Onderzoek doen in de biologie

Welke statistische toetsen moet ik gebruiken?

Met de focus op het gebruik van Excel en de interpretatie van de waarden uit Excel.

Door Anne Flier & Karin Koens

16-12-2018

Inhoud

Inleiding

Zijn jongens nou echt beter in wiskunde dan meisjes? Je zou hiervoor alle cijfers van jongens en meisjes met elkaar kunnen vergelijken, maar hoe weet je zeker dat het resultaat dat jij vindt niet door toeval komt? Er komt dus heel veel bij kijken voordat je een zinnige uitspraak kunt doen over de resultaten van een onderzoek. Allereerst formuleer je een onderzoeksvraag. Aan de hand van de onderzoekvraag verzamel je data om je onderzoeksvraag te beantwoorden. Maar welke uitspraken mag je dan over deze data doen? Als je de ene groep mensen met de andere groep vergelijkt dan kunnen daar zeker verschillen uit komen, maar zijn deze verschillen **significant?** Of berusten ze op toeval?
In deze module zou ik je graag willen leren hoe je data kunt analyseren en hoe je vervolgens een conclusie trekt. Ik zal hier een tipje van de sluier oplichten als het gaat om het doen van statistisch onderzoek.

Bij een onderzoek heb je meestal drie type onderzoeksvragen die je kunt kiezen:

1. Je gaat zoeken naar een **verschil in gemiddelden.** Een onderzoeksvraag die in deze categorie past is bijvoorbeeld: *is er een verschil in het gemiddelde antibioticagebruik tussen kinderen tot tien jaar en mensen ouder dan 65 jaar?* Je hebt hierbij dus heel duidelijk twee *discrete* groepen: een groep kinderen tot tien jaar en een groep mensen ouder dan 65 jaar. Je doet een T-toets (bladzijde 7).
2. Je kan ook zoeken naar een **verband.** Een vraag die hierbij past is: *is er een verband tussen het antibioticagebruik en de leeftijd van mensen*. Je hebt hier *non-discrete groepen*, ook wel een continue groep genoemd. Je hebt namelijk een groep die continue doorloopt: je onderzoekt maar 1 groep mensen van alle leeftijden. Bij dit type onderzoek doe je een Pearson correlatiecoëfficiënt (bladzijde 10).
3. Tenslotte heb je nog een andere manier waarbij je kunt zoeken naar een **verschil in aantal.** Bij dit type onderzoek ben je bij elke categorie aan het turven hoeveel waarnemingen in een categorie vallen. Je turft bijvoorbeeld hoeveel larven van een bepaalde soort vlieg je vindt op bepaalde hoogten. Als je dit onderzoek uitvoert, zie je dat je geen gemiddelden kunt berekenen. Een larve zit namelijk of wel of niet op een bepaalde hoogte. Zo heb je een absoluut getal per categorie en geen gemiddelde. Vaak zeg je bij dit type onderzoek dat je zoekt naar een **correlatie**. Belangrijk is hierbij dat je geen continue groep hebt, maar categorieën. Je doet een Chi-kwadraat toets (bladzijde 13).

Als je data statistisch wilt analyseren, is het belangrijk dat je de gevonden data overzichtelijk verzamelt. Het handigst is om je data in een tabel in Excel te zetten. Als je werkt met twee discrete groepen, dan kun je Excel het gemiddelde van je metingen bij elke groep laten bepalen om een indruk te krijgen van de resultaten. Als je werkt met non-discrete groepen dan kun je Excel vanuit je data een grafiek laten maken. Op deze manier krijg je al een indruk van de resultaten. Nu komen we natuurlijk wel weer terug bij de vraag: is je verschil of verband significant of gaat het hier om toeval? Om hier achter te komen, kun je een aantal statistische toetsen uitvoeren. In deze module laat ik je kennis maken met drie statistische toetsen. Elk geschikt voor één van de bovengenoemde manieren van onderzoek doen. Maar voor ik hierop verder ga, is het belangrijk dat je eerst de basis van statistiek snapt.

Als je geïnteresseerd bent in hoe je de statistische toetsen met de hand kunt uitrekenen, kun je een kijkje nemen in de bijlage. Hier staat in detail uitgelegd hoe je alles met de hand kunt uitrekenen.

De onderzoeksvraag en de hypothesen

Een essentieel onderdeel bij het doen van onderzoek is de onderzoeksvraag. Tot nu toe heb je misschien vooral geleerd dat je onderzoeksvraag een open vraag moet zijn, maar als je onderzoeken gaat lezen kom je heel vaak de volgende type vraag tegen: *heeft medicijn A gemiddeld meer genezingen tot gevolg dan medicijn B?* Zo'n vraag is te beantwoorden met "ja" of "nee". Op deze vraag mag je echter alleen antwoord geven als je ook daadwerkelijk een statistische analyse op je data uitvoert. Dit is om te controleren of jouw antwoord berust op toeval of niet.

Bij een onderzoeksvraag horen ook hypothesen. In de statistiek zijn er twee type hypothesen:

- **H0 = Nul hypothese.** Deze hypothese zegt altijd dat er *geen* verschil is of *geen* verband tussen twee variabelen.
- **HA = Alternatieve hypothese.** De alternatieve hypothese wordt pas aangenomen als duidelijk is dat de nulhypothese mag worden verworpen. De alternatieve hypothese zegt namelijk dat er wel duidelijk een verschil of een verband is.

H0 en HA hypothesen kun je ook wel vergelijken met de rechtspraak in Nederland. Iemand is onschuldig tot het tegendeel is bewezen. H0, er is geen verschil, wordt behouden tot het tegendeel wordt bewezen. Zie tabel 1.

Tabel 1: rechtszaak en H0/HA

Voorbeeld rechtszaak

	Vrijspraak	Veroordeeld
Verdachte onschuldig	Goed	Fout type 1

Algemeen voorbeeld

	H0 wordt behouden	H0 wordt verworpen
H0 is waar	Goed	A

Maar net als in een rechtszaak kun je in onderzoek ook fouten maken. Het mag niet gebeuren dat H0 wordt verworpen terwijl hij eigenlijk waar is (dus als een onschuldig iemand wordt veroordeeld). Daarom wordt van tevoren bepaald hoe groot de kans dat je onterecht de H0 verwerpt, mag zijn. Deze is vastgesteld op de waarde **α = 5% = 0,05**. In de statistiek nemen we dan het risico dat we onterecht de H0 verwerpen. Dit risico is 5%. Ter illustratie: als je bij 100 vrouwen die zwanger zijn een zwangerschapstest afneemt, en deze zwangerschapstest is het resultaat van klinische testen waarvan de geldigheid statistisch getoetst is, dan kan het zijn dat bij vijf vrouwen de zwangerschap niet aangetoond wordt met deze test.

P-waarde

In de statistiek kun je duidelijk voor elke soort test het toetsingscriterium uitrekenen en de daarbij behorende p-waarde opzoeken. De waarde α wordt ook wel een **significante p-waarde** genoemd. Internationaal is afgesproken dat een p-waarde van 0,05 als significant wordt gezien.
De p-waarde geeft de kans aan dat de H0 waar is. Als je een p-waarde hebt van 0,8, is dus 80% kans dat H0 waar is en 20% kans dan HA waar is. Als je een p-waarde hebt van 0,05, is er dus 5% kans dat H0 waar is en een 95% kans dat HA waar is. Internationaal is afgesproken dat we bij een **p-waarde van 0,05 of kleiner** mogen zeggen dat HA waar is.
Op dit punt wordt er geaccepteerd dat er een verschil is: H0 wordt verworpen en HA aangenomen. Er wordt nu gezegd dat er een **significant** verschil is tussen de gemiddelden van twee groepen of dat het verband significant is. Als je een p-waarde hebt gevonden van 0,06, zeg je dat er *geen* verschil is. Je zit namelijk boven de grens van 0,05. Met een statistische toets, die onder andere Excel kan uitvoeren, kan bepaald worden wat de p-waarde is bij een bepaalde steekproef.
Als je wilt weten hoe statistici tot een grenswaarde van 0,05 zijn gekomen, dan raad ik je aan om de bijlage te bekijken op bladzijde 18.

Gemiddelde, standaardafwijking en standaardfout in Excel

Excel is tegenwoordig echt een uitkomst voor het doen van basisstatistiek. Met de hand hoef je bijna niets meer uit te rekenen, dus haal daar ook je voordeel uit. Als we teruggaan naar het onderzoek naar het antibioticagebruik, onderzoeken we twee groepen mensen. De groep kinderen jonger dan 10 jaar en de groep mensen ouder dan 65 jaar. We doen eerst een steekproef bij de groep kinderen jonger dan 10 jaar. Stel dat je na een kleine steekproef de gegevens verzameld hebt, zoals in tabel 2 te zien is.

Tabel 2: gegevens antibioticagebruik bij steekproef van kinderen jonger dan 10 jaar

Ondervraagde patiënt nummer	Aantal kuren genomen in afgelopen jaar door kinderen tot 10 jaar
1	1
2	4
3	2
4	1
5	2
6	3
7	5
8	0
9	2
10	3
11	1
12	4
13	2
14	1

Tegenwoordig kun je Excel alles laten uitrekenen, daarom heb ik hier op een rijtje gezet welke commando's je Excel kunt geven om het gemiddelde, de standaardafwijking en de standaardfout uit te laten rekenen. De waarden die cursief gedrukt staan, kunnen in jouw bestand verschillen. Ze geven alleen aan welke cellen Excel moet lezen.

- Gemiddelde berekenen: =GEMIDDELDE(*A2:A15*) = 2,21
- Standaardafwijking berekenen: =STDEV.S(*A2:A15*) = 1,43
- Standaardfout berekenen: =STDEV.S(*A2:A16*)/WORTEL(aantal(*A2:A15*)) = 0,38

Maar wat heb je dan aan het gemiddelde, de standaardafwijking en de standaardfout?
Het gemiddelde is altijd handig om te berekenen als je zoekt naar een verschil in gemiddelden. Het geeft je een eerste indruk over je resultaten.
De **standaardafwijking** geeft weer wat de **spreiding** is van je gegevens. Als je een grote standaardafwijking hebt, dan liggen je gegevens niet dicht bij elkaar. Als je een kleinere standaardafwijking hebt in je gegevens, dan liggen je waarnemingen wel dicht bij elkaar. Een kleinere spreiding geeft vaak sneller een significant resultaat. De spreiding kun je ook weergeven in een interval. Deze is gelijk aan het gemiddelde ± de standaardafwijking. In bovenstaand voorbeeld heb je het volgende interval: 2,21 ± 1,43 → [0,78; 3,64].
Verder heb je nog de standaardfout. De standaardfout wordt gebruikt om het 95% **betrouwbaarheidsinterval** te berekenen. Als je de steekproef zou herhalen, dan zou 95% van de gemiddelde meetwaarden in dit gebied liggen. Het **95% betrouwbaarheidsinterval** is te berekenen door de volgende formule: $1{,}96 \cdot SE$. Wij zagen dat SE gelijk is aan 0,38, daarom krijg je volgende berekening: $1{,}96 \cdot 0{,}38 = 0{,}74$. Dit geeft het volgende interval 2,21 ± 0,74 → [1,47; 2,95]. Dit betekent dat, mocht je de proef een aantal keer herhalen, 95% van de nieuwe gemiddelden in dit interval zullen liggen.

Twee-zijdig/tailed en Een-zijdig/tailed

In het vervolg van deze module zal je nog een aantal termen tegenkomen: **two-tailed** en **one-tailed.** Oftewel **tweezijdig** en **eenzijdig**. Deze begrippen hebben te maken met de manier waarop je onderzoek doet en waarop je vraag is geformuleerd. Je zult deze begrippen namelijk tegenkomen in Excel als je een statistische analyse gaat uitvoeren en als je tabellen gebruikt om te bepalen of de analyse in je onderzoek een significant verschil/verband oplevert.
Bij een tweezijdige onderzoeksvraag heb je nog geen idee welke richting je antwoord op zal gaan. Je zoekt alleen naar een verschil of een verband. Je vermoedt niet dat de ene groep meer antibioticakuren gebruikt dan de andere groep. Of in het geval van het verband: je weet nog niet of er een positief of negatief verband zal zijn.
Als je een eenzijdige onderzoeksvraag hebt, dan heb je al wel een vermoeden. Je denkt nu wel dat het gebruik van bijvoorbeeld antibioticakuren hoger zal liggen in de ene groep. Of je vermoedt dat er een positief verband is tussen twee variabelen. Je hebt een richting gegeven aan je vraag.
Het doen van een tweezijdige of eenzijdige test heeft invloed op de grens waarop je resultaat **significant** wordt.

T-toets

Een t-toets doe je als je zoekt naar een **verschil** tussen de **gemiddelde waarden** van twee discrete groepen.
Bij het doen van t-toetsen zijn er twee mogelijkheden waarop je data is verzameld:

- **Gepaarde data.** Als je gepaarde data hebt verzameld dan onderzoek je *dezelfde groep* alleen dan op een ander moment of onder andere condities. Bijvoorbeeld: je onderzoekt de reactiesnelheid van mensen voordat ze een mok koffie hebben gedronken (groep A) en bij dezelfde mensen onderzoek je ook hun reactiesnelheid nadat ze een mok koffie hebben gedronken (groep B). Groep A en B bevatten exact dezelfde mensen.
- **Ongepaarde data.** Bij ongepaarde data doe je onderzoek bij verschillende groepen mensen. Het voorbeeld waarbij we het gemiddelde antibioticagebruik van kinderen tot tien jaar (groep A) en mensen ouder dan 65 jaar (groep B) onderzoeken, is een voorbeeld met ongepaarde data. In dit geval bestaat groep A niet uit dezelfde mensen als groep B.

Het onderzoek

Het onderzoek waar wij een statistische analyse op gaan loslaten, is het eerdergenoemde onderzoek over het antibioticagebruik tussen twee discrete groepen mensen. Zoals je net gelezen hebt, zijn dit twee verschillende groepen en bevatten deze groepen niet dezelfde mensen. Het gaat hier dus om ongepaarde data.

- De onderzoeksvraag: *is er een verschil in het gemiddelde aantal antibioticakuren van minstens een week in het afgelopen jaar tussen kinderen tot tien jaar en mensen ouder dan 65 jaar?*
    - H0 = er is geen verschil in het *gemiddelde* aantal antibioticakuren van minstens een week in het afgelopen jaar tussen kinderen tot tien jaar en mensen ouder dan 65 jaar.
    - HA = er is wel een verschil in het *gemiddelde* aantal antibioticakuren van minstens een week in het afgelopen jaar tussen kinderen tot tien jaar en mensen ouder dan 65 jaar.

De gevonden data ziet er zo uit:

Tabel 3: gegevens onderzoek antibioticakuren steekproef 1 en 2

Ondervraagde patiënt nummer	Aantal antibioticakuren van steekproef 1: tot 10 jaar	Ondervraagde patiënt nummer	Aantal antibioticakuren van steekproef 2: vanaf 65 jaar
1	1	15	5
2	4	16	3
3	2	17	6
4	1	18	4
5	2	19	1
6	3	20	3
7	5	21	1
8	0	22	2
9	2	23	0
10	3	24	1
11	1	25	4
12	4	26	3
13	2	27	3
14	1	28	2
	$\mu = 2{,}214286$		$\mu = 2{,}714286$

We zien een verschil in de gemiddelden: de ouderen gebruiken gemiddeld meer antibioticakuren dan de kinderen, maar is dit een toevallig verschil of een significant verschil?

Excel

Excel kan de T-toets voor jou uitvoeren. Belangrijk hierin is, dat je snapt welke code je moet intikken en wat de waarde die Excel geeft, vervolgens betekent.

=T.TOETS(matrix1;matrix2;zijden;type_getal)

- **matrix1** = gegevens van steekproef 1
- **matrix2** = gegevens van steekproef 2
- **zijden** = hier gaat het om tweezijdig of eenzijdig. In het geval dat hierboven is beschreven, hebben we geen richting gegeven aan onze vraag. We zochten alleen naar een verschil en verwachtten niet dat het gebruik in de ene groep hoger is dan in de andere. Daarom kies je voor tweezijdig.
    - 1 = eenzijdig
    - 2 = tweezijdig.
- **type_getal** = gepaard of ongepaard. Waarbij je eigenlijk altijd optie 1 of 3 kiest. Met optie 2 hoef je nog niet te werken. In het geval van ons voorbeeld hebben we twee verschillende groepen, dus “ongepaard”, “3” is het beste om in te vullen.
    - 1 = gepaard.
    - ~~2 = ongepaard met gelijke variantie~~
    - 3 = ongepaard met ongelijke variantie.

Als je de t.toets laat uitvoeren door Exel dan krijg je direct een p-waarde als antwoord. In Excel ziet dat er zo uit:

ALS | =T.TOETS(B2:B15;D2:D15;2;3)

	A	B	C	D	E
1	Ondervraagde patiënt nummer	Aantal antibioticakuren van steekproef 1: tot 10 jaar	Ondervraagde patiënt nummer	Aantal antibioticakuren van steekproef 2: vanaf 65 jaar	
2	1	1	15	5	
3	2	4	16	3	
4	3	2	17	6	
5	4	1	18	4	
6	5	2	19	1	
7	6	3	20	3	
8	7	5	21	1	
9	8	0	22	2	
10	9	2	23	0	
11	10	3	24	1	
12	11	1	25	4	
13	12	4	26	3	
14	13	2	27	3	
15	14	1	28	2	
16					
17	p-waarde	;D2:D15;2;3)			

Figuur 1: uitwerking van de ongepaarde T-toets in Excel

De p-waarde die daaruit komt rollen is: **p = 0,404.** Dus ondanks dat de gemiddelden duidelijk verschillend waren, verschilden ze niet genoeg van elkaar om te concluderen dat er een significant verschil is. De p-waarde is namelijk een stuk hoger dan de grenswaarde van 0,05.

Een goed geformuleerde conclusie is dan: *er is geen (significant) verschil in het gemiddelde aantal kuren van minstens een week in het afgelopen jaar tussen kinderen tot tien jaar en mensen ouder dan 65 jaar (p=0,40).* Je ziet dat deze conclusie heel erg lijkt op de onderzoeksvraag, maar dat er een paar aanpassingen zijn gedaan “geen (significant) verschil” en de p-waarde is tussen haakjes toegevoegd.
u
Als je een grotere steekproef uitvoert, kunnen de resultaten leiden tot een verschil dat wel significant blijkt te zijn. Dit kun je simuleren door de bovenstaande tabel een aantal keer onder elkaar te kopieren, zodat je 7 keer onder elkaar dezelfde tabel hebt en je steekproef dus 7 keer zo groot is maar de gemiddelden hetzelfde. Dan krijg je de volgende p-waarde na het uitvoeren van de t.toets: p = 0,02. Er is dan **wel een significant verschil**, terwijl de gemiddelden nog evenveel van elkaar verschillen. H0 mag nu verworpen worden en de volgende conclusie mag worden getrokken: *er is wel een verschil in het gemiddelde aantal antibioticakuren van minstens een week in het afgelopen jaar tussen kinderen tot tien jaar en mensen ouder dan 65 jaar (p=0,02).* Als je deze conclusie in je onderzoek trekt, zet je er altijd tussenhaakjes achter om welke p-waarde het ging.

Pearson's correlatiecoëfficiënt

Het onderzoek

De Pearson's correlatiecoëfficiënt wordt gebruikt om te zoeken naar een verband. Hierbij is de onafhankelijke variabele een numerieke continue reeks data. Zoals altijd is een goede onderzoeksvraag essentieel. Een voorbeeld van een vraag waarbij je de Pearson's correlatiecoëfficiënt gaat gebruiken zou zijn:

- *Is er een verband tussen de leeftijd van een mens en het aantal antibioticakuren van minstens een week die hij of zij in het afgelopen jaar heeft gehad?*
    - H0 = er is geen verband tussen de leeftijd van een mens en het aantal antibioticakuren van minstens een week die hij of zij in het afgelopen jaar heeft gehad.
    - HA = er is wel een verband tussen de leeftijd van een mens en het aantal antibioticakuren van minstens een week die hij of zij in het afgelopen jaar heeft gehad.

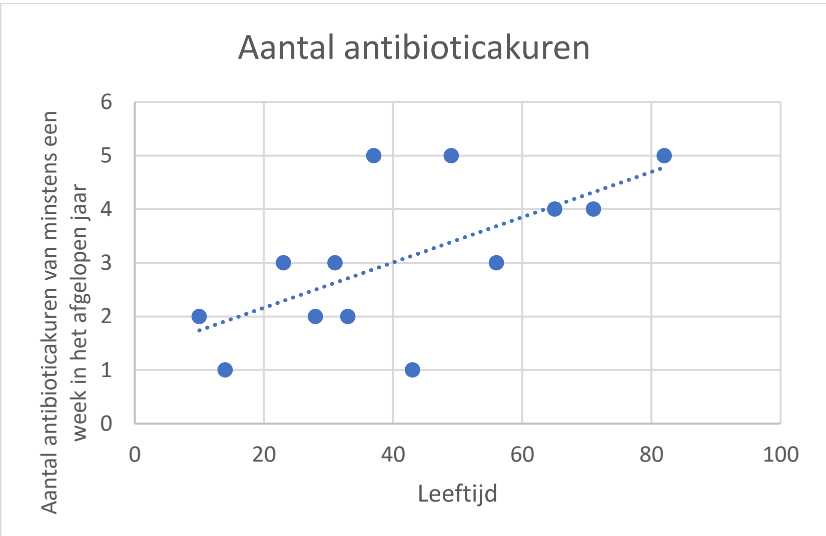
Het uitvoeren van een Pearson's correlatiecoëfficiënt wordt ook wel een **regressieanalyse** genoemd. De waarde die je krijgt wanneer je een regressieanalyse uitvoert, wordt de r-waarde genoemd. Deze kan tussen -1 en 1 liggen. Waarbij -1 een perfect negatief verband laat zien (als x groter wordt dan wordt y kleiner) en +1 een perfect positief verband laat zien (als x groter wordt dan wordt y groter). Aan de hand van de r-waarde kun je zeggen of je een positief verband of een negatief verband hebt:

- r-waarde < 0: negatief verband als je een **p-waarde hebt van 0,05 of kleiner!**
- r-waarde > 0: positief verband als je een **p-waarde hebt van 0,05 of kleiner!**

Bij het uitvoeren van de steekproef die bij de bovengenoemde vraag hoort, krijg je de gegevens die te zien zijn in tabel 4. Zoals je ziet, hebben we een steekproef genomen van 13 mensen en aan elk persoon de leeftijd gevraagd en vervolgens het aantal antibioticakuren van minstens een week in het afgelopen jaar genoteerd. Met behulp van Excel kun je bij deze tabel een grafiek maken. Die ziet eruit zoals in figuur 2 te zien is.

Tabel 4: gegevens antibioticakuren

Leeftijd van patiënt	Aantal antibioticakuren …
10	2
23	3
14	1
65	4
49	5
33	2
56	3
43	1
37	5
28	2
31	3
71	4
82	5



Figuur 2: grafische weergave van tabel 4

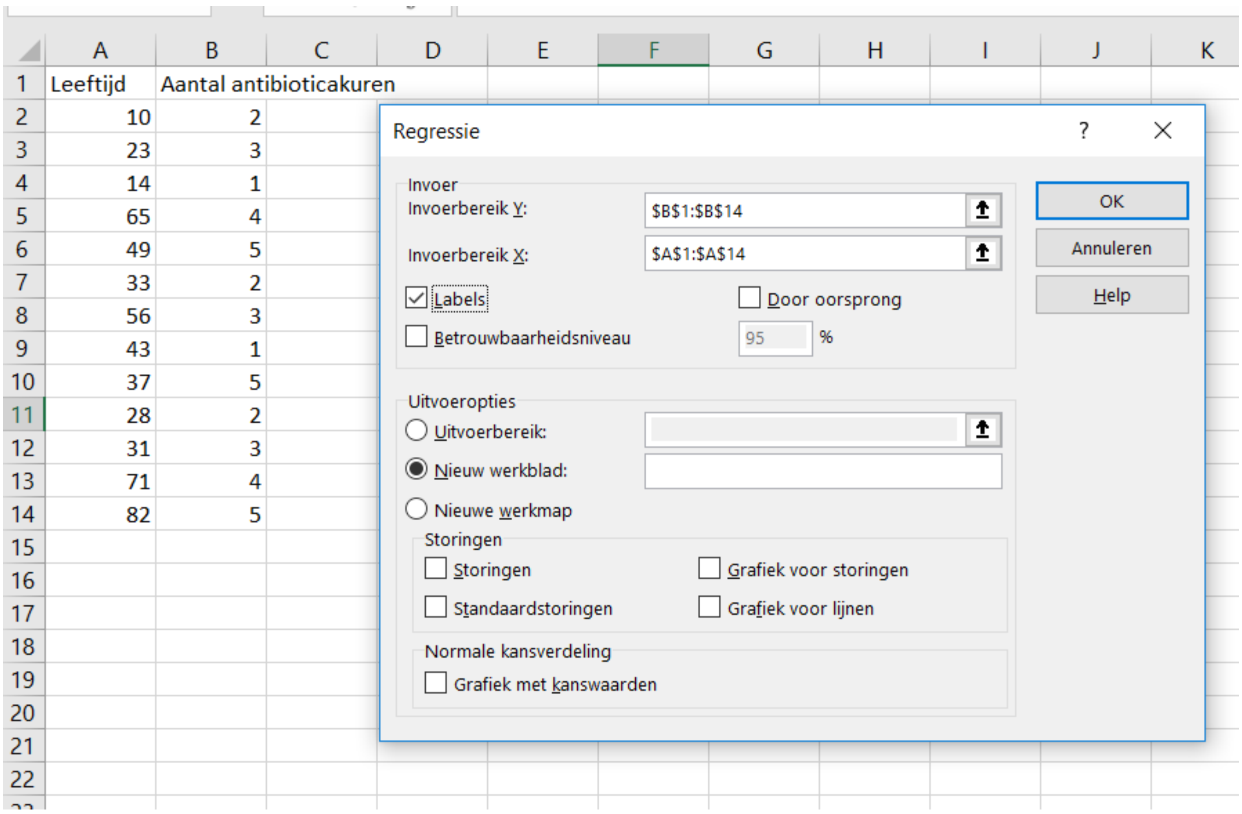
Er lijkt dus een positief verband aanwezig te zijn, want als de leeftijd omhooggaat dan wordt het aantal antibioticakuren ook hoger. Met Excel kun je vervolgens uitrekenen over dit verband ook significant is en niet berust op toeval.

Excel

In Excel kun je geen commando voor de Pearson's correlatiecoëfficiënt typen, zoals bij de T.toets. Je moet een *gegevensanalyse* uitvoeren, maar voordat je dit kunt, moet je deze toepassing eerst zichtbaar maken.

Dit doe je door: Bestand → Opties → Invoegtoepassingen → onderin klik je op "beheren" → Excel invoegtoepassingen → uit de lijst van inactieve invoegtoepassingen klik je op "Analysis Toolpak" → Start.

Nu staat onder het tabblad "gegevens" helemaal rechts in het blok "analyse", "gegevensanalyse". Klik hierop. In het boxje functies scrol je door totdat je "regressie" ziet staan. Dan "ok".
Bij "Invoerbereik Y" selecteer je in dit voorbeeld "aantal antibioticakuren" en bij "Invoerbereik X" selecteer je "Leeftijd". Dit doe ik met de kopjes "Leeftijd" en "Aantal antibioticakuren" inbegrepen en omdat ik dat doe, moet ik het vakje labels aanklikken. Dit is te zien in figuur 3.



Figuur 3: regressieanalyse uitvoeren Excel

In een nieuw werkblad komt dan de samenvatting van de uitvoer van de analyse. Deze bevat veel informatie, waarvan wij maar een klein deel nodig hebben. De informatie die we nodig hebben, heb ik gekleurd in figuur 4. Het gaat om de **correlatiecoëfficiënt R** en natuurlijk de **p-waarde.**

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,644826763
R-kwadraat	0,415801555
Aangepaste kleinste kwadraat	0,362692605
Standaardfout	1,150494865
Waarnemingen	13

Variantie-analyse

	Vrijheidsgraden	*Kwadratensom*	*Gemiddelde kwadraten*	*F*	*Significantie F*
Regressie	1	10,36305414	10,36305414	7,829218	0,017334384
Storing	11	14,56002279	1,323638435		
Totaal	12	24,92307692			

	Coëfficiënten	*Standaardfout*	*T- statistische gegevens*	*P-waarde*	*Laagste 95%*	*Hoogste 95%*	*Laagste 95,0%*	*Hoogste 95,0%*
Snijpunt	1,315626325	0,705724959	1,864219634	0,089178	-0,237663837	2,868916486	-0,237663837	2,868916486
Leeftijd	0,042245125	0,01509793	2,798074018	0,017334	0,009014806	0,075475444	0,009014806	0,075475444

Figuur 4: resultaat van regressieanalyse Excel

De regressieanalyse geeft een Pearson’s correlatiecoëfficiënt, oftewel een r-waarde van 0,64, r = 0,64. Deze is dus boven 0. Er lijkt dus zeker een positief verband te zitten tussen de leeftijd en het aantal antibioticakuren. Om hier zeker van te zijn kijken we vervolgens nog naar de p-waarde. Deze p-waarde = 0,017. Dit is duidelijk lager dan 0,05. Het resultaat is **significant**. H0 wordt verworpen en de conclusie luidt: *er is wel een verband tussen de leeftijd van een mens en het aantal antibioticakuren van minstens een week die hij of zij in het afgelopen jaar heeft gehad (p=0,017).*
Als je deze conclusie trekt over je onderzoek, zet je er altijd tussen haakjes achter om welke p-waarde het gaat.

Chi-kwadraat toets

De laatste toets die ik met je wil bespreken, is de Chi-kwadraat toets. De onafhankelijke variabele of de beide variabelen is of zijn nominaal en kwalitatief. Je kunt de chi-kwadraat toets in twee situaties gebruiken:

1. Als je onderzoekt doet naar een verschil in de frequentieverdeling van aantallen.
2. Als je onderzoekt doet naar een correlatie.

In beide gevallen zul je aantallen turven en kun je geen gemiddelden berekenen. Beide wil ik uitleggen aan de hand van een voorbeeld.

Verschil in de frequentieverdelingen

Het onderzoek

We doen onderzoek naar een vliegensoort die zijn eieren legt in de bladeren van bomen. We verwachten dat de hoogte waarop deze bladeren hangen niet uitmaakt en dat er dus een gelijke verdeling zal zijn in het aantal larven dat we gaan vinden. Van tevoren bepalen we welke categorieën er zijn: 0 tot en met 2 meter boven de grond, 2 tot en met 4 meter boven de grond en 4 tot en met 6 meter boven de grond. We verwachten nu 1/3 van de larven te vinden in de categorie 0 tot 2 meter, 1/3 in de categorie 2 tot 4 meter en 1/3 in de categorie 4 tot 6 meter.

- Onze onderzoeksvraag: *is er een verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van de bladeren?*
    - H0 = er is geen verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van bladeren.
    - HA = er is wel een verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van bladeren.

Na een dag te hebben rondgelopen in het bos, heb je geturfd hoeveel larven je vindt op welke hoogte. Je hebt de volgende gegevens verzameld:

Tabel 5: verzamelde gegevens

	0 tot en met 2 meter	2 meter tot en met 4 meter	4 meter tot en met 6 meter	Totaal
Waarnemingen	93	86	54	243
Verwachting	81	81	81	243

Omdat je 243 waarnemingen hebt gedaan, kun je ook de rij verwachtingen in vullen, want dat is in elke kolom 243/3= 81.

De waarden voldoen niet aan onze verwachting, maar nu rest nog de vraag: is dit toeval of niet? Is het verschil groot genoeg?

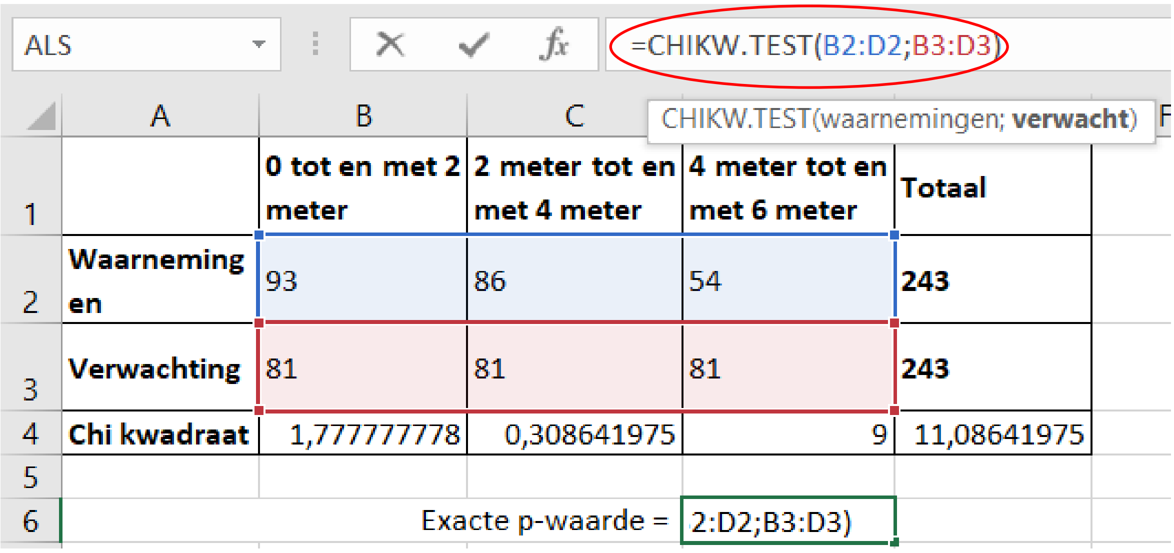
Excel

De exacte p-waarde is nu te bepalen met Excel. Het uitrekenen van de p-waarde kan Excel heel simpel met de volgende code:

```
=CHIKW.TEST(waarnemingen;verwacht)
```

- Waarbij “waarnemingen” de cellen zijn met waarnemingen die je hebt gedaan.
- En “verwacht” de cellen zijn met de verwachting.

We nemen in figuur 5 weer het voorbeeld met de larven van de vliegen, waarbij je roodomrand de code voor de chi-kwadraat toets ziet. De uitkomst (en dus de p-waarde) zal in cel D6 te zien zijn, als je op enter drukt.


ALS =CHIKW.TEST(B2:D2;B3:D3)

CHIKW.TEST(waarnemingen; **verwacht**)

	A	B	C	D	E
1		**0 tot en met 2 meter**	**2 meter tot en met 4 meter**	**4 meter tot en met 6 meter**	**Totaal**
2	**Waarnemingen**	93	86	54	**243**
3	**Verwachting**	81	81	81	**243**
4	**Chi kwadraat**	1,777777778	0,308641975	9	11,08641975
5					
6			Exacte p-waarde =	2:D2;B3:D3)	

Figuur 5: chi-kwadraat in Excel met voorbeeld van de larven van de vliegen

De **exacte p-waarde** is als je op enter drukt **0,0039.**
Deze p-waarde is kleiner dan 0,05 en is dus significant. Een goed geformuleerde conclusie bij deze vraag zou zijn: *er is een significant verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van de bladeren (p=0,0039).*

Correlatie

De chi-kwadraat test kun je ook goed gebruiken als je een **correlatie** onderzoekt. Het verschil met onderzoek waarbij het verband tussen een onafhankelijke en een afhankelijke variabele wordt onderzocht is dat in dat tweede geval (verband) rekening wordt gehouden met een oorzaak-gevolg situatie terwijl dat bij een correlatie niet het geval is. De onderzoeksvraag is dus anders geformuleerd. Dit probeer ik uit te leggen aan de hand van het volgende voorbeeld.

Het onderzoek

Je bent onderzoeker in Afrika en je bent benieuwd of het drinkgedrag tussen twee groepen giraffen die elk in een ander soort gebied leven verschillend is. Een groep giraffen leef op een open savanne en de andere groep giraffen leeft op een bosrijke savanne. Je wilt weten of er een correlatie is tussen de leefomgeving en het aantal keer dat ze op een dag gaan drinken.

- Onderzoeksvraag: *is er een correlatie tussen het leefgebied van giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken?*
    - H0 = er is geen correlatie tussen het leefgebied van giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken.
    - HA = er is wel een correlatie tussen het leefgebied van giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken.

Je volgt de giraffen gedurende een dag en krijgt de volgende waarnemingen:

Tabel 6: waargenomen gegevens op de savanne

		Aantal keer drinken per dag		
		meer dan 5 keer	minder dan 5 keer	Totaal
Leef-gebied	Giraffen in open savanne	6	14	**20**
	Giraffen in bosrijke savanne	18	12	**30**
	Totaal	**24**	**26**	**50**

Op basis van de totale waarden die dikgedrukt zijn weergeven in de tabel, kun je een verwachting opstellen volgens een kruistabel. Let goed op dat het totaal niet verandert.

Tabel 7: verwachte gegevens op basis van het totale aantal

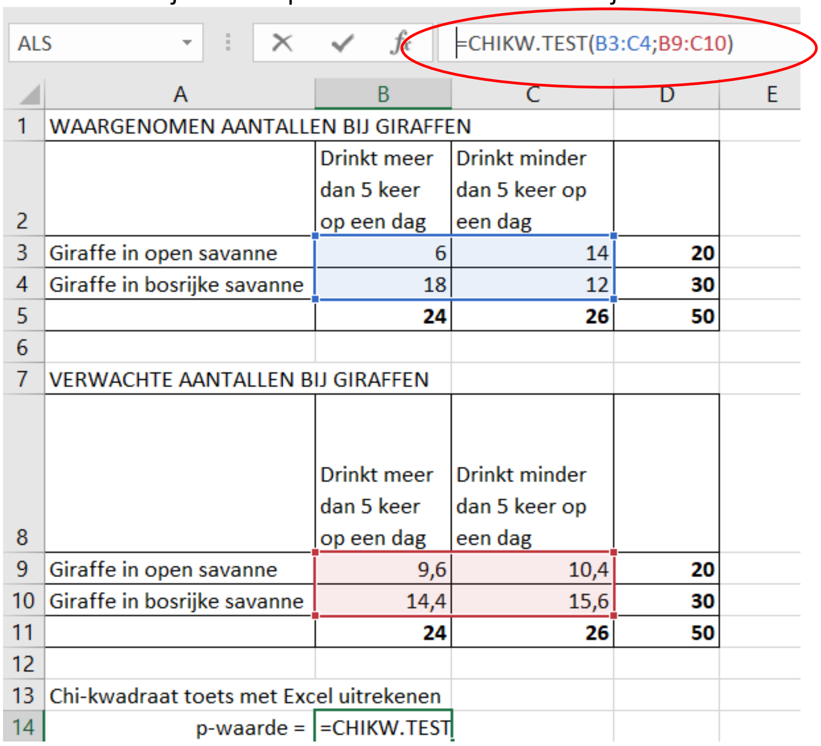
		Aantal keer drinken per dag		
		meer dan 5 keer	minder dan 5 keer	Totaal
Leef-gebied	Giraffen in open savanne	24*20/50=9,6	26*20/50=10,4	**20**
	Giraffen in bosrijke savanne	24*30/50 = 14,4	26*30/50 = 15,6	**30**
	Totaal	**24**	**26**	**50**

Excel

Als je al deze gegevens hebt verzameld dan zet je zoals je in figuur 6 is voor gedaan. Dan neem je een lege cel en begin je daar de volgende code te toetsen:

=CHIKW.TEST(waarnemingen;verwacht)

Op de plekken van "waarnemingen" en "verwacht" moeten cellen worden geselecteerd. In dit voorbeeld zie je de complete code in het rood omlijnde vak.


ALS | =CHIKW.TEST(B3:C4;B9:C10)

	A	B	C	D	E
1	WAARGENOMEN AANTALLEN BIJ GIRAFFEN				
2		Drinkt meer dan 5 keer op een dag	Drinkt minder dan 5 keer op een dag		
3	Giraffe in open savanne	6	14	**20**	
4	Giraffe in bosrijke savanne	18	12	**30**	
5		**24**	**26**	**50**	
6					
7	VERWACHTE AANTALLEN BIJ GIRAFFEN				
8		Drinkt meer dan 5 keer op een dag	Drinkt minder dan 5 keer op een dag		
9	Giraffe in open savanne	9,6	10,4	**20**	
10	Giraffe in bosrijke savanne	14,4	15,6	**30**	
11		**24**	**26**	**50**	
12					
13	Chi-kwadraat toets met Excel uitrekenen				
14	p-waarde =	=CHIKW.TEST			

Figuur 6: uitwerking van chi-kwadraat toets Excel met giraffen voorbeeld

Als je dan op enter drukt, dan komt de volgende p-waarde tevoorschijn: p = 0,037514. Deze p-waarde is kleiner dan 0,05 en daarom mag HA worden aangenomen.

Een goed geformuleerde conclusie die je hieruit kunt trekken is: *er is een correlatie tussen het leefgebied van de giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken (p=0,037).*

Opdrachten: Hoe werk ik in Excel?

Via een aantal opdrachten ga je nu proberen al deze theorie zelf te ontdekken en uiteindelijk onder de knie te krijgen. Het doel van deze opdrachten is vooral om je inzicht te geven in de mogelijke toetsen en hoe je de uitkomsten uit Excel interpreteert. Uiteindelijk is het de bedoeling dat je de statistiek kunt gebruiken bij het analyseren van je eigen onderzoekresultaten en je daar een juiste conclusie uit kunt trekken.
Bij het maken van onderstaande vragen, mag je terugkijken in de module.

Vraag 1

a) Bij welke p-waarde zeggen we normaal dat iets significant is? (Blz. 5)
b) Als je een p-waarde hebt van 0,051, hoe groot is de kans dat H0 dan waar is? Is er een significant resultaat? (Blz. 5)
c) Als je zoekt naar een verband. Welke toets ga je dan doen? Waarom deze toets? (Blz. 3)

Vraag 2

a) Je vraagt aan 21 leerlingen van je school hoe ze het vak biologie vorig jaar vonden en aan dezelfde leerlingen vraag je hoe ze het vak dit jaar vinden. Ze doen dit aan de hand van twee keer ongeveer dezelfde stelling: het vak biologie was vorig jaar uiterst interessant/het vak biologie is dit jaar uiterst interessant. De antwoorden worden gegeven op een schaal van 1 tot 5, waarbij 1 zeer oneens is en 5 zeer eens.
Wat is je onderzoeksvraag?
Wat is dan je nulhypothese en wat is je alternatieve hypothese? (Blz. 4)
b) Nienke berekent de gemiddelde beoordeling van de leerlingen per schooljaar. Welke toets gaat ze gebruiken? Waarom deze toets? (Blz. 3)
c) Hamza denkt dat de gemiddelde beoordeling niet representatief is en hij wil een andere toets doen waarbij hij turft hoeveel mensen “zeer oneens” waren. Hoeveel mensen “oneens” waren enz. Welke toets zou hij kunnen gebruiken? Waarom deze toets? (Blz. 3)

Vraag 3

Het gemiddelde aantal jongen dat na een jaar nog leeft van een groep huiskatten, wordt vergeleken met het gemiddelde aantal jongen dat na een jaar nog leeft van een groep wilde katten. Je denkt dat het gemiddelde aantal bij wilde katten lager zal liggen. Je neemt een steekproef van 30 bij zowel de groep huiskatten als bij de groep wilde katten. Je vindt voor de groep huiskatten een gemiddelde van 4,5. Voor de groep wilde katten vindt je het gemiddelde van 2,8.

a) Voor je een eenzijdige of een tweezijdige toets uit? Waarom? (Blz. 6)
b) Welke toets voer je uit? Waarom deze toets? (Blz. 3)
c) Je vindt na deze toets een p-waarde die kleiner is dan 0,0005. Geef een goed geformuleerde conclusie.

Vraag 4

Je hebt ergens gehoord dat je oren blijven groeien, ook als je volgroeid bent. Je neemt de proef op de som en doet een klein onderzoek. Je vindt de volgende waarden:

Leeftijd in jaren	Oorlengte in mm
38	65
57	75
75	74
92	81

a) Wat is de onderzoeksvraag? Wat is je nulhypothese en wat is je alternatieve hypothese? (Blz. 4)
b) Welke statistische toets ga je doen? Waarom deze toets? (Blz. 3)

c) Je vindt na deze toets een r-waarde van 0,922, maar een p-waarde van 0,08. Welke goed geformuleerde conclusie kun je nu trekken?
d) Hoe zou je het onderzoek beter kunnen maken?

Vraag 5

Tijdens de examens wordt aan een aantal leerlingen van jouw school gevraagd of ze koffiedrinken om ’s nachts wakker te blijven en door te kunnen leren. Van de 87 jongens die worden ondervraagd, blijkt 44 dit te doen. Van de 112 meisjes die zijn ondervraagd, blijkt 45 dit te doen.

a) Maak een tabel van je waarnemingen. Wat was je verwachting? Maak hier ook een tabel van. (Blz. 14)
b) Welke toets doe je nu? Waarom deze toets? Je vindt een p-waarde van 0,14. Is het resultaat significant? Geef een goed geformuleerde conclusie.

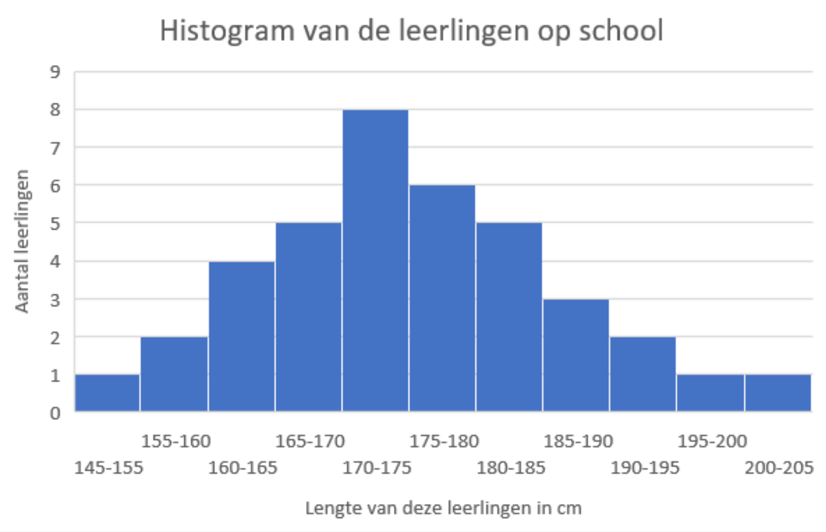
Bijlage 1

Waar komt de p-waarde grens van 0,05 vandaan?

Waar komt deze 0,05 eigenlijk vandaan? Hoe ziet dat eruit in een grafiek?
In het dagelijks leven blijkt dat heel veel gegevens zich normaal verdelen in een grafiek. Deze termen klinken misschien een beetje raar, maar ik zal ze uitleggen. We nemen het voorbeeld waarbij je de lengte van de leerlingen uit je jaarlaag gaat opmeten. Je meet bij 38 mensen uit de bovenbouw van vwo de lichaamslengte op en deelt ze onder in categorieën. Het zou dan goed kunnen dat je een tabel krijgt die erg lijkt op tabel 8. Als je van deze tabel een staafdiagram maakt, een zogenaamd **histogram,** dan ziet die eruit als figuur 7.

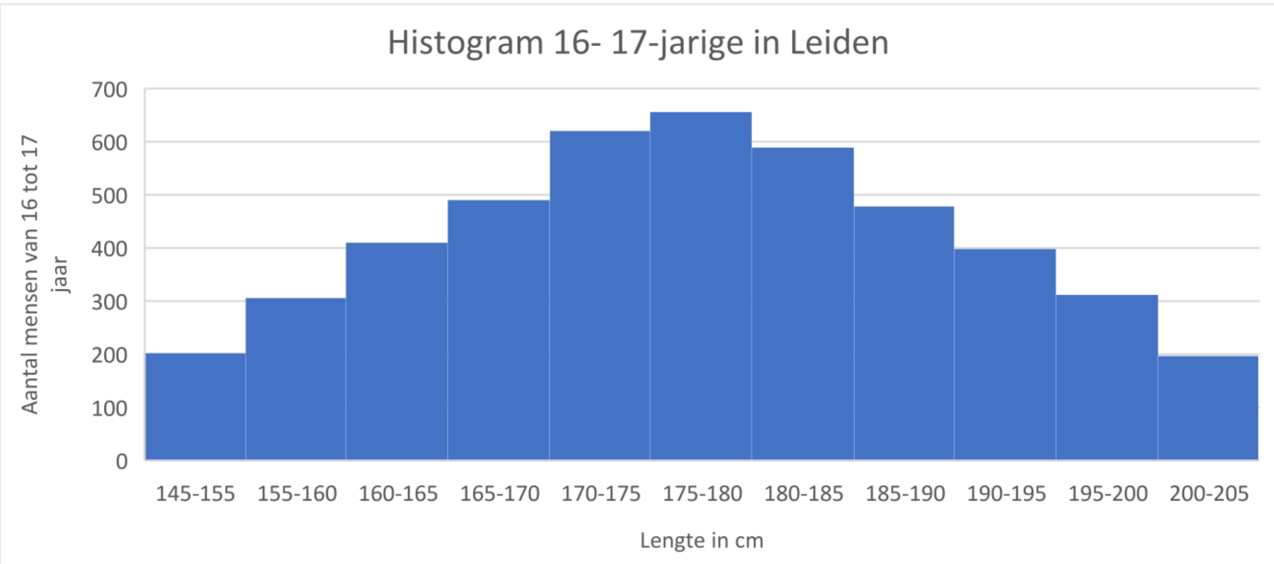
Tabel 8: gegevens van leerlingen op school

Lichaams-lengte	Aantal leerlingen
145-155	1
155-160	2
160-165	4
165-170	5
170-175	8
175-180	6
180-185	5
185-190	3
190-195	2
195-200	1
200-205	1



Figuur 7: histogram van de leerlingen op school

Je ziet hier duidelijk dat de meeste leerlingen een lengte hebben van 170 cm tot 175 cm. Grotere en kleinere leerlingen komen ook zeker voor, maar minder hoog in frequentie. Het feit dat er geen uitschieters zijn en de grafiek vanaf de categorie met de meeste leerlingen naar beiden kanten langzaam afloopt, doet ons sterk vermoeden dat deze gegevens **normaal verdeeld** zijn.
Als je jouw doelgroep, of eigenlijk je **steekproef** uitbreidt naar iedereen in Leiden van 16 tot 17 jaar dan krijg je misschien figuur 8.

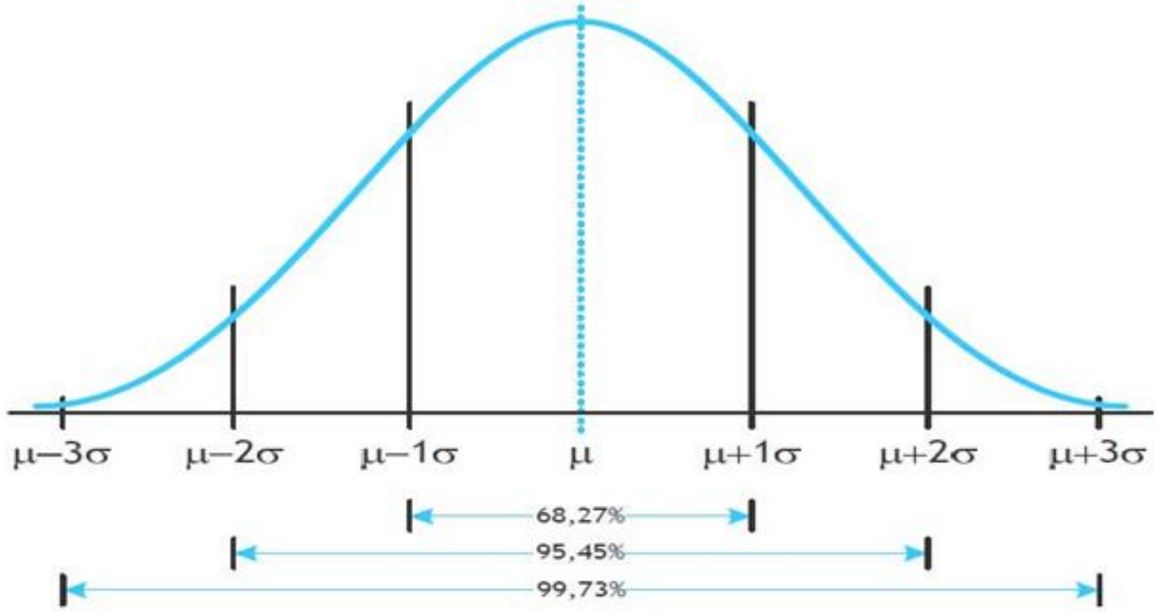


Figuur 8: histogram van 16- 17-jarige in Leiden

Deze figuur zegt dat de gemiddelde lengte tussen 175 cm en 180 cm ligt. Hij is dus verschoven toen je de steekproef vergrootte. Verder oogt dit histogram erg symmetrisch. Net als bij figuur 7 loopt de grafiek in figuur 8 vanaf de categorie met de meeste personen langzaam af naar beneden. We zeggen ook nu dat de waarden **normaal** verdeeld zijn.

Maar wat zegt een normaalverdeling?
Met een normaalverdeling kun je heel veel uitspraken doen over de kans dat je waarden wel of niet in deze grafiek liggen. Ook het begrip **standaardafwijking** heeft hier veel betrekking op. De standaardafwijking ligt precies op het buigpunt van de grafiek en geeft een maat over de spreiding van je waarnemingen. Als je een grote standaardafwijking hebt, dan heb je waarnemingen die veel van elkaar verschillen. Als je een kleine standaardafwijking hebt, dan liggen de waarnemingen dichter bij elkaar.

Een normaalverdeling zonder ook maar enige getallen ziet er zo uit:



Figuur 9: normaalverdeling (HBO statistiek[1])

Waarin **μ** het gemiddelde is en **σ** de standaardafwijking. Vanuit hier zijn er een paar regels opgesteld die altijd gelden:

- Ongeveer 68% van de waarden liggen tussen μ ± σ
- Ongeveer 95% (!) van de waarden liggen tussen μ ± 2σ
- Ongeveer 100% van de waarden liggen tussen μ ± 3σ

Het uitroepteken is je vast wel opgevallen. Deze heb ik hier geplaatst vanwege onze p-waarde van 0,05 oftewel 5%. 5% van alle gegevens valt dus buiten μ ± 2σ. Het gebied hierbuiten wordt ook wel het **kritieke gebied** genoemd. Links van het gemiddelde ligt dus 2,5% kritiek gebied en rechts van het gemiddelde ligt 2,5% kritiek gebied.

[1] HBO statitiek, Normale verdeling n z-score. http://www.hbostatistiek.nl/normale-verdeling/normale-verdeling-en-z-score/ Opgeroepen op 23 november 2018

Gemiddelde, standaardafwijking en standaarderror met de hand

We pakken het voorbeeld er weer bij dat wordt gegeven op bladzijde 3. We hebben een steekproef gedaan bij een groep kinderen tot tien jaar en hen gevraagd hoeveel antibioticakuren van minstens een week zij het afgelopen jaren hebben gehad. We verzamelen de volgende gegevens:

Tabel 9: gegevens antibioticagebruik bij steekproef van kinderen jonger dan 10 jaar

Ondervraagde patiënt nummer	Aantal kuren genomen in afgelopen jaar door kinderen tot 10 jaar
1	1
2	4
3	2
4	1
5	2
6	3
7	5
8	0
9	2
10	3
11	1
12	4
13	2
14	1

Bij zo'n kleine steekproef kan je het gemiddelde, de variantie, standaardafwijking en standaardfout nog met de hand berekenen, daarom zal ik dat ook kort toelichten.

- Het gemiddelde bereken je door alle waarden bij elkaar op te tellen en te delen door het aantal waarden. Het universele teken voor gemiddelde is µ. Bij tabel 9 is het gemiddelde 2,21.
- De **standaardafwijking** (σ) is de wortel van de variantie (s^2). De standaardafwijking is een maat voor de **spreiding** van al je gevonden waarden. Als de standaardafwijking klein is, dan heb je veel waarnemingen die erg op elkaar lijken. Bij een grote standaardafwijking zijn je waarnemingen erg verschillend van elkaar.

  $$\sigma = \sqrt{s^2}$$
- De variantie (s^2) berekenen, is vervolgens iets lastiger. Deze gebruik je als je een **steekproef** hebt genomen. En dat is in de biologie vrijwel altijd het geval. Voor s^2 (de variantie) geldt dan:

  $s^2 = \frac{\sum(x - \mu)^2}{n-1}$, dus de standaardafwijking: $\sigma = \sqrt{\frac{\sum(x - \mu)^2}{n-1}}$
  - Hierbij staat s^2 dus voor de variantie.
  - Het ∑ teken betekent "som van". Je moet hier dus voor elke meting apart (x) de afwijking van het gemiddelde (µ) berekenen. Voor de eerste meting geldt dan 2 – 1,35 = 0,65 en dit moet je daarna in het kwadraat doen ($0{,}65^2 = 0{,}42$). Al deze metingen die al in het kwadraat zijn moet je bij elkaar optellen.
  - Tenslotte moet je wat je net hebt berekend, delen door het aantal waarnemingen – 1, dus in dit geval delen door 13.

    Als je al deze stappen uitvoert voor tabel 9 krijg je de volgende variantie (s^2): 2,03. Als je hier de wortel van trekt heb je dus een standaardafwijking (σ) van 1,43.
- Daarna kun je hier nog de standaardfout van uitrekenen. De standaardfout wordt ook wel de standaard error (SE) genoemd. Waarbij *n* het aantal waarnemingen is.

$$SE = \frac{\sigma}{\sqrt{n}}$$

Als je dit uitrekent voor bovenstaand voorbeeld krijg je: $\frac{1{,}43}{\sqrt{15}} = 0{,}37$.

SE heb je nodig om het 95% betrouwbaarheidsinterval te berekenen. Als je de steekproef zou herhalen, dan zou 95% van de gemiddelde meetwaarden in dit gebied liggen. Het **95% betrouwbaarheidsinterval** is te berekenen door de volgende formule: $1{,}96 \cdot SE$. Wij zagen dat SE gelijk is aan 0,37, daarom krijg je volgende formule: $1{,}96 \cdot 0{,}37 = 0{,}73$.

In het geval van bovenstaand voorbeeld krijg je dan: 2,21 ± 0,73. Dit betekent dat, mocht je de proef een aantal keer herhalen, 95% van de nieuwe gemiddelden in het interval [1,48; 2,94] ligt. Als je het aantal waarnemingen in je steekproef vergroot (dus *n*), dan wordt SE kleiner. Je betrouwbaarheidsinterval wordt steeds nauwer en daarmee steeds betrouwbaarder.

Vrijheidsgraden/*degrees of freedom*

Bij het uitvoeren van statistische toetsen met de hand is het verder belangrijk om te weten hoeveel vrijheidsgraden (of *degrees of freedom*) je hebt. Dit heb je namelijk nodig als je tabellen gebruikt om te bepalen of de analyse in je onderzoek een significant verschil/verband oplevert. Het begrip “vrijheidsgraden” wordt gebruikt om aan te geven hoeveel getallen je vrij kunt invullen als je het totale aantal waarnemingen weet. Om dit duidelijker te maken, heb ik hier een voorbeeld.

Tabel 10: voorbeeldtabel voor het aantal vrijheidsgraden (1.1)

?	?	**10**
?	?	**20**
16	**14**	**30**

De dikgedrukte getalen zijn de totale waarden. Deze staan vast en kunnen niet veranderen.
De vraag is nu, hoeveel van de vraagtekens kun je vrij invullen?

Tabel 11: voorbeeldtabel voor het aantal vrijheidsgraden (1.2)

4	? → 6	**10**
? → 12	? → 8	**20**
16	**14**	**30**

Als je alleen de “4” invult, kun je met behulp van de totale waarden (dikgedrukt) bepalen welk getal de vraagtekens moeten aannemen.

Tabel 12: voorbeeldtabel voor het aantal vrijheidsgraden (1.3)

? → 3	7	**10**
? → 13	? → 7	**20**
16	**14**	**30**

Hetzelfde geldt als je alleen het getal “7” laat staan. Je kunt dan ook met behulp van de totale waarden (dikgedrukt) bepalen welk getal de vraagtekens moeten aannemen.

In <u>tabel 10, 11 en 12</u> kun je dus maar **1 getal vrij invullen**. Het aantal vrijheidsgraden/*degrees of freedom* is gelijk aan 1.

Bij een grotere tabel met je iets beter opletten, maar het wordt niet veel lastiger.

Tabel 13: voorbeeldtabel voor het aantal vrijheidsgraden (2.1)

?	?	?	?	**60**
?	?	?	?	**51**
42	**32**	**12**	**25**	**111**

Je hebt hier nu 8 onbekende getallen. Als je er één invult, dan krijg je maar één extra getal te weten. Als je er twee invult, krijg je maar twee extra getallen te weten. Als je er dan drie invult, dan krijg je ze alle acht te weten.

Tabel 14: voorbeeldtabel voor het aantal vrijheidsgraden (2.2)

16	20	10	? → 14	**60**
? → 26	? → 12	? → 2	? → 11	**51**
42	**32**	**12**	**25**	**111**

Het **aantal vrijheidsgraden** bij tabel 13 en 14 is dus **3.** Je kunt hier namelijk 3 getallen vrij invullen, dan worden de overige 5 bepaald.

T-toets met de hand

Ongepaard

Als je niet heel veel data hebt, dan kun je de gegevens met de hand uit rekenen. We gaan hier verder met dezelfde tabel waarnemingen als bij het voorbeeld in Excel. De onderzoeksvraag was: *is er een verschil in het gemiddelde aantal antibioticakuren van minstens een week tussen kinderen tot tien jaar en mensen ouder dan 65 jaar?* De data die we verzamelen is ongepaard, want de steekproef onder de kinderen tot tien jaar bevat niet dezelfde mensen als de steekproef onder de mensen ouder dan 65 jaar. Tabel 15 geeft de resultaten van je twee steekproeven.

Tabel 15: gegevens onderzoek antibioticakuren steekproef 1 en 2

Ondervraagde patiënt nummer	Aantal antibioticakuren van steekproef 1: tot 10 jaar	Ondervraagde patiënt nummer	Aantal antibioticakuren van steekproef 2: vanaf 65 jaar
1	1	15	5
2	4	16	3
3	2	17	6
4	1	18	4
5	2	19	1
6	3	20	3
7	5	21	1
8	0	22	2
9	2	23	0
10	3	24	1
11	1	25	4
12	4	26	3
13	2	27	3
14	1	28	2
	μ = 2,214286		μ = 2,714286

De hoeveelheid data die we hier hebben is al aan de grote kant om het nog met de hand te doen, maar ik kan je wel laten zien welke formules een rol spelen bij het doen van t-toetsen.

$$t = \frac{\mu_a - \mu_b}{SE_D} \text{ waarbij } SE_D = \sqrt{(SE_1)^2 + (SE_2)^2}$$

- *t* is hier de toetsingsgrootheid. Als je *t* hebt uitgerekend, dan kun je in een tabel (tabel 17) opzoeken wat de bijbehorende p-waarde is.

- μ_a is het gemiddelde van steekproef A *en* μ_b het gemiddelde van de steekproef B.
- Voor SE_D voert het te ver om uit te leggen waar dit precies voor staat. Je rekent dit apart uit volgens de formule hierboven.
  Hiervoor heb je de eerdergenoemde formule voor standaardfout nodig. $SE = \frac{\sigma}{\sqrt{n}}$
  Voor het uitrekenen van de standaardafwijking (σ) gebruik je deze formule $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n-1}}$ die op bladzijde 20 beschreven staat.

Tabel 16: T-toets uitgevoerd op gegevens uit tabel 15

Aantal antibioticakuren steekproef 1: tot 10 jaar	$(x - \mu)^2$			Aantal antibioticakuren steekproef 2: vanaf 65 jaar	$(x - \mu)^2$	
1	1,47449	**Variantie =**		5	5,22449	**Variantie =**
4	3,188776	**2,027473**		3	0,081633	**2,835165**
2	0,045918			6	10,79592	
1	1,47449	**Standaard**		4	1,653061	**Standaard**
2	0,045918	**deviatie =**		1	2,938776	**deviatie =**
3	0,617347	**1,423893**		3	0,081633	**1,683795**
5	7,760204			1	2,938776	
0	4,903061	**Standaard**		2	0,510204	**Standaard**
2	0,045918	**error =**		0	7,367347	**Error =**
3	0,617347	**0,380552**		1	2,938776	**0,450013**
1	1,47449			4	1,653061	
4	3,188776			3	0,081633	
2	0,045918			3	0,081633	
1	1,47449			2	0,510204	
μ = 2,214286	Totaal: 26,35714			μ = 2,714286	Totaal: 36,85714	
$SE_D = \sqrt{(SE_1)^2 + (SE_2)^2} = \sqrt{(0{,}38..)^2 + (0{,}45..)^2}$ = 0,589348						
$t = \frac{\mu_a - \mu_b}{SE_D} = \frac{2{,}71.. - 2{,}21..}{0{,}58..}$ = 0,848395						

De gevonden waarde die je ziet in de onderste rij, is de t-waarde. Hij is dus **0,84.** Deze t-waarde zegt op zichzelf niets. Hiervoor moet je hem vergelijken met tabel 17. Deze tabel ziet er heel vol uit, maar ik zal kort toelichten hoe het precies werkt.

1. Eerst bepaal je hoeveel vrijheidsgraden/*degrees of freedom* er zijn. Dan weet je in welke rij je je getallen moet aflezen.
   Bij zo veel getallen is het misschien lastig te zien hoeveel vrijheidsgraden je hebt. Je hebt in dit geval maar één totaalwaarde, dus je kunt bijna alle waarden vrij invullen tot aan de laatste waarde. Die wordt bepaald door het totale aantal. Kort voorbeeld: als je 3 + 4 + 6 = 13 hebt, dan kun je de "3" en "4" vrij invullen, maar dan is de "6" bepaald. Voor het aantal vrijheidsgraden is er nu een makkelijke formule: **vrijheidsgraden = (n1 – 1)(n2 – 1).** Waarin *n1* het aantal waarnemingen is in steekproef 1 en *n2* het aantal waarnemingen is in steekproef 2. In het geval van dit onderzoek dus: **df = (14 -1)(14 -1) = 169.** Dit getal staat niet precies in tabel 17, daarom zoek je naar het dichtstbijzijnde getal. In dit geval 200 (extra: of rekent het gemiddelde uit. Dit is nauwkeuriger).

2. Vervolgens kijk je in de tabel bij de altijd terugkomkende 5% in de "two tailed" rij. Je moet hier kijken bij de "two-tailed" rij, omdat je zoekt naar een verschil. We hebben niet gezegd dat we verwachten dat het antibioticagebruik in de ene groep hoger zal zijn dan in de andere groep. We hebben dus geen richting gegeven aan onze vraag, daarom is het een tweezijdige toets.
3. De waarde die daar staat wordt de **kritieke** waarde genoemd. Als jouw t-waarde **hoger** is dan de kritieke waarde, is je p-waarde 0,05 of kleiner, dus mag je H0 verwerpen en de HA aannemen. Onze t-waarde is 0,84. Hij is dus veel lager dan de 1,9… Daarom mogen we H0 niet verwerpen en mogen we **niet zeggen dat er een verschil** is ondanks het duidelijke verschil in gemiddelde.

Gepaard

Als je werkt met gepaarde data, dan volg je bijna hetzelfde stappenplan. Ook de formules die je gebruikt zijn hetzelfde:

$t = \frac{\mu_a - \mu_b}{SE_D}$ waarbij $SE_D = \sqrt{(SE_1)^2 + (SE_2)^2}$

Het enige verschil zijn de vrijheidsgraden, omdat je nu twee keer dezelfde groep hebt, is de formule voor het aantal vrijheidsgraden: **vrijheidsgraden = n – 1.**
Je gebruikt hier nog wel dezelfde tabel om af te lezen wat je kritieke t-waarde is.

Tabel 17: t-toets tabel (C. Dougherty, 2002)[2]

		Significance level					
Degrees of freedom	*Two-tailed test:*	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323
400		1.649	1.966	2.336	2.588	3.111	3.315
500		1.648	1.965	2.334	2.586	3.107	3.310
600		1.647	1.964	2.333	2.584	3.104	3.307
∞		1.645	1.960	2.326	2.576	3.090	3.291

[2] Statistical Tables, C. Dougherty, 2002.

Pearson's correlatiecoëfficiënt met de hand

Bij een kleine tabel is de correlatiecoëfficiënt (de r-waarde) nog met de hand te berekenen. Dit gaat als volgt:

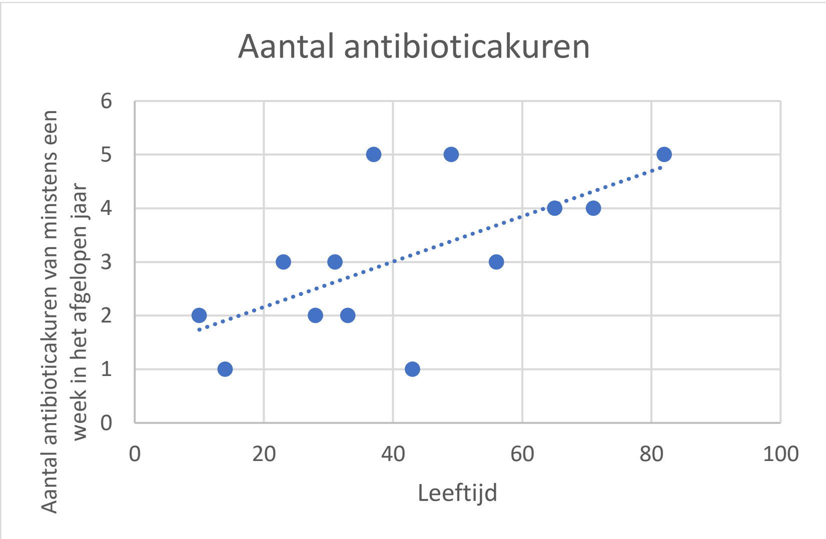
$$r = \frac{SP(xy)}{\sqrt{SS(x) \cdot SS(y)}}$$

- $SS(x) = \sum(x - \bar{x})^2$, waarin x met een dakje staat voor het gemiddelde van x. In dit geval is "x" de leeftijd.
- $SS(y) = \sum(y - \bar{y})^2$, waarin y met een dakje staat voor het gemiddelde van y. In dit geval is "y" het aantal antibioticakuren van minstens een week in het afgelopen jaar.
- $SP(xy) = \sum(x - \bar{x})(y - \bar{y})$

We gaan met de hand antwoord geven op de eerdergenoemde onderzoeksvraag: *is er een verband tussen de leeftijd en het aantal antibioticakuren in het afgelopen jaar*.
De verzamelde gegevens kun je in tabel 18 zien en de bijbehorende grafiek staat in figuur 10.

Tabel 18: gegevens antibioticakuren

Leeftijd van patiënt	Aantal antibioticakuren …
10	2
23	3
14	1
65	4
49	5
33	2
56	3
43	1
37	5
28	2
31	3
71	4
82	5



Figuur 10: grafische weergave van tabel 18

De r-waarde, de SS(x) waarde, de SS(y) waarde en de SP(xy) waarde, kun je allemaal met de hand uitrekenen. Dit heb ik voor gedaan in tabel 19 op de volgende bladzijde.
Omdat deze tabel al 13 waarnemingen bevat, zou ik het niet aanraden om dit met de hand te berekenen, maar als je maar 5 waarnemingen hebt, is het wel te doen.

Tabel 19: uitwerking van Pearson's correlatiecoëfficiënt met de hand

Leeftijd	Aantal antibiotica kuren	$(x - \bar{x})$	$(y - \bar{\bar{y}})$	$(x - \bar{x})^2$	$(y - \bar{\bar{y}})^2$	$(x - \bar{x}) \cdot (y - \bar{\bar{y}})$
10	2	10 – 41,69 = -31,69	2 – 3,08 = -1,08	$(-13,69)^2$ = 1004,40	$(-1,08)^2$ = 1,16	-31,69 x -1,08 = 34,13
23	3	-18,69	-0,08	349,40	0,01	1,44
14	1	-27,69	-2,08	766,86	4,31	57,51
65	4	23,31	0,92	543,25	0,85	21,51
49	5	7,31	1,92	53,40	3,70	14,05
33	2	-8,69	-1,08	75,56	1,16	9,36
56	3	14,31	-0,08	204,71	0,01	-1,10
43	1	1,31	-2,08	1,71	4,31	-2,72
37	5	-4,69	1,92	22,02	3,70	-9,02
28	2	-13,69	-1,08	187,48	1,16	14,75
31	3	-10,69	-0,08	114,33	0,01	0,82
71	4	29,31	0,92	858,94	0,85	27,05
82	5	40,31	1,92	1624,71	3,70	77,51
$\bar{x} = 41{,}69$	$\bar{y} = 3{,}08$			SS(x) = 5806,77	SS(y) = 24,92	SP(xy)= 245,31

Uiteindelijk heb je dan de volgende formule:

$$r = \frac{SP(xy)}{\sqrt{SS(x) \cdot SS(y)}} = \frac{245{,}31}{\sqrt{5806{,}77 \cdot 24{,}92}} = 0{,}64$$

Met deze r-waarde op zich kun je niet zo veel. Het enige wat je nu weet, is dat je te maken hebt met een positief verband: als x groter wordt, dan wordt y ook groter. Of deze relatie significant is, weet je nog niet. Hiervoor moet je er eerst weer een tabel bij pakken. Deze tabel is te vinden op de volgende bladzijde: tabel 20.

1. In de vrijheidsgraden/*degrees of freedom* (df) kolom zoek je naar de juiste vrijheidsgraden behorende bij je experiment. Hierbij hoort altijd de volgende formule: **vrijheidsgraden = n – 2**, waarin *n* het aantal gepaarde waarnemingen is. In dit geval gaat het iets anders dan hiervoor is vermeld, omdat je nu met een "x" en een bijbehorende "y" werkt. Je hebt dus waarden die bij elkaar horen. Ze vormen paren. Alle paren kun je vrij invullen, behalve het laatste paar. Een paar bestaat uit twee waarnemingen, daarom reken je met *n – 2.* In dit geval hadden we 13 paar waarnemingen, dus 13 – 2 = 11.
2. Verder kijk je daarna in de "two tailed" kolom, want we zochten alleen naar een verband en wisten niet of deze positief of negatief zou zijn. Onze vraag heeft geen richting en is dus tweezijdig. We kijken in de "two tailed" kolom bij p = 0,05. Dit is dan de 4^e kolom van links. De **kritieke r-waarde** die daar staat is de **0,553.**
3. De H0 mag worden verworpen als je gevonden r-waarde **hoger** is dan de kritieke r-waarde. Mijn r-waarde was 0,64 en is dus **hoger**, daarom mag H0 worden verworpen. Je hebt bij deze analyse een p-waarde gevonden die kleiner is dan 0,05. Het resultaat is **significant.**
   er is een positief verband tussen de leeftijd van een mens en het aantal antibioticakuren van minstens een week die hij of zij in het afgelopen jaar heeft gehad (p<0,05).
   Extra: als je rechts kijkt van de kritieke r-waarde in de "two tailed kolom" met een p-waarde van 0,02 zie je dat de kritieke r-waarde 0,63 is. De gevonden r-waarde is 0,64 en dus nog steeds hoger. Je mag nu zelfs zeggen dat H0 wordt verworpen met een significantie van een p-waarde die kleiner is dan 0,02. Het resultaat is dus nog **significanter.** Er is een nog kleinere kans dat dit berust op toeval.

Tabel 20: Pearson tabel (Radford)[3]

Table of Critical Values for Pearson's *r*

	Level of Significance for a One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for a Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	0.951	0.988	0.997	0.9995	0.9999	0.99999
2	0.800	0.900	0.950	0.980	0.990	0.999
3	0.687	0.805	0.878	0.934	0.959	0.991
4	0.608	0.729	0.811	0.882	0.917	0.974
5	0.551	0.669	0.755	0.833	0.875	0.951
6	0.507	0.621	0.707	0.789	0.834	0.925
7	0.472	0.582	0.666	0.750	0.798	0.898
8	0.443	0.549	0.632	0.715	0.765	0.872
9	0.419	0.521	0.602	0.685	0.735	0.847
10	0.398	0.497	0.576	0.658	0.708	0.823
11	0.380	0.476	0.553	0.634	0.684	0.801
12	0.365	0.457	0.532	0.612	0.661	0.780
13	0.351	0.441	0.514	0.592	0.641	0.760
14	0.338	0.426	0.497	0.574	0.623	0.742
15	0.327	0.412	0.482	0.558	0.606	0.725
16	0.317	0.400	0.468	0.542	0.590	0.708
17	0.308	0.389	0.456	0.529	0.575	0.693
18	0.299	0.378	0.444	0.515	0.561	0.679
19	0.291	0.369	0.433	0.503	0.549	0.665
20	0.284	0.360	0.423	0.492	0.537	0.652
21	0.277	0.352	0.413	0.482	0.526	0.640
22	0.271	0.344	0.404	0.472	0.515	0.629
23	0.265	0.337	0.396	0.462	0.505	0.618
24	0.260	0.330	0.388	0.453	0.496	0.607
25	0.255	0.323	0.381	0.445	0.487	0.597
26	0.250	0.317	0.374	0.437	0.479	0.588
27	0.245	0.311	0.367	0.430	0.471	0.579
28	0.241	0.306	0.361	0.423	0.463	0.570
29	0.237	0.301	0.355	0.416	0.456	0.562
30	0.233	0.296	0.349	0.409	0.449	0.554
40	0.202	0.257	0.304	0.358	0.393	0.490
60	0.165	0.211	0.250	0.295	0.325	0.408
120	0.117	0.150	0.178	0.210	0.232	0.294
∞	0.057	0.073	0.087	0.103	0.114	0.146

Adapted from Appendix 2 (Critical Values of *t*) using the square root of $[t^2/(t^2 + df)]$
Note: Critical values for Infinite *df* actually calculated for *df*= 500.

[3] Radford, Statsbook, chapter 20. Table of Critical values.

Chi-kwadraat toets

De chi-kwadraat toets is de minst lastige toets om met de hand uit te voeren. In de inleiding heb ik al genoemd dat je de chi-kwadraat toets kunt gebruiken als je zoekt naar een verschil in aantal, maar ook als je een correlatie zoekt. Voor beide onderzoeksvragen heb ik een voorbeeld.

Verschil in aantal

Het uitleggen van het gebruik van de chi-kwadraat toets bij een onderzoek naar het verschil in aantal, doe ik hier aan de hand van hetzelfde voorbeeld genoemd op bladzijde 13.

Het onderzoek

We doen onderzoek naar een vliegensoort die zijn eieren legt in de bladeren van bomen. We verwachten dat de hoogte waarop deze bladeren hangen niet uitmaakt en dat er dus een gelijke verdeling zal zijn in het aantal larven dat we gaan vinden. Van tevoren bepalen we welke categorieën er zijn: 0 tot en met 2 meter boven de grond, 2 tot en met 4 meter boven de grond en 4 tot en met 6 meter boven de grond. We verwachten nu 1/3 van de larven te vinden in de categorie 0 tot 2 meter, 1/3 in de categorie 2 tot 4 meter en 1/3 in de categorie 4 tot 6 meter.

- Onze onderzoeksvraag: *is er een verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van de bladeren?*
  - H0 = er is geen verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van bladeren.
  - HA = er is wel een verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van bladeren.

Na een dag te hebben rondgelopen in het bos, heb je geturfd hoeveel larven je vindt op welke hoogte. Je hebt de volgende gegevens verzameld:

Tabel 21: verzamelde gegevens

	0 tot en met 2 meter	2 meter tot en met 4 meter	4 meter tot en met 6 meter	Totaal
Waarnemingen	93	86	54	243
Verwachting	81	81	81	243

Omdat je 243 waarnemingen hebt gedaan, kun je ook de rij verwachtingen in vullen, want dat is in elke kolom 243/3= 81. De waarden voldoen niet aan onze verwachting, maar nu rest nog de vraag: is dit toeval of niet? Is het verschil groot genoeg?

De bijbehorende formule is deze keer niet zo lastig en vergt niet heel veel rekenwerk. Je moet alleen goed weten hoe je de verwachting uitrekent.

$$\chi^2 = \sum \left[\frac{(waargenomen - verwacht)^2}{verwacht}\right]$$

- Waarin χ^2 staat voor chi-kwadraat. De chi-kwadraat is in feite een maat voor het verschil in de verwachte waarden en de gevonden/waargenomen waarden.
- *Waargenomen* is wat je daadwerkelijk gevonden hebt in je onderzoek.
- *Verwacht* is wat je had verwacht op basis van het totaal aantal waarnemingen. In dit geval is de verwachting makkelijk te berekenen. Je hebt 243 waarnemingen en 3 categorieën, dus 243/3 = 81.

Tabel 22: berekening van de totale chi-kwadraat

	0 tot en met 2 meter	2 meter tot en met 4 meter	4 meter tot en met 6 meter	Totaal
Waarnemingen	93	86	54	243
Verwachting	81	81	81	243
Chi kwadraat	$(93\text{-}81)^2)/81$ = 1,777777778	$(86\text{-}81)^2)/81$ = 0,308642	$(54\text{-}81)^2)/81$ = 9	1,77.. + 0,30.. + 9 =11,08642

De chi-kwadraat waarde die we hebben gevonden is: 11,08…

Zoals ook bij de voorgaande twee statistische toetsen moeten we de kritieke waarde opzoeken in een tabel (tabel 23) en hem vergelijken met de door ons gevonden waarde. Als onze waarde **hoger** is dan de **kritieke waarde** dan, dan mogen we de H0 verwerpen en de HA aannemen.

Tabel 23 wordt als volgt gebruikt:

1. Ook nu moet je eerst het aantal vrijheidsgraden uitrekenen. Meestal gaat het hier om het **aantal categorieën – 1**, maar dit moet je **altijd controleren**! Het kan namelijk nogal verschillen wat je een categorie noemt en wat niet. In ons geval hebben we drie categorieën, namelijk drie verschillende hoogten waar we naar kijken, dus het aantal vrijheidsgraden = 3 – 1 = 2. Dit klopt, want als je 1 getal invult, kun je nog 1 getal vrij invullen, maar de laatste niet meer. Je kunt dus 2 getallen vrij invullen. Voor meer uitleg over het bepalen van het aantal vrijheidsgraden kijk je op bladzijde 21.
2. In de tabel kijk je dan in de kolom van 5%, in de rij van het aantal vrijheidsgraden = *degrees of freedom* = 2. Daar zie je een kritieke waarde staan van 5,991. De chi-kwadraat-waarde die wij hadden gevonden was 11,08… en dus fors hoger dan de 5,991. We mogen de H0 verwerpen en de HA aannemen. *Er is dus wel een verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van bladeren (p<0,05).*
3. Extra: omdat onze waarde zo veel hoger is dan de gevonden kritieke chi-kwadraat waarde, kunnen we ook in de kolommen rechts ervan kijken. Bij een p-waarde van 0,01 (1%) zie je dat onze gevonden chi-kwadraat waarde van 11,08... nog steeds boven 9,210 ligt, maar bij een p-waarde van 0,001 (0,1%) hebben wij een waarde die lager is dan de kritieke waarde van 13,816. We zouden onze conclusie daarom mogen aanpassen naar: *er is een verschil tussen het gemiddelde aantal larven dat we vinden op verschillende hoogte van bladeren (p<0,01).* En niet naar p<0,001, want onze waarde is lager dan deze kritieke waarde.

Tabel 23: kritieke chi-kwadraat waarden (C. Dougherty, 2002)[4]

χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

	Significance level		
Degrees of freedom	5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588

Correlatie onderzoek

De chi-kwadraat test kun je ook goed gebruiken als je een **correlatie** onderzoekt.
Je hebt twee groepen giraffen. Een groep giraffen die op de open savanne leeft en een groep giraffen die op een bosrijke savanne leeft. Je wilt weten of er een correlatie is tussen de leefomgeving en het aantal keer dat ze op een dag gaan drinken.

- Onderzoeksvraag: *is er een correlatie tussen het leefgebied van giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken?*
    - H0 = er is geen correlatie tussen het leefgebied van giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken.
    - HA = er is wel een correlatie tussen het leefgebied van giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken.

Je volgt de giraffen gedurende een dag en krijgt de volgende waarnemingen:

Tabel 24: waargenomen gegevens op de savanne

		Hoeveelheid drinken		
		Drinkt meer dan 5 keer op een dag	Drinkt minder dan 5 keer op een dag	Totaal
Lee fge- bie	Giraffen in open savanne	6	14	**20**
	Giraffen in bosrijke savanne	18	12	**30**
	Totaal	**24**	**26**	**50**

[4] Statistical Tables, C. Dougherty, 2002.

Op basis van het totale aantal giraffen dat meer dan 5 keer op een dag drinkt en het totaal van de grootte van de steekproeven, kun je met behulp van een kruistabel een verwachting opstellen. Let goed op dat het totaal niet verandert.

Tabel 25: verwachte gegevens op basis van het totale aantal

		Hoeveelheid drinken		
		Drinkt meer dan 5 keer op een dag	Drinkt minder dan 5 keer op een dag	Totaal
Leefge-bie	Giraffen in open savanne	24*20/50=9,6	26*20/50=10,4	**20**
	Giraffen in bosrijke savanne	24*30/50 = 14,4	26*30/50 = 15,6	**30**
	Totaal	**24**	**26**	**50**

Nu je weet wat je waarnemingen (tabel 24) zijn en wat je op basis van het totaal eigenlijk had verwacht (tabel 25), kun je de chi-kwadraat uitrekenen. $\chi^2 = \sum \left[\frac{(waargenomen - verwacht)^2}{verwacht}\right]$

Tabel 26: totale chi-kwadraat uitrekenen

		Hoeveelheid drinken	
		Drinkt meer dan 5 keer op een dag	Drinkt minder dan 5 keer op een dag
Leefge-bied	Giraffen in open savanne	$\chi^2 = (6-9,6)^2/9,6 = 1,35$	$\chi^2 = (18-14,4)^2/14,4 = 0,9$
	Giraffen in bosrijke savanne	$\chi^2 = (14-10,4)^2 /10,4 = 1,25$	$\chi^2 = (12-15,6)^2 /15,6 = 0,83$
Totale $\chi^2 = 1,35 + 0,9 + 1,25 + 0,83 = 4,23$			

Het vinden van de kritieke waarde en het bepalen van de significantie:

1. Om nu de kritieke chi-kwadraat waarde te vinden in tabel 23, moet je weten hoeveel vrijheidsgraden je hebt. Het lijkt alsof je hier 4 categorieën hebt: giraffen op open savanne, giraffen in bosrijkgebied, drinkt meer dan 5 keer op een dag, drinkt minder dan 5 keer op een dag. Als je nu weer de formule "aantal categorieën – 1" gebruikt, zou je uitkomen op 3 vrijheidsgraden. Als je dit uitprobeert in de tabel, dan zie je dat je maar 1 vrijheidsgraad hebt. Dus *degrees of freedom = 1.* Zie voor meer uitleg over het bepalen van de vrijheidsgraden bladzijde 21.
2. De gevonden chi-kwadraat waarde is 4,23.
3. Bij een vrijheidsgraad van 1 heb je een kritieke chi-kwadraat waarde van 3,84. Onze waarde is hoger dan de kritieke waarde, dus ons resultaat is significant. De conclusie die je nu kunt trekken is: *er is een significante correlatie tussen het leefgebied van de giraffen en het aantal keer dat de giraffen meer dan 5 keer of minder dan 5 keer per dag drinken (p<0,05).*

Opdrachten: Kan ik het ook met de hand?

Via een aantal opdrachten ga je nu proberen al deze theorie zelf te ontdekken en uiteindelijk onder de knie te krijgen. Het doel van deze opdrachten is vooral om je inzicht te geven in de mogelijke toetsen en hoe je de uitkomsten zelf zou kunnen berekenen. Uiteindelijk is het de bedoeling dat je de statistiek kunt gebruiken bij het analyseren van je eigen onderzoekresultaten en je daar een juiste conclusie uit kunt trekken. Alles wat ik hier heb beschreven, zal uiteindelijk terugkomen op het universitaire onderwijs. Dan zal je nog veel meer over statistische toetsen leren.

Bij het maken van onderstaande vragen, mag je terugkijken in de module. Als het niet is aangegeven gaat het hier altijd om een **tweezijdige toets**. Deze vragen zijn iets moeilijker dan de voorgaande vragen, omdat je nu de r-waarde, t-waarde, en je chi-kwaraat-waarde zelf met de hand moet berekenen.

Vraag 1

a) Bij welke p-waarde zeggen we normaal dat iets significant is? (Blz. 5)

b) Je hebt onderstaande tabel. De dik gedrukte getallen zijn totale waarden.

20	10	30	**60**
5	30	5	**40**
5	40	5	**50**
30	**80**	**40**	**150**

Hoeveel vrijheidsgraden heb je hier? (Blz. 21)

c) Je hebt een μ van 3 en een σ 1,5. Tussen welk interval liggen de meeste waarnemingen? (Blz. 20)

Vraag 2

Je hebt ergens gehoord dat je oren blijven groeien, ook als je volgroeid bent. Je neemt de proef op de som en doet een klein onderzoek. Je vindt de volgende waarden:

Leeftijd in jaren	Oorlengte in mm
38	65
57	72
75	74
92	81

a) Wat is je nulhypothese en wat is je alternatieve hypothese? (Blz. 4)

b) Welke statistische toets ga je doen? (Blz. 3)

c) Voer de statistische toets uit en trek een goed geformuleerde conclusie.

d) Hoe zou je het onderzoek beter kunnen maken?

Vraag 3

Het aantal jongen dat na een jaar nog leeft van een groep huiskatten wordt vergeleken met het aantal jongen dat na een jaar nog leeft van een groep wilde katten. Je denkt dat dit aantal bij wilde katten lager zal liggen. Je neemt een steekproef van 30 bij zowel de groep huiskatten als bij de groep wilde katten. Je vindt voor de groep huiskatten een gemiddelde van 4,5. Voor de groep wilde katten vindt je het gemiddelde van 2,8. Verder vindt je voor de groep huiskatten een standaardafwijking van 1,2 en voor de groep wilde katten een standaardafwijking van 0,7.

a) Bepaal de standaard error van de groep huiskatten en van de groep wilde katten. (Blz. 20)

b) Bereken de t-waarde. (Blz. 22)

c) Hoeveel vrijheidsgraden heb je hier? (Blz. 21)

d) Je voert een eenzijdige toets uit. Is je resultaat significant? (Blz. 25).

Vraag 4
Tijdens de examens wordt aan een aantal leerlingen van jouw school gevraagd of ze koffiedrinken om ’s nachts wakker te blijven en door te kunnen leren. Van de 70 jongens die worden ondervraagd, blijkt 44 dit te doen. Van de 96 meisjes die zijn ondervraagt, blijkt 22 dit te doen.

a) Maak een tabel van je waarnemingen. Wat was je verwachting? Maak hier ook een tabel van. (Blz. 31)
b) Welke toets doe je nu? Voer deze toets uit. Is het resultaat significant? (Blz. 31)

Bijlage 2

Antwoorden: Hoe werk ik in Excel?

Vraag 1

a) p-waarde is 0,05.
b) Een p-waarde van 0,051 geeft aan dat er 5,1% kans is dat H0 waar is. Dit is groter dan 5% (een p-waarde van 0,05) en daarom wordt hier gezegd dat het resultaat niet significant is.
c) Pearson's correlatiecoëfficiënt, omdat je naar een verband zoekt.

Vraag 2

a) Onderzoeksvraag: "Is er een verschil in de gemiddelde beoordeling van het vak biologie tussen dit jaar en vorig jaar?" Ook goed: "Is er een verschil in het aantal beoordelingen per categorie van het vak biologie tussen dit jaar en vorig jaar?"
Nulhypothese: er is geen verschil in de *gemiddelde* beoordeling van het vak biologie tussen dit jaar en vorig jaar.
Alternatieve hypothese: er is wel een verschil in de *gemiddelde* beoordeling van het vak biologie tussen dit jaar en vorig jaar.
b) Een *gepaarde* T-toets.
   a. Je kiest voor een T-toets omdat je onderzoek gaat doen naar het verschil in gemiddelden.
   b. Dit moet gepaard, omdat je dezelfde leerlingen zowel het vak biologie van vorig jaar als van dit jaar laat beoordelen.
c) Chi-kwadraat. We zijn per categorie aan het turven hoeveel mensen het hier mee eens waren. Uiteindelijk kun je dan de verwachting bepalen en een statistische toets uitvoeren. Deze manier is beter dan de T-toets manier, want als je kiest voor "zeer oneens" en dit het cijfer 1 krijgt, dan zou "oneens" twee keer zo goed moeten zijn als "zeer oneens" omdat dit het cijfer 2 krijgt. Dit staat echter niet in verhouding tot de werkelijkheid.

Vraag 3

a) Eenzijdig, want we hebben duidelijk een richting gegeven in onze vraag. We vroegen ons namelijk af of de jongen van huiskatten vaker na een jaar nog leefden dan de jongen van wilde katten.
b) Een *ongepaarde* T.toets. Want je gaat zoeken naar het verschil tussen *gemiddelde*. Dit is ongepaard omdat je het bij twee groepen dieren doet en niet bij dezelfde dieren.
c) Een p-waarde van 0,0005 is kleiner dan $p = 0,05$. Daarom spreken we hier van een significant antwoord. Een goede conclusie: het gemiddelde aantal jongen dat na een jaar nog leeft bij wilde katten is **significant lager** dan bij huiskatten ($p<0,05$).
Extra: omdat de p-waarde duidelijk veel lager is dan 0,05, mag je de conclusie ook anders formuleren. Namelijk: het gemiddelde aantal jongen dat na een jaar nog leeft bij wilde katten is significant lager dan bij huiskatten ($p<0,0005$).

Vraag 4

a) Is er een verband tussen de leeftijd van mensen en de oorlengte van mensen?
Nulhypothese: er is geen verband tussen de leeftijd van mensen en de oorlengte van mensen.
Alternatieve hypothese: er is wel een verband tussen de leeftijd van mensen en de oorlengte van mensen.
b) Pearson's correlatiecoëfficiënt, omdat je zoekt naar een verband tussen de leeftijd en de oorlengte.
c) De gevonden r-waarde van 0,922 laat je denken dat er een postieve correlatie is, maar als je de p-waarde ziet, is hij 0,08. 0,08 is groter dan 0,05 en daarom hebben we hier **niet** te maken

met een significant resultaat. Conclusie: er is geen verband tussen de leeftijd van mensen en de oorlengte van mensen (p=0,08).

d) Door meer mensen te onderzoeken. Hierdoor krijg je een betrouwbaardere uitkomst. Mogelijk was ons resultaat dan wel significant geworden.
Extra: in dit onderzoek wordt nu geen onderscheid gemaakt tussen mannen en vrouwen. Om deze variabele weg te nemen, zou je alleen mannen of alleen vrouwen moeten onderzoeken. Dit maakt het onderzoek betrouwbaarder.

Vraag 5

a) Tabel met waarnemingen en tabel met verwachting.

Waarnemingen			
	Drinkt koffie	Drinkt niet	Totaal
Mannen	44	43	**87**
Vrouwen	45	67	**112**
Totaal	**89**	**110**	**199**

Verwachting			
	Drinkt koffie	Drinkt niet	Totaal
Mannen	38,9095477	48,0904523	**87**
Vrouwen	50,0904523	61,9095477	**112**
Totaal	**89**	**110**	**199**

b) Chi-kwadraat toets, omdat je aantallen hebt geturfd. Je hebt gekeken naar hoeveel leerlingen er koffie drinken en niet.
De gevonden p-waarde = 0,14. 0,14 is groter dan 0,05 en niet kleiner. Daarom hebben we hier **niet** te maken met een significant verschil. Conclusie: er is geen associatie tussen het geslacht en het wel of niet koffiedrinken om 's nachts wakker te blijven en te leren (p=0,14).

Antwoorden: kan ik het ook met de hand?

Vraag 1

a) p-waarde is 0,05.
b) 4 vrijheidsgraden. (3-1)(3-1)=4. Of als je 4 getallen invult, dan kun je de rest vrij invullen.
c) [1,5;4,5], want het gaat om 3 – 1,5 of 3 + 1,5.

Vraag 2

a) Nulhypothese: er is geen verband tussen de leeftijd van mensen en de oorlengte van mensen. Alternatieve hypothese: er is wel een verband tussen de leeftijd van mensen en de oorlengte van mensen.
b) Pearson's correlatiecoëfficiënt
c) Het is handig als je veel moet berekenen om er een overzichtelijke tabel van te maken:

Leeftijd	Oorlengte	$(x - \bar{x})$	$(y - \bar{\bar{y}})$	$(x - \bar{x})^2$	$(y - \bar{\bar{y}})^2$	$(x - \bar{x}) \cdot (y - \bar{y})$
38	65	(38 – 65,5) = -27,5	-8	$(-27,5)^2$ = 756,25	64	-27,5 x -8 = 220
57	72	-8,5	-1	72,25	1	8,5
75	74	9,5	1	90,25	1	9,5
92	81	26,5	8	702,25	64	212
$\bar{x}$ = 65,5	$\bar{y}$ = 73			SS(x) = 1564,75	SS(y) = 130	SP(xy) = 450

$r = \frac{SP(xy)}{\sqrt{SS(x) \cdot SS(y)}}$. Dus r = 450/(wortel van 1564,75 x 130) = 0,997

Aantal vrijheidsgraden is n – 2 = 4 – 2 = 2. Door te kijken in tabel 13 zien we dat de kritieke r-waarde bij p = 0,05 en een tweezijdige toets 0,950 is. De gevonden r-waarde zit daar net boven, daarom kunnen we zeggen dat we hier een significant verschil hebben. Een juiste conclusie is: er is wel een verband tussen de leeftijd van mensen en de oorlengte van mensen (p<0,05).

Extra: bij dit onderzoek zie je zelfs dat de gevonden r-waarde ook nog boven de kritieke r-waarde r=0,980 ligt als je uit gaat van een p=0,02. Hij ligt ook nog boven de kritieke r-waarde 0,990 bij een p=0,01. Maar hij ligt niet meer boven de kritieke r-waarde van 0,999 bij een p van 0,001. Je kunt dus maximaal de volgende conclusie trekken: er is wel een verband tussen de leeftijd van mensen en de oorlengte van mensen (p<0,01)

d) Door meer mensen te onderzoeken. Extra: in dit onderzoek wordt nu geen onderscheid gemaakt tussen mannen en vrouwen. Om deze variabele weg te nemen, zou je alleen mannen of alleen vrouwen moeten onderzoeken.

Vraag 3

a) Voor de standaard error geldt: $SE = \frac{\sigma}{\sqrt{n}}$. We hebben een standaarddeviatie van 1,3 en 0,7.
De standaard error voor de huiskatten => 1,3/wortel30 = 0,24
De standaard error voor de wilde katten => 0,7/wortel30 = 0,13

b) Voor het berekenen van de t-waarde geldt het volgende: $t = \frac{\mu_a - \mu_b}{SE_D}$ waarbij $SE_D = \sqrt{(SE_1)^2 + (SE_2)^2}$. Dus voor $SE_D = \sqrt{(0{,}24)^2 + (0{,}13)^2} = 0{,}27$ en dan geldt voor $t = \frac{4{,}5 - 2{,}8}{0{,}27} = 6{,}3$

c) Aantal vrijheidsgraden = 30 – 1 = 29.

d) Kritieke t-waarde (gevonden in tabel 10 onder de one-tailed kolom) = 1,699. De gevonden t-waarde ligt hier boven, dus: het gemiddelde aantal jongen dat na een jaar nog leeft bij wilde katten is **significant lager** dan bij huiskatten (p<0,05).
Extra: om nauwkeuriger te kunnen aangeven hoe significant het verschil is, kun je ook kijken naar de andere kritieke waarden rechts van de eerste kritieke waarde. Je ziet hier de volgende kritieke waarden staan: 2,045; 2,462; 2,756; 3,396; 3,659. Al deze kritieke waarden zijn lager

dan de gevonden t-waarde van 6,3. Je kunt nu dus zeggen dat: het gemiddelde aantal jongen dat na een jaar nog leeft bij wilde katten significant lager is dan bij huiskatten ($p<0{,}0005$).

Vraag 4

a) Tabel met waarnemingen en tabel met verwachting.

Waarnemingen			
	Drinkt koffie	Drinkt niet	Totaal
Mannen	44	26	**70**
Vrouwen	22	74	**96**
Totaal	**66**	**100**	**166**

Verwachting			
	Drinkt koffie	Drinkt niet	Totaal
Mannen	27,8313253	42,1686747	**70**
Vrouwen	38,1686747	57,8313253	**96**
Totaal	**66**	**100**	**166**

b) Chi-kwadraat toets.

$(44\text{-}27{,}8)^2/27{,}8 = 9{,}4$	$(22\text{-}38{,}2)^2/38{,}2 = 6{,}9$	$(26/42{,}2)^2/42{,}2 = 6{,}2$	$(74\text{-}57{,}83)^2/57{,}8 = 4{,}5$	9,4 + 6,9 + 6,2 + 4,5 = 27

Chi-kwadraat waarde (afhankelijk van je afronding) om en nabij 27.
Het aantal vrijheidsgraden is hier 1. Je kunt hier een getal vrij invullen, daarom kijken we in tabel 16 bij een vrijheidsgraad van 1. De kritieke chi-kwadraat waarde is 3,84. De gevonden waarde ligt hier ver boven. Het resultaat is dus significant.
Extra: de gevonden waarde ligt daarnaast ook boven 6,6 en 10,8 en daarom kan gezegd worden dat we hier te maken hebben met een significantie van $p < 0{,}001$.

Bijlage 3

Informatie over data (soms getallen)

Nominale data:
Een variabele die aangeduid wordt met een naam. Altijd kwalitatief. De naam van de variabele kan (deels) een getal zijn.
Bijvoorbeeld: blond (een variabele van haarkleur), AB (een variabele van bloedgroep), baardagame (een variabele van diersoort), p53 (een variabele van eiwitten), 06 12 34 56 78 (een variabele van telefoonnummer).
Je kunt met de getallen niet rekenen.

Ordinale data:
Een variabele die je ordent op een schaal. Altijd kwalitatief. De variabele kan eventueel weergegeven worden als een getal.
Voorbeeld van ordinale data: keuzemogelijkheden die de voorkeur aangeven voor een bepaald gegeven.
Voorbeelden van een ordinale schaal zijn: hoger of lager, beter of slechter.
Je kunt met de getallen niet rekenen.

Interval data:
Een variabele die je ordent op een schaal, waarbij het verschil tussen opeenvolgende variabelen even groot is en iets zegt over het verschil in absolute hoeveelheid, maar niet in relatieve hoeveelheid. Je kunt daardoor (op een zinvolle manier) met deze verschillen rekenen. Je kunt daarbij optellen en aftrekken, maar niet vermenigvuldigen of delen.
Voorbeeld van een intervalschaal:
De tijdsaanduiding op een dag. Om 10.00 uur is het een uur later dan om 9.00 uur. Maar om 8 uur 's avonds (20.00 uur) is het niet twee keer zo laat als om 10.00 uur.

Ratio data:
De variabelen zijn kwantitatieve data die numeriek zijn. De ratioschaal heeft een absoluut nulpunt. Dat betekent dat je kunt rekenen met de getallen, waarbij je kunt optellen, aftrekken, delen en vermenigvuldigen.
Voorbeeld van ratiodata:
Tijdsduur, lengte.

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

Nog wat begrippen:

Niet-numeriek en numeriek:

Niet-numerieke data kunnen niet gemeten worden in hoeveelheden. nominale data en ordinale data behoren tot dit type data.

Numerieke data kunnen gemeten worden in hoeveelheden. Intervaldata en ratiodata behoren tot dit type.

Continue data en discrete data:

Numerieke data kunnen continu of discreet zijn. Continue data kunnen alle waarden aannemen (vaak wel binnen een bepaald gebied), discrete data kunnen alleen sommige waarden aannemen binnen een bepaald gebied. Voorbeeld van continue data zijn temperatuur, lichaamslengte. Een voorbeeld van discrete data zijn: schoenmaat.

Bronnen:

https://www.verenigingnlt.nl/site/assets/files/1067/v125_meten_en_interpreteren_ev_ll_22022010.pdf

https://www.scriptiemaster.nl/scriptie-kennisbank/meetlevel-variabelen/