

Kwantitatieve gegevensverwerking met Jamovi

Onderzoeksskills 2

dr. Willem De Keyzer

2023

Aan het einde van deze module zul je de volgende tools en functies van Jamovi kennen:

- Navigeren door en gebruikmaken van de Jamovi-omgeving
- Variabelen en cases verkennen en maken
- Beschrijvende gegevensverwerking
- Eenvoudige grafieken en boxplots
- Parametrisch testen

Doorheen de module zijn er leercontroles (genummerd Q1, Q2, enz.) die je vragen stellen over wat er zojuist is behandeld. Deze vragen worden niet gesteld ter beoordeling, maar om je te helpen Jamovi volledig te leren gebruiken. De antwoorden op alle vragen staan aan het einde van deze module zodat je kunt controleren of je het goed hebt gedaan en je kunt op de vraagnummers klikken om naar dat antwoord te gaan.

Inhoud

Vertrouwd raken met Jamovi	3
Jamovi op je eigen computer installeren	3
Hulp met Jamovi	4
Download de gegevensbestanden	4
Eerste kennismaking met Jamovi	4
Cases en variabelen	7
Variabelen instellen	8
Variabele gegevenstypen	10
Coding value labels	12
Een nieuwe variabele berekenen	14
Descriptives, plots, and parametric testing	19
Descriptive data	19
Frequency tables	22
Basic plots	24
Histogram	24
Box plots	26
Parametric testing and assumptions	28
Tests of normality (normal distribution)	28
Homogeneity of variance	34
Antwoorden op vragen	37

Vertrouwd raken met Jamovi

In dit gedeelte zullen we vooral wennen aan het gebruik van Jamovi en het instellen van onze gegevens.

Jamovi op je eigen computer installeren

Omdat Jamovi een gratis en open-source applicatie is, kan je deze eenvoudig op je eigen computer installeren. Download het Jamovi-installatieprogramma voor jouw systeem van:

<https://www.jamovi.org/download.html>

Als je de optie krijgt, kies dan de versie 'Solid'. Zodra het installatieprogramma is gedownload, dubbelklik je erop om Jamovi op je computer te installeren. Voor een volledige handleiding en instructies voor het installeren van Jamovi, zie sectie 1 van de Jamovi-gebruikershandleiding hier:

<https://www.jamovi.org/user-manual.html>

Jamovi is een gewone applicatie-installatie voor Windows en Mac, maar Jamovi is ook beschikbaar voor Linux en Chromebooks met enkele extra stappen (zie de bovenstaande gebruikershandleiding voor meer informatie). Als je Jamovi om de een of andere reden niet op je computer kunt installeren, kun je hier een online versie van Jamovi gebruiken:

<https://cloud.jamovi.org/>

Houd er rekening mee dat de online versie van Jamovi slechts een demover-sie is en niet altijd beschikbaar is. Het wordt sterk aanbevolen om Jamovi op je eigen computer te installeren als dat mogelijk is.

Hulp met Jamovi

Jamovi heeft veel online hulpbronnen om je te helpen bij het gebruik! Naast de gebruikershandleiding waarnaar hierboven wordt verwezen, kan je hier een aantal nuttige gidsen vinden op de pagina met hulpmiddelen van de Jamovi-gemeenschap:

<https://www.jamovi.org/community.html>

Er is met name een YouTube-afspeellijst met nuttige Jamovi-videogidsen ([klik hier](#)) en een volledig handboek over statistiek met Jamovi ([klik hier](#)) gelinkt aan de pagina met hulpmiddelen van de community die handig kunnen zijn als je hulp nodig hebt buiten een tutorial om, of als je nog meer details wilt weten over statistiek met Jamovi.

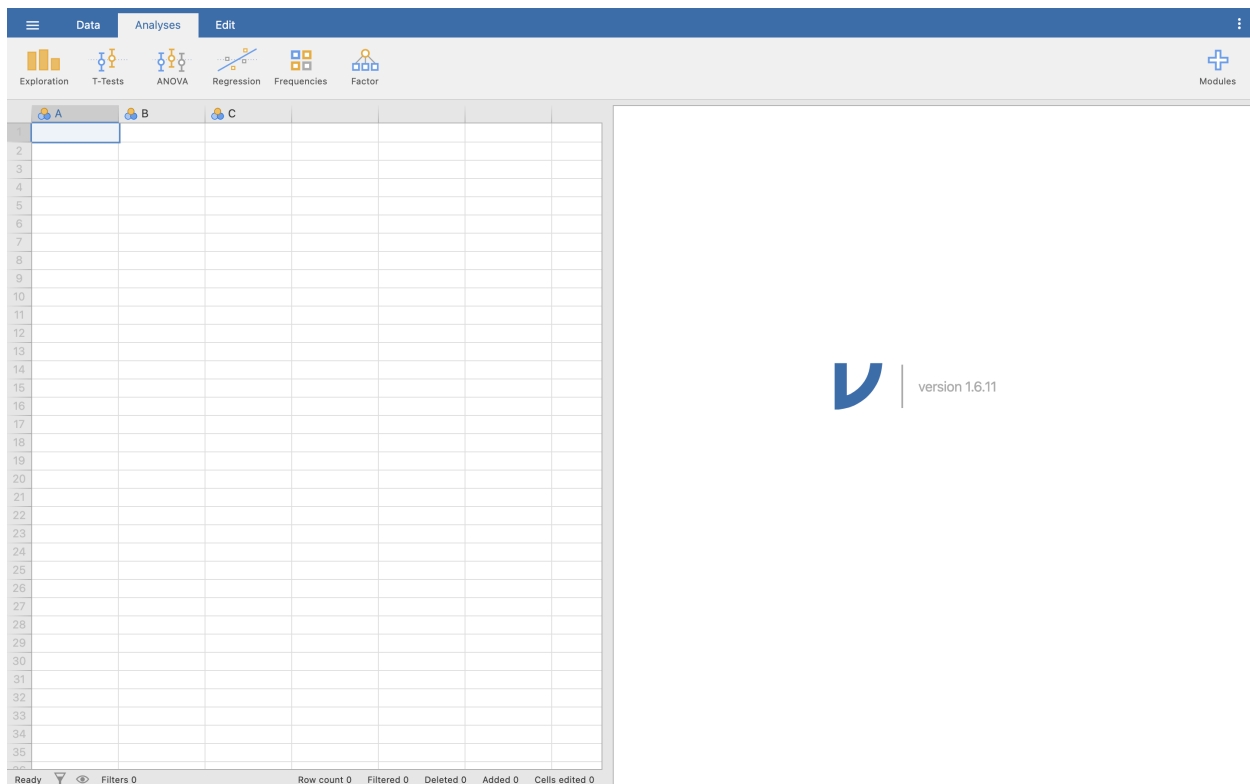
Download de gegevensbestanden

Download alle Jamovi-gegevensbestanden van [GitHub](#) en sla ze ergens op waar je er gemakkelijk bij kunt.

Jamovi gebruikt de '.omv' bestandsextensie, wat een aangepast bestandsformaat is. In tegenstelling tot andere bestandsformaten kun je deze niet dubbelklikken als snelkoppeling om ze in Jamovi te openen - maak je geen zorgen, je moet gewoon eerst Jamovi openen en ze dan vanuit Jamovi openen.

Eerste kennismaking met Jamovi

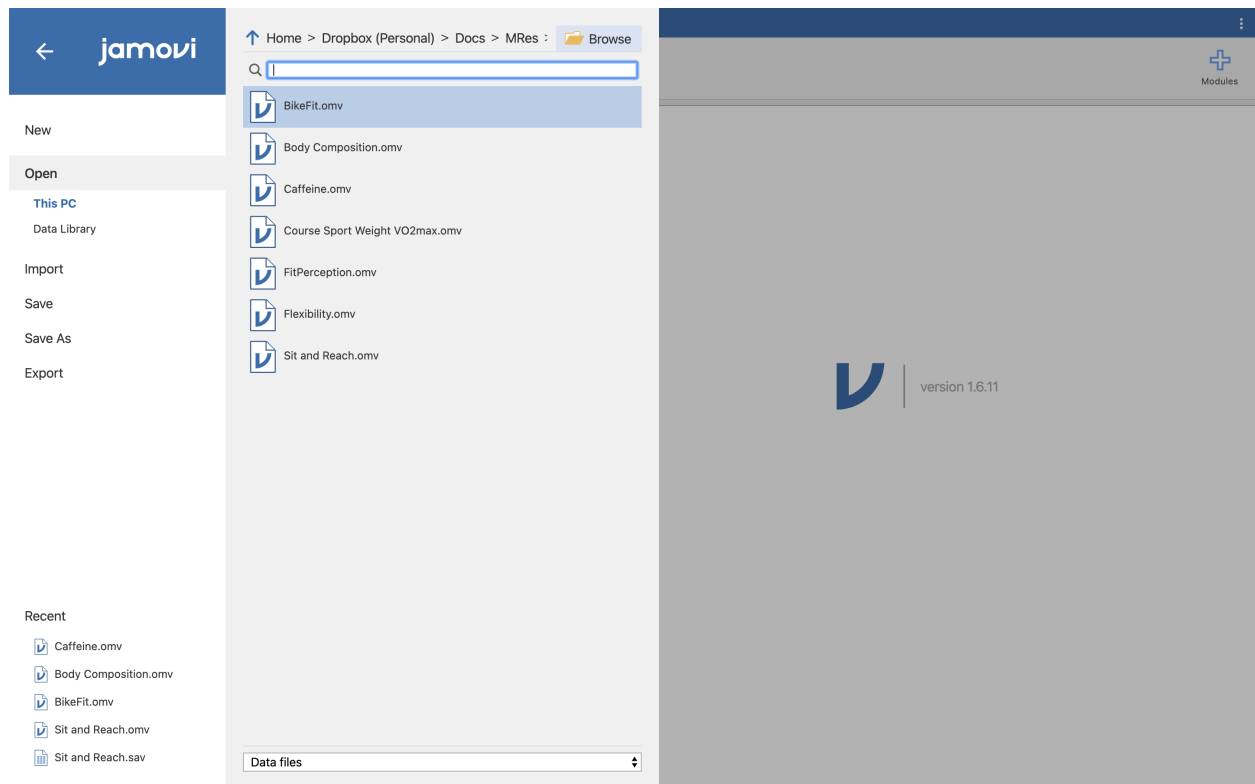
Open na de installatie de Jamovi-toepassing. Je zou een lege Jamovi-werkruimte moeten zien:



De werkruimte in Jamovi is verdeeld in een aantal secties.

- Het grote rasterachtige gebied aan de linkerkant (dat eruit ziet als Excel) is het Spreadsheet. Hier komen je gegevens te staan!
- Het lege gebied aan de rechterkant is de Results Viewer. Dit is een blanco vel papier waarop Jamovi de resultaten afdrukt van alle tests die je uitvoert. Je zult binnenkort zien hoe dit werkt.
- Bovenaan staat het Menu. Dit heeft een paar opties - Data, Analyses en Edit. We zullen vooral Data en Analyses gebruiken. Als je op een van deze opties klikt, verandert het lint direct eronder om te laten zien wat je met elke optie kunt doen - in de schermafbeelding hierboven hebben we Analyses geselecteerd, dus het lint toont ons alle verschillende Analysetests die we kunnen uitvoeren.

Het menu heeft ook drie gestapelde witte lijnen in de linkerhoek - dit opent het File menu. Klik daar nu op, klik dan op Open, vervolgens op Browse en navigeer naar de plaats waar je alle gegevensbestanden van hebt opgeslagen:



Open het bestand Body Composition.omv.

Je zou nu gegevens moeten zien in de Spreadsheet-weergave. Als je meer gegevens wilt zien, kan je de grootte van de Spreadsheet en Results Viewer aanpassen door de dikke grijze lijn in het midden van het scherm aan te klikken en te verslepen.

	PartID	Age	Gender	Course	Weight	Height	BIA1	BIA2	FitPercept
1	1	34	2	Sports Ther...	73.0	176.0	23.6	23.4	Neither fit or...
2	2	21	2	Sports Ther...	79.5	178.0	16.3	16.4	Somewhat ...
3	3	28	2	Sport and E...	72.7	175.0	15.0	14.3	Neither fit or...
4	4	19	2	Sport and D...		184.0	17.7	17.5	Neither fit or...
5	5	21	2	Sports Ther...	78.0	180.5	17.8	17.6	Very unfit
6	6	21	2	Sport Psych...	77.0	168.6	23.5	24.4	Somewhat ...
7	7	33	1	Sports Ther...	51.0	162.0	20.0	20.2	Somewhat fit
8	8	19	1	Sport and E...	52.2	160.0	25.1	24.3	Somewhat fit
9	9	24	2	Sports Ther...	117.9	164.0	38.4	37.9	Very unfit
10	10	19	1	Sports Ther...	44.6	158.0	16.5	17.5	Somewhat ...
11	11	30	2	Sport Psych...	86.0	189.0	9.6	9.9	Somewhat ...
12	13	27	2	Sport and D...	34.0	188.0	27.5	27.9	Very unfit
13	14	22	2	Sports Ther...	68.5	173.6	16.0	15.7	Neither fit or...
14	15	24	1	Sport Psych...	85.0	167.5	29.2	28.7	Neither fit or...
15	16	19	1	Sports Ther...	65.0	171.0			Very fit
16	17	21	1	Sports Ther...	54.0	167.0	19.9	20.1	Somewhat fit
17	18	21	2	Sports Ther...	98.9	188.4	17.9	17.6	Neither fit or...
18	19	19	2	Sport Psych...	81.0	179.3	17.7	17.5	Very unfit
19	20	21	1	Sports Ther...	78.5	173.0	40.0	39.5	Somewhat fit
20	21	28	2	Sport Psych...	67.6	180.6	12.8	12.8	Very unfit
21	22	18	2	Sports Ther...	82.2	167.0	31.8	31.3	Somewhat ...
22	23	19	1	Sport and D...	93.2	166.0	46.6	44.5	Very fit
23	24	25	2	Sport and E...	71.1	175.0	13.1	13.1	Somewhat fit
24	25	19	1	Sports Ther...	49.7	167.0	19.4	19.6	Neither fit or...
25	26	20	2	Sports Ther...	103.0	187.5	20.5	19.7	Somewhat ...
26	27	25	2	Sports Ther...	72.4	177.0			Neither fit or...
27	28	26	2	Sport and E...	66.0	158.0	19.1	19.7	Somewhat fit
28	29	24	1	Sport Psych...	58.6	163.0	25.9	24.4	Neither fit or...
29	30	25	2	Sports Ther...	62.0	170.0	5.8	4.1	Somewhat fit
30	31	18	2	Sports Ther...	107.0	171.0	33.3	34.7	Very fit
31	32	19	2	Sport and E...	95.0	178.0	26.4	26.7	Neither fit or...
32	33	22	1	Sports Ther...	60.1	165.4	26.3	25.6	Somewhat ...
33	34	24	2	Sport Psych...	93.0	183.2	13.5	13.9	Neither fit or...
34	35	18	2	Sports Ther...	63.0	164.0	2.0	9.2	Very unfit
35	36	32	2	Sport and D...	66.6	165.2	15.0	14.6	Neither fit or...

Cases en variabelen

In Jamovi staat elke kolom voor een variabele, net als in Excel, en elke rij voor een case (eenheid). Vergeet niet om persoonlijke gegevens minstens te pseudonimiseren door een anoniem nummer te gebruiken. In dit bestand wordt deze variabele 'PartID' genoemd.

- Q1. Hoeveel kolommen met gegevens zijn er?
- Q2. Hoeveel rijen gegevens zijn er?
- Q3. Waarom een deelnemers-ID als er een recordnummer is?
- Q4. Waarom hebben we hoe dan ook een deelnemers-ID nodig? Waarom gebruiken we niet gewoon de naam van de deelnemer?

Bekijk de kolom PartID - merk op dat sommige ID's ontbreken, zodat het rijnummer

niet altijd overeenkomt met het nummer Participant ID (PartID).

Voeg nog een eenheid toe aan de gegevens, typ de volgende nummers onderaan het blad. Je kunt rechtstreeks in elke cel typen (vergelijkbaar met Excel).

- PartID: 53
- Age: 21
- Gender: 2
- Course: [laat leeg]
- Weight: 41.0
- Height: 160.0
- BIA1: 18.8
- BIA2: 18.8
- FitPercept: [laat leeg]

Variabelen instellen

Tot nu toe gedraagt het gegevensblad zich ongeveer zoals Excel wat betreft wat we kunnen invoeren. Maar in tegenstelling tot Excel kunnen we in Jamovi meer informatie invoeren over wat er in elk van onze variabelen (de kolommen in het gegevensblad) staat. Dit staat bekend als metadata. Om deze te bewerken, klik je op Data en vervolgens op Setup (het pictogram dat eruitziet als een tandwiel bovenop een mini gegevensblad). Er wordt een extra menu geopend bovenaan het Jamovi-scherm:

DATA VARIABLE

PartID

Participant ID

Measure type Continuous

Data type Decimal

Missing values

Levels

Retain unused levels ☐

	PartID	Age	Gender	Course	Weight	Height	BIA1	BIA2	FitPercept
46	48	24	2	Sports Ther...	64.2	164.5	19.7	19.1	Very unfit
47	49	34	1	Sport Psych...	69.3	161.0	33.9	37.0	Neither fit or...
48	50	19	2	Sport and D...	71.0	174.6	20.1	20.9	Very fit
49	51	21	1	Sports Ther...	73.1	177.2	31.8	31.7	Neither fit or...
50	52	24	2	Sports Ther...	64.0	175.0	16.1	15.3	Somewhat ...
51	53	21	2		41.0	160.0	18.8	18.8	
52									
53									
54									
55									
56									
57									
58									
59									
60									
61									
62									
63									
64									
65									
66									

Ready Filters 0

Row count 51 Filtered 0 Deleted 0 Added 1 Cells edited 9

version 1.6.11

Bovenaan zie je in grote tekst de naam van de variabele die we aan het bewerken zijn. Dit is de variabele Label. Klik op een cel in een andere kolom om naar die variabele te gaan - zie hoe de tekst bovenaan het menu voor het instellen van variabelen verandert.

Direct onder het grote Label-gebied staat nog een tekstvak - dit is het Beschrijvingsgebied, waar je wat extra tekst kunt toevoegen om duidelijker te maken wat er in onze variabele staat. Klik op de kolom PartID en zie hoe er nu 'Participant ID' staat - dit is de volledige naam van de variabele, in plaats van het kortere label 'PartID'.

Waarom gebruiken we aparte labels en beschrijvingen? Wanneer je met een variabele werkt, kan het soms vervelend zijn om de volledige naam steeds opnieuw in te typen - 'PartID' is veel sneller in te typen dan 'Deelnemer-ID'. Maar als we onze resultaten willen zien, willen we de volledige beschrijving van de variabele, zodat het makkelijker en duidelijker is om te zien waar onze

resultaten naar verwijzen.

Wijzig nu een paar variabele eigenschappen:

- Je ziet dat de meeste van onze variabelen een Naam en Beschrijving hebben, maar er ontbreken er twee. Voer voor de kolom Age de beschrijving 'Age (years)' in en voer voor de kolom Gender de beschrijving 'Gender' in.

Variabele gegevenstypen

Onder de beschrijving staan nog enkele opties. Klik op het dropdownmenu naast Measure type:

The screenshot shows the 'DATA VARIABLE' editor for a variable named 'PartID' (Participant ID). The 'Measure type' dropdown menu is open, showing four options: Nominal, Ordinal, Continuous (which is highlighted), and ID. The 'Data type' is set to 'Decimal'. The 'Missing values' field is empty. To the right of the dropdown is a 'Levels' column with up and down arrows. At the bottom right, there is a 'Retain unused levels' toggle switch.

Q5. Hoeveel datatypes zijn er beschikbaar in Jamovi?

Jamovi heeft drie hoofdtypen gegevens:

- Nominaal
- Ordinaal
- Continu

Nominaal en Ordinaal zijn precies hetzelfde als de meetniveaus die je elders vindt. Continu' is gewoon de combinatie van Interval en Ratio - Jamovi rekent achter de schermen uit of de gegevens die je invoert Interval of Ratio zijn, op basis van de getallen in de gegevens.

Jamovi heeft ook een extra gegevenstype, genaamd 'ID'. Dit wordt gewoon gebruikt door Jamovi om een Deelnemer ID kolom te identificeren - het is precies hetzelfde als een Nominale kolom, maar laat Jamovi weten dat het zich geen zorgen hoeft te maken over het labelen van niveaus en dat het gewoon moet gebruiken wat er in de cellen wordt getypt. Maak je niet te veel zorgen over het verschil - weet alleen dat je je PartID kolom alleen op 'ID' hoeft in te stellen.

Q6. Waarom zou PartID van het type ID of Nominal zijn, ook al lijkt het een geordende lijst van getallen te zijn?

Klik onder de opties voor het meetniveau op het vervolgkeuzemenu voor Data type:

The screenshot shows the 'DATA VARIABLE' dialog box in Jamovi. The variable name is 'PartID' and the label is 'Participant ID'. The 'Measure type' is set to 'Nominal'. The 'Data type' dropdown menu is open, showing options: 'Integer', 'Decimal' (selected with a checkmark), and 'Text'. The 'Missing value' field is empty. To the right, there is a 'Levels' section with a list box and up/down arrows. At the bottom, there is a toggle switch for 'Retain unused levels' which is currently turned off. A circular arrow icon is in the top right corner, and a right arrow icon is on the right side of the dialog.

Gegevenstypen vertellen Jamovi wat voor soort gegevens we gaan invoeren in de cellen van het gegevensblad. Deze zouden vrij duidelijk moeten zijn:

- Integer: gehele getallen zonder decimaalteken (1, 2, 3, ...)
- Decimal: getallen met een decimaalteken (1.5, 2.3, 3.0, 4.6, ...)
- Text: woorden (Ja, Nee, John Smith, voetbal, ...)

We zullen het gedeelte 'Missing values' voorlopig negeren.

Codering value labels


Tot slot staat rechtsonder in de variabelenviewer de instelling Levels. Hier kunnen we, als we categorische gegevens gebruiken (d.w.z. nominale of ordinale gegevenstypen), de labels instellen die moeten verschijnen voor elk niveau van die categorie.

Klik bijvoorbeeld op de kolom Gender terwijl het venster Variable setup geopend staat. Merk op hoe de gegevens worden ingevoerd als de getallen 1 en 2, in plaats van als tekst. Verander het Measure type in 'Nominal' en je zou de getallen 1 en 2 moeten zien verschijnen in het vak Levels.

DATA VARIABLE

Gender

Description

Measure type  Nominal

Data type Text

Missing values

Levels
1
2

Retain unused levels ☐

Klik op het getal 1 in het vak Levels om het label voor deze waarde in te voeren. Verander dit van een '1' in het woord 'Male'. Klik op het nummer 2 en verander het in

'Female'.

The screenshot shows the 'DATA VARIABLE' settings for a variable named 'Gender'. The 'Measure type' is set to 'Nominal' and the 'Data type' is set to 'Text'. The 'Missing values' field is empty. A 'Levels' dialog box is open, showing a list of levels: 'Male' (level 1) and 'Female' (level 2). The 'Female' level is currently selected and highlighted in blue. To the right of the list are up and down arrow icons. At the bottom of the dialog, there is a toggle switch for 'Retain unused levels' which is currently turned off.

Merk op hoe de nummers 1 en 2 naar rechtsonder de tekst die je typt verschuiven, zodat je nog steeds kunt zien welk label naar welk nummer verwijst. Klik ergens buiten het labelsvak om deze labels op te slaan.

Merk nu op dat de cellen in de kolom Gender in het gegevensblad veranderen van '1' en '2' in 'Male' en 'Female'. De onderliggende gegevens zijn nog steeds hetzelfde - we hebben ze alleen gelabeld.

Waarom getallen en labels gebruiken voor nominale gegevens? Net als bij de namen en labels van onze variabelen, maakt het het makkelijker om met de gegevens te werken en toch onze resultaten duidelijk af te lezen. Als we bijvoorbeeld alleen de mannelijke deelnemers aan dit onderzoek willen selecteren, kunnen we filteren op 'Gender = 1', maar onze resultaten zouden nog steeds het volledige label 'Male' bevatten.

Ga nu door en controleer of alle variabelen in ons gegevensblad correct zijn

ingesteld. Zorg ervoor dat de variabelen overeenkomen met de volgende:

- Age: Age (years), Continuous, Integer
- Gender: Gender, Nominal, Integer, with 2 levels
- Course: Course Name, Nominal, Integer, with 4 levels
- Weight: Weight (kg), Continuous, Decimal
- Height: Height (cm), Continuous, Decimal
- BIA1: BIA Machine 1 Body Fat (%), Continuous, Decimal
- BIA2: BIA Machine 2 Body Fat (%), Continuous, Decimal
- FitPercept: Fitness Perception, Nominal, Integer, with 5 levels

Als je al deze variabelen hebt ingesteld, klik je op het cirkelpictogram met een pijl omhoog rechtsboven in het menu voor het instellen van variabelen om het menu te verbergen en terug te keren naar de hoofdweergave.

Zorg ervoor dat je je gegevens regelmatig opslaat - klik op de drie witte lijnen linksboven om het menu File weer te openen en klik vervolgens op Save.

Gebruik Ctrl + S in Windows of Cmd + S op Mac als sneltoets om je werk op te slaan. Zorg ervoor dat je regelmatig opslaat om te voorkomen dat je iets kwijtraakt! Al je gegevens, tests en resultaten worden opgeslagen in hetzelfde .omv-bestand, zodat je het later weer kunt openen en verder kunt gaan waar je gebleven was.

Een nieuwe variabele berekenen

We ontbreken nog een BMI-variabele, dus gaan we er een maken. Je zou alle gegevens in Excel kunnen maken en ze dan kopiëren en plakken in Jamovi, maar Jamovi heeft een functie Compute variable die de berekening voor ons kan uitvoeren.

Klik op de kolom FitPercept, klik vervolgens op Data en vervolgens op Compute (het

pictogram dat eruitziet als een rekenmachine). Dit voegt een nieuwe kolom toe na de kolom waar onze cursor staat (daarom hebben we eerst op FitPercept geklikt) en opent het menu Computed variabele bovenaan, vergelijkbaar met het menu voor het instellen van variabelen:

The screenshot shows the Jamovi software interface. At the top, there's a menu bar with 'Data', 'Analyses', and 'Edit'. Below it is a toolbar with icons for 'Paste', 'Edit', 'Setup', 'Compute', 'Transform', 'Add', 'Delete', 'Filters', and 'Rows'. The main area displays the 'COMPUTED VARIABLE' dialog box. The 'Name' field contains 'J', the 'Description' field is empty, and the 'Formula' field contains '= eg: ROW() <= 100'. Below the dialog, a data table is visible with columns: ID, Age, Gender, Course, Weight, Height, BIA1, BIA2, FitPercept, and J. The 'J' column is highlighted. The bottom status bar shows 'Row count 51', 'Filtered 0', 'Deleted 0', 'Added 1', and 'Cells edited 10'.

Jamovi geeft onze variabele automatisch een letter van het alfabet als naam - 'j' in dit geval. Verander de nieuwe variabelenaam in BMI en voer vervolgens Body Mass Index in als beschrijving.

Onder de naam en beschrijving staat het vak formule box. Hier voeren we een formule in om Jamovi te vertellen wat hij voor ons moet berekenen - heel vergelijkbaar met wanneer we functies in Excel typen, maar in plaats van ze in één cel te typen en de formule in alle andere cellen te kopiëren en plakken, voeren we in Jamovi gewoon de functie bovenaan in en wordt deze automatisch voor de hele kolom voor ons ingevuld.

We kunnen direct in het formulevak typen of Jamovi sommige delen voor ons laten invullen. Klik op de fx knop om alle opties te zien die we in het formulevak kunnen

invoeren:

COMPUTED VARIABLE

BMI

Body Mass Index

Formula

f_x = eg: ROW() <= 100

Functions	Variables
MAX	PartID
MAXABSIQR	Age
MAXABSZ	Gender
MEAN	Course
MIN	Weight
RANK	Height

MEAN(number 1, number 2, ..., ignore_missing=0, min_valid=0)
Returns the mean of a set of numbers.

Height	BIA1
176.0	
178.0	
175.0	15.0

15.0 14.3 Neither fit or...

Jamovi vertelt ons wat elke functie doet als we er één keer op klikken - zie hier dat ik op de MEAN-functie heb geklikt, en Jamovi geeft er een korte beschrijving onder. Naast de functies staan onze variabelen. Dubbelklik op deze variabelen om ze in te voegen in het functievak.

Typ of kopieer en plak nu het volgende in het functievak:

`Weight / ((Height/100) * (Height / 100))`

COMPUTED VARIABLE

BMI

Body Mass Index

Formula

f_x = Weight / ((Height/100) * (Height / 100))

Retain unused levels ☐

Druk op enter en je zou moeten zien dat de BMI kolom automatisch wordt ingevuld met onze BMI waarden. Merk op dat alle cases (rijen) die geen gewichts- of lengtegegevens hebben, ook leeg blijven in de nieuwe BMI-kolom.

Even samenvatten: we hebben gekeken naar het gegevensblad dat onze gegevens (cellen) bevat gerangschikt als Variabelen (kolommen) en Cases (rijen). We hebben in het menu Variable Setup gekeken naar de metadata-informatie over onze variabelen en de waarden die ze bevatten. Dit zijn de twee hoofdcomponenten van wat we gegevens zouden noemen.

De variabelen en de gegevens worden allemaal opgeslagen in één Jamovi-gegevensbestand (.omv). Zorg ervoor dat je je gegevensbestand opslaat wanneer je wijzigingen hebt aangebracht in zowel de variabelen als de cases. We gaan nu verder met het onderzoeken van enkele van onze gegevens.

Gebruik het File menu van eerder om je werk op te slaan - klik op de drie gestapelde witte lijnen in de hoek en kies Save of Save as. U kunt ook de snel-

toetsen Ctrl + S (Windows) of Cmd + S (Mac) gebruiken om op te slaan.

Descriptives, plots, and parametric testing

Now that we have our data setup in Jamovi, we can start investigating it with some descriptives, some basic plots, and some parametric testing.

Descriptive data

In the Excel class we took a lot of time looking at descriptive data. This is information about your data, for instance, the total number of records, the minimum, maximum, range, standard deviation, etc. Jamovi can provide this information a lot quicker than us calculating it ourselves each time. Here we are going to use some the Exploration analyses to help describe our data.

In the Analyses menu, click on the Explore icon (that looks like a bar graph), then click Descriptives in the dropdown menu that appears. A new middle area should open up between our Spreadsheet and Results Viewer:

The screenshot shows the Jamovi software interface. The top menu bar includes 'Data', 'Analyses', and 'Edit'. The 'Analyses' menu is open, showing options like 'Exploration', 'T-Tests', 'ANOVA', 'Regression', 'Frequencies', and 'Factor'. The 'Exploration' option is selected, and a dropdown menu shows 'Descriptives'. The main workspace is divided into three panels: a 'Spreadsheet' on the left, a 'Descriptives' panel in the middle, and a 'Results' panel on the right. The 'Spreadsheet' panel shows a table with columns 'ID', 'Age', 'Gender', and 'Course'. The 'Descriptives' panel has a search bar and a list of variables: 'PartID', 'Age', 'Gender', 'Course', 'Weight', 'Height', 'BIA1', 'BIA2', and 'FitPercent'. The 'Results' panel shows the 'Descriptives' output, including N, Missing, Mean, Median, Standard deviation, Minimum, and Maximum.

Notice that something has also showed up on our Results Viewer! This is how

Jamovi works - when you run an Analyses, the results are printed out onto the viewer on the right. The Descriptives Analyses prints out a table of information by default, but at the moment, we haven't told Jamovi exactly what we want Descriptives of yet, so the table is empty.

Click on the Weight variable in the left hand box, and click the arrow to move it into the Variables box on the right. Then, click on the Gender variable, and use the arrow to move it across to the Split By section:

The screenshot shows the Jamovi software interface. On the left, a data table is visible with columns for ID, Age, and Gender. The 'Analyses' menu is open, and the 'Descriptives' option is selected. In the 'Descriptives' panel, the 'Variables' box contains 'Weight' and the 'Split by' box contains 'Gender'. The 'Results' panel on the right displays the 'Descriptives' table, which is automatically updated with the following data:

	Gender	Weight
N	Male	17
	Female	33
Missing	Male	0
	Female	1
Mean	Male	64.2
	Female	76.9
Median	Male	60.1
	Female	72.7
Standard deviation	Male	13.8
	Female	18.1
Minimum	Male	44.6
	Female	34.0
Maximum	Male	93.2
	Female	118

This is a simple way of looking at all the weights of our participants (dependent list) broken down by gender (factor list).

Notice our Results table on the right has now automatically updated to show the Descriptives we have asked for! You should be able to see the mean, median, standard deviation, minimum, and maximum for both Male and Female groups. All of these require separate equations in Excel, so you can see Jamovi is a lot quicker to summarise data.

- Q7. In your descriptive data who, on average is heavier, male or female?
- Q8. How many males and females are there? And how big is our total sample size?

Let's look at some more of our data. Previously I mentioned we had two columns, BIA1 and BIA2, where we took the body fat percentage using two 'identical' machines. Here we can briefly examine these two fields to compare any differences.

Click on Weight in the Variables area and notice the arrow icon changes to point left - click this to put it back into our variables list. Do the same for Gender. The table in our Results viewer should go empty again.

Now select BIA1 and BIA2 and move them into the Variables box. This will show the same information as before, but now we want a little bit extra. Click on the Statistics dropdown menu below our variables selector to see all the various statistics we can ask Jamovi to show us. Click on Range to add it to our table in the Results viewer:

The screenshot shows the Jamovi software interface. On the left, a data table is visible with columns PartID, Age, and FitPercept. The main window is titled 'Descriptives' and shows a list of variables on the left: PartID, Age, Gender, Course, Weight, Height, FitPercept, and BMI. The 'Variables' box on the right contains BIA1 and BIA2. Below the variables selector, the 'Statistics' dropdown menu is open, showing options for Sample Size, Percentile Values, Dispersion, Central Tendency, Distribution, and Normality. The 'Range' option under Dispersion is selected. On the right, the 'Results' panel displays a table titled 'Descriptives' with columns BIA1 and BIA2.

	BIA1	BIA2
N	49	48
Missing	2	3
Mean	22.5	22.7
Median	20.1	20.1
Standard deviation	9.35	9.12
Range	44.6	40.4
Minimum	2.00	4.10
Maximum	46.6	44.5

Feel free to click around any other Statistics options you might be interested in - you can add and remove as many as you like at any time in Jamovi, and the Results viewer will auto-update to show what you have changed.

Q9. From the descriptive data for the two machines, would you say the machines show the 'same' results?

Frequency tables

Let's also see how many students are on each of the four courses offered. Move BIA1 and BIA2 out of the variables selected, then move Course into the variables selection box.

Click the Frequency tables checkbox underneath our variables selection to add a frequency table to our Results viewer (this will only work for Nominal or Ordinal data types). If you want to make the Results viewer a bit tidier, untick all the other Statistics options, but leave in the N and Missing options:

The screenshot shows the Jamovi software interface. On the left, a list of variables is visible, including PartID, Age, Gender, Weight, Height, BIA1, BIA2, FitPercept, and BMI. The 'Descriptives' panel is active, showing the 'Course' variable selected for analysis. The 'Statistics' section is expanded, showing options for Sample Size, Central Tendency, Percentile Values, Dispersion, Distribution, and Normality. The 'Results' panel on the right displays the 'Descriptives' table and the 'Frequencies of Course' table.

Descriptives

Course	
N	50
Missing	1

Frequencies of Course

Levels	Counts	% of Total	Cumulative %
Sport Therapy	27	54.0%	54.0%
Sport and Dance Therapy	8	16.0%	70.0%
Sport and Exercise Science	5	10.0%	80.0%
Sport Psychology, Coaching and Physical Education	10	20.0%	100.0%

Here you are presented with frequency as a count, as an overall percentage, and as a cumulative percentage. Top table also tells us about the total number of values and how many are missing values - that is, any cases we have that do not have a valid Course ID.

Q10. How many students actually provided a course id?

Q11. Which course has the greatest percentage of students?

Q12. What percentage of students takes Sport and Dance Therapy?

Notice that our counts in the Frequency table add up to the number of valid Course IDs we have in our data. You need to make sure you always check for any missing values in your data - in this case, you might need to go back and find that missing Course ID, or else exclude the case from any further analyses if Course ID is required.

Basic plots

Histogram

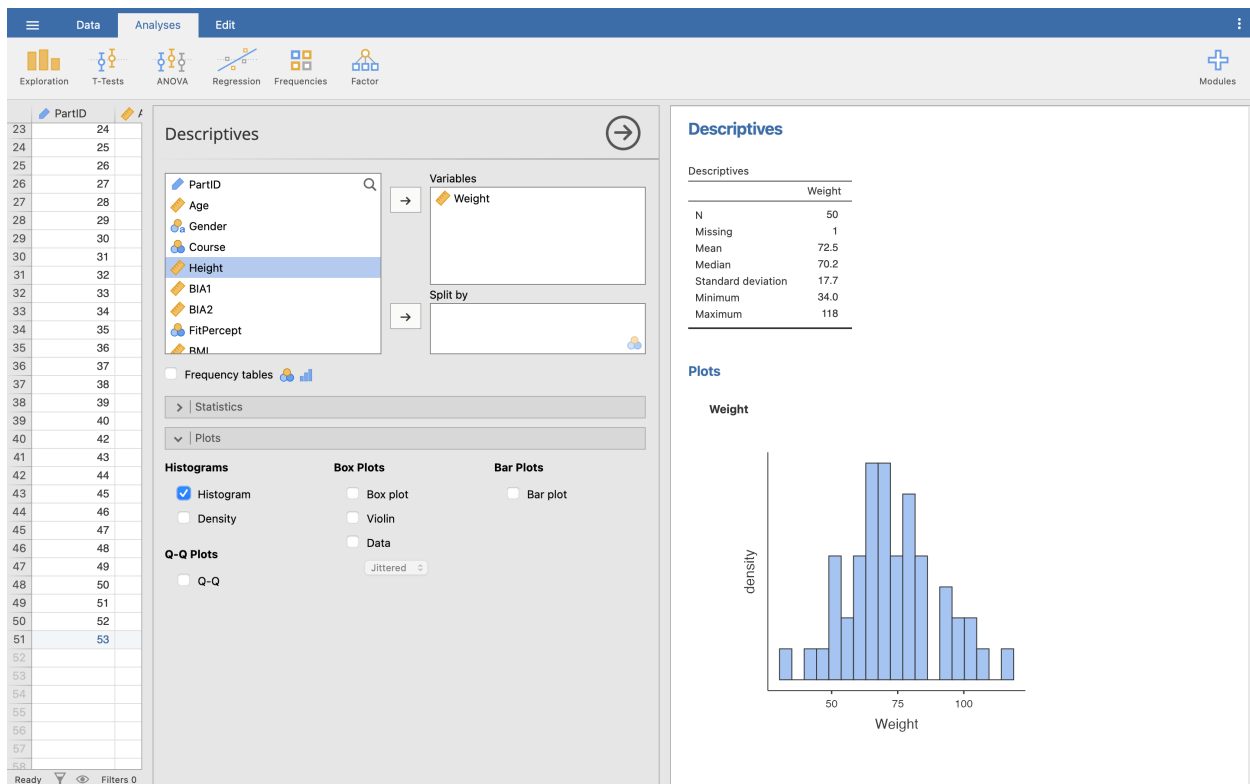
While we expect you to generate graphs in Excel, and produce tables in MS-Word, you can still generate graphs in Jamovi for a quick view of data before formatting them in Excel. Remember, for any coursework, it is generally best to create your final graphs in Excel so you can completely customise how they look.

We will look at two graphs:

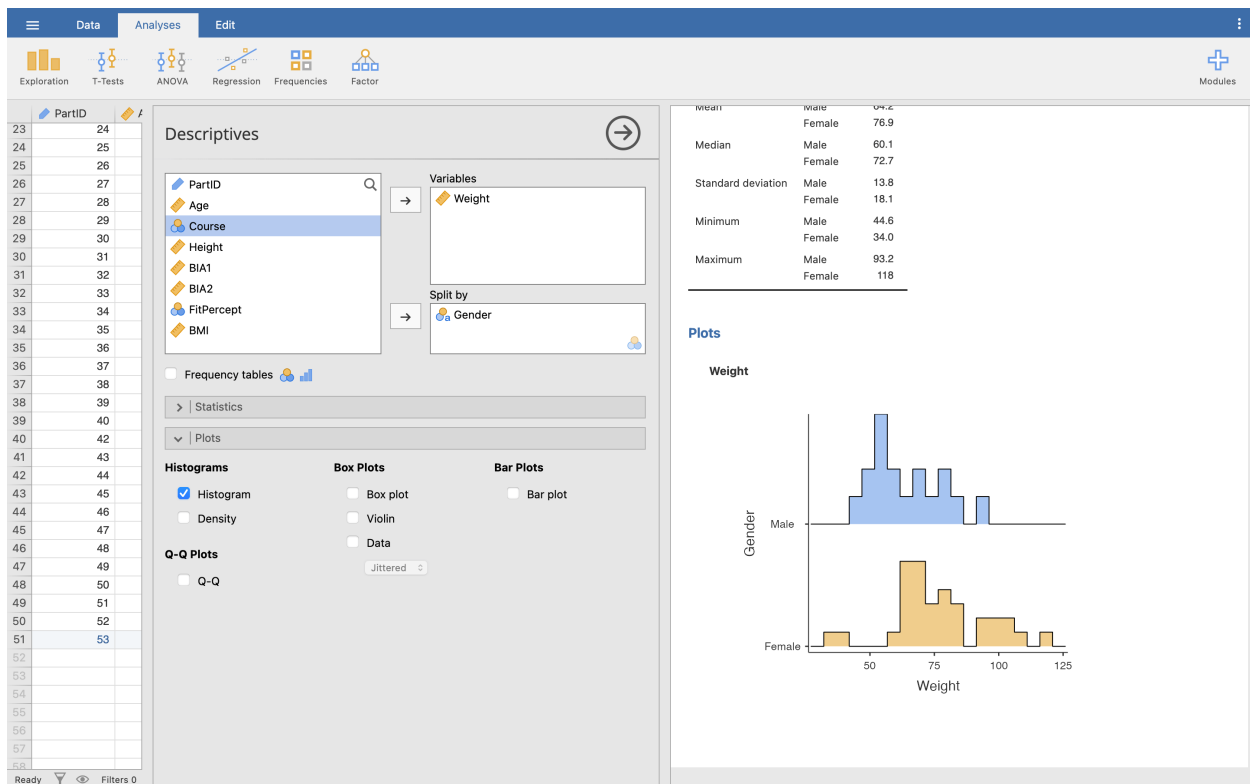
- Histogram
- Box plot

We will continue to use our Body Composition.omv data set. Still in the Descriptive menu, clear all the variables from our variable selection, and untick the Frequency table option. Move Weight into the variables selection and turn back on the Default descriptive statistics - tick the boxes for N, Missing, Mean, Median, Std. Deviation, Minimum, and Maximum.

Then, you can click the Statistic menu bar again to fold it out of the way, and click on the Plots menu bar. Click on Histogram to generate a simple histogram in our Results viewer.



This is a histogram for our entire data set, but we can split this by Gender exactly as we did for the Descriptives. Move Gender into the Split By box and see what happens to our histogram:

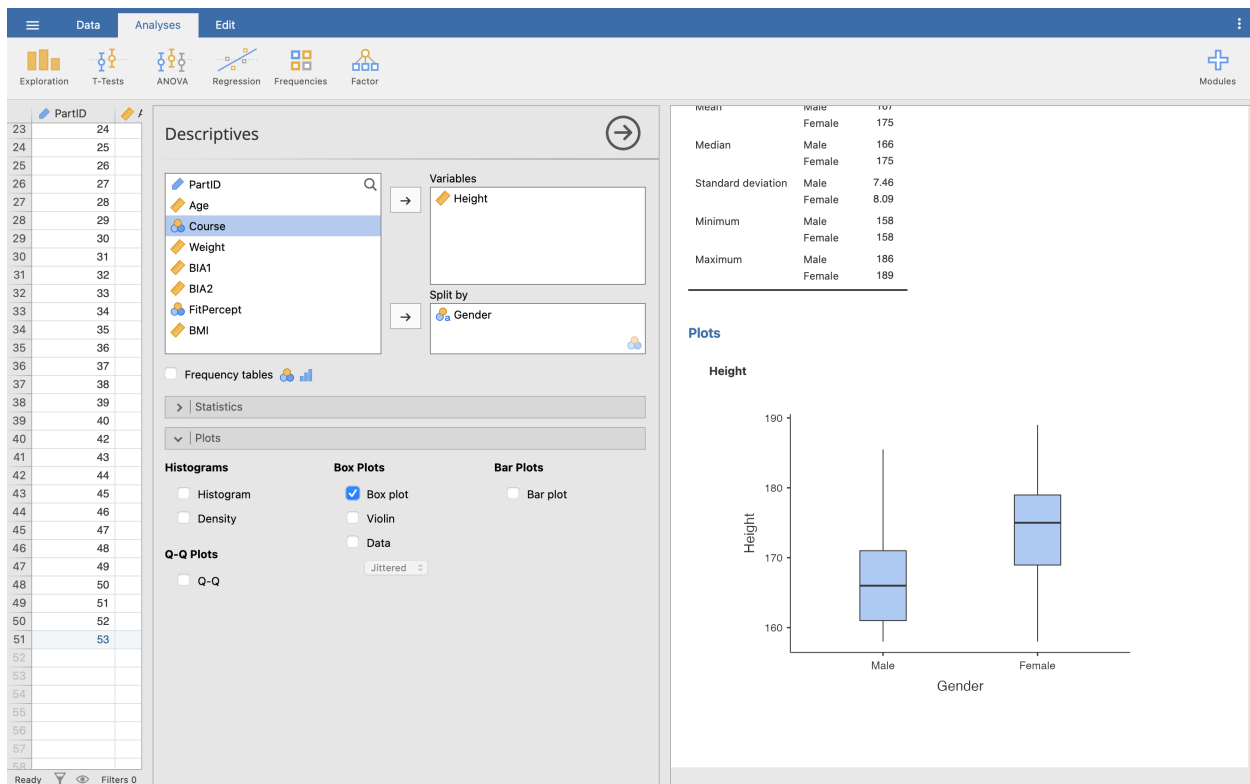


You can split your data using any Nominal or Ordinal variable - try moving other variables in and out of the Split By box to see how the histogram changes.

Q13. Based on the histograms, which Gender has a greater variability of weight? Why?

Box plots

Boxplots may also be useful to visually inspect your data, especially if you are looking for outliers. We will use a box plot to check for any outliers in our height data. Clear the selection boxes, then move Height into the variables selection, and Split By Gender again. Then, click the Box plot tick box to show a box plot in the Results viewer:



Remember in a boxplot:

- Top whisker = 1.5 x interquartile range
- Top of box = 3rd quartile
- Line in the middle = median
- Bottom of box = 1st quartile
- Bottom whisker = 1.5 x interquartile range

Any outliers are shown with circles beyond the whiskers of the box plot. There are no recognised outliers in our Height data.

Q14. On the boxplot, what is the approximate value for median male height and median female height?

Feel free to click around now on any other Plot or Statistics options you think look interesting, and try exploring different variables and different ways of splitting the

data. Remember, if you want to remove something from the Results viewer, just untick it.

Parametric testing and assumptions

Before we can decide which test to use we need to look at the 'quality' of our data.

There are 4 assumptions we need to look for:

- Level of data
- Independence of data (often called random allocation of data)
- Normal distribution
- Homogeneity of variance

These have all been discussed in the lecture so we will not examine what they mean here, but we do need to look at the third and fourth assumptions and test them in Jamovi.

Remember, if you are comparing multiple variables, then you must check each variable (or segmented variable) against your assumptions.

Tests of normality (normal distribution)

While we could simply look at a histogram to visually inspect for a normal distribution, there are a number of statistical tests we can also perform. The main statistical test for normality in Jamovi is the Shapiro-Wilk (SW) test.

Let us examine the distribution of male and female height data. To compare the two sets of independent data we need to make sure that both male and female data do not deviate significantly from a normal distribution. (Note: testing repeated measures data requires a slightly different test, but here we will examine normal distribution for independent samples and look at repeated measures later)

Still in the Descriptives menu, move Height into the variables selection, and split the

data by Gender. Then, under the Statistics menu, click the Shapiro-Wilk tick box:

The screenshot shows the Jamovi software interface. On the left, the 'Data' tab is active, displaying a list of variables: PartID, Age, Course, Weight, BIA1, BIA2, FitPercept, and BMI. The 'Analyses' tab is selected, and the 'Descriptives' menu is open. In the 'Variables' box, 'Height' is selected. In the 'Split by' box, 'Gender' is selected. The 'Statistics' section is expanded, showing various options. Under 'Sample Size', 'N' and 'Missing' are checked. Under 'Percentile Values', 'Cut points for 4 equal groups' is selected. Under 'Dispersion', 'Std. deviation', 'Variance', 'Range', and 'IQR' are checked. Under 'Central Tendency', 'Mean', 'Median', and 'Sum' are checked. Under 'Distribution', 'Skewness' and 'Kurtosis' are checked. Under 'Normality', 'Shapiro-Wilk' is checked. On the right, the 'Results' tab is active, displaying a table of Descriptives.

Descriptives		
	Gender	Height
N	Male	17
	Female	34
Missing	Male	0
	Female	0
Mean	Male	167
	Female	175
Median	Male	166
	Female	175
Standard deviation	Male	7.46
	Female	8.09
Minimum	Male	158
	Female	158
Maximum	Male	186
	Female	189
Shapiro-Wilk W	Male	0.919
	Female	0.972
Shapiro-Wilk p	Male	0.142
	Female	0.522

This will add rows for Shapiro-Wilk W and Shapiro-Wilk p values to our descriptives table in the Results viewer.

Here we are testing whether the distribution of Height (by gender) is normally distributed so we can use parametric tests later. We want to see whether there is a significant deviation from normality, i.e. whether the p value result is less than 0.05 for Height for either Male or Female groups. Both groups have $p > 0.05$. This means there is no significant deviation from a normal distribution, so the test of normality has been satisfied.

Now you try: Perform a test of normality for Weight by Gender using the Body Composition.omv data set.

Q15. What test would you run?

Q16. What p-value for the test of normality do you get for the Weight data of the Females group? What does this p-value mean?

A simple SW test is ok to start, but you should still perform further checks on your data. The first two we will examine are Skewness and Kurtosis.

Still in the Descriptives menu, choose the Height variable split by Gender. Under Statistics, tick Skewness and Kurtosis.

The screenshot shows the Jamovi software interface. On the left, the 'Data' tab is active, displaying a list of variables: PartID, Age, Course, Weight, BIA1, BIA2, FitPercept, and BMI. The 'Analyses' tab is selected, and the 'Descriptives' menu is open. In the 'Descriptives' panel, the 'Variables' list contains 'Height', and the 'Split by' list contains 'Gender'. The 'Statistics' section is expanded, showing options for Sample Size, Central Tendency, Dispersion, Distribution, and Normality. The 'Distribution' section has 'Skewness' and 'Kurtosis' checked. The 'Results' panel on the right displays the 'Descriptives' table, which shows the mean, standard deviation, minimum, maximum, skewness, and kurtosis for the 'Height' variable, split by 'Gender'.

	Gender	Height
N	Male	17
	Female	34
Missing	Male	0
	Female	0
Mean	Male	167
	Female	175
Median	Male	166
	Female	175
Standard deviation	Male	7.46
	Female	8.09
Minimum	Male	158
	Female	158
Maximum	Male	186
	Female	189
Skewness	Male	1.03
	Female	-0.0365
Std. error skewness	Male	0.550
	Female	0.403
Kurtosis	Male	0.959
	Female	-0.481
Std. error kurtosis	Male	1.06
	Female	0.788

The first way we can examine this data is to look at the absolute values for Skewness and Kurtosis.

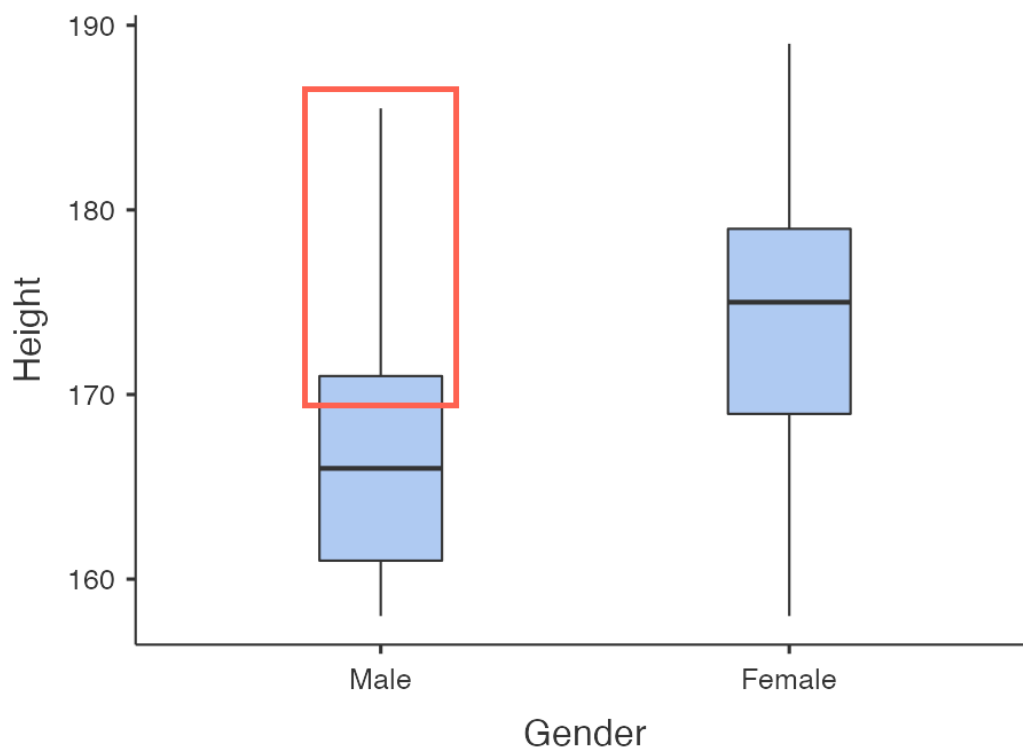
- Male skewness = 1.028
- Male kurtosis = 0.959
- Female skewness = -0.036
- Female kurtosis = -0.481

Q17. Based on the raw Skewness and Kurtosis values, describe the shape of your male data distribution.

A rule of thumb suggests that -0.8 to $+0.8$ is acceptable for skewness and -3 to $+3$ is acceptable for kurtosis. Note that the skewness for males falls outside the -0.8 to 0.8 , but all other values are well within their ranges. Strictly speaking we could suggest that the male height data is not normally distributed, and has a slight positive skew (i.e. a pile up of data on the left hand side).

Turn back on the box plots for this data. You can see the slight positive skew for the male data with a longer positive tail:

Height



A further way of examining this data is to convert the height skewness and kurtosis

to a 'standardised' value. We do this by taking the value and dividing it by the standard error of its value. You can find the standard error values in the Descriptives table just underneath the Skewness and Kurtosis.

Before we work out our standardised values though, let's change the number of decimal places Jamovi is using to show our Results, so we can be more accurate. In the far top right-hand corner of Jamovi is three white dots - click these to open the Preferences panel. Here, change the Number format dropdown menu to 3 dp (3 decimal places). Click anywhere outside the Preferences panel to close it again.

The screenshot shows the Jamovi software interface. On the left, a data table with columns 'PartID' and 'Height' is visible. The main panel displays the 'Descriptives' results for 'Height', split by 'Gender'. The results table shows various statistics for Male and Female groups. On the right, the 'Preferences' panel is open, showing settings for 'Results', 'Plots', 'Import', and 'Updates'. The 'Number format' is set to '3 dp'.

Descriptives Results Table:

	Gender	Height
N	Male	17
	Female	34
Missing	Male	0
	Female	0
Mean	Male	166.782
	Female	174.524
Median	Male	166.000
	Female	175.000
Standard deviation	Male	7.460
	Female	8.085
Minimum	Male	158.000
	Female	158.000
Maximum	Male	185.500
	Female	189.000
Skewness	Male	1.028
	Female	-0.036
Std. error skewness	Male	0.550
	Female	0.403
Kurtosis	Male	0.959
	Female	-0.481
Std. error kurtosis	Male	1.063
	Female	0.788

Preferences Panel Settings:

- Results:** Number format: 3 dp, p-value format: 3 dp, References: Hidden, Syntax mode: ☐
- Plots:** Plot theme: Default, Color palette: jmv
- Import:** Default missings: NA
- Updates:** jamovi is up-to-date: ☐ Check again, Automatically install updates: ☒
- Screen Capture Tool:** ☐
- Developer mode:** ☐

Now we can calculate our standardised skewness and kurtosis values:

Male	Female
$S. Skew = \frac{Skewness}{SE (Skewness)} = \frac{1.028}{0.550} = 1.869$	$S. Skew = \frac{Skewness}{SE (Skewness)} = \frac{-0.036}{0.403} = -0.090$
$S. Kurtosis = \frac{Kurtosis}{SE (Kurtosis)} = \frac{0.959}{1.063} = 0.902$	$S. Kurtosis = \frac{Kurtosis}{SE (Kurtosis)} = \frac{-0.481}{0.788} = -0.610$

These values can now be compared to the values we would expect to get by chance alone, so a value less than -1.96 or greater than 1.96 would suggest that this distribution is significantly skewed or shows significant kurtosis. Both of our values are within this range so we can be happy that our data satisfies the condition of normality.

Note: You cannot use these tests in very large samples as the standard skewness and kurtosis values are likely to be significant even if the skew and kurtosis are only slightly different to normal.

Even though we would like to think that statistics should be clear-cut in giving us answers, you can now see that it can often requires careful interpretation of your results. Here we have seen two tests that show our data is normally distributed, with one test that shows it is not! This is where your experience of working with numbers is going to be very useful.

Repeated measures normality: With independent samples data, or two different groups, you need to check for normality of each separate group. If you have a repeated measures test, e.g. your group performs a test under two dif-

ferent conditions, then rather than testing the normality of each group, you would test for the normality of the difference between the two conditions.

Homogeneity of variance

This is a check to see that the samples we are testing come from populations with a similar variance.

As with the normality test, there is a statistical test and a rule of thumb. Starting with the rule of thumb, we want to make sure that the two sample variances differ by a factor of less than 3.

Imagine we wished to compare male and female height - we therefore need to compare the sample variance of the male height and the female height.

We can view these variances in Jamovi - still with Height split by Gender selected, tick the variance check box. This will display the variances for Male and Female groups in our Descriptives table in the Results viewer.

To test our rule of thumb, take whichever of the two is the larger variance and divide it by the smaller variance. In our case, that would be:

$$65.368 / 55.649 = 1.17$$

1.17 is well within our rule of thumb of 3, so we can assume homogeneity of variance. But, let's also test it properly. The statistical test we will use for this is Levene's Test of Homogeneity.

To do this, we actually need to go to a new menu option. Click the right-arrow in the circle at the top right of the Descriptives menu to close it. Then, still in the Analyse menu, click the T-Tests icon, and choose Independent Samples T-Test from the dropdown menu.

This will open the T-Test menu. You'll see that it looks very similar to the Descrip-

tives menu - we choose our variables of interest and how we want to split them (here called Grouping Variable) at the top, and just tick the things we want Jamovi to show us underneath.

Move Height into our variables selection and split by Gender again. Don't worry about any of the other options, but just tick the Homogeneity test option under Assumptions.

The screenshot shows the Jamovi software interface. On the left, a list of variables is visible, including PartID, Age, Course, Weight, BIA1, BIA2, FitPercept, BMI, and PartID. The 'Independent Samples T-Test' configuration window is open in the center. Under 'Dependent Variables', 'Height' is selected. Under 'Grouping Variable', 'Gender' is selected. The 'Tests' section has 'Student's' checked. The 'Assumption Checks' section has 'Homogeneity test' checked. The 'Missing values' section has 'Exclude cases analysis by analysis' selected. On the right, the 'Missing' table is displayed, showing the distribution of 'Height' for 'Male' and 'Female'. Below this, the 'Independent Samples T-Test' results table is shown, displaying the Student's t statistic, degrees of freedom, and p-value. Finally, the 'Assumptions' section shows the 'Homogeneity of Variances Test (Levene's)' results, including the F statistic, degrees of freedom, and p-value.

	Male	Female
Mean	166.782	174.524
Median	166.000	175.000
Standard deviation	7.460	8.085
Variance	55.649	65.368
Minimum	158.000	158.000
Maximum	185.500	189.000

	Statistic	df	p	
Height	Student's t	-3.305	49.000	0.002

Homogeneity of Variances Test (Levene's)				
	F	df	df2	p
Height	0.178	1	49	0.675

Note. A low p-value suggests a violation of the assumption of equal variances

Notice how this gives us a new area in the Results viewer, under the heading Independent Samples T-Test. But, our Descriptives table is still displayed above it (you might need to scroll back up). If you click on the descriptives table in the Results, the descriptive menu will open back up in the middle, and you can make any changes. This is how your results are stored and edited in Jamovi - just click them in the viewer to open the menu for that particular result. If you want to remove a result from your viewer entirely, right-click on it's title and select 'Remove' from the pop-up menu.

We won't worry about the T-Test results for now - just look at the table titled Homogeneity of Variances Test (Levene's). Here, we are looking to see if the test result is significant, just as we did for the normality test. The non-significant value of $p = 0.675$ says that the variances are not significantly different to each other and we can continue with our parametric tests.

Now you try: Repeat the test of Homogeneity of Variance for Weight by Gender.

Q18. What is the Levene statistic and p value, and what does this mean?

Q19. Would you be happy running a parametric test on this data? And why?

Antwoorden op vragen

A1. In dit gegevensbestand staan 9 kolommen met gegevens: deelnemer ID, leeftijd, geslacht, cursus, gewicht, lengte, BIA1, BIA2 en FitPercept.

A2. Er zijn 50 rijen met gegevens, oftewel 50 'cases' (eenheden). Gebruik PartId niet om het aantal records weer te geven.

A3. Ook al is er een rijnummer, dit identificeert geen specifieke eenheid uit de gegevens, alleen de volgorde van de rijen (of cases) die momenteel worden weergegeven. Met de PartId kan de gebruiker gemakkelijk een bepaalde case identificeren. Als je de gegevens sorteert of filtert, kunnen de rijnummers veranderen, maar PartId blijft consistent.

A4. Ten eerste zal de PartId uniek zijn, maar het maakt het ook mogelijk om de gegevens te anonimiseren.

A5. Jamovi heeft vier hoofdtypen gegevens: Nominaal, Ordinaal, Continu en ID. Continu is een combinatie van de gegevenstypen Interval en Ratio. ID is een uniek gegevenstype voor Jamovi, dat alleen wordt gebruikt voor deelnemers-ID-variabelen, dat vergelijkbaar is met een nominaal gegevenstype, maar zonder categorieniveaus of labels.

A6. Het veld is puur een 'naam' voor elk record, je kunt er niet uit afleiden dat er een volgorde is, of dat het ene voor of na het andere moet komen.

A7. Vrouwen zijn gemiddeld zwaarder. Gemiddelde massa vrouw = 76.855 kg, gemiddelde massa man = 64.153 kg.

A8. Er zijn 17 mannen en 33 vrouwen in de steekproef.

A9. Het is moeilijk definitief te zeggen of de machines 'hetzelfde' laten zien. De beschrijvende informatie zou suggereren dat de waarden 'gemiddeld' hetzelfde zijn, maar afzonderlijke machines binnen de steekproef zouden heel verschillende resultaten kunnen laten zien, bijv. machine 1 geeft 15% en 25% voor twee deelnemers, maar machine 2 geeft 25% en 15% voor

dezelfde twee. Gemiddeld' zouden beide machines 20% aangeven. Het feit dat getallen op elkaar 'lijken' betekent niet dat ze statistisch gezien hetzelfde zijn.

A10. 50 studenten hebben een cursus opgegeven.

A11. Sporttherapie, met 54%.

A12. 16% van de studenten volgt Sport- en Danstherapie.

A13. Uit de grafiek blijkt dat vrouwen variabler zijn en zich links en rechts van de mannen in de grafiek uitstrekken. De standaarddeviatie is ook groter voor vrouwen (18,08 kg) vergeleken met mannen (13,84 kg).

A14. In de boxplot is de mediaan de dikke zwarte lijn in het midden van elk vak. De geschatte waarde voor de mediane lengte bij mannen is 166 cm en bij vrouwen 175 cm.

A15. Voor een normaliteitstest zou je de Shapiro-Wilk test uitvoeren.

A16. Voor vrouwen is Shapiro-Wilk $p = 0,291$. Dit is vergelijkbaar met de p -waarde voor mannen. Een niet-significante waarde suggereert dat de steekproefverdeling voor het gewicht van vrouwen niet significant verschilt van een normale verdeling, zodat aan een van de aannames voor het uitvoeren van parametrische tests is voldaan.

A17. De positieve scheefheidswaarde (1,028) suggereert dat de grafiek een iets langere rechterstaart heeft, en de positieve kurtosiswaarde (0,959) suggereert dat de curve iets dunner is dan een normale verdeling (ook wel leptokurtisch genoemd).

A18. De Levene teststatistiek, $F = 0,642$, $p = 0,427$ (gebaseerd op het gemiddelde). Dit betekent dat de varianties niet significant verschillen, dus we kunnen homogeniteit van variantie aannemen.

A19. Ja, we hebben voldaan aan de vier voorwaarden, 1. Het zijn gegevens van hoog meetniveau, 2. Ze zijn willekeurig toegewezen/selecteerd, 3. Ze zijn normaal verdeeld, en 4. De varianties zijn gelijk.