Beknopt maar significant* handboek voor M5 statistiek

Hotel Management School Maastricht ZUYD

* p = 0.01

Welk meetniveau?

Variabele
data

Categorisch **ABC**
Labels, groepen, onderscheidende categorieën

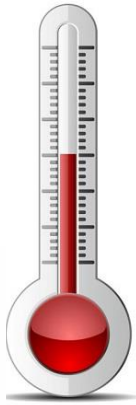
Nominaal
Categorieën zijn evenwaardig (= en ≠)

Ordinaal
Er is een volgorde in de categorieën (> en <)

Continu **123**
Gemeten, schaal, intervallen, waarden

Interval
Gelijke afstanden tussen waarden (+ en -)

Ratio
*Gelijke afstanden tussen waarden én een absoluut 0-punt (* en /)*



Welk diagram?

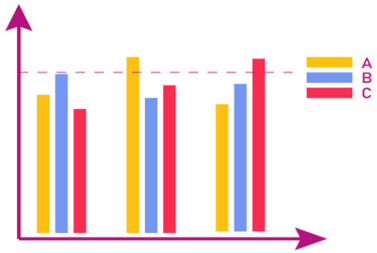
Variabele
data

Categorisch **ABC**
Labels, groepen, onderscheidende categorieën

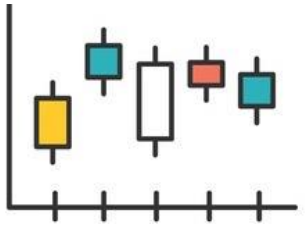
Continu **123**
Gemeten, schaal, intervallen, waarden

1 variabele
Staafdiagram of taartdiagram

2 variabelen
Geclusterd staafdiagram



1 van elk
Opties: staafdiagram, error bar plot, boxplot



1 variabele
Histogram

2 variabelen
Scatterplot, of lijndiagram (tijdsverloop)

Hypotheses opstellen

Een hypothese is een beknopt statement over het verwachte resultaat van een experiment of project. Er bestaan twee typen:

- de **nulhypothese** H_0 Gezien als standaard, normaal, ´wat er is´ in de huidige wereld
- de **alternatieve hypothese** H_1 / H_a Wat zou kunnen zijn, iets buitengewoons, een verschil

Voorbeeld : Een onderzoeker test de werkzaamheid van een nieuw medicijn. Normaal gesproken helpt een pil of behandeling niet, dus de normale realiteit H_0 is: het medicijn heeft geen voordelen en werkt niet. Als het medicijn geen verschil maakt, zijn er geen verschillen te verwachten in de gemiddelden (μ´s) van testgroep en controlegroep, dus:

H_0 = het medicijn heeft geen effect OFWEL $\mu_{test} = \mu_{control}$

In een ´alternatieve´ wereld zou het nieuwe medicijn echter wel iets kunnen doen, dus H_1 is ´het medicijn maakt een verschil´:

H_1 = het medicijn doet iets OFWEL $\mu_{test} \neq \mu_{control}$

De bovenstaande hypothesen hebben geen richting. Ze gaan simpelweg over een mogelijk verschil in μ´s. Deze test is tweezijdig: elk resultaat kan meer of minder, beter of slechter zijn. We kunnen echter ook een hypothese met een richting formuleren, in welk geval we kunnen zeggen: het medicijn verbetert de gezondheid:

H_1 = het medicijn heeft een positief effect OFWEL $\mu_{test} > \mu_{control}$

Nu testen we éénzijdig: het resultaat van de testgroep zou hoger/beter moeten zijn.

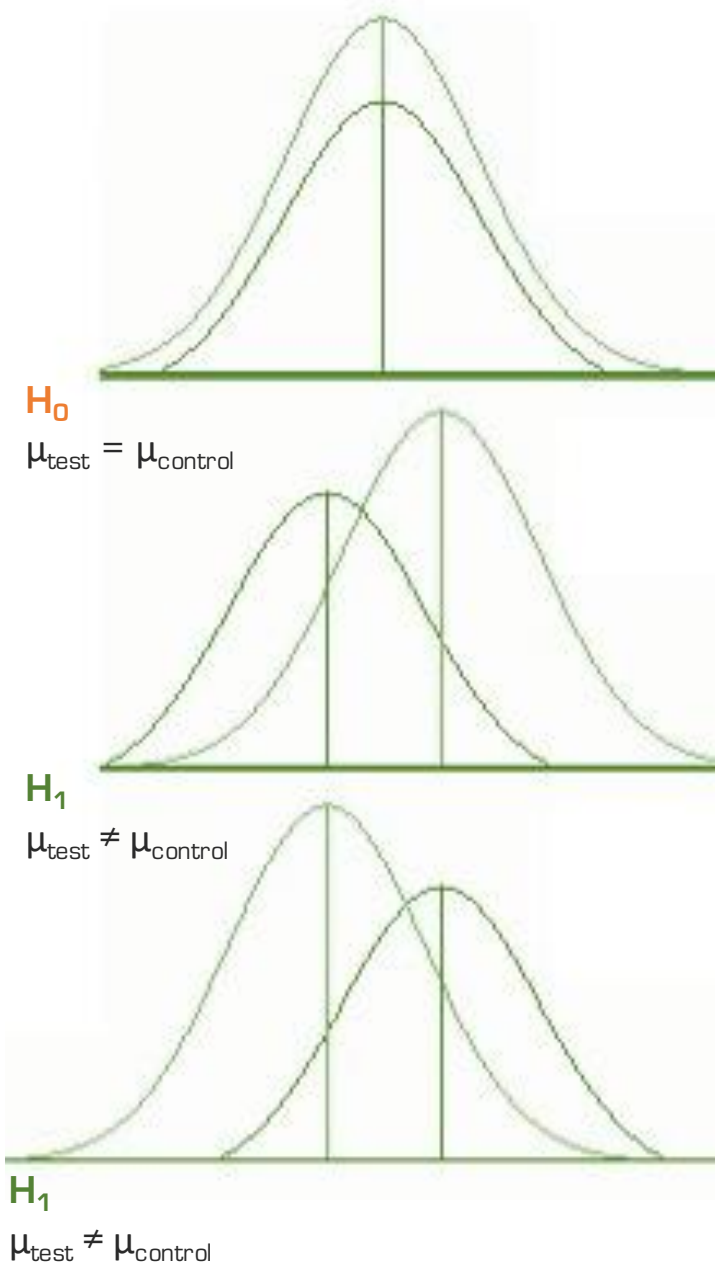
Hypotheses, μ, p, en distributie

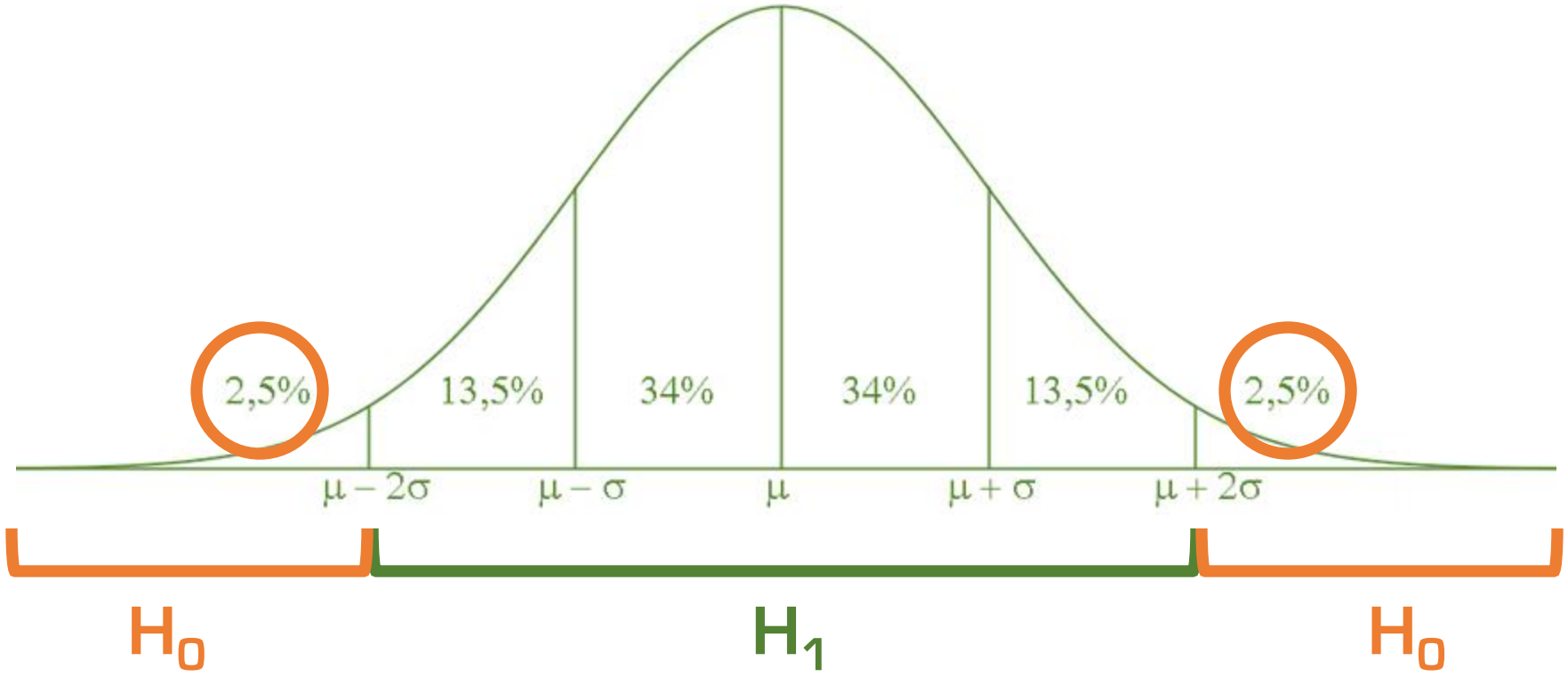
Statistische tests berekenen – simpel gezegd – de waarschijnlijkheid dat de H_0 waar is – dat wil zeggen: de waarschijnlijkheid dat je de observaties van je studie zou vinden als H_0 waar zou zijn.

Waarom? Omdat als we willen dat er iets ´speciaals en anders´ is voor H_1, dan willen we dat de μ´s van de groepen voldoende verschillen en we willen niet te veel observaties die geen verschil in μ´s suggereren.

Normaal gesproken gebruiken we een kritische p-waarde van 0.05, wat betekent dat we willen dat ten minste 95% van de geobserveerde data binnen het betrouwbaarheidsinterval van 95% valt om een significant verschil in μ´s aan te nemen en dus H_1 aan te nemen. Dat betekent ook dat we een kans van max. 5% toestaan voor de mogelijkheid dat je de geobserveerde data hoe dan ook zou vinden (bijvoorbeeld door toeval), en dat de μ´s helemaal niet zo verschillend zijn - wat zou betekenen dat we bij H_0 zouden blijven.

Als de p-waarde van H_0 lager is dan de gekozen kritische p-waarde (bijvoorbeeld 0.05), dan verwerpen we H_0 en accepteren we H_1.





Als de p-waarde van een toets lager is dan de gekozen kritische p-waarde (bijv. 0.05), dan zijn er beperkte observaties in het H_0-gebied en kunnen we H_0 verwerpen

&

nemen we aan dat er genoeg observaties zijn in het H_1-gebied om een verschil in μ's aan te nemen en H_1 te ondersteunen, dus accepteren we H_1.

! **Als je éénzijdig toetst, moet de p-waarde van de test worden gedeeld door 2 (je kunt de andere kant van de normaalverdeling niet 'gebruiken'). In Jamovi kun je tweezijdig of éénzijdig aanvinken; SPSS rapporteert in de resultatentabel beide.**

Error-types

Als je statistiek gebruikt om beslissingen te nemen over hypothesen, bestaat de kans dat je foute conclusies trekt. Opties zijn:

Type 1-fout, of vals-positief

H_0 verwerpen terwijl deze eigenlijk waar is en aangenomen zou moeten worden

Oftewel: H_1 accepteren terwijl je dat niet zou moeten doen

Type 2-fout, of vals-negatief

H_0 niet verwerpen terwijl deze eigenlijk onjuist is en verworpen zou moeten worden

Oftewel: H_1 niet accepteren terwijl je dat wel zou moeten doen

Type 3-fout, voor de enthousiastelingen

H_0 verwerpen en H_1 aannemen omdat de cijfers en toets dat zeggen, maar gebaseerd op foute redenering/methodologie

Voorbeeld1. Conclusie 'het medicijn verbetert de gezondheid' bij tweezijdige toetsing

Voorbeeld2. Het medicijn heeft een positief effect en mensen voelen zich beter volgens de p-waarde, maar de toets toetst niet correct het effect en de ziekte maar bijvoorbeeld meer het algemeen humeur.

Validiteit en Betrouwbaarheid

Validiteit heeft betrekking op de nauwkeurigheid van metingen: meet je wat je moet meten?

e.g . Een nieuwe vragenlijst wordt gebruikt en de resultaten komen overeen met die van eerder gebruikte vragenlijsten 👍

e.g. Een thermometer wordt *naast de deur* geplaatst in een gecontroleerde laboratoriumkoelruimte 👎

Onbetrouwbaar & niet valide

Onbetrouwbaar maar wel valide

Betrouwbaarheid heeft betrekking op de consistentie van metingen en gegevens: kunnen de resultaten onder dezelfde omstandigheden worden gereproduceerd? En als de N toeneemt, wordt de uitkomst van de toets dan steeds sterker en duidelijker?

Voorbeeld1. Een respondent antwoordt ´ja´ op *Eet je vlees?* en ´nee´ op *Ben je vegetariër ?* 👍

Voorbeeld2. Hoe meer respondenten (hogere N), hoe meer significant een gevonden verschil, hetgeen het resultaat dus versterkt

→ Interne consistentie wordt doorgaans uitgedrukt in Cronbach's alpha α

Betrouwbaar maar niet valide

Betrouwbaar en valide

Validiteit

Indruksvaliditeit (face validity)

Is het geheel op het eerste gezicht logisch, ook voor deelnemers?

Inhoudsvaliditeit (content validity)

Dekt de test alles wat het zou moeten?

Criteriumvaliditeit (criterion validity)

Voorspelt of correleert een test nauwkeurig met de criteriumuitkomst?

Constructvaliditeit (Construct validity)

Is een set indicatoren representatief voor een concept dat niet direct meetbaar is?

Interne validiteit (Internal validity)

Beantwoord je met onderzoeksontwerp, uitvoering en analyse de onderzoeksvraag zonder vooroordelen / aannames?

Externe validiteit (External validity)

Kunnen de onderzoeksresultaten worden gegeneraliseerd naar andere contexten?

Betrouwbaarheid

- Steekproefgrootte (N)
- Test - hertest

- Standaardisatie
- Pilots (proefdraaien)
- Peer feedback
- Rapporteren en verantwoorden

Kwantitatief

- Intersubjectiviteit
- Triangulatie (diversificatie)
- Iteratie (herhaling)

Kwalitatief

Steekproefvorming (sampling)

→ Bepaal criteria voor de doelpopulatie, bijvoorbeeld leeftijdsgrenzen of woongebied
→ Vorm een steekproef van onderzoekseenheden* die zo willekeurig én representatief mogelijk is

Willekeurig, oftewel random: Ga niet zelf selectief uitzoeken, of alleen vrienden of familie vragen.
Representatief: De steekproef moet de optimale mix van kenmerken *binnen de doelpopulatie* bevatten.

A-SELECTE STEEKPROEF Ook: **WILLEKEURIGE SELECTIE** (probability sampling)	gebaseerd op het principe van willekeur, dus *random:* willekeurige selectie en toeval binnen een populatie of subpopulatie
SELECTE STEEKPROEF (non-probability sampling)	Selectie wordt beïnvloed door subjectieve grenzen en/of doelgerichte, handige selectie uit de populatie of subpopulatie

*Onderzoekseenheden zijn vaak mensen, maar kunnen bijvoorbeeld ook dieren zijn.

Steekproefvorming: opties

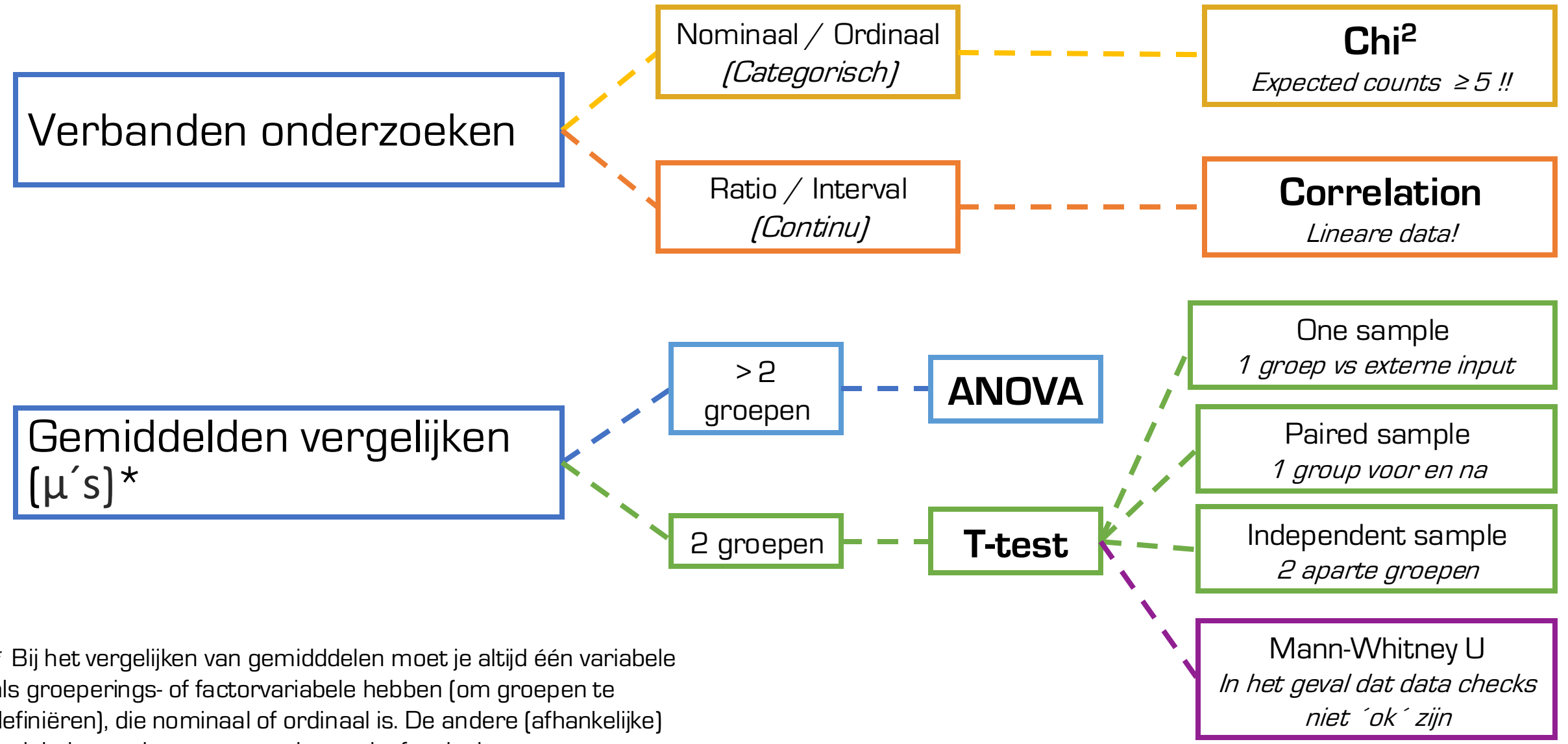
A-SELECTE STEEKPROEF

- (SIMPLE) RANDOM — Elke willekeurige onderzoekseenheid binnen de doelgroep (random vanuit een database/lijst)
- STRATIFIED — Willekeurige sampling binnen ***relatieve subpopulatie (strata)***
  bijv. als een studentenpopulatie 50% Nederlandse, 20% Duitse, 20% Belgiche en 10% Franse studenten heeft, selecteer je willekeurig mensen in diezelfde verhoudingen (dus bijv. Van de 1000: 500 NL, 200 DE, 200 BE en 100 FR)
- CLUSTERED — Een grote populatie verdelen in ***representatieve clusters*** en vervolgens clusters selecteren
  bijv. je wil een steekproef van alle taalstudenten in Nederland, maar kunt niet naar alle uni's en hbo's gaan, dus maakt je clusters (namelijk: elke taalafdeling van elke instelling is een cluster), en kies je één cluster van elke instelling en praat met die studenten
- SYSTEMATIC — Op basis van een patroon, zoals elke 10^{e} gast

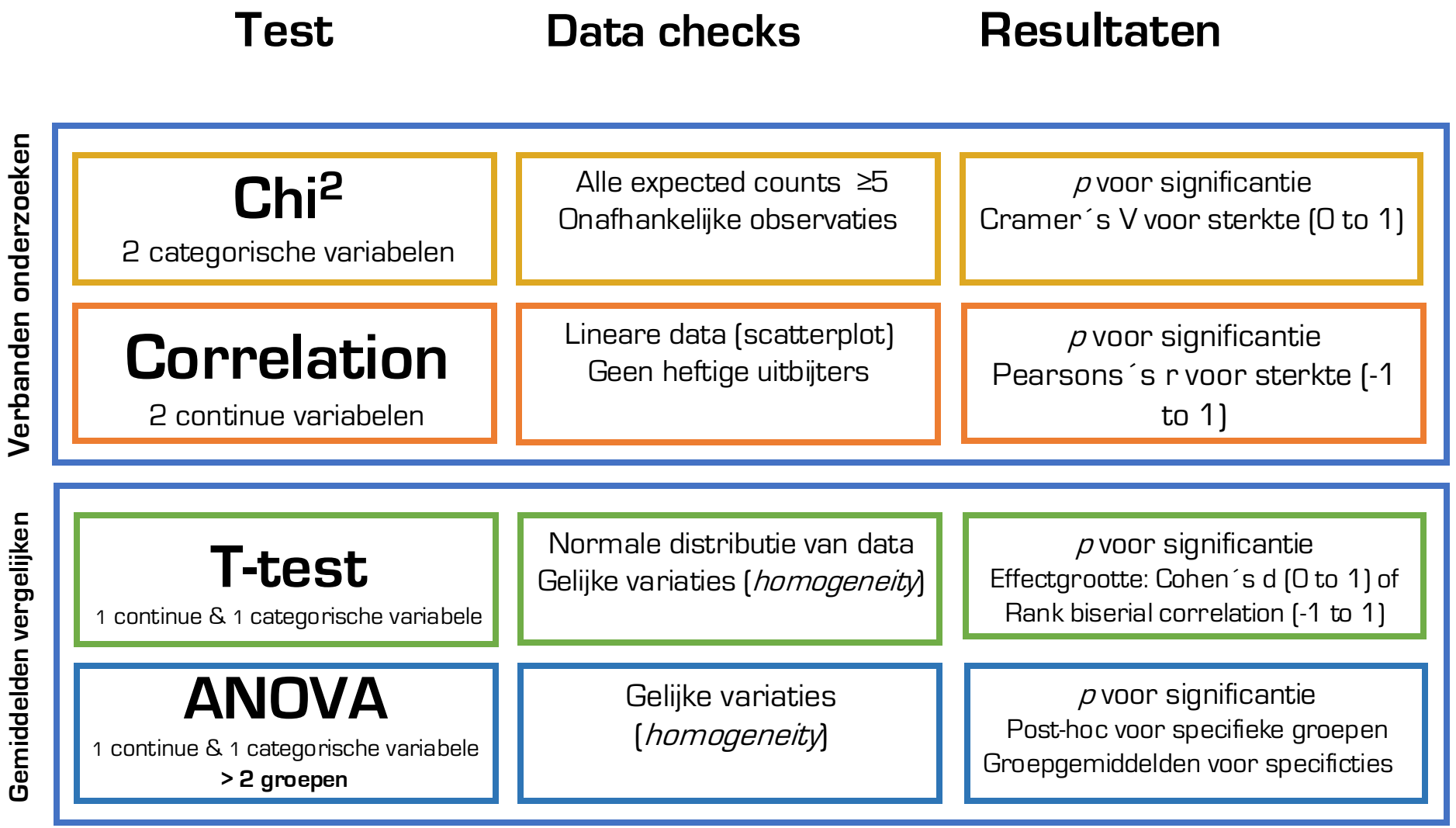
SELECTE STEEKPROEF

- SNEEUWBAL — Elke willekeurige benaderde onderzoekseenheid mag andere onderzoekseenheden benaderen
- QUOTUM — Willekeurige sampling binnen (relatief) vaste subpopulaties, *maar niet random*
  bijv. als een hotel 80% basickamers en 20% suites heeft, 40 basickamer-personen en 10 suitepersonen. Je stopt pas als je de quota haalt, dus als je die 40 en 10 hebt.
- ZELF-SELECTIE — Vrijwillige aanmelding voor deelname
- GEMAKSSTEEKPROEF — Snel en gemakkelijk: elke onderzoekseenheid die je binnen een bepaalde periode of gebied kunt vinden. Erg 'niet willekeurig'.

Welke toets?



* Bij het vergelijken van gemidddelen moet je altijd één variabele als groeperings- of factorvariabele hebben (om groepen te definiëren), die nominaal of ordinaal is. De andere (afhankelijke) variabele wordt gemeten op interval- of rationiveau.



* Bij het vergelijken van gemiddelen moet je altijd één variabele als groeperings- of factorvariabele hebben (om groepen te definiëren), die nominaal of ordinaal is. De andere (afhankelijke) variabele wordt gemeten op interval- of rationiveau.

CHI²

- De Chi²-toets, of χ^2-toets, wordt gebruikt om te kijken of er een verband bestaat tussen twee categorische variabelen.
- Omdat categorische variabelen geen numerieke metingen hebben en daarom niet met behulp van gemiddelden kunnen worden geanalyseerd, werkt Chi² met frequenties om te zien of er verschillen zijn tussen geobseveerde data (de metingen) en de verwachte metingen (*expected counts)*.
- De *expected counts* voor alle cellen moeten minimaal 5 zijn en wordt berekend met behulp van de formule (rowtotal x columntotal) / N, bijvoorbeeld (28x32) /106 voor cel **1-NO** in de tabel hieronder
- Je vindt de Chi2-test in Jamovi onder Analyses > Frequencies > Independent Samples $\chi 2$ of association
- Jamovi maakt een kruistabel met de χ^2-testresultaten en Cramer's V (indien geselecteerd onder *Statistics*)

Contingency Tables

qualification		healthplan No	healthplan Yes	Total
1	Observed	12	16	28
	Expected	8.45	19.55	28.00
2	Observed	18	27	45
	Expected	13.58	31.42	45.00
3	Observed	0	8	8
	Expected	2.42	5.58	8.00
4	Observed	2	18	20
	Expected	6.04	13.96	20.00
5	Observed	0	5	5
	Expected	1.51	3.49	5.00
Total	Observed	32	74	106
	Expected	32	74	106

χ^2 Tests

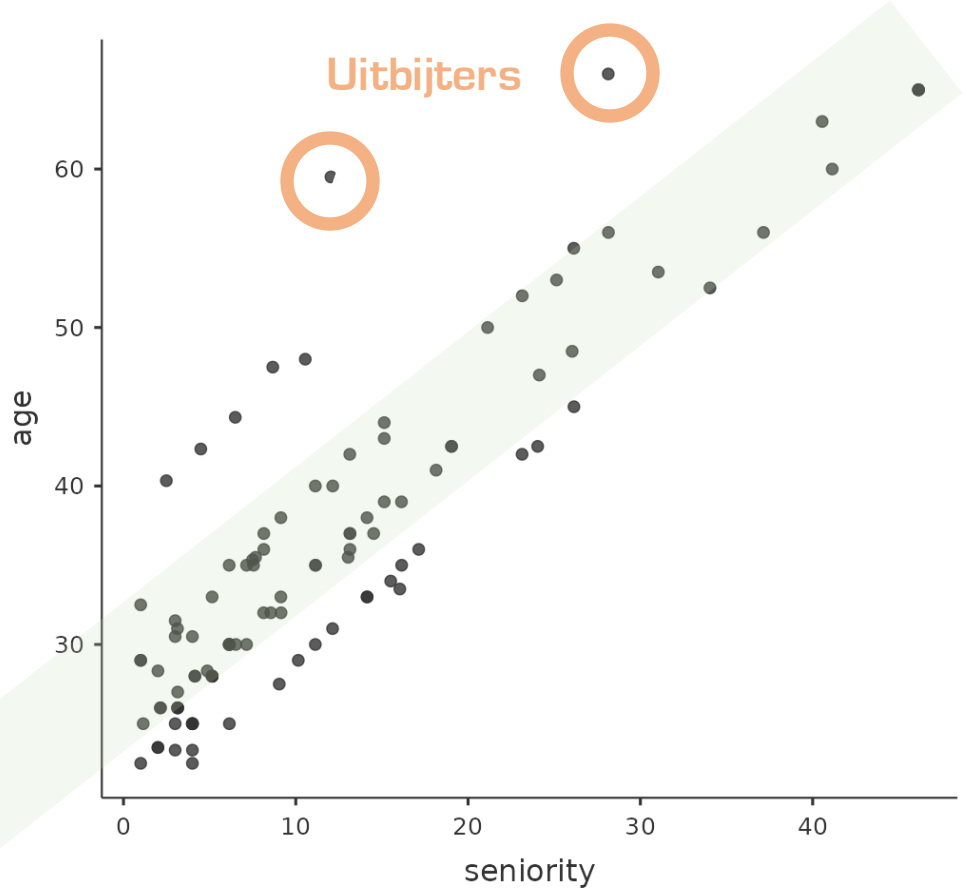
	Value	df	p
χ^2	13.7	4	0.008
N	106		

Nominal

	Value
Phi-coefficient	NaN
Cramer's V	0.359

CORRELATIE

- De correlatietest wordt gebruikt om te zien of er een relatie is tussen twee continue variabelen
- Voor data-checks ga je in Jamovi naar Exploration > Scatterplot om te zien of de data redelijkerwijs lineair zijn
- Je vindt de correlatietest in Jamovi onder Analyses > Regressie > Correlatiematrix
- Jamovi maakt een correlatiematrix met de uitkomsten, waaronder Pearson's r



Correlation Matrix

		age	seniority
age	Pearson's r	—	
	df	—	
	p-value	—	
seniority	Pearson's r	0.882	—
	df	104	—
	p-value	<.001	—

T-TEST

- De T-test wordt gebruikt om de gemiddelden (μ's) van 2 groepen voor een continue afhankelijke variabele te vergelijken, om te zien of er een significant verschil is tussen de groepen. Er zijn 3 soorten T-test:
  - **One sample t-test**. Wordt gebruikt om de μ van je verzamelde data te vergelijken met een al bestaande μ (bijvoorbeeld uit de literatuur)
  - **Independent t-test**. Wordt gebruikt om de μ's van twee groepen die je zelf hebt geobserveerd te vergelijken
  - **Paired samples t-test**. Wordt gebruikt om de μ's van gegevens die binnen één groep zijn verzameld, maar op twee verschillende momenten te vergelijken (bijv. μ hartslag vóór en μ ná het drinken van koffie). ¡Hiervoor zijn dezelfde mensen in beide groepen nodig!
- Hier moet je duidelijk éénzijdige of tweezijdige hypotheses opstellen om de juiste instellingen te kiezen en resultaten correct te interpreteren. Bij éénzijdig testen geldt slechts de helft van de berekende p-waarde. Jamovi verandert automatisch de p wanneer je éénzijdig aanvinkt; in SPSS zou je kolom ´ éénzijdig´ moeten uitlezen.
- T-tests vindt je in Jamovi onder Analyses > T-tests.
- Om de data-checks te doen, kies je in Jamovi de *Homogeneity* (Levene's) en de *Normality* (Shapiro-Wilk).
  - Als beide p-waarden **hoger** zijn dan 0.05 kun je doorgaan met een normale T-test.
  - Als één van beide niet ok is, of als beide niet ok zijn, moet je de optie *Mann-Whitney U* aanvinken en doorgaan (De Mann-Whitney U gebruikt medianen in plaats van gemiddelden)
- Voor effectgrootte vink je aan *Effect Size* onder *Additional Statistics*. Hiermee worden Cohen's d (van toepassing bij het uitvoeren van een normale t-test) en Rank biserial correlation (van toepassing bij het uitvoeren van de Mann-Whitney U) toegevoegd.

Normality Test (Shapiro-Wilk)

	W	p
perf_quantity	0.979	0.094

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

Homogeneity of Variances Test (Levene's)

	F	df	df2	p
perf_quantity	0.256	1	104	0.614

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

Independent Samples T-Test

		Statistic	df	p
perf_quantity	Student's t	1.43	104	0.157

Note. $H_a\ \mu_2 \neq \mu_1$

	Effect Size
Cohen's d	0.383
Rank biserial correlation	−0.206

ANOVA

- ANOVA wordt gebruikt om gemiddelden (μ's) van 3 of meer groepen voor een continue afhankelijke variabele te vergelijken, om te zien of er een significant verschil is tussen die groepen. Je moet 2 opties voor ANOVA kennen, waarvoor de keuze is gebaseerd op Levene's test van homogeniteit (data-check).
  - **ANOVA Fisher´s test**. Wordt gebruikt als er aan homogeniteit wordt voldaan (dus: Levene´s $p > 0.05$)
  - **ANOVA Welch test**. Wordt gebruikt als er **niet** aan homogeniteit wordt voldaan (dus: Levene´s $p < 0.05$)
- Je vindt ANOVA in Jamovi onder Analyses > ANOVA > One-way ANOVA.
- Om de homogeniteit te controleren vink je in Jamovi binnen de ANOVA aan *Homogeneity test* (Levene´s)

Maar.... als er een significant verschil wordt gevonden, moeten we nog steeds kijken tussen welke groepen dat verschil bestaat, waarvoor we post-hoc-tests gebruiken. Deze post-hoc-tests tonen p-waarden voor specifieke combinaties van groepen, zodat je de resultaten nauwkeuriger kan rapporteren. Om ook te kunnen rapporteren welke groepen hoger scoren dan anderen, bekijk je beschrijvende statistieken (*descriptives*) om de gemiddelen van de groepen te zien. Post-hoc-opties zijn:

- Voor de ANOVA Welch test gebruiken we de Games-Howell
- Voor de ANOVA Fisher's test gebruiken we de Bonferroni of Tukey, waarbij de keuze afhankelijk is van aanvaardbare risico's:
  - Bonferroni is vrij strikt, wat betekent dat de kans op een type I-fout kleiner is, maar de kans op een type II-fout groter is
  - Tukey is minder strikt, wat betekent dat de kans op een type I-fout groter is, maar de kans op een type II-fout kleiner. Jamovi heeft alleen Tukey.

Assumption Checks

Homogeneity of Variances Test (Levene's)

	F	df1	df2	p
perf_quantity	1.68	4	101	0.160

One-Way ANOVA

		F	df1	df2	p
perf_quantity	Welch's	8.80	4	28.6	<.001
	Fisher's	4.18	4	101	0.004

Tukey Post-Hoc Test – perf_quantity

		1	2	3	4	5
1	Mean difference	—	0.365	1.132	−0.439	−1.136
	p-value	—	0.734	0.156	0.742	0.324
2	Mean difference		—	0.767	−0.804	−1.501
	p-value		—	0.486	0.117	0.081
3	Mean difference			—	−1.571	−2.268
	p-value			—	0.024	0.014
4	Mean difference				—	−0.697
	p-value				—	0.789
5	Mean difference					—
	p-value					—

NOTITIES

Hotel Management School Maastricht

HMSM Team Research Skills (2025). *Beknopt maar significant handboek voor M5 statistiek* . Version 0725NL.