estadística descriptiva i els gràfics de caixa per si hi hagués algun valor atípic extrem.

Aneu a «ANOVA» → «Repeated measures ANOVA». Com s'ha comentat abans, la variable independent (el factor de mesures repetides) té nivells, en aquest cas 3. Canvieu el nom de la variable RM Factor 1 a Time post intervention i feu el mateix amb els 3 nivells posant Week 0, Week 3 i Week 6, respectivament.

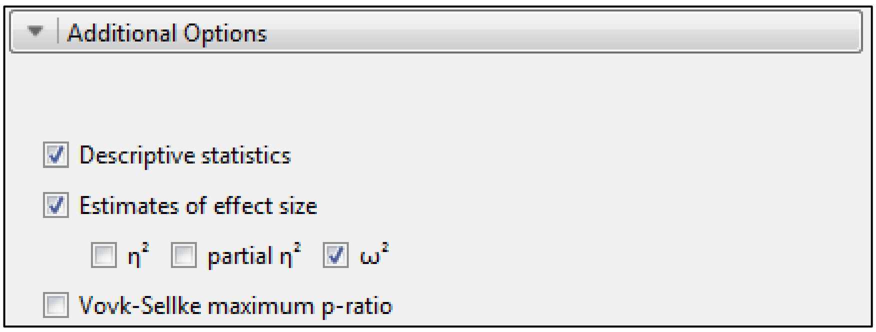
Un cop fet, apareixeran a la caixa «Repeated Measures Cells». Ara, afegiu les dades adients al nivell corresponent.



A «Assumption Checks», seleccioneu «Sphericity tests» i totes les opcions incloses a «Sphericity corrections».



A «Additional Options», marqueu «Descriptive statistics», «Estimates of effect size» i «ω^2».



El resultat hauria d'incloure 4 taules. La tercera taula, la qual correspon als efectes inter-subjectes, es pot ignorar en aquesta anàlisi.

ENTENENT EL RESULTAT

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	ω²
Time post intervention	None	4.320ᵃ	2.000ᵃ	2.160ᵃ	212.321ᵃ	< .001ᵃ	0.058
	Greenhouse-Geisser	4.320ᵃ	1.235ᵃ	3.497ᵃ	212.321ᵃ	< .001ᵃ	0.058
	Huynh-Feldt	4.320ᵃ	1.284ᵃ	3.365ᵃ	212.321ᵃ	< .001ᵃ	0.058
Residual	None	0.346	34.000	0.010			
	Greenhouse-Geisser	0.346	21.001	0.016			
	Huynh-Feldt	0.346	21.822	0.016			

Note. Type III Sum of Squares
ᵃ Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

La taula d'efectes intra-subjectes ("Within Subjects Effects") mostra un estadístic F gran, altament significatiu (p < 0,001) i amb una mida de l'efecte entre petita i mitjana (0,058). Aquesta taula presenta els estadístics que assumeixen l'esfericitat ("None") i els dos mètodes de correcció. Les principals diferències estan en els graus de llibertat ("df", per l'anglès *degrees of freedom*) i el valor dels quadrats mitjans. Sota la taula s'indica que el supòsit d'esfericitat ha estat violat.

La taula següent ofereix els resultats del test d'esfericitat de Mauchly. Com es pot veure, hi ha una diferència significativa (p < 0,001) en les variàncies de les diferències entre els grups. Els valors èpsilon (ε) de Greenhouse-Geisser i de Huynh-Feldt estan per sota de 0,75. Per tant, el resultat de l'ANOVA s'hauria de basar en la correcció de Greenhouse-Geisser:

Test of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
Time post intervention	0.381	< .001	0.618	0.642

Per a obtenir una taula més clara, torneu a «Assumption Checks» i seleccioneu únicament «Greenhouse-Geisser» com a correcció de l'esfericitat.

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	ω²
Time post intervention	Greenhouse-Geisser	4.320ᵃ	1.235ᵃ	3.497ᵃ	212.321ᵃ	< .001ᵃ	0.058
Residual	Greenhouse-Geisser	0.346	21.001	0.016			

Note. Type III Sum of Squares
ᵃ Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Hi ha una diferència significativa entre les mitjanes de les diferències entre tots els grups: **F (1,235, 21,0) = 212,3, p < 0,001, ω^2 = 0,058.**

Descriptives

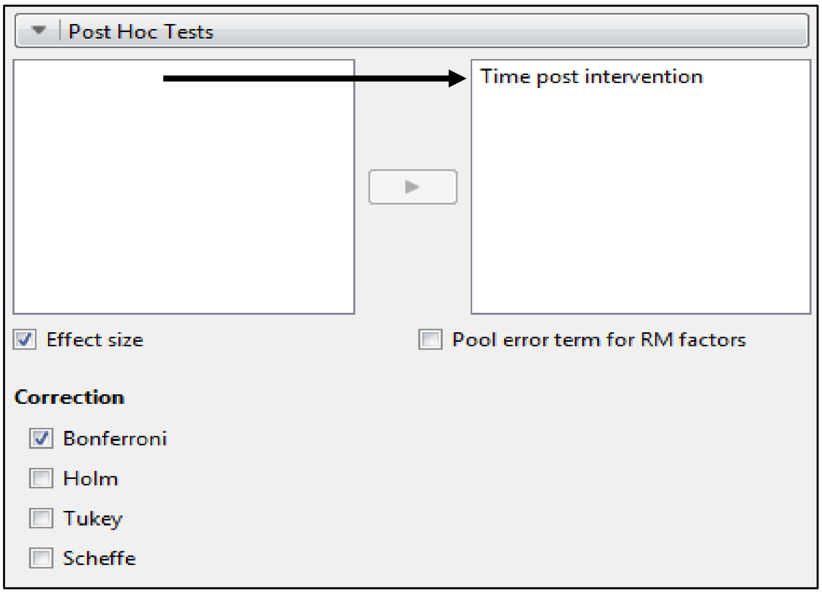
Time post intervention	Mean	SD	N
Week 0	6.408	1.191	18.000
Week 3	5.842	1.123	18.000
Week 6	5.779	1.102	18.000

L'anàlisi descriptiva suggereix que els nivells de colesterol en sang van ser més alts en la setmana 0, comparats amb els de les setmanes 3 i 6.

No obstant això, si l'ANOVA no reporta diferències significatives, no es pot prosseguir amb l'anàlisi.

PROVES POST HOC

Si l'ANOVA és significativa, es pot dur a terme l'anàlisi post hoc. A «Post Hoc Tests» afegiu Time post intervention a la caixa d'anàlisi de la dreta, seleccioneu «Effect size» i, en aquest cas, utilitzeu Bonferroni per a la correcció post hoc.



Afegiu també, a «Descriptive plots», el factor Time post intervention a la caixa «Horizontal axis» i marqueu «Display error bars».

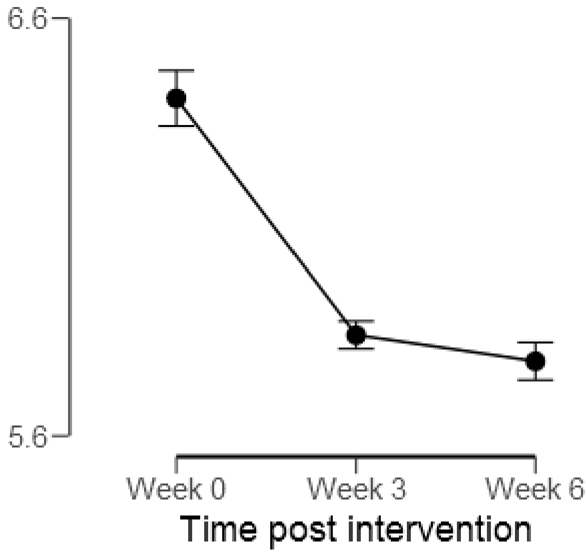
Post Hoc Comparisons - Time post intervention

		Mean Difference	SE	t	Cohen's d	p_{bonf}
Week 0	Week 3	0.566	0.037	15.439	3.639	< .001
	Week 6	0.629	0.042	14.946	3.523	< .001
Week 3	Week 6	0.063	0.017	3.781	0.891	0.004

Note. Cohen's d does not correct for multiple comparisons.

L’anàlisi post hoc mostra que hi ha diferències significatives en els nivells de colesterol en sang entre totes les combinacions de valors de temps i que estan associades amb mides de l’efecte grans.

REPORTANT ELS RESULTATS



Es va utilitzar la correcció de Greenhouse-Geisser, atès que el test d’esfericitat de Mauchly va ser significatiu. L’anàlisi va mostrar que els nivells de colesterol van diferir significativament: F (1,235, 21,0) = 212,3, p < 0,001, ω^2 = 0,058.

L’anàlisi post hoc utilitzant la correcció de Bonferroni va revelar que els nivells de colesterol van disminuir significativament a mesura que va anar passant el temps entre les setmanes 0 i 3 (diferència de les mitjanes = 0,566 unitats, p < 0,001) i entre les setmanes 3 i 6 (diferència de les mitjanes = 0,063 unitats, p = 0,004).

ANOVA DE MESURES REPETIDES DE FRIEDMAN

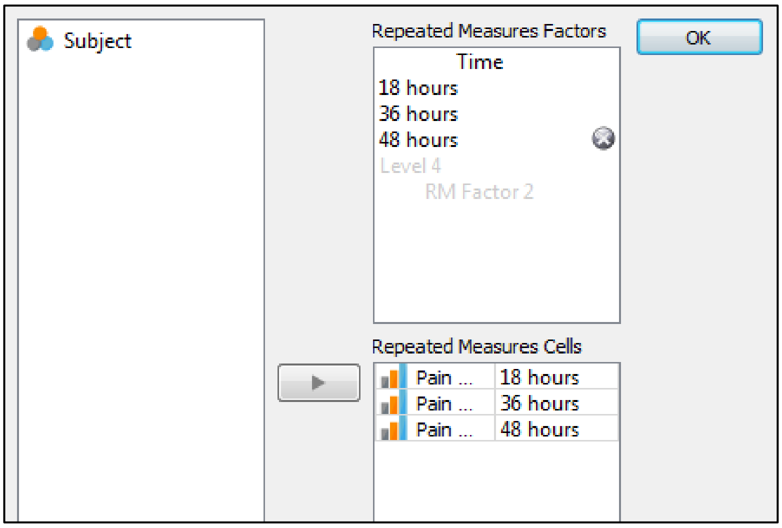
Si els supòsits paramètrics han estat violats o les dades són ordinals, s'hauria de considerar l'ús de l'alternativa no paramètrica, el test de Friedman. Com el test de Kruskal-Wallis, la prova de Friedman s'utilitza per a l'anàlisi de la variància de mesures repetides d'un sol factor per rangs, i no suposa que les dades provinguin d'una distribució en particular. Es tracta d'una altra "prova òmnibus" (global) que no especifica quins grups de la variable independent són significativament diferents entre si. Si el test de Friedman és significatiu, JASP proporciona l'opció d'executar la prova post hoc de Conover.

Carregueu **Friedman RMANOVA.csv** a JASP. Aquest arxiu conté 3 columnes amb les puntuacions de dolor subjectiu registrades a les 18, 36 i 48 hores després d'haver realitzat exercici. Comproveu que les puntuacions de dolor estan assignades com a variables ordinals.

EXECUTANT EL TEST DE FRIEDMAN

Aneu a «ANOVA» → «Repeated Measures ANOVA». La variable independent (factor de mesures repetides) té 3 nivells. Canvieu el nom de la variable RM Factor 1 a Time i feu el mateix amb els 3 nivells posant 18 hours, 36 hours i 48 hours, respectivament.

Després de fer això, apareixeran a la caixa «Repeated Measures Cells». Afegiu ara les dades adients al nivell corresponent.



D'aquesta manera, obtindreu la taula d'ANOVA estàndard de mesures repetides intra-subjectes. Per a executar el test de Friedman, expandiu la pestanya «Nonparametrics», moveu Time a la caixa «RM Factor» i marqueu «Connover's post hoc test».

ENTENENT EL RESULTAT

S'haurien d'haver obtingut dues taules.

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Time	26.772	2	< .001	0.764

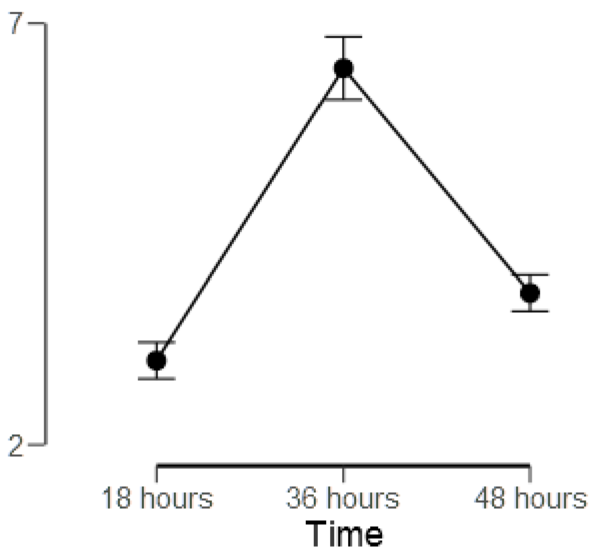
Connover's Post Hoc Comparisons - Time

		T-Stat	df	W_i	W_j	p	p_{bonf}	p_{holm}
18 hours	36 hours	15.171	28	17.000	44.500	< .001	< .001	< .001
	48 hours	6.344	28	17.000	28.500	< .001	< .001	< .001
36 hours	48 hours	8.827	28	44.500	28.500	< .001	< .001	< .001

La prova de Friedman mostra que el temps té un efecte significatiu sobre la percepció del dolor. Les comparacions post hoc dos a dos de Connor mostren que les percepcions de dolor són significativament diferents en cada moment.

REPORTANT ELS RESULTATS

El temps té un efecte significatiu en les puntuacions de dolor subjectiu χ^2 (2) = 26,77, p < 0,001. Les comparacions dos a dos mostren que la percepció del dolor és significativament diferent en cada moment (tots els valors de p < 0,001).



ANOVA DE MESURES INDEPENDENTS DE DOS FACTORS

Si l'ANOVA d'un sol factor avalua situacions en què només es manipula una variable independent, l'ANOVA de dos factors s'utilitza quan s'ha manipulat més d'1 variable independent. En aquest cas, les variables independents es coneixen com a factors.

FACTOR 1	FACTOR 2	
CONDICIÓ 1	Grup 1	Variable dependent
	Grup 2	Variable dependent
CONDICIÓ 2	Grup 1	Variable dependent
	Grup 2	Variable dependent
CONDICIÓ 3	Grup 1	Variable dependent
	Grup 2	Variable dependent

Els factors estan dividits en nivells, de manera que, en aquest cas, el factor 1 té 3 nivells i hi ha 2 nivells per al factor 2.

L'"efecte principal" (*Main effect*) és l'efecte d'una de les variables independents sobre la variable dependent, ignorant els efectes de qualsevol altra variable independent. Hi ha 2 efectes principals que es posen a prova, ambdós "inter-subjectes" (*Between-subjects*): en aquest cas, comparant les diferències en el factor 1 (és a dir, la condició), i les diferències en el factor 2 (els grups). Es produeix una **interacció** quan un factor influeix en l'altre.

L'ANOVA de dos factors de mesures independents és una altra "prova òmnibus" (global) que s'utilitza per provar 2 hipòtesis nul·les:

1. **No hi ha cap efecte significatiu inter-subjectes, és a dir, no hi ha diferències significatives entre les mitjanes dels grups en qualsevol dels factors.**
2. **No hi ha cap efecte d'interacció significatiu, és a dir, no hi ha diferències de grup significatives entre les condicions.**

SUPÒSITS

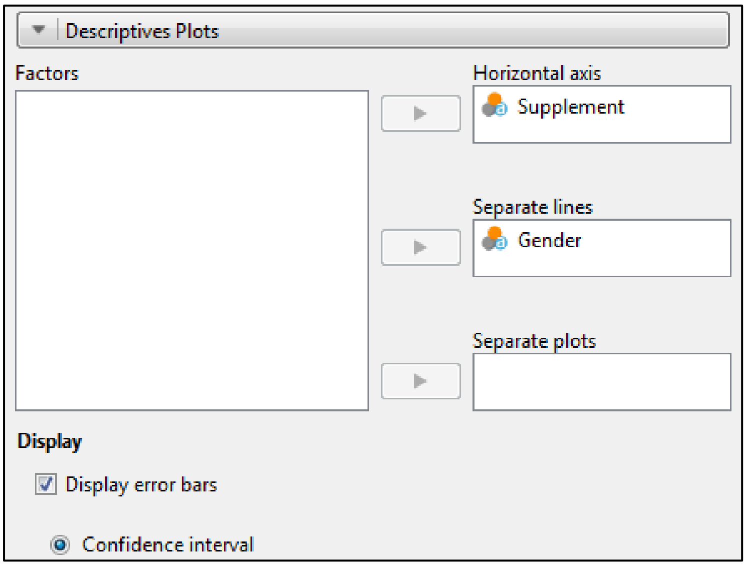
Com la resta de proves paramètriques, l'ANOVA de dos factors de mesures independents realitza una sèrie d'assumpcions que s'haurien d'abordar en el disseny de la investigació o que haurien de ser comprovades.

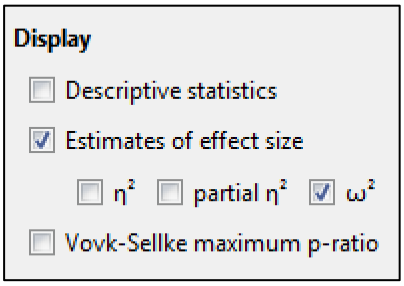
- Les variables independents (factors) haurien de tenir almenys dos grups independents categòrics (nivells).
- La variable dependent hauria de ser contínua i mostrar una distribució aproximadament normal al llarg de totes les combinacions dels factors.
- Hi hauria d'haver homogeneïtat de la variància en cadascuna de les combinacions dels factors.
- No hi hauria d'haver valors atípics significatius.

EXECUTANT L'ANOVA DE DOS FACTORS DE MESURES INDEPENDENTS

Obriu **2-way independent ANOVA.csv** a JASP. Aquest arxiu inclou 3 columnes de dades: Factor 1 – Gender, amb 2 nivells (home i dona); Factor 2 – Supplement, amb 3 nivells (control, carbohidrats CHO i proteïnes) i la variable dependent (Jump power). A «Descriptive statistics», comproveu les dades per si hi hagués valors atípics significatius. Aneu a «ANOVA» → «ANOVA», afegiu Jump power a la caixa «Dependent Variable», i Gender i Supplement a la caixa «Fixed Factors».

A «Descriptives Plots», afegiu Supplement a «Horizontal axis» i Gender a «Separate lines». A «Additional Options», marqueu «Descriptive statistics», «Estimates of effect size» i «ω²».





ENTENENT EL RESULTAT

El resultat hauria de contenir 2 taules i un gràfic.

ANOVA - Jump power

Cases	Sum of Squares	df	Mean Square	F	p	ω²
Gender	119108.037	1.000	119108.037	9.589	0.003	0.058
Supplement	896116.137	2.000	448058.068	36.071	< .001	0.477
Gender ∗ Supplement	275806.438	2.000	137903.219	11.102	< .001	0.138
Residual	521712.054	42.000	12421.716			

Note. Type III Sum of Squares

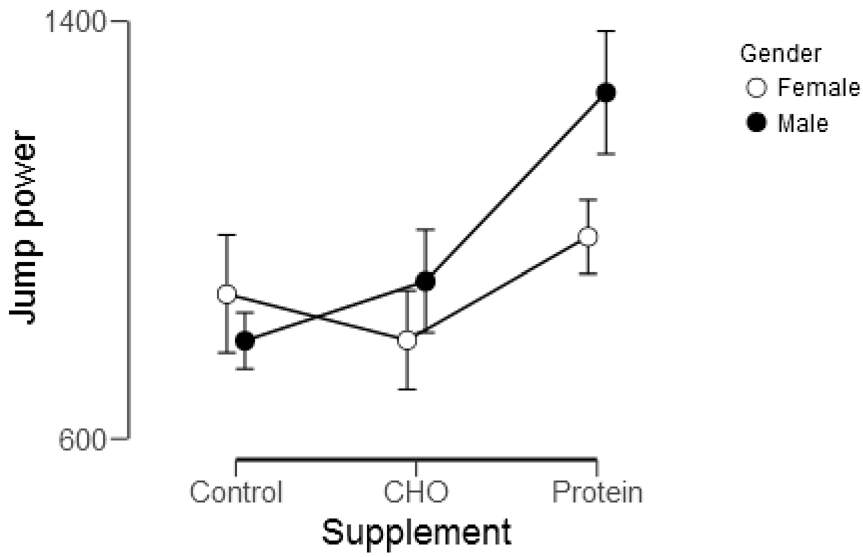
La taula de l'ANOVA mostra que hi ha efectes principals significatius per a Gender i Supplement ($p = 0{,}003$ i $p < 0{,}001$, respectivament), amb mides de l'efecte mitjana i gran, respectivament. Això suggereix que hi ha una diferència significativa entre la potència de salt de cada gènere, amb independència del suplement, i diferències significatives entre suplements, independentment del gènere.

També hi ha una interacció significativa entre Gender i Supplement ($p < 0{,}001$), que també té una mida de l'efecte entre mitjana i gran (0,138). Això suggereix que les diferències en la potència de salt entre els gèneres estan afectades d'alguna manera pel tipus de suplement utilitzat.

L'estadística descriptiva i el gràfic suggereixen que les diferències principals es donen entre gèneres quan s'utilitza un suplement proteic.

Descriptives - Jump power ▼

Gender	Supplement	Mean	SD	N
Female	Control	877.500	134.563	8.000
	CHO	789.286	102.283	7.000
	Protein	986.667	91.924	9.000
Male	Control	788.125	64.417	8.000
	CHO	901.875	117.502	8.000
	Protein	1263.125	140.863	8.000



COMPROVACIÓ DE SUPÒSITS

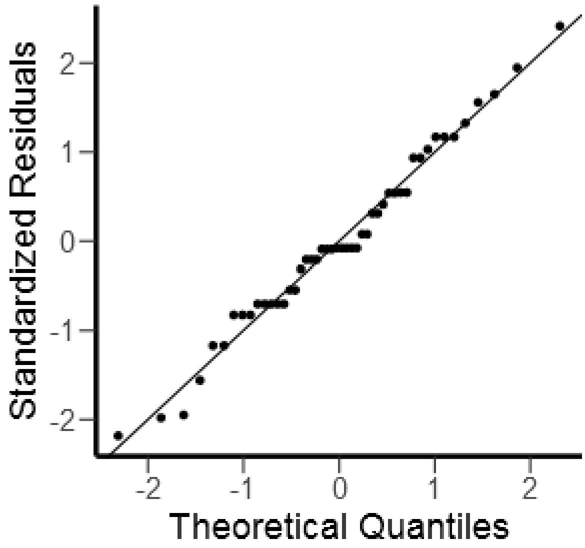
A «Assumption Checks», marqueu «Homogeneity tests» i «Q-Q plot of residuals».

Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
1.100	5.000	42.000	0.375

El test de Levene no mostra diferències significatives de la variància, per la qual cosa l'homogeneïtat de la variància no ha estat violada.



El gràfic Q-Q mostra que les dades semblen estar distribuïdes normalment i ser lineals. Es pot, per tant, acceptar el resultat de l'ANOVA ja que cap d'aquests supòsits no ha estat violat.

No obstant això, si l'ANOVA no mostra cap diferència significativa, no es pot prosseguir amb l'anàlisi.

PROVES POST HOC

Si l'ANOVA és significativa, es pot realitzar una anàlisi post hoc. A «Post Hoc Tests» afegiu Supplement a la caixa d'anàlisi de la dreta, marqueu «Effect size» i, en aquest cas, marqueu «Tukey» per a la correcció post hoc.

El test post hoc no es realitza per a Gender perquè només hi ha 2 nivells.

Post Hoc Comparisons - Supplement

		Mean Difference	SE	t	Cohen's d	p_{tukey}
Control	CHO	−12.768	40.102	−0.318	−0.109	0.946
	Protein	−292.083	38.853	−7.518	−1.919	< .001
CHO	Protein	−279.315	39.561	−7.060	−1.782	< .001

Note. Cohen's d does not correct for multiple comparisons.

El test post hoc no mostra diferències significatives entre el grup de control i el de suplement CHO, independentment del gènere, però sí que mostra diferències significatives entre el grup de control i el de proteïnes ($p < 0,001$) i entre el de CHO i el de proteïnes ($p < 0,001$).

Aneu ara a les opcions «Simple Main Effects». Aquí, afegiu Gender a la caixa «Simple effect factor» i Supplement a la caixa «Moderator factor 1». Efectivament, els efectes principals simples són comparacions dos a dos.

Simple Main Effects - Gender

Level of Supplement	Sum of Squares	df	Mean Square	F	p
Control	31951.563	1	31951.563	2.572	0.116
CHO	47325.030	1	47325.030	3.810	0.058
Protein	323700.184	1	323700.184	26.059	< .001

Aquesta taula mostra que no hi ha diferències de gènere en la potència de salt entre els grups control i CHO ($p = 0,116$ i $p = 0,058$, respectivament). No obstant això, hi ha una diferència significativa ($p < 0,001$) de potència de salt entre gèneres en el grup de suplement proteic.

REPORTANT ELS RESULTATS

Es va utilitzar una ANOVA de dos factors per examinar l'efecte del gènere i el tipus de suplement sobre la potència de salt. Es van trobar efectes principals per als dos gèneres ($F (1, 42) = 9,59$, $p = 0,003$, $\omega^2 = 0,058$) i el suplement ($F (2, 42) = 30,07$, $p < 0,001$, $\omega^2 = 0,477$). Hi va haver una interacció estadísticament significativa entre els efectes del gènere i el suplement en la potència de salt ($F (2, 42) = 11,1$, $p < 0,001$, $\omega^2 = 0,138$).

La correcció post hoc de Tukey va mostrar que la potència de salt va ser significativament superior en el grup de proteïnes comparat amb els grups control i CHO ($t = -1,919$, $p < 0,001$ i $t = -1,782$, $p < 0,001$, respectivament).

Els efectes principals simples van mostrar que la potència de salt va ser significativament més gran entre els homes que entre les dones en el grup dels que van utilitzar un suplement de proteïnes ($F (1) = 28,06$, $p < 0,001$).

ANOVA MIXTA AMB JASP

L'ANOVA mixta (una altra ANOVA de dos factors) és una combinació de l'ANOVA de mesures repetides i la de mesures independents, en la qual es troben involucrades més d'1 variable independent (conegudes com a factors).

Variable independent (factor 2)	*Variable independent (factor 1) = temps o condició*		
	Temps/condició 1	**Temps/condició 2**	**Temps/condició 3**
Grup 1	Variable dependent	Variable dependent	Variable dependent
Grup 2	Variable dependent	Variable dependent	Variable dependent

Els factors estan dividits en nivells, en aquest cas, el factor 1 té 3 nivells i el factor 2 posseeix 2 nivells. Això resulta en 6 combinacions possibles.

Un "efecte principal" és l'efecte d'una de les variables independents sobre la variable dependent, ignorant els efectes de qualsevol altra variable independent. Es posen a prova 2 efectes principals: en aquest cas, la comparació de les dades al llarg del factor 1 (és a dir, el temps) es coneix com a factor "**intra-subjectes**", mentre que la comparació de les diferències en el factor 2 (és a dir, els grups) es denomina factor "**inter-subjectes**". Es dona una interacció quan un factor influeix sobre l'altre factor.

L'efecte principal del temps o la condició (factor intra-subjectes) posa a prova, amb independència del grup al qual pertanyen les dades:

Variable independent (factor 2)	*Variable independent (factor 1) = temps o condició*		
	Temps/condició 1	**Temps/condició 2**	**Temps/condició 3**
Grup 1 *Grup 2*	Totes les dades	Totes les dades	Totes les dades

L'efecte principal del grup (factor inter-subjectes) posa a prova que, amb independència de la condició a la qual pertanyin les dades:

Variable independent (factor 2)	*Variable independent (factor 1) = temps o condició*		
	Temps/condició 1	**Temps/condició 2**	**Temps/condició 3**
Grup 1	Totes les dades		*
Grup 2	Totes les dades		

Els efectes principals simples són, efectivament, comparacions dos a dos:

Variable independent (factor 2)	*Variable independent (factor 1) = temps o condició*		
	Temps/condició 1	**Temps/condició 2**	**Temps/condició 3**
Grup 1	Dades *	Dades *	Dades *
Grup 2	Dades	Dades	Dades

Una ANOVA mixta és una altra "prova òmnibus" (global) emprada per posar a prova 3 hipòtesis nul·les:

1. **No hi ha efecte significatiu intra-subjectes, és a dir, no hi ha diferències significatives entre les mitjanes de les diferències entre totes les condicions / els temps**.
2. **No hi ha efecte significatiu inter-subjectes, és a dir, no hi ha diferències significatives entre les mitjanes dels grups**.
3. **No hi ha efecte d'interacció significatiu, és a dir, no hi ha diferències significatives dels grups a través de les condicions / el temps**.

SUPÒSITS

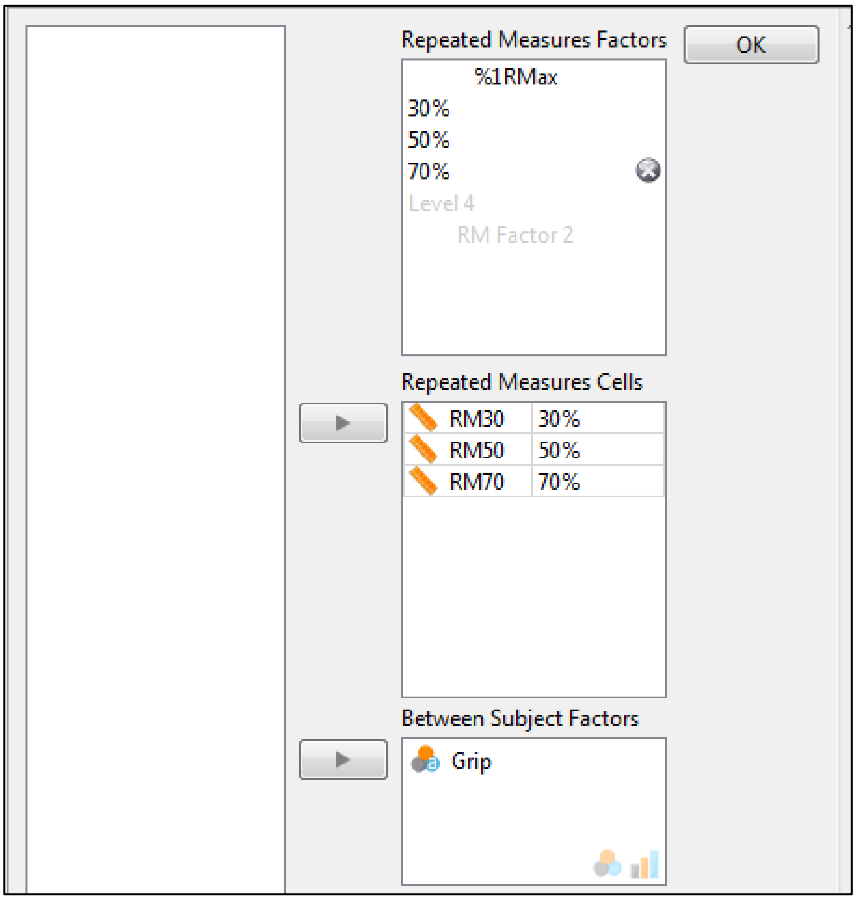
Com la resta de proves paramètriques, l'ANOVA mixta realitza una sèrie de supòsits que s'haurien de tenir en compte en el disseny de la investigació o que es podrien provar.

- El factor "**intra-subjectes**" hauria de contenir almenys dos grups categòrics (nivells) relacionats (mesures repetides).
- El factor "**inter-subjectes**" hauria de contenir almenys dos grups categòrics (nivells) no relacionats (mesures independents).
- La variable independent hauria de ser contínua i tenir una distribució aproximadament normal per a totes les combinacions de factors.
- Hi hauria d'haver homogeneïtat de variància per a cadascun dels grups i, si hi hagués més de 2 nivells, esfericitat entre els grups relacionats.
- No hi hauria d'haver valors atípics significatius.

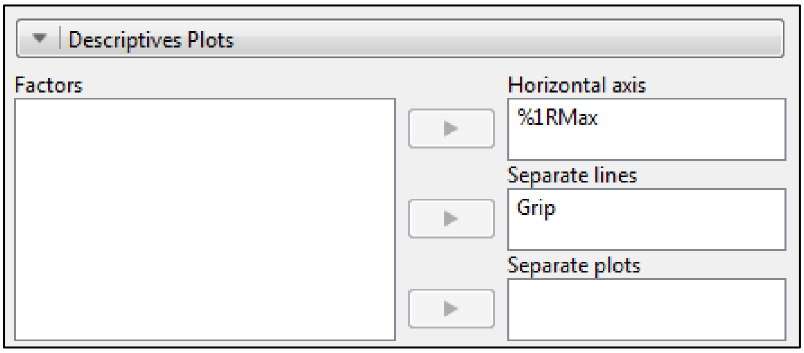
EXECUTANT L'ANOVA MIXTA

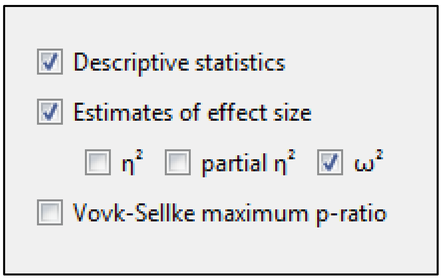
Obriu **2-way Mixed ANOVA.csv** a JASP. Aquest arxiu conté 4 columnes de dades relatives a les empunyadures en l'aixecament de peses i a la velocitat de l'aixecament amb 3 càrregues de pes diferents (%1RM). La columna 1 conté el tipus d'empunyadura, les columnes 2-4 contenen les 3 mesures repetides (30%, 50% i 70%). Comproveu si hi ha valors atípics significatius mitjançant els gràfics de caixa i aneu a «ANOVA» → «Repeated measures ANOVA».

Definiu el factor de mesures repetides introduint %1RMax a la caixa «Repeated Measures Factor» i afegiu 3 nivells (30%, 50% i 70%). Afegiu la variable adient a la caixa «Repeated Measures Cells» i afegiu Grip a la caixa «Between-Subjects Factors»:



A «Descriptive plots», moveu %1RM a l'eix horitzontal i Grip a línies separades. A «Additional Options», marqueu «Descriptive statistics», «Estimates of effect size» i «ω^2».





ENTENENT EL RESULTAT

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	ω^2
%1RM	5.605[a]	2[a]	2.803[a]	115.450[a]	< .001[a]	0.744
%1RM ∗ Grip	0.583[a]	2[a]	0.291[a]	12.003[a]	< .001[a]	0.218
Residual	0.874	36	0.024			

Note. Type III Sum of Squares
[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

El resultat hauria d'incloure 3 taules i un gràfic.

Per a l'efecte principal respecte a %1RM, la taula d'efectes intra-subjectes ("Within Subjects Effects") reporta un estadístic F gran, que és altament significatiu ($p < 0,001$) i a més mostra una mida de l'efecte gran (0,744). Així, independentment del tipus d'empunyadura, hi ha una diferència significativa entre les tres càrregues.

Finalment, hi ha una interacció significativa entre %1RM i Grip ($p < 0,001$), que també mostra una mida de l'efecte gran (0,218). Això suggereix que les diferències entre les càrregues estan d'alguna manera afectades pel tipus d'empunyadura utilitzada.

No obstant això, JASP informa, sota la taula, que el supòsit d'esfericitat ha estat violat. Tractarem l'assumpte en la següent secció.

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	ω^2
Grip	1.095	1	1.095	20.925	< .001	0.499
Residual	0.942	18	0.052			

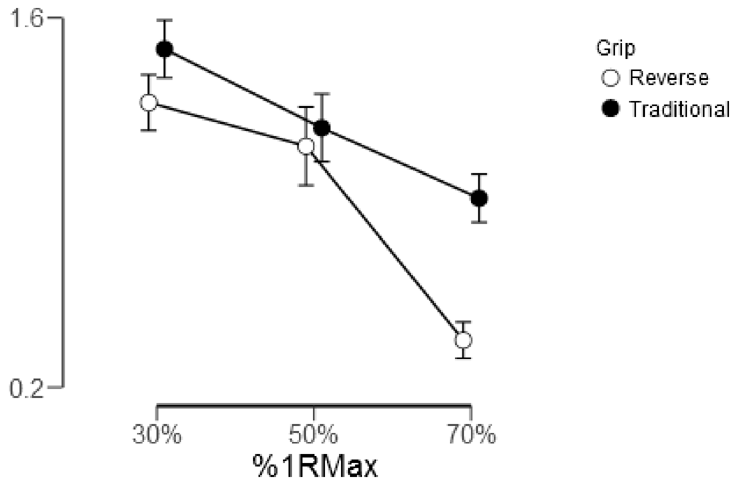
Note. Type III Sum of Squares

Per a l'efecte principal respecte a Grip, la taula d'efectes inter-subjectes ("Between Subjects Effects") mostra una diferència significativa entre els diferents tipus d'empunyadures ($p < 0,001$), amb independència de les càrregues.

A partir de l'estadística descriptiva i del gràfic, sembla que hi ha una diferència més gran entre els dos tipus d'empunyadures amb la càrrega de pes més alta del 70%.

Descriptives

%1RMax	Grip	Mean	SD	N
30%	Reverse	1.279	0.178	10.000
	Traditional	1.482	0.217	10.000
50%	Reverse	1.114	0.198	10.000
	Traditional	1.183	0.256	10.000
70%	Reverse	0.379	0.105	10.000
	Traditional	0.917	0.086	10.000



COMPROVACIÓ DE SUPÒSITS

A «Assumptions Checks», marqueu «Sphericity tests», «Sphericity corrections» i «Homogeneity tests».

Test of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
%1RMax	0.649	0.025	0.740	0.791

La prova d'esfericitat de Mauchly és significativa, per la qual cosa el supòsit ha estat violat. Per tant, s'hauria d'utilitzar la correcció de Greenhouse-Geisser ja que èpsilon és < 0,75. Torneu a «Assumption Checks» i, a «Sphericity corrections», deixeu marcat únicament «Greenhouse-Geisser». Això donarà com a resultat una taula actualitzada d'efectes intra-subjectes ("Within Subjects Effects"):

Within Subjects Effects ▼

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	ω^2
%1RM	Greenhouse-Geisser	5.605 [a]	1.480 [a]	3.787 [a]	115.450 [a]	< .001 [a]	0.744
%1RM ∗ Grip	Greenhouse-Geisser	0.583 [a]	1.480 [a]	0.394 [a]	12.003 [a]	< .001 [a]	0.218
Residual	Greenhouse-Geisser	0.874	26.639	0.033			

Note. Type III Sum of Squares
[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Test for Equality of Variances (Levene's)

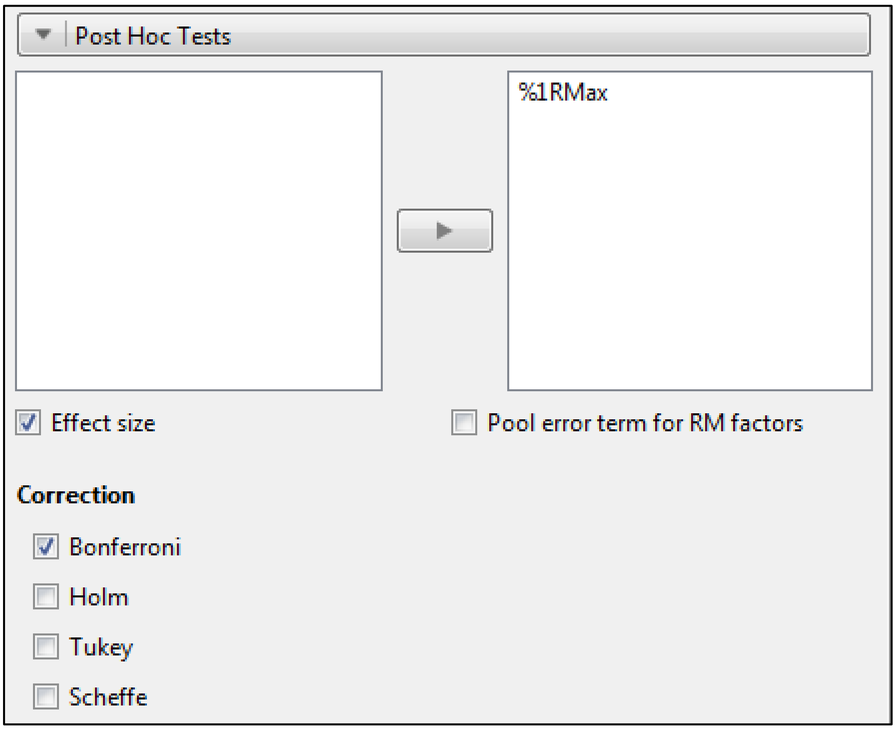
	F	df1	df2	p
RM30	0.523	1.000	18.000	0.479
RM50	0.346	1.000	18.000	0.564
RM70	0.183	1.000	18.000	0.674

La prova de Levene mostra que no hi ha diferència significativa en la variància de la variable dependent a través dels dos tipus d'empunyadura.

No obstant això, si l'ANOVA no reporta diferències significatives, no podeu anar més enllà amb l'anàlisi.

PROVES POST HOC

Si l'ANOVA és significativa, es pot dur a terme l'anàlisi post hoc. A «Post Hoc Tests», afegiu %1RM a la caixa d'anàlisi de la dreta, marqueu «Effect size» i, en aquest cas, utilitzeu Bonferroni per a la correcció post hoc. En l'anàlisi de mesures repetides només hi ha disponibles les correccions de Bonferroni i de Holm.

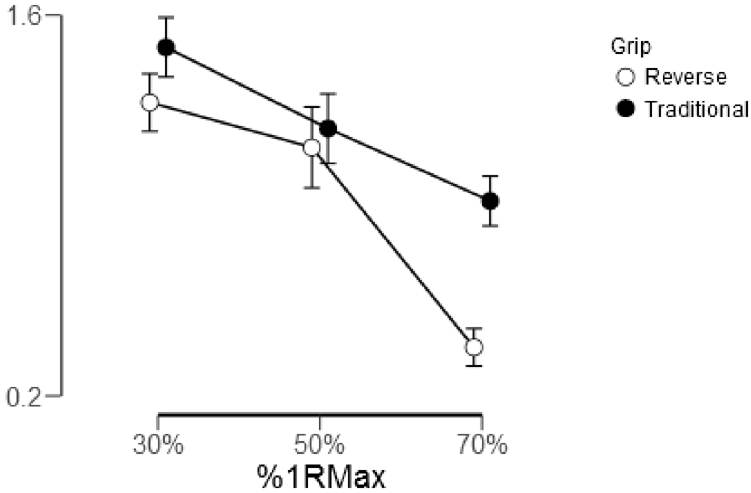


Post Hoc Comparisons - %1RMax

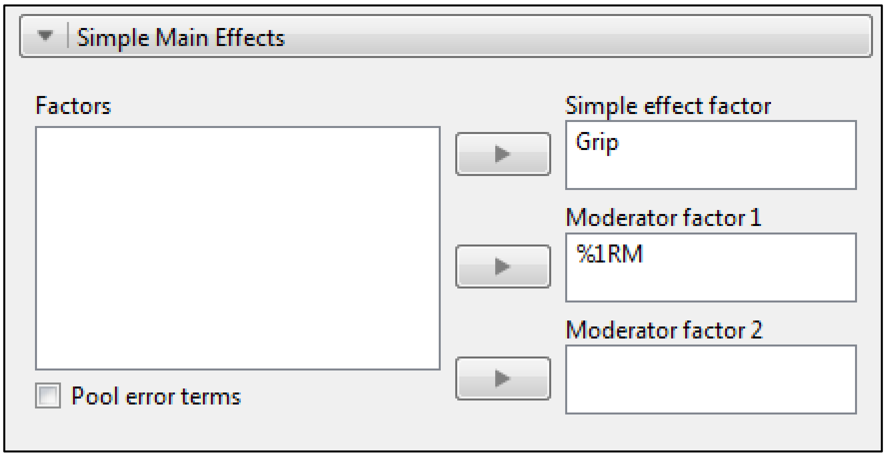
		Mean Difference	SE	t	Cohen's d	p_{bonf}
30%	50%	0.232	0.060	3.856	0.862	0.003
	70%	0.733	0.050	14.583	3.261	< .001
50%	70%	0.500	0.073	6.839	1.529	< .001

Note. Cohen's d does not correct for multiple comparisons.

L'anàlisi post hoc mostra que, amb independència del tipus d'empunyadura utilitzada, cada càrrega de pes és significativament diferent de la resta i, com es veu en el gràfic, la velocitat d'aixecament decreix a mesura que augmenta el pes.



Finalment, a «Simple Main Effects», afegiu Grip a la caixa «Simple effect factor» i %1RM a la caixa «Moderator factor 1».



Simple Main Effects - Grip

Level of %1RM	Sum of Squares	df	Mean Square	F	p
30%	0.206	1	0.206	5.229	0.035
50%	0.024	1	0.024	0.461	0.506
70%	1.447	1	1.447	157.212	< .001

Els resultats mostren que hi ha una diferència significativa en la velocitat d'aixecament al llarg dels dos tipus d'empunyadura en la càrrega inferior del 30%, així com en la càrrega més gran del 70% ($p = 0{,}035$ i $p < 0{,}001$, respectivament).

REPORTANT ELS RESULTATS

Amb la correcció de Greenhouse-Geisser, hi va haver un efecte principal significatiu de la càrrega (F= (1,48, 26,64) = 115,45, p < 0,001). L'anàlisi post hoc amb la correcció de Bonferroni va mostrar una disminució seqüencial significativa en la velocitat d'aixecament entre el 30% i el 50% de càrrega (p = 0,035) i entre el 50% i el 70% de càrrega (p < 0,001).

Hi va haver un efecte principal significatiu per al tipus d'empunyadura (F (1, 18) = 20,925, p < 0,001) mostrant una velocitat global més alta d'aixecament amb l'empunyadura tradicional que amb la reversible.

Utilitzant la correcció de Greenhouse-Geisser, hi va haver una interacció significativa entre la càrrega i el tipus d'empunyadura (F (1,48, 26,64) = 12,00, p < 0,001), mostrant que el tipus d'empunyadura va afectar la velocitat d'aixecament a través de les diferents càrregues.

PROVA DE CHI QUADRAT PER A L'ASSOCIACIÓ

La prova de chi quadrat (χ^2) d'independència (també coneguda com a prova χ^2 de Pearson o prova χ^2 d'associació) es pot utilitzar per determinar si hi ha relació entre dues o més variables categòriques. El test produeix una taula de contingència, o taula de doble entrada, que mostra les agrupacions creuades de les variables categòriques.

El test χ^2 posa a prova la hipòtesi nul·la que no hi ha associació entre dues variables categòriques. Compara les freqüències observades de les dades amb les freqüències que s'haurien d'esperar si no hi hagués relació entre les dues variables.

L'anàlisi requereix complir amb dos supòsits:

1. Les dues variables han de ser categòriques (nominals o ordinals).
2. Cada variable hauria de tenir dos o més grups categòrics independents.

La majoria dels tests estadístics ajusten un model a les dades observades assumint la hipòtesi nul·la que no hi ha diferència entre les dades observades i les modelades (esperades). L'error o la desviació del model es calcula com a:

$$\text{Desviació} = \sum (observat - model)^2$$

La majoria dels models paramètrics es basen en mitjanes i desviacions estàndard poblacionals. El model χ^2, en canvi, es basa en freqüències esperades.

Com es calculen les freqüències esperades? Per exemple, hem categoritzat 100 persones entre homes i dones i entre persones altes i baixes. Si existís una distribució homogènia entre les 4 categories, la freqüència esperada = 100/4 o 25%. No obstant això, les dades reals observades no presenten una distribució de la freqüència homogènia.

Distribució homogènia	Homes	Dones	Total fila
Alt/a	25	25	50
Baix/a	25	25	50
Total columna	50	50	

Distribució observada	Homes	Dones	Total fila
Alt/a	57	24	**81**
Baix/a	14	5	**19**
Total columna	**71**	**29**	

El model basat en els valors esperats es pot calcular de la manera següent:

Model (valors esperats) **= (*total de fila x total de columna*) / 100**

Model – home alt = (81 x 71) / 100 = 57,5
Model – dona alta = (81 x 29) / 100 = 23,5
Model – home baix = (19 x 71) / 100 = 13,5
Model – dona baixa = (19 x 29) / 100 = 5,5

Aquests valors es poden afegir a la taula de contingència:

	Home (H)	Dona (M)	Total fila
Alt/a (A)	57	24	81
Esperat	**57,5**	**23,5**	
Baix/a (B)	14	5	19
Esperat	**13,5**	**5,5**	
Total columna	71	29	

L'estadístic χ^2 es deriva de $\sum \frac{(\mathbf{observat}-\mathbf{esperat})^2}{\mathbf{esperat}}$

Validesa

La prova de χ^2 només és vàlida quan es disposa d'una mida de mostra raonable, és a dir, menys del 20% de les cel·les tenen un valor esperat inferior a 5 i cap d'elles inferior a 1.

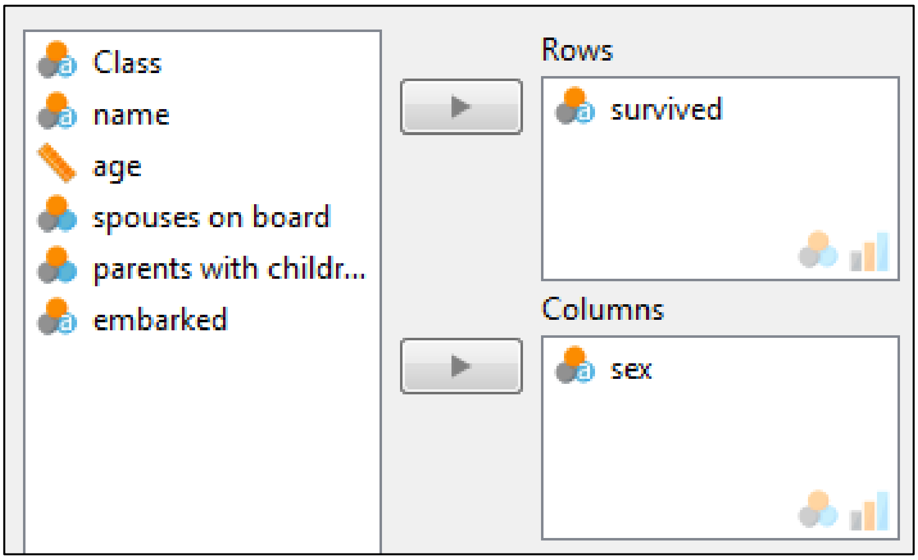
EXECUTANT L'ANÀLISI

El conjunt de dades **Titanic survival** és un conjunt clàssic de dades utilitzat en *machine learning* que conté dades sobre 1.309 passatgers i la tripulació que eren a bord del Titanic quan es va enfonsar el 1912. Podem fer-lo servir per veure les relacions entre la seva supervivència i altres factors. La variable dependent és Survived i les variables independents possibles són la resta de variables disponibles.

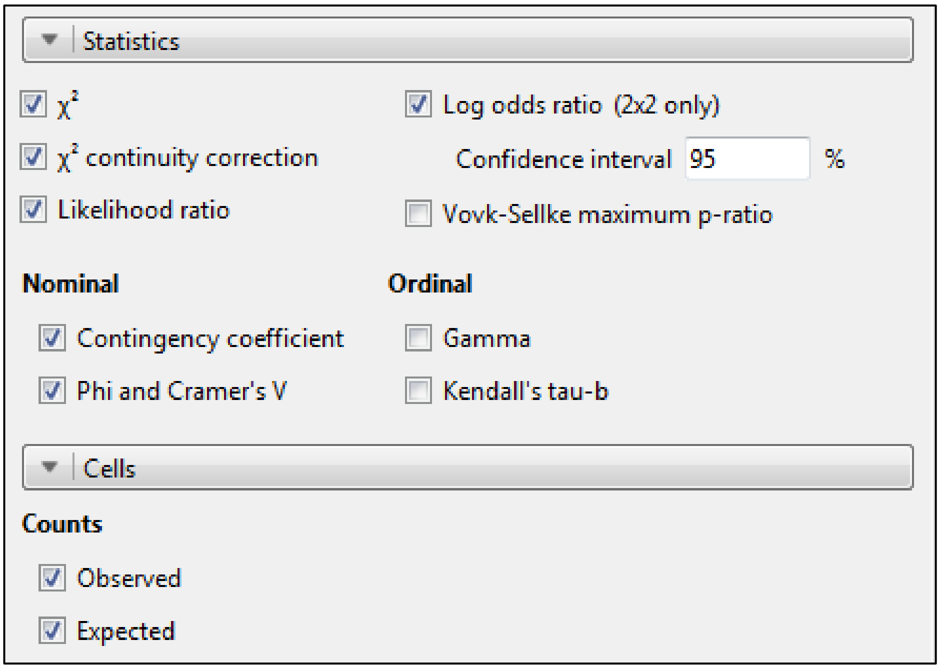
Class	survived	name	sex	age
Third	No	Abbing, Mr. Anthony	male	42
Third	No	Abbott, Master. Eugene Joseph	male	13
Third	No	Abbott, Mr. Rossmore Edward	male	16
Third	Yes	Abbott, Mrs. Stanton (Rosa Hunt)	female	35
Third	Yes	Abelseth, Miss. Karen Marie	female	16
Third	Yes	Abelseth, Mr. Olaus Jorgensen	male	25
Second	No	Abelson, Mr. Samuel	male	30
Second	Yes	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28
Third	Yes	Abrahamsson, Mr. Abraham August Johannes	male	20
Third	Yes	Abrahim, Mrs. Joseph (Sophie Halaut Easu)	female	18
Third	No	Adahl, Mr. Mauritz Nils Martin	male	30
Third	No	Adams, Mr. John	male	26

Per convenció,* la variable independent se sol situar en les columnes de la taula de contingència i la variable dependent en les files.

Obriu **Titanic survival.csv** a JASP, afegiu Survived a la caixa «Rows» (files) com a variable dependent i Sex a la caixa «Colums» (columnes) com a variable independent.



Després, marqueu les següents opcions:



* En realitat, l'anàlisi produeix el mateix resultat independentment de la convenció utilitzada. Alguns autors, de fet, recomanen utilitzar la contrària: les variables independents en les files i les variables dependents en les columnes. En aquest text seguim la convenció utilitzada per l'autor. (Nota del revisor.)

ENTENENT EL RESULTAT

En primer lloc, feu una ullada a la taula de contingència generada.

Contingency Tables

survived		sex: female	sex: male	Total
No	Count	127.0	682.0	809.0
	Expected count	288.0	521.0	809.0
	% within row	15.7 %	84.3 %	100.0 %
	% within column	27.3 %	80.9 %	61.8 %
	% of Total	9.7 %	52.1 %	61.8 %
Yes	Count	339.0	161.0	500.0
	Expected count	178.0	322.0	500.0
	% within row	67.8 %	32.2 %	100.0 %
	% within column	72.7 %	19.1 %	38.2 %
	% of Total	25.9 %	12.3 %	38.2 %
Total	Count	466.0	843.0	1309.0
	Expected count	466.0	843.0	1309.0
	% within row	35.6 %	64.4 %	100.0 %
	% within column	100.0 %	100.0 %	100.0 %
	% of Total	35.6 %	64.4 %	100.0 %

Recordeu que la prova de χ^2 només és vàlida quan disposem d'una mida de mostra raonable, és a dir, menys del 20% de les cel·les amb un valor esperat inferior a 5 i cap d'elles inferior a 1.

A la taula, si ens fixem en el % de fila ("% within row"), podem observar que en el Titanic van morir més homes que dones, i que van sobreviure més dones que homes. No obstant això, hi ha una relació significativa entre el gènere i la supervivència?

Els resultats es mostren aquí:

Chi-Squared Tests

	Value	df	p
X^2	365.9	1	< .001
X^2 continuity correction	363.6	1	< .001
Likelihood ratio	372.9	1	< .001
N	1309		

L'estadístic χ^2 (χ^2 (1) = 365,9, p < 0,001) suggereix que hi ha una relació significativa entre el gènere i la supervivència.

La correcció per continuïtat de χ^2 ("χ^2 continuity correction") es pot utilitzar per prevenir una sobreestimació de la significació estadística en el cas de disposar de conjunts de dades petits. Principalment s'utilitza quan almenys una cel·la de la taula té un valor esperat inferior a 5.

Com a precaució, tingueu en compte que aquesta correcció pot sobrecorregir el resultat de l'anàlisi i resultar massa conservadora, fins al punt que pot no rebutjar la hipòtesi nul·la quan ho hauria de fer (un error Tipus II).

La raó de versemblança ("Likelihood ratio") és una alternativa al chi quadrat de Pearson. Es basa en la teoria de la màxima versemblança. Per a mostres grans, produeix el mateix resultat que l'χ^2 de Pearson. Es recomana especialment per a mostres de mida petita, és a dir, < 30.

En el cas de les variables nominals, el coeficient Phi ("Phi-coefficient"; només per a taules de contingència de 2 x 2) i la V de Cramér ("Cramér's V"; la més popular), són proves de la magnitud de l'associació (és a dir, mides de l'efecte). Ambdós valors es troben en un rang d'entre 0 (no hi ha relació) y 1 (relació perfecta). Es pot veure que la magnitud de la relació entre les variables mostra una mida de l'efecte gran.

Nominal

	Value
Contingency coefficient	0.5
Phi-coefficient	0.5
Cramer's V	0.5

El coeficient de contingència ("Contingency coeficient") produeix un valor ajustat de Phi i només es recomana en el cas de disposar de taules de contingència de grans dimensions, les taules de 5 x 5 o superiors.

Mida de l'efecte[4]	**df**	**Petit**	**Moderat**	**Gran**
Phi i V de Cramér (solo 2 x 2)	1	0,1	0,3	0,5
V de Cramér	2	0,07	0,21	0,35
V de Cramér	3	0,06	0,17	0,29
V de Cramér	4	0,05	0,15	0,25
V de Cramér	5	0,04	0,13	0,22

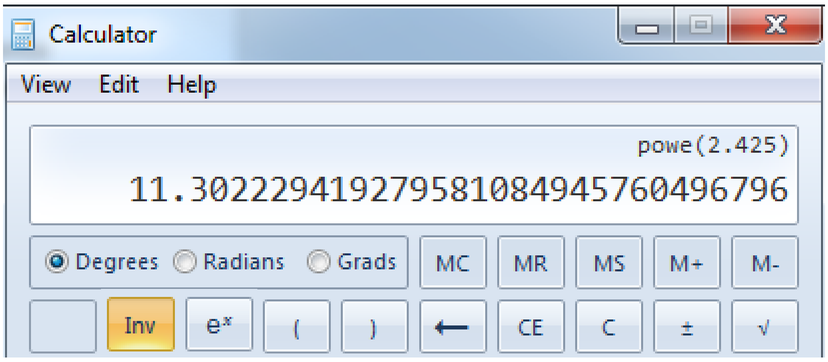
JASP també proporciona la raó de probabilitats (OR, de l'anglès *odds ratio*), utilitzada per a comparar la probabilitat relativa d'ocurrència del resultat d'interès (supervivència), donada l'exposició a la variable d'interès (en aquest cas, el gènere).

[4] Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor. Dent. Endod. 2017; 42(2):152-155.

Log Odds Ratio ▼

	Log Odds Ratio	95% Confidence Intervals Lower	Upper
Odds ratio	−2.425	−2.692	−2.159
Fisher's exact test	−2.423	−2.701	−2.150

Per algun motiu, JASP calcula les OR com un logaritme natural. Per convertir aquets valors, calculeu l'antilogaritme natural (per exemple, utilitzant la calculadora de Windows: introduïu el valor i després cliqueu **Inv** seguit de **e**^{x}), que en aquest cas és 11,3. Això suggereix que els homes van tenir 11,3 vegades més probabilitats de morir que les dones.



Com es calcula? S'han d'utilitzar els valors de la taula de contingència en les fórmules següents:

Probabilitat[homes]	= Van morir / Van sobreviure	= 682 / 162	= 4,209.
Probabilitat[dones]	= Van morir / Van sobreviure	= 127 / 339	= 0,374.

OR = Probabilitat[homes] / Probabilitat[dones] = 11,3.

ANANT UN PAS MÉS ENLLÀ

També es pot descompondre encara més la taula de contingència a manera d'anàlisi post hoc, convertint els recomptes i els recomptes esperats de cada cel·la en un residu estandarditzat. Això pot revelar si les freqüències observades i les freqüències esperades són significativament diferents en cada cel·la.

El residu estandarditzat per a cada cel·la d'una taula és una versió de la puntuació z estandarditzada, calculada com a:

$$z = \frac{\text{observat} - \text{esperat}}{\sqrt{\text{esperat}}}$$

En el cas especial que df = 1, el càlcul del residu estandarditzat inclou un factor de correcció:

$$z = \frac{|\text{observat} - \text{esperat}| - 0{,}5}{\sqrt{\text{esperat}}}$$

El valor resultant de la z té un signe positiu si observat > estimat, i un de negatiu si observat < estimat. Les significacions de les puntuacions z es mostren a continuació.

Puntuació z	Valor p
< -1,96 o > 1,96	< 0,05
< -2,58 o > 2,58	< 0,01
< -3,29 o > 3,29	< 0,001

Dones no z= -9,5	Homes no z = 7,0
Dones sí z = 12,0	Homes sí z = -8,9

Contingency Tables

		sex		
survived		female	male	Total
No	Count	127.0	682.0	809.0
	Expected count	288.0	521.0	809.0
	% within row	15.7 %	84.3 %	100.0 %
	% within column	27.3 %	80.9 %	61.8 %
	% of Total	9.7 %	52.1 %	61.8 %
Yes	Count	339.0	161.0	500.0
	Expected count	178.0	322.0	500.0
	% within row	67.8 %	32.2 %	100.0 %
	% within column	72.7 %	19.1 %	38.2 %
	% of Total	25.9 %	12.3 %	38.2 %
Total	Count	466.0	843.0	1309.0
	Expected count	466.0	843.0	1309.0
	% within row	35.6 %	64.4 %	100.0 %
	% within column	100.0 %	100.0 %	100.0 %
	% of Total	35.6 %	64.4 %	100.0 %

Quan calculem les puntuacions z per a cada cel·la de la taula de contingència, es pot observar que van morir significativament menys dones i més homes de l'esperat (p < 0,001).

DISSENY EXPERIMENTAL I ORGANITZACIÓ DE LES DADES EN EXCEL PER IMPORTAR A JASP

Prova t per a dues mostres independents

Exemple de disseny:

Variable independent	**Grup 1**	**Grup 2**
Variable dependent	Dada	Dada

Variable independent **Variable dependent**

Categòrica **Contínua**

	A	B
1	Group	Data
2	1	0
3	1	0
4	1	3.8
5	1	6
6	1	0.7
7	1	2.9
8	1	2.8
9	1	2
10	1	2
11	1	8.5
12	1	1.9
13	1	3.1
14	1	1.5
15	1	3
16	1	3.6
17	1	0.9
18	1	-2.1
19	2	2
20	2	1.7
21	2	4.3
22	2	7
23	2	0.6
24	2	2.7
25	2	3.6

Si cal, es poden afegir més variables dependents.

Prova t per a dues mostres aparellades

Exemple de disseny:

Variable independent	Pretest	Postest
Participant	**Variable dependent**	
1	Dada	Dada
2	Dada	Dada
3	Dada	Dada
...n	Dada	Dada

Pretest **Postest**

	A	B
1	Pre-test	Post-test
2	60	60
3	103	103
4	58	54
5	60	54
6	64	63
7	64	61
8	65	62
9	66	64
10	67	65
11	69	61
12	70	68
13	70	67
14	72	71
15	72	69
16	72	68
17	82	81
18	58	60
19	58	56
20	59	57
21	61	57
22	62	55
23	63	62
24	63	60
25	63	59

Correlació

Exemple de disseny:

Correlació simple

Participant	**Variable 1**	**Variable 2**	**Variable 3**	**Variable 4**	**Variable ...n**
1	Dada	Dada	Dada	Dada	Dada
2	Dada	Dada	Dada	Dada	Dada
3	Dada	Dada	Dada	Dada	Dada
...n	Dada	Dada	Dada	Dada	Dada

Correlació múltiple

	A	B	C	D	E	F
1	Participant	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
2	1	533	77	77	106	106
3	2	472	63	59	92	93
4	3	484	82	77	93	78
5	4	536	72	72	103	93
6	5	630	77	68	104	93
7	6	563	68	68	101	87
8	7	531	77	82	108	106
9	8	344	50	50	86	92
10	9	346	54	50	90	86
11	10	386	59	54	85	80
12	11	460	54	63	89	83
13	12	492	63	59	92	94

Regressió

Exemple de disseny:

Regressió simple

Participant	Resultat	Predictor 1	Predictor 2	Predictor 3	Predictor ...n
1	Dada	Dada	Dada	Dada	Dada
2	Dada	Dada	Dada	Dada	Dada
3	Dada	Dada	Dada	Dada	Dada
...n	Dada	Dada	Dada	Dada	Dada

Regressió múltiple

	A	B	C	D	E	F
1	Participant	Outcome	Predictor 1	Predictor 2	Predictor 3	Predictor 4
2	1	533	77	77	106	106
3	2	472	63	59	92	93
4	3	484	82	77	93	78
5	4	536	72	72	103	93
6	5	630	77	68	104	93
7	6	563	68	68	101	87
8	7	531	77	82	108	106
9	8	344	50	50	86	92
10	9	346	54	50	90	86
11	10	386	59	54	85	80
12	11	460	54	63	89	83
13	12	492	63	59	92	94

Regressió logística
Exemple de disseny:

	Variable dependent (categòrica)	Factor (categòric)	Covariable (contínua)
Participant	**Resultat**	**Predictor 1**	**Predictor 2**
1	Dada	Dada	Dada
2	Dada	Dada	Dada
3	Dada	Dada	Dada
…n	Dada	Dada	Dada

	A	B	C	D
1	ID	Outcome	Factor	Covariate
2	1	Yes	Yes	70
3	2	Yes	No	80
4	3	Yes	Yes	50
5	4	Yes	No	60
6	5	Yes	No	40
7	6	Yes	No	65
8	7	Yes	No	75
9	8	Yes	No	80
10	9	Yes	No	70
11	10	Yes	No	60
12	11	No	Yes	65
13	12	No	Yes	50
14	13	No	Yes	45
15	14	No	Yes	35
16	15	No	Yes	40
17	16	No	Yes	50
18	17	No	No	55
19	17	Yes	No	65
20	18	No	Yes	45

Si cal, es poden afegir més factors i covariables.

ANOVA de mesures independents d'un factor

Exemple de disseny:

Variable independent	**Grup 1**	**Grup 2**	**Grup 3**	**Grup…n**
Variable dependent	Dada	Dada	Dada	Dada

Variable independent **Variable dependent**

(Categòrica) **(Contínua)**

	A	B
1	Group	Dependent variable
2	Group 1	3.8
3	Group 1	6
4	Group 1	0.7
5	Group 1	2.9
6	Group 1	2.8
7	Group 1	2
8	Group 1	2
9	Group 1	3.5
10	Group 2	1.9
11	Group 2	3.1
12	Group 2	1.5
13	Group 2	3
14	Group 2	3.6
15	Group 2	0.9
16	Group 2	-0.6
17	Group 3	1.1
18	Group 3	4.5
19	Group 3	6.1
20	Group 3	5
21	Group 3	2.4
22	Group 3	3.9
23	Group 3	3.5
24	Group 3	5.1
25	Group 3	3.5

Si cal, es poden afegir més variables dependents.

ANOVA de mesures repetides d'un factor

Exemple de disseny:

	Variable independent (factor)			
Participant	**Nivell 1**	**Nivell 2**	**Nivell 3**	**Nivell ...n**
1	Dada	Dada	Dada	Dada
2	Dada	Dada	Dada	Dada
3	Dada	Dada	Dada	Dada
4	Dada	Dada	Dada	Dada
...n	Dada	Dada	Dada	Dada

Factor (temps)

Nivells

(Grups relacionats)

	A	B	C	D
1	Participant	Week 0	Week 3	Week 6
2	1	6.42	5.83	5.75
3	2	6.76	6.2	6.13
4	3	6.56	5.83	5.71
5	4	4.8	4.27	4.15
6	5	8.43	7.71	7.67
7	6	7.49	7.12	7.05
8	7	8.05	7.25	7.1
9	8	5.05	4.63	4.67
10	9	5.77	5.31	5.33
11	10	3.91	3.7	3.66
12	11	6.77	6.15	5.96
13	12	6.44	5.59	5.64
14	13	6.17	5.56	5.51
15	14	7.67	7.11	6.96
16	15	7.34	6.84	6.82
17	16	6.85	6.4	6.29
18	17	5.13	4.52	4.45
19	18	5.73	5.13	5.17

Si cal, es poden afegir més nivells.

ANOVA de mesures independents de dos factors

Exemple de disseny:

Factor 1	**Suplement 1**			**Suplement 2**		
Factor 2	**Dosi 1**	**Dosi 2**	**Dosi 3**	**Dosi 1**	**Dosi 2**	**Dosi 3**
Variable dependent	Dada	Dada	Dada	Dada	Dada	Dada

Factor 1 **Factor 2** **Variable dependent**

	A	B	C
1	supp	dose	len
2	OJ	1000	19.7
3	OJ	1000	23.3
4	OJ	1000	23.6
5	OJ	1000	26.4
6	OJ	1000	20
7	OJ	1000	25.2
8	OJ	1000	25.8
9	OJ	1000	21.2
10	OJ	1000	14.5
11	OJ	1000	27.3
12	OJ	2000	25.5
13	OJ	2000	26.4
14	OJ	2000	22.4
15	OJ	2000	24.5
16	OJ	2000	24.8
17	OJ	2000	30.9
18	OJ	2000	26.4
19	OJ	2000	27.3
20	OJ	2000	29.4
21	OJ	2000	23
22	VC	1000	16.5
23	VC	1000	16.5
24	VC	1000	15.2
25	VC	1000	17.3

Si cal, es poden afegir més factors i variables dependents.

ANOVA mixta

Exemple de disseny:

Factor 1 (Inter-subjectes)	**Grup 1**			**Grup 2**		
Nivells del factor 2 (Mesures repetides)	**Prova 1**	**Prova 2**	**Prova 3**	**Prova 1**	**Prova 2**	**Prova 3**
1	Dada	Dada	Dada	Dada	Dada	Dada
2	Dada	Dada	Dada	Dada	Dada	Dada
3	Dada	Dada	Dada	Dada	Dada	Dada
...n	Dada	Dada	Dada	Dada	Dada	Dada

Factor 1

(Categòric)

Nivells del factor 2

(Continu)

	A	B	C	D
1	Group	Level 1	Level 2	Level 3
2	Group 1	1.31	0.9	0.9
3	Group 1	1.29	0.89	0.72
4	Group 1	1.8	0.9	0.96
5	Group 1	1.4	1.26	0.97
6	Group 1	1.49	1.18	0.88
7	Group 1	1.35	1.15	0.92
8	Group 1	1.45	1.19	1
9	Group 1	1.21	1.2	0.85
10	Group 1	1.79	1.48	0.99
11	Group 1	1.73	1.68	0.98
12	Group 2	1.55	0.9	0.55
13	Group 2	1.27	0.95	0.41
14	Group 2	1.53	0.87	0.42
15	Group 2	1.26	1.15	0.44
16	Group 2	1.14	1.12	0.38
17	Group 2	1.11	1.08	0.34
18	Group 2	1.1	1.0758	0.18
19	Group 2	1.08	1.18	0.24
20	Group 2	1.3	1.26	0.39
21	Group 2	1.45	1.55	0.44

Chi quadrat: taules de contingència

Exemple de disseny:

Participant	Resposta 1	Resposta 2	Resposta 3	Resposta ...n
1	Dada	Dada	Dada	Dada
2	Dada	Dada	Dada	Dada
3	Dada	Dada	Dada	Dada
...n	Dada	Dada	Dada	Dada

Totes les dades haurien de ser categòriques.

	A	B	C	D	E
1	Respondant	Response 1	Response 2	Response 3	Response 4
2	1	Female	clay	Morning	yes
3	2	Male	astro	Morning	No
4	3	Female	grass	Evening	No
5	4	Male	clay	Afternoon	No
6	5	Male	clay	Morning	No
7	6	Male	grass	Evening	No
8	7	Female	grass	Evening	yes
9	8	Male	clay	Morning	yes
10	9	Female	grass	Morning	No
11	10	Male	clay	Afternoon	No
12	11	Female	clay	Afternoon	No
13	12	Male	astro	Afternoon	No
14	13	Male	astro	Afternoon	No
15	14	Male	astro	Afternoon	yes
16	15	Female	clay	Morning	No
17	16	Male	astro	Afternoon	yes
18	17	Female	astro	Afternoon	yes
19	18	Male	grass	Morning	No
20	19	Male	clay	Afternoon	No

ALGUNS CONCEPTES EN ESTADÍSTICA FREQÜENTISTA

L'aproximació freqüentista és la metodologia estadística més freqüentment ensenyada i utilitzada. Descriu els resultats obtinguts a partir d'una mostra basats en la freqüència o la proporció de les dades a partir d'estudis repetits amb què es defineix la probabilitat dels successos.

L'estadística freqüentista utilitza marcs de referencia rígids que inclouen la prova d'hipòtesi, els valors de p, els intervals de confiança, etc.

Prova d'hipòtesi

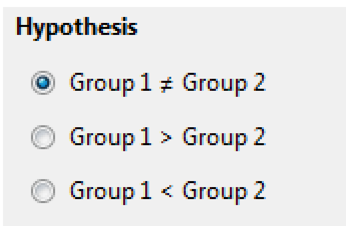
Una hipòtesi pot ser definida com "**una explicació provisional basada en evidències limitades com a punt de partida per a investigacions addicionals**".

Hi ha dos tipus basics d'hipòtesis, una **hipòtesi nul·la** (H_0) i una **hipòtesi alternativa o experimental** (H_1). La **hipòtesi nul·la** és la posició per defecte per a la majoria de les anàlisis estadístiques en les quals s'ha establert que no hi ha relació ni dependència entre grups. La **hipòtesi alternativa** estableix que hi ha una relació o una diferència entre els grups i la direcció d'aquesta diferència o relació. Per exemple, si es porta a terme un estudi per observar els efectes d'un suplement sobre el temps d'esprint en un grup de participants comparat amb un grup placebo:

1) H_0 = **no** hi ha diferències en els temps d'esprint entre els dos grups.
2) H_1 = hi ha diferències en els temps d'esprint entre els dos grups.
3) H_2 = el grup 1 és millor que el grup 2.
4) H_3 = el grup 1 és pitjor que el grup 2.

La prova d'hipòtesis es refereix als procediments estrictament predefinits que s'utilitzen per acceptar o rebutjar les hipòtesis i la probabilitat que pogués ser el resultat de la mera casualitat. La confiança amb què s'accepta o rebutja una hipòtesi nul·la es denomina nivell de significació. El nivell de significació es denota per α, normalment 0,05 (5%). Aquesta es la probabilitat d'acceptar un efecte com a vertader (95%) i que només hi hagi un 5% de probabilitat que el resultat es doni per mera casualitat.

En JASP es poden seleccionar fàcilment diferents tipus d'hipòtesis. No obstant això, la hipòtesi nul·la sempre apareix marcada per defecte.



Errors de Tipus I i Tipus II

La probabilitat de rebutjar la hipòtesi nul·la quan en realitat és vertadera es diu error de Tipus I, mentre que la probabilitat d'acceptar la hipòtesi nul·la quan no és vertadera es coneix com a error de Tipus II.

		La veritat	
		No culpable (H_0)	**Culpable (H_1)**
El veredicte	**Culpable (H_1)**	**Error de Tipus I** Un innocent és empresonat	Decisió correcta
	No culpable (H_0)	Decisió correcta	**Error de Tipus II** Un culpable queda lliure

L'error de Tipus I es considera el pitjor error que es pot cometre en l'anàlisi estadística.

La potència d'una prova estadística es defineix com la probabilitat que el test rebutgi la hipòtesi nul·la quan la hipòtesi alternativa és vertadera. Per a un determinat nivell de significació, si la mida de la mostra augmenta, la probabilitat de cometre errors de Tipus II disminueix, per la qual cosa s'incrementa la potència estadística.

Prova d'hipòtesi

L'essència de la prova d'hipòtesi és definir en primer lloc la **hipòtesi nul·la (o l'alternativa),** establir el nivell de **α**, normalment 0,05 (5%), i recopilar i analitzar dades d'una mostra. Utilitzem un **estadístic** per determinar a quina distància (o el nombre de desviacions estàndard) es troba la mitjana observada en la mostra en relació amb la mitjana de la població establerta en la hipòtesi nul·la. El valor de l'estadístic es compara amb un valor crític. Aquest és un valor de tall que defineix el límit en què es poden obtenir menys del 5% de les mesures de diferents mostres si la hipòtesi nul·la és vertadera.

Si la probabilitat d'obtenir per casualitat una diferència entre les mitjanes és inferior al 5% quan es defineix la hipòtesi nul·la, es pot rebutjar la hipòtesi nul·la i acceptar la hipòtesi alternativa.

El **valor de p** és la probabilitat d'obtenir un resultat en una mostra, suposant que el valor definit en la hipòtesi nul·la és vertader. Si el valor de p és inferior al 5% ($p < 0,05$), es rebutja la hipòtesi nul·la. Quan el valor de p és superior al 5% ($p > 0,05$), acceptem la hipòtesi nul·la.

Mida de l'efecte

La mida de l'efecte és una mesura estàndard que es pot calcular en molts tipus d'anàlisi estadística. Si la hipòtesi nul·la és rebutjada, el resultat és significatiu. Aquesta significació només avalua la probabilitat d'obtenir el resultat en la mostra per casualitat, però no indica com de gran és la diferència (significació pràctica), ni es pot utilitzar per comparar entre diferents estudis.

La mida de l'efecte indica la magnitud de la diferència entre els grups. Així, per exemple, si és donés una disminució significativa dels temps en l'esprint en distancies de 100 m en un grup que pren suplements alimentaris en comparació amb un altre grup placebo, la mida de l'efecte indicaria quant més efectiva va ser la intervenció amb aquests suplements. A continuació, es mostren algunes de les mides de l'efecte més comunes.

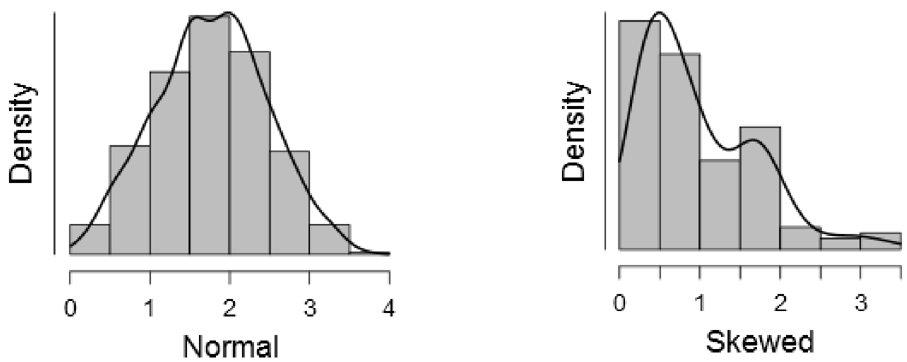
Test	Mesura	Irrellevant	Petit	Mitjà	Gran
Entre mitjanes	d de Cohen	< 0,2	0,2	0,5	0,8
Correlació	Coeficient de correlació (r) Rang biserial (r_B) Rho de Spearman	< 0,1 < 0,1 < 0,1	0,1 0,1 0,1	0,3 0,3 0,3	0,5 0,5 0,5
Regressió múltiple	Coeficient de correlació múltiple (R)	< 0,10	0,1	0,3	0,5
ANOVA	Eta Eta parcial Omega quadrat	< 0,1 < 0,01 < 0,01	0,1 0,01 0,01	0,25 0,06 0,06	0,37 0,14 0,14
Chi quadrat	Phi (només en taules 2x2) V de Cramér Raó de probabilitats (només en taules 2x2)	< 0,1 < 0,1 < 1,5	0,1 0,1 1,5	0,3 0,3 3,5	0,5 0,5 9,0

En conjunts de dades petits, hi pot haver una mida de l'efecte de moderada a gran però no haver-hi diferències significatives. Això pot suggerir que l'anàlisi no va tenir suficient potència estadística i que l'augment del nombre de punts de dades podria mostrar un resultat significatiu. Per contra, quan s'utilitzen conjunts de dades grans, les proves significatives poden ser enganyoses, ja que efectes petits o irrellevants poden produir resultats estadísticament significatius.

PROVA PARAMÈTRICA *vs.* PROVA NO PARAMÈTRICA

La majoria de les investigacions recullen informació a partir d'una mostra de la població d'interès ja que normalment resulta impossible recopilar dades de tota la població. No obstant això, volem saber fins a quin punt les dades recollides reflecteixen adequadament la mitjana, la desviació estàndard, la proporció, etc., de la població basant-nos en la distribució paramètrica d'aquestes funcions. Aquestes mesures són els **paràmetres poblacionals**. Les estimacions d'aquests paràmetres en la mostra són els estadístics. L'estadística paramètrica requereix que s'estableixin supòsits sobre les dades que inclouen la distribució normal i l'homogeneïtat de la variància.

En alguns casos es poden violar aquests supòsits, en el sentit que les dades poden ser notablement asimètriques:

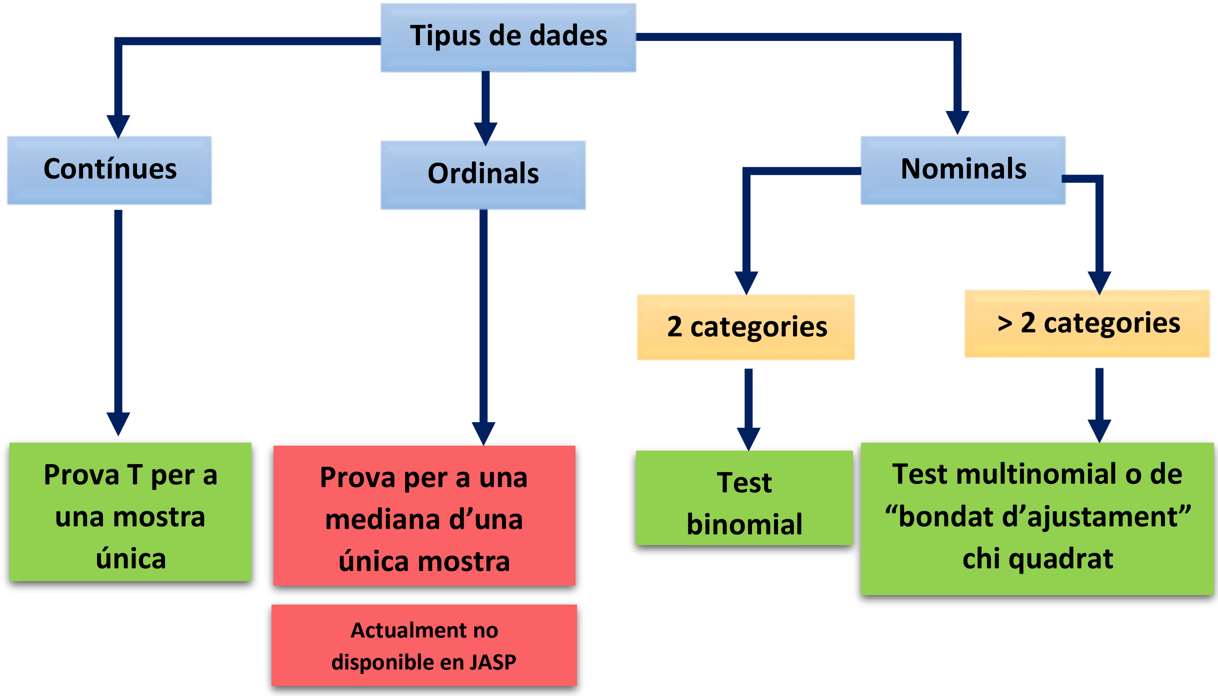


De vegades, transformar les dades serveix per rectificar aquesta situació, però no sempre és així. També és comú recollir dades ordinals (per exemple, puntuacions en escales Likert) per a les quals alguns termes com mitjana i desviació estàndard no tenen sentit. Com a tal, no hi ha paràmetres associats amb dades ordinals (**no paramètriques**). Les alternatives no paramètriques inclouen, entre altres, la mediana i els quartils.

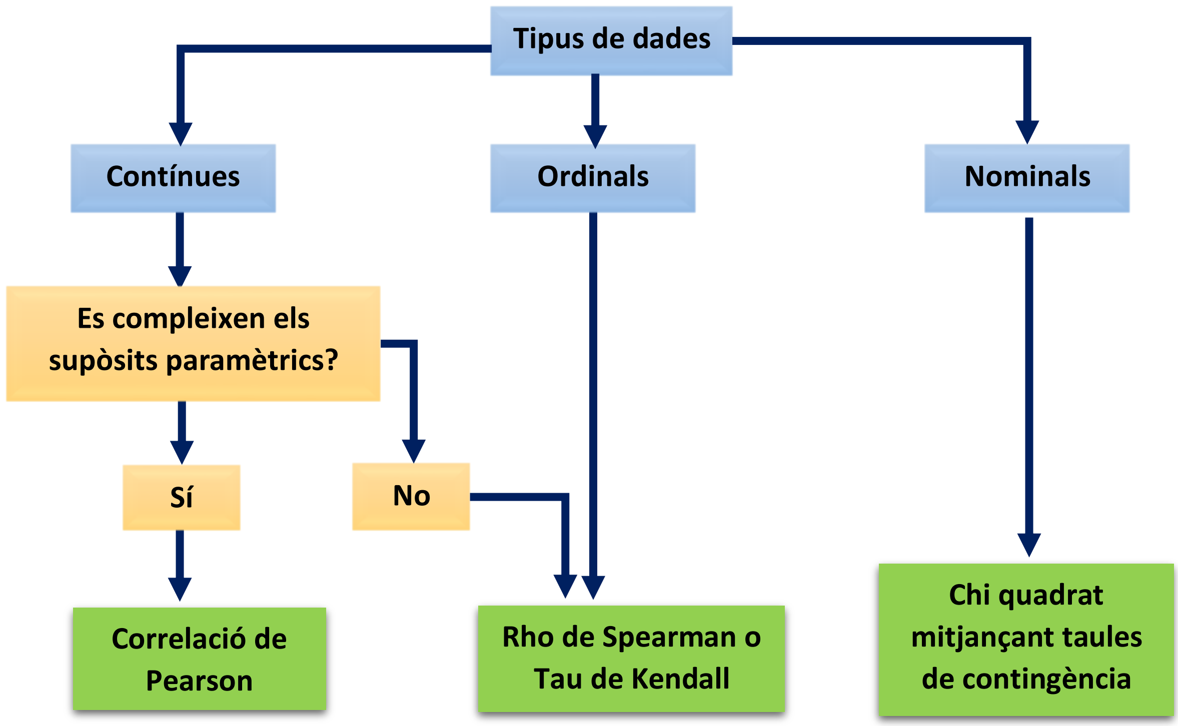
Per a aquests dos casos disposem de proves estadístiques no paramètriques. Hi ha equivalents per a la majoria de les proves paramètriques clàssiques més comunes. Aquestes proves no assumeixen una distribució normal de les dades o l'existència de paràmetres en la població, i es basen en l'ordenació de les dades en rangs, dels valors més baixos als més alts. Tots els càlculs posteriors es realitzen amb aquests rangs en comptes de fer-ho amb els valors de les dades reals.

QUINA PROVA HAURÍEU D'UTILITZAR?

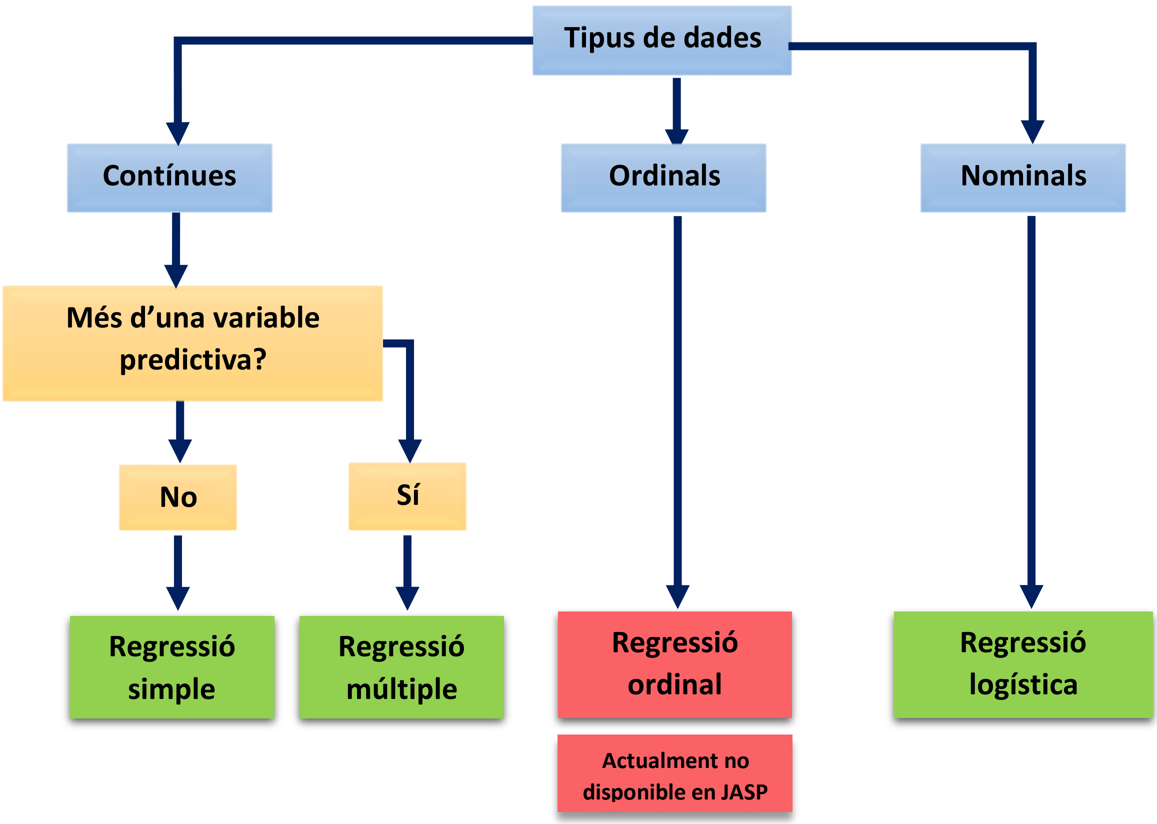
Comparació d'una mesura mostral amb la mitjana coneguda o hipotètica poblacional



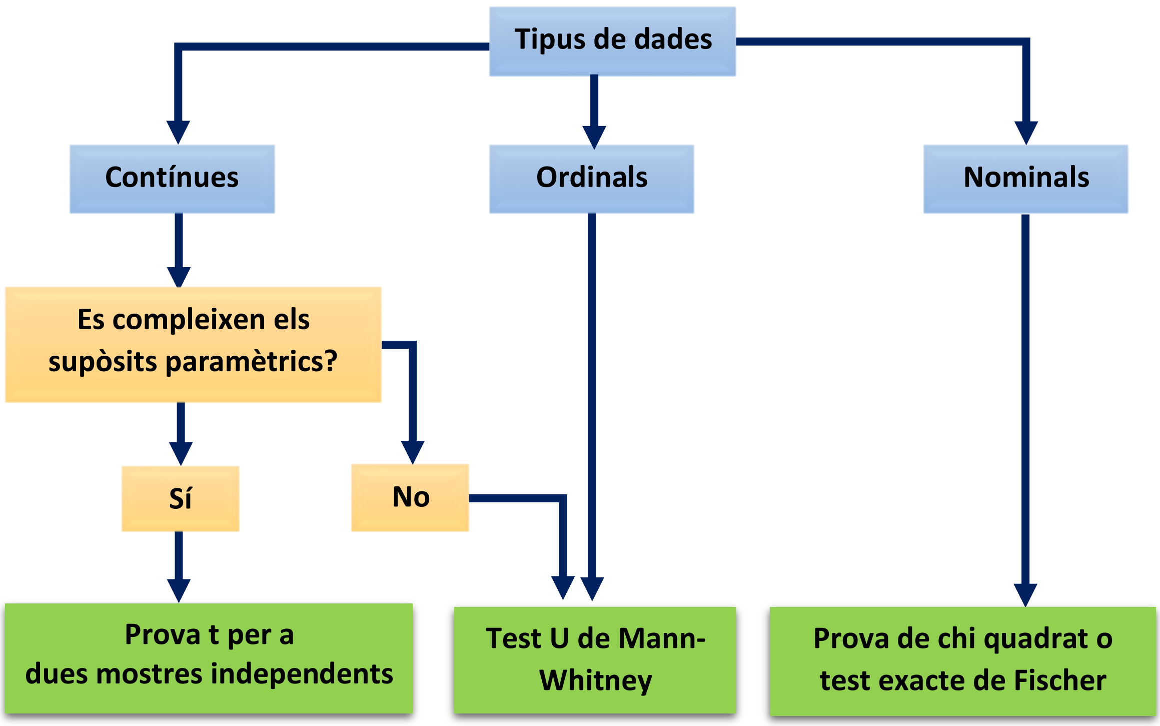
Prova per a la relació entre dues o més variables



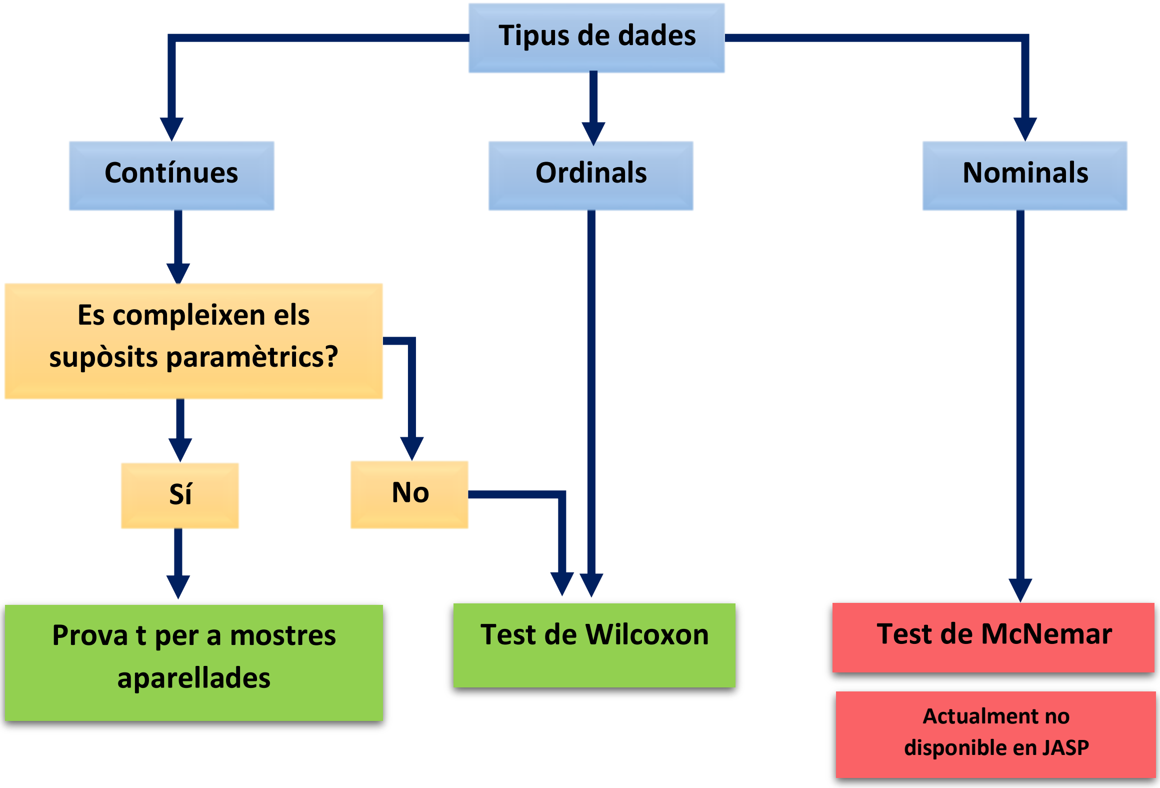
Predicció de resultats



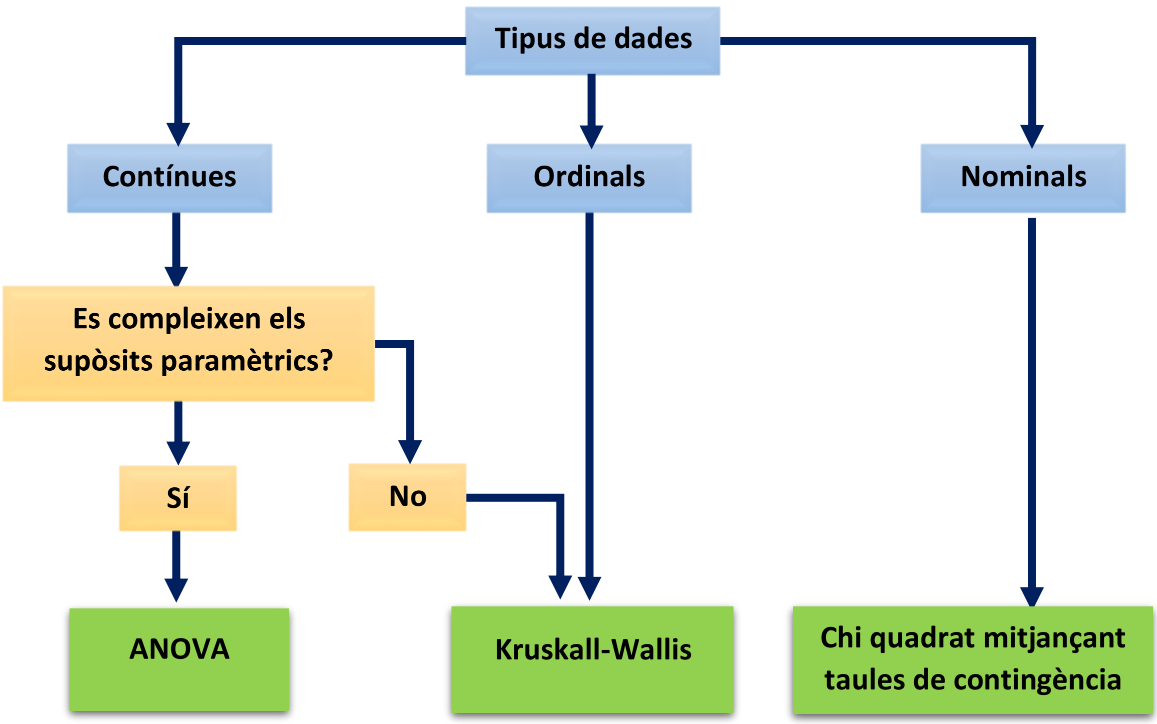
Prova per a les diferències entre dos grups independents



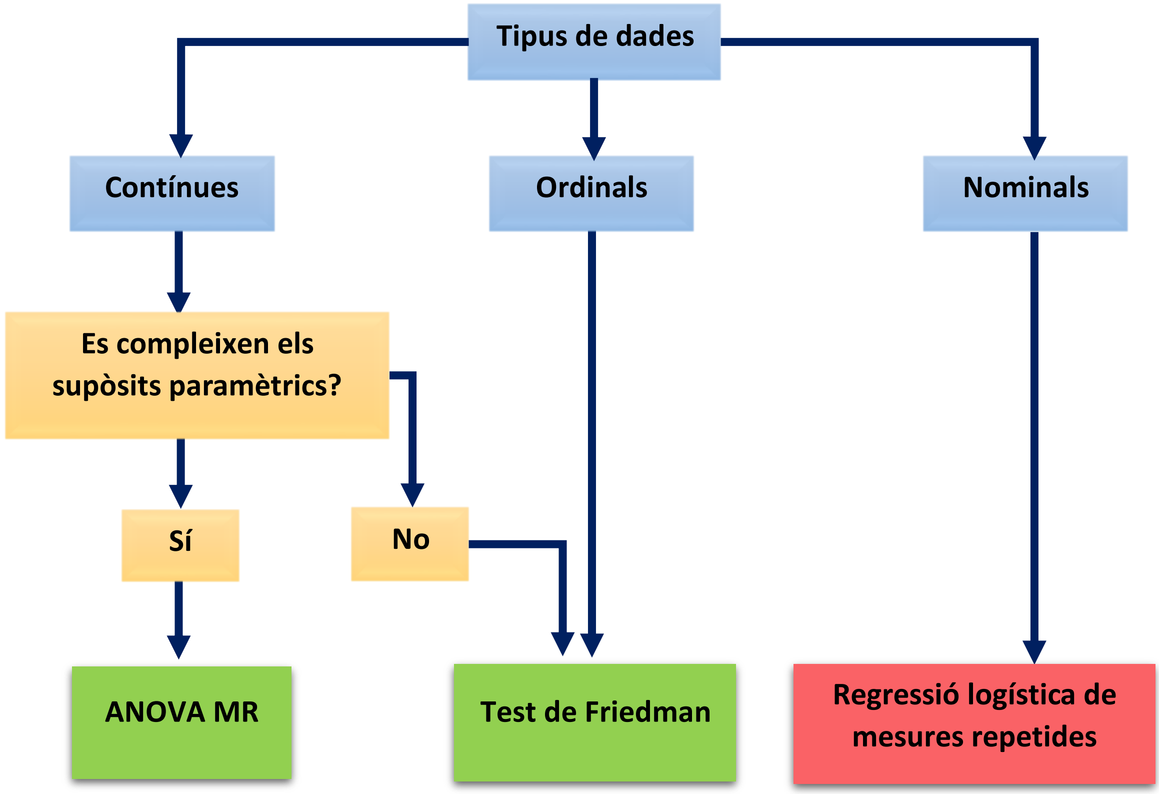
Prova per a dos grups relacionats



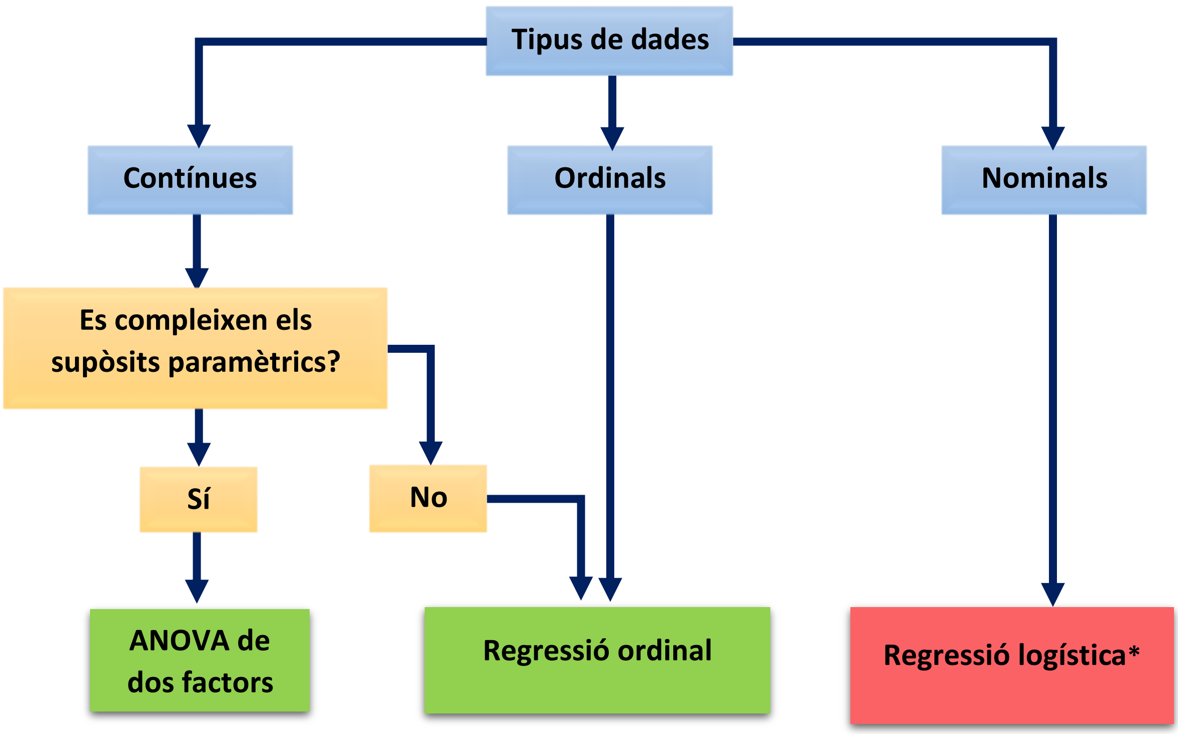
Prova per a les diferències entre tres o més grups independents



Prova per a les diferències entre tres o més grups relacionats



Prova per a interaccions entre dues o més variables independents



* Tot i que apareix com a no disponible en aquest diagrama, la regressió logística és un procediment disponible en la versió 0.9.2 de JASP. (Nota del revisor.)