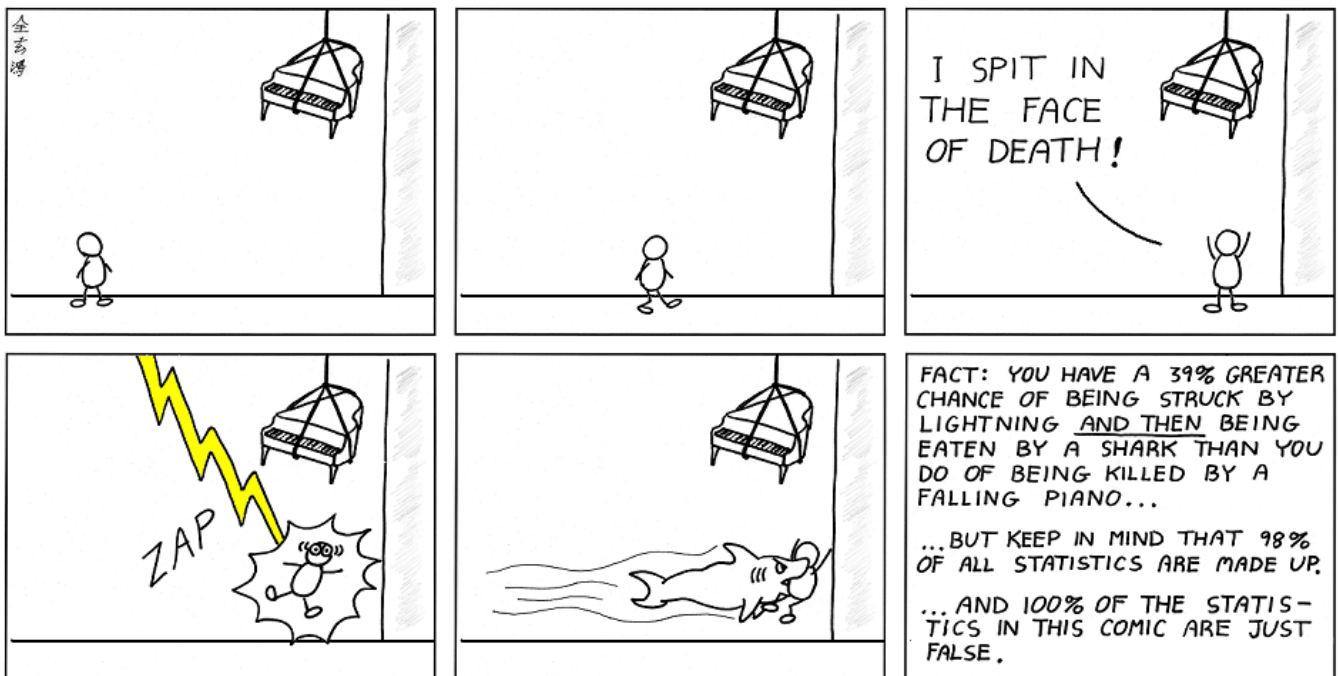


# Statistische toetsen

Een handleiding

voor elke leerling die worstelt  
met het toetsen van zijn gegevens  
bij het PWS



Hanna Bodde en Annalie Koerts  
Karla Thie

# Inhoudsopgave

<b>1.</b>	<b>Inleiding</b>	<b>3</b>
<b>2.</b>	<b>Criteria voor de keuze van de juiste statistische toets</b>	<b>4</b>
2.1	Aantal steekproeven: 1 of meer	4
2.2	Manier van steekproeftrekking: afhankelijk of onafhankelijk	4
2.3	Meetniveau	4
2.4	Verdeling van de populatie: parametrische en niet-parametrische toetsen	5
<b>3.</b>	<b>Schema voor de keuze van de juiste statistische toets</b>	<b>7</b>
<b>4.</b>	<b>Uitwerking statistische toetsen</b>	<b>8</b>
4.1	T-toets	8
4.2	Tekentoets	9
4.3	Chi-kwadraattoets	11
4.4	Wilcoxontoets	12
4.5	Mann-Whitneytoets	14
<b>5.</b>	<b>Nawoord</b>	<b>15</b>
<b>6.</b>	<b>Literatuurlijst</b>	<b>15</b>

# 1. Inleiding

Beste PWS-er,

Nu je bent begonnen voor de voorbereidingen voor je profielwerkstuk heb je hoogstwaarschijnlijk al je onderwerp gekozen en een onderzoeksvraag geformuleerd. Er is een grote kans dat je bij deze onderzoeksvraag zelf een onderzoekje gaat uitvoeren. Dit kan zijn in de vorm van enquêtes houden bij je jaargenoten over hun alcohol gebruik, onderzoeken welke leermethode de beste resultaten oplevert, of het gehoor van je klasgenoten testen na een avondje stappen. Onderzoek doen kan dus op veel manieren. Wat deze manieren echter allemaal met elkaar gemeen hebben is dat je de significantie van je gegevens zult moeten gaan berekenen. Wat houdt die significantie nou precies in? En op welke manier kan jij nou die significantie berekenen bij jouw onderzoek? Voor deze vragen hebben wij nou een handleiding gemaakt.

## *Een onderzoeksvraag, wat nu?*

Het is erg handig om van te voren al te bedenken hoe je precies wilt gaan onderzoeken, maar vooral ook: wat voor gegevens je wilt hebben en hoe je vervolgens deze gegevens met elkaar wilt vergelijken. Het gebeurt nogal eens dat PWS-ers aan hun onderzoek beginnen en vervolgens een hoop gegevens hebben waarmee ze niet kunnen rekenen. Dat is zonde. Niet alleen is het jammer dat je de helft van je gegevens niet kunt gebruiken om je onderzoeksvraag te beantwoorden, nog meer is het natuurlijk zonde van je tijd. Bedenk dus van te voren welke gegevens je precies nodig hebt en op welke manier je met deze gegevens wilt gaan rekenen.

- |         |  |
|---------|--|
| Stap 1: | Onderzoeksvraag bedenken   |
| Stap 2: | Bedenken op wat voor manier je dit zou kunnen onderzoeken  |
| Stap 3: | Een nulhypothese en alternatieve hypothese opstellen   |
| Stap 4: | Kijken hoe je deze nulhypothese kunt toetsen en opzoeken hoe je daarbij de significantie van je gegevens kunt berekenen. |

## *Wat betekent het als je je onderzoek significant bewezen hebt?*

Nu heb je aan de hand van je gegevens uitgerekend dat je onderzoek significant bewezen is. Leuk, maar wat houdt dit nou eigenlijk precies in? Letterlijk betekent het dat wanneer je je hypothese significant bewezen hebt, je bewezen hebt dat een bepaald verschijnsel (datgene wat je onderzocht hebt) zo vaak voorkomt dat het (normaal gesproken) niet aan toeval toe te schrijven is. Dit maakt je onderzoek dus betekenisvol. Je hebt bewezen dat datgene wat jij onderzocht hebt zo vaak voorkomt (of juist van iets anders afwijkt) dat dit geen toeval meer kan zijn, en dus waar is. Bij een significantie niveau van 10% is de kans dat je uitkomst toeval is klein, maar groter dan bij een significantie niveau van 5%. In het geval dat je onderzoek significant bewezen is bij een significantie niveau van 1% is de kans op een toevallige uitkomst heel erg klein geworden en kun je je uitkomst erg nauwkeurig noemen.

## 2. Criteria voor de keuze van de juiste statistische toets

Bij het testen van een hypothese moet gebruik worden gemaakt van een geschikte statistische toets.

De keuze van deze statistische toets hangt af van vier criteria:

- het aantal steekproeven
- onafhankelijk of afhankelijke steekproef
- meetniveau
- verdeling van de populatie

In de paragrafen hieronder volgt een uitleg bij elk van deze vier criteria.

### 2.1 Aantal steekproeven: 1 of meer

Het begrip steekproef moet je ruim interpreteren. Als je de beste verkoop van product X wilt onderzoeken onder twee verschillende promotiecampagnes, gebruik je twee steekproeven. Als je het gebruik van frisdrank in vijf leeftijdscategorieën wilt onderzoeken gebruik je vijf steekproeven. Wil je het gemiddelde gebruik van frisdrank in een steekproef wilt vergelijken met het gemiddelde gebruik door de totale populatie trek je één steekproef. Het is dus mogelijk om voor meer dan twee steekproeven te kiezen. Dit is echter af te raden voor een profielwerkstuk waar limieten van tijd en geld aan zijn gebonden. Bovendien vraagt een groter aantal steekproeven om veel ingewikkeldere statistische toetsen. Mocht je dus van plan zijn met drie of meer steekproeven te werken, probeer je vraagstelling dan zo aan te passen dat je dit aantal kunt reduceren naar twee.

### 2.2 De manier van steekproeftrekking: onafhankelijk of afhankelijk

De onafhankelijkheid van steekproeven houdt in dat de proefpersonen van de verschillende groepen **onafhankelijk** geselecteerd zijn. De selectie van proefpersonen in de ene steekproef heeft dus geen invloed op de selectie van proefpersonen in de andere steekproef. De proefpersonen zijn volledig willekeurig in een bepaalde groep gekomen en hebben dezelfde kans geselecteerd te worden (dit heet een aselechte steekproef). Je kunt een (redelijk) aselechte steekproef nemen door bijvoorbeeld aparte klassen te nemen. Bij **afhankelijke** steekproeven houd je het lot in eigen handen. In een onderzoek waarbij bijvoorbeeld twee groepen van respondenten met elkaar worden vergeleken, en in de ene steekproef alleen mannen zitten, en in de andere steekproef alleen de partners van die mannen, dan heb je een afhankelijke steekproef; de selectie van de proefpersonen in de tweede steekproef is immers afhankelijk van de proefpersonen in de eerste steekproef, alleen de partners mogen in de tweede steekproef, en niet random personen. Andere voorbeelden zijn: het afnemen van een enquête bij een aantal personen en na een maand dezelfde mensen dezelfde lijst voorleggen om bijvoorbeeld hun veranderingen te meten (de selectie van de personen in de tweede steekproef is weer afhankelijk van de personen in de eerste steekproef); of het indelen van studenten in twee groepen, namelijk een testgroep en een controlegroep. Als een bepaalde student wordt ingedeeld in de testgroep, wordt een student met zoveel mogelijk dezelfde eigenschappen in de controlegroep gezet.

### 2.3 Meetniveau

In de statistiek zijn er vier verschillende meetschalen waarmee je je gegevens meet.

Je hebt dit onderscheid nodig voor de juiste keuze van een statistische toets.

**Ratio** Een ratioschaal betreft getallen en heeft een absoluut nulpunt. Voorbeelden hiervan zijn verkopen in euro's, lengtes in centimeters en gewicht in kilo's. Je kunt met deze meetschalen verhoudingen tussen metingen berekenen. 10 kilometer is bijvoorbeeld precies twee keer zo ver als 5 kilometer.

**Interval** In sommige gevallen zijn de intervallen tussen getallen op een numerieke (dus met cijfers) schaal gelijk van grootte. Je kunt hierbij geen verhoudingen berekenen. 10°C is niet twee keer zo warm als 5°C. Dit geldt wel bij temperatuurmeting in Kelvin, dat is dan ook op rationiveau.

Voor het kiezen van een statistische toets is de intervalmeting echter vergelijkbaar met een ratiometing.

**Nominaal** Nominaal verwijst naar namen, de waarden worden ingedeeld in categorieën met een naam. Je kunt deze niet rangschikken. Voorbeelden zijn godsdiensten, diersoorten en provincies.

**Ordinaal** Metingen op ordinaal niveau hebben een natuurlijke ordening, je kunt ze dus rangschikken. Voorbeelden zijn opleidingsniveaus, de 5-puntsschaal bij enquêtes (zeer mee oneens - mee -oneens - neutraal - mee eens - zeer mee eens), inkomensklasse. Let op: ook cijfers kunnen op ordinaal niveau zijn, je kunt bepaalde waarden namelijk nummers geven op volgorde van belangrijkheid of hoeveelheid, dan zijn het rangnummers. Verwar dit niet met een ratio- of intervalschaal.

## 2.4 Verdeling van de populatie: parametrische en niet-parametrische toetsen

Heb je geconstateerd dat je een ratio- of intervalschaal gebruikt (verreweg het meest gebruikt), moet je beslissen of je een parametrische of een non-parametrische toets gaat gebruiken. Het verschil is niet moeilijk. Voor veel statistische methoden is het nodig dat de details van de kenmerken van de populatie bekend zijn of geschat kunnen worden. Dit worden **parametrische toetsen** genoemd (een parameter is een kenmerk van een populatie). In het algemeen moeten de testresultaten min of meer de normale (klokvormige) verdeling benaderen om nauwkeurige conclusies te trekken. Het gemiddelde zit bij een normale verdeling in het midden, en de verdeling neemt aan beide kanten gelijkmatig af. Heb je bij je gegevens zelfs gebruik gemaakt van meer dan één steekproef, dan moeten de populaties dezelfde variantie hebben. Variantie is de standaardafwijking (ook wel standaarddeviatie genoemd), die jullie kennen van wiskunde, in het kwadraat. Het jullie bekende symbool voor de standaardafwijking is  $\sigma$ , dus die voor de variantie is  $\sigma^2$ .

Het kan zijn dat er niet wordt voldaan aan de voorwaarden van een normale verdeling waardoor je niet de standaard toets kunt gebruiken die we allemaal met wiskunde statistiek hebben geleerd. In dit geval mag je geen parametrische toets gebruiken en heb je een **niet-parametrische toets** (parametervrije toets) nodig.

Om dus te bepalen of je een parametrische of niet-parametrische toets je moet gebruiken, stel je dus vast of je met een normale verdeling hebt te maken of niet. Een ruwe voorwaarde is dat je steekproef groter moet zijn dan 40 personen. Een andere methode is kijken of het gemiddelde van je scores hetzelfde is als de mediaan (het middelste getal als je alle getallen rangschikt) en de modus (het getal dat het vaakst voorkomt). Is dit niet het geval, dan heb je geen normale verdeling. Deze simpele methode brengt wel problemen met zich mee. Het komt namelijk zelden voor dat deze precies hetzelfde zijn, en welke afwijking mag je dan permitteren?

Een meer nauwkeurige methode is het vaststellen van de **scheefheid** (de afwijking van de normale verdeling). Aan de hand van een simpele voorbeeldsteekproef zullen we de formules die hiervoor nodig zijn verduidelijken.

Je wilt de gemiddelde leeftijd van studenten onderzoeken. Je hebt hiervoor een steekproef van negen studenten genomen. In de onderstaande tabel zie je de gegevens.

Responsnr.	Leeftijd	Afwijking gem. (d)	Afwijking gem. <sup>3</sup> (d <sup>3</sup> )
1	22	1	1
2	18	5	125
3	24	1	1
4	19	4	64
5	25	2	8
6	27	4	64
7	23	0	0
8	26	3	27
9	23	0	0
Totaal	207	20	290

Het gemiddelde bereken je door 207 te delen door 9, de afwijking van het gemiddelde is de absolute waarde (een positief getal) van de leeftijd min het gemiddelde. Hieronder zie je formule voor de scheefheid.

$$\text{Scheefheid} = \frac{\sum d^3 \cdot n}{\sigma^3 \cdot (n-1)(n-2)}$$

Hierin is  $\sum d^3$  de som van de afwijking van het gemiddelde van elke score tot de derde macht,  $n$  het aantal scores (dus de steekproefomvang),  $\sigma$  de geschatte standaardafwijking. Deze laatste kun je met je GR berekenen:

- Voer de leeftijden in bij een lijst (STAT -> 1: Edit...)
- Klik op (STAT -> Calc -> 1: 1-Var Stats)
- Heb je je gegevens in lijst 1, dan hoeft je er niks achter te zetten, heb je ze bij lijst 2, zet er dan achter  $L_2$  (List ->  $L_2$ ) en dan op ENTER
- Het eerste symbool is het steekproefgemiddelde (23), verder naar beneden zie je  $\sigma x$  (2.828..) staan: de standaardafwijking. Deze ga je gebruiken voor in de formule.

$$\text{Scheefheid} = \frac{\sum d^3 \cdot n}{\sigma^3 \cdot (n-1)(n-2)} = \frac{\sqrt{290} \cdot 9}{2,83^3 \cdot 8 \cdot 7} = 0,12$$

Hoe dichterbij 0 zit, hoe kleiner de scheefheid. Is de waarde negatief, heb je een links-scheve verdeling, is de waarde positief, dan heb je een rechts-scheve verdeling. Dan heb je nu nog de standaardfout van de scheefheid nodig, zie onderstaande formule.

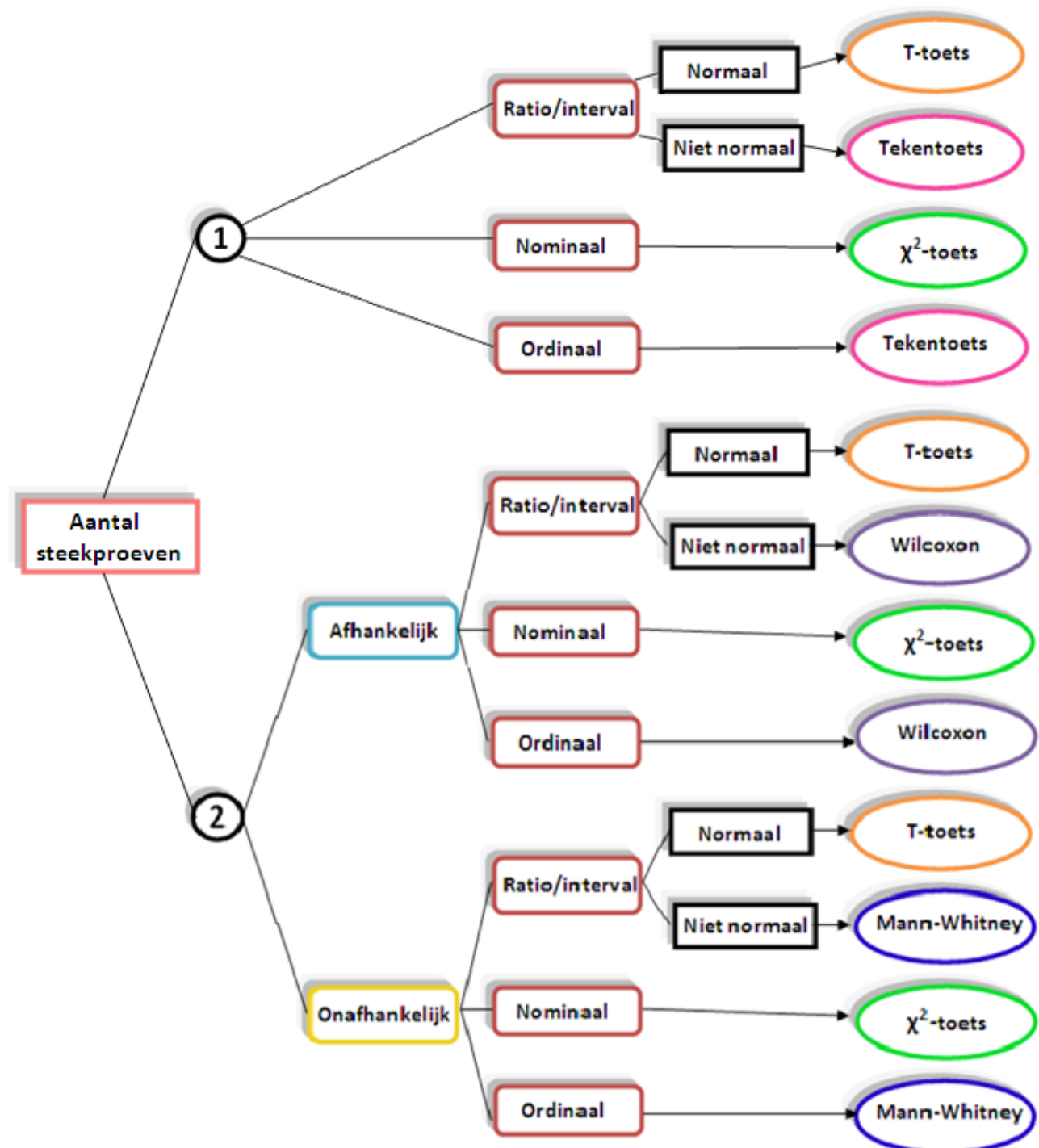
$$\text{Standaardfout van de scheefheid} = \sqrt{\frac{6 \cdot n \cdot (n-1)}{(n-1) \cdot (n+1) \cdot (n+3)}} = \sqrt{\frac{6 \cdot 9 \cdot 8}{7 \cdot 10 \cdot 12}} = 0,72$$

En voor de significantie van de scheefheid deel je de scheefheid door de standaardfout.

$$\text{Significantie} = \frac{\text{scheefheid}}{\text{standaardfout}} = \frac{0,12}{0,72} = 0,17$$

Deze waarde vergelijk je met de minimale waarde van  $z$  voor een significantieniveau van 5% ( $\alpha = 0,05$ ) bij een tweezijdige toets, namelijk 1,96. In het voorbeeld is de significantie voor de scheefheid lager dan 1,96, in dit geval hebben we te maken met een normale verdeling en mogen we dus een parametrische toets gebruiken. Is de waarde hoger dan 1,96, dan moet je een parametervrijetoets gebruiken.

### 3. Schema voor de keuze van de juiste statistische toets



Schema 1. Het kiezen van de juiste statistische toets

NB. Hierin betekent 'Normaal' dat de steekproef een normale verdeling betreft.

## 4. Uitwerking statistische toetsen

In dit hoofdstuk volgt een overzicht van de statistische toetsen die in het schema staan, met voorbeelden ter verduidelijking.

### 4.1 T-toets

De T-toets is een heel belangrijke toets, omdat het de enige toets is die werkt met gegevens uit een normale verdeling. Je kunt deze toets op je GR uitvoeren. Er zijn twee soorten T-toetsen: één waarbij de standaardafwijking van de populatie bekend is, en één waarbij die niet bekend is. Een voorbeeld waarbij dat wel zo is, is als je een onderzoek hebt gelezen waarbij het gemiddelde en de standaarddeviatie gegeven is, en jij wilt onderzoeken of dat gemiddelde ook klopt bij ons op school. Als je zelf een onderzoek hebt bedacht, is de standaardafwijking nog niet bekend.

#### 4.1.1 De standaardafwijking is bekend

##### Voorbeeld

In een onderzoek naar examens lees je dat het gemiddelde cijfer voor VWO een 6,6 is, met een standaardafwijking van 0,4. Maar, denk je, ons examencijfer zal vast veel hoger liggen. Je besluit dit te onderzoeken.

##### Stap 1. De vraagstelling formuleren.

Is het gemiddelde examencijfer op het Praedinius hoger dan het landelijk VWO-gemiddelde?

##### Stap 2. De hypothesen opstellen.

$H_0: \mu = 6,6$ .

$H_1: \mu > 6,6$

##### Stap 3. Gegevens vergaren.

Je vraagt aan 30 oud-leerlingen hun examencijfer. De gegevens zie je in de tabel.

6,5	8,1	6,8	6,1	7,5	6,6	6,4	6,3	7,5	6,8
6,8	6,2	6,3	6,6	6,4	6,8	7,8	7,3	7,9	7,0
7,3	6,7	7,1	7,4	7,3	7,7	7,6	6,7	7,0	6,9

##### Stap 4. Gegevens verwerken.

Aangezien je te maken hebt met een normale verdeling, en je een ratioschaal hebt, mag je de T-toets doen. Omdat de standaarddeviatie bekend is, gebruik je een variant erop: de Z-toets.

Noteer het gemiddelde door middel van 1-Var Stats (STAT -> CALC -> 1: 1-Var Stats). Je ziet dat het gemiddelde 6,98 is. Deze gegevens gaan we nu toetsen:

$\mu = 6,6$

$\sigma = 0,4/\sqrt{30}$

Het steekproefgemiddelde reken je immers uit met de  $\sqrt{n}$ -wet: de standaardafwijking van de populatie deel je door de wortel van de grootte van je steekproef, omdat de kans dat het steekproefgemiddelde ver afwijkt van het ware gemiddelde is veel kleiner dan alle afwijkingen van de hele populatie.



De kans dat ons steekproefgemiddelde (6,98) afwijkt van het gemiddelde (6,6), reken je uit door middel van de p-waarde. De p-waarde is de kans, indien de nulhypothese waar is, om een waarde te krijgen van de toetsingsgrootte die minstens even extreem is als de geobserveerde waarde. Met extreem bedoelen we waarden die nog meer zouden wijzen in de richting van de alternatieve hypothese. Deze p-waarde ga je vergelijken met het significantieniveau  $\alpha$ . Je kunt meestal het beste kiezen voor  $\alpha = 0,05\%$ .

De p-waarde rekenen we uit:

$$P(X \geq 6,6) = \text{normalcdf}(6,98, 10^{99}, 6,6, 0,4/\sqrt{30})$$

Je krijgt hier  $9,80 \dots \cdot 10^{-8}$  uit. Dus de kans is  $0,000000098 = 0,0000098\%$ .

Deze waarde is veel kleiner dan ons gekozen significantieniveau 5%. Daarom kunnen we stellen dat we de nulhypothese verwerpen.

#### Stap 5. Toetsen.

Je kunt de p-waarde ook uitrekenen door middel van de Z-test op je GR. Dit is nauwkeuriger, omdat het alle gegevens erbij betreft en niet alleen het gemiddelde.

Vul de waarden van de tabel in  $L_1$  (LIST -> 1: Edit...).

Kies (STAT -> TESTS -> 1: Z-Test). Zorg ervoor dat je het volgende op je scherm krijgt:

Inpt: **Data** Stats

$\mu_0$ : 6.6 (het gemiddelde van het examenonderzoek)

$\sigma$ : 0.4 (de standaardafwijking het examenonderzoek)

List:  $L_1$

Freq: 1

$\mu$ :  $\neq \mu_0$   $< \mu_0$   **$> \mu_0$**  (onze alternatieve hypothese was immers dat het gemiddelde van onze steekproef meer was dan het gemiddelde van het examenonderzoek)

Klik op Calculate. Je krijgt nu een aantal gegevens, waaronder de p-waarde. Deze is  $9,80 \dots \cdot 10^{-8}$ .

#### Stap 6. Gegevens analyseren

Onze p-waarde is kleiner dan 0,05, onze vooraf gekozen significantieniveau.

Daarom kan de nulhypothese verworpen worden.

In ons voorbeeldje was de p-waarde toevallig gelijk met wat we hadden uitgerekend door middel van de normale verdeling, maar het kan net zo goed erg verschillen. Doe dus altijd stap 5.

#### 4.1.2 De standaardafwijking is niet bekend

Het kan heel goed zijn dat je een onderzoek gaat doen waarbij de standaardafwijking niet bekend is.

In dat geval gebruik je de T-toets. Deze gaat bijna precies hetzelfde als de Z-test:

Vul de waarden van de tabel in.

Kies nu de T-test (STAT -> TESTS -> 2: T-Test).

In het scherm vul je hetzelfde in als bij de T-toets, maar merk op dat je geen standaardafwijking hoeft in te vullen, en klik op Calculate. Bij de p-waarde vind je nu  $3,06 \dots \cdot 10^{-4} = 0,000306 = 0,031\%$ .

Dit is ook minder dan 5%, en ook nu weer kunnen we de nulhypothese verwerpen.

#### 4.2 Tekentoets

De tekentoets wordt gebruikt bij één steekproef die afhankelijk is genomen.

Het heet een tekentoets omdat je de tekens telt die aangeven of de score boven of onder de mediaan ligt.

De mediaan van een rij waarnemingsgetallen is het middelste waarnemingsgetal in het geval de getallen in volgorde van grootte staan. Ruwweg kan gesteld worden dat het aantal waarnemingsgetallen dat groter is dan de mediaan gelijk is aan het aantal waarnemingsgetallen dat kleiner is dan de mediaan.

### Voorbeeld

De eigenaar van een fietsenzaak beweert dat de mediaan van het aantal verkochte sportfietsen per maand gelijk is aan 18. Op grond van een aselechte steekproef met de verkoopcijfers van twaalf maanden zal beslist worden of de bewering van de eigenaar geaccepteerd wordt. De normale toets is hierbij niet te gebruiken, want het is niet bekend of de toevalsvariabele  $X = \text{het aantal verkochte sportfietsen per maand}$  normaal verdeeld is.

#### Stap 1. Vraagstelling formuleren.

Is de mediaan van het aantal verkochte sportfietsen per maand 18?

#### Stap 2. Hypothesen opstellen.

$H_0: P = 0,5$

$H_1: P \neq 0,5$ .

De nulhypothese is dat de mediaan 18 is. Bij deze veronderstelling is de kans op een plusteken even groot als de kans op een minteken. Dus  $P(+) = 0,5$ .  $P(\text{waarneming} > \text{mediaan}) = 0,5$ , want de mediaan splitst de populatie ruwweg in twee even grote delen.

#### Stap 3. Gegevens verzamelen.

Maandverkoop	waarneming – mediaan	teken
10 20 28 8	----->	- + + -
19 2 24 20		+ - + +
19 20 26 20		+ + + +

#### Stap 4. Gegevens verwerken.

Het omzetten van de steekproefresultaten in plus- en mintekens levert drie mintekens op en negen plustekens. Datgene wat we willen toetsen ( $X$ ) is het aantal plustekens in de steekproef van lengte 12. Bij  $H_0$  is  $X$  binomiaal verdeeld met  $n = 12$  en  $p = 0,5$ . We komen op een steekproefresultaat van  $X = 9$ . We nemen  $\alpha = 0,05$ .

#### Stap 5. Toetsen

Omdat  $9 > 0,5 * 12$  is de overschrijdingskans van 9 gelijk aan  $P(X \geq 9)$ .

Deze kans is gelijk aan  $1 - P(X \leq 8) = 1 - \text{binomcdf}(12, 0,5, 8) = 0,073....$

#### Stap 6. Gegevens analyseren.

0,073 is meer dan 0,05. Daarom kunnen we  $H_0$  niet verwerpen, er is dus geen aanleiding om de mediaan 18 in twijfel te trekken.

Bij het toepassen van de tekentoets spreek je eerst af of je mintekens of plustekens telt.

Afhankelijk van de probleemstelling volgt hieruit bij een eenzijdige toets  $H_1: p < 0,5$  of  $H_1: p > 0,5$ .

### 4.3 Chi-kwadraat-toets

De chi-kwadraat-toets is een toets die je gebruikt als je bepaalde eigenschappen van een populatie wilt vergelijken met de verwachtingswaarde of met die van een andere populatie. Je kunt bijvoorbeeld testen of eerste klassers een voorkeur voor een bepaalde sport hebben. Over een steekproef van 100 eerste klassers vergelijkt je hun scores voor elke sport met die van de verwachtingswaarde 25 (want als er geen voorkeur bestaat is de kans dat ze per toeval voor een bepaalde sport kiezen 0,25). Je kunt ook een bepaalde populatie vergelijken met een andere populatie om te kijken of ze dezelfde verdeling hebben voor een bepaalde eigenschap (in dit voorbeeld voorkeur voor sport). Je eerste populatie wordt dan de verwachtingswaarde. Is het aantal proefpersonen van de twee populaties niet gelijk, deel of vermenigvuldig alle scores van je verwachtingswaarde dan zo dat de aantallen gelijk zijn.

### Voorbeeld

Je bent benieuwd of mensen in de 6<sup>e</sup> klas heel andere sporten doen dan de 1<sup>e</sup> klas.  
Hiervoor hou je een enquête onder 90 leerlingen.

#### Stap 1. Vraagstelling formuleren.

Is de meest beoefende sport in de zesde klas gelijk als in de eerste klas?

#### Stap 2. Hypothesen opstellen.

H<sub>0</sub>: De meest beoefende.

H<sub>1</sub>: De meest beoefende sport is in de zesde klas verschillend aan die in de eerste klas

#### Stap 3. Gegevens verzamelen.

	Tennis	Voetbal	Hockey	Fitness	Totaal
1 <sup>e</sup> klas	20	30	40	10	100
6 <sup>e</sup> klas	15	20	35	30	100
Vershil	5	10	5	20	
Vershil <sup>2</sup>	25	100	25	400	

#### Stap 4. Gegevens verwerken.

In de tabel zie je al meteen het verschil in het kwadraat. De 1<sup>e</sup> klas was in dit geval de verwachtingswaarde.

#### Stap 5. Toetsen

Voor de chi-kwadraattoets gebruik je de onderstaande formule.

$$\chi^2 = \frac{\sum \text{verschil}^2}{\text{verwachtingswaarde}} = \frac{25}{20} + \frac{100}{30} + \frac{25}{40} + \frac{400}{10} = 45,2$$

Deze waarde moet je vervolgens vergelijken met de waarde in de tabel die je onderaan op deze site vindt:

<http://www.bioplek.org/techniekkartenbovenbouw/techniek98x2.html>

Je moet kijken naar het aantal vrijheidsgraden (n-1), in ons geval is dat 4 – 1 = 3.

In de tabel vind je een waarde van 7.815.

#### Stap 6. Gegevens analyseren

45.2 > 7.815, en de nulhypothese wordt verworpen. 6<sup>e</sup> klassers hebben dus andere favoriete sporten dan 1<sup>e</sup> klassers.

### 4.4 Wilcoxontoets

De Wilcoxon wordt gebruikt bij twee afhankelijke steekproeven. Deze toets gebruikt iets meer informatie dan bijvoorbeeld de tekentoets, waardoor hij eerder significante verschillen kan ontdekken.

Je kunt met deze toets berekenen of er een significant verschil is tussen resultaten van groep X en groep Y.

Je kunt van deze toets gebruik maken in het geval dat je 5 of meer waarnemingen van groep X hebt en 10 of meer van groep Y.

### Voorbeeld

Twee rastafara's willen graag weten of hun wietplantjes sneller of juist langzamer gaan groeien onder invloed van muziek. De ene helft van de zaadjes planten ze in een stille kamer, de andere helft in een kamer met een gettoblaster waarop reggea wordt afgespeeld. Na een week zetten ze de plantjes één voor één op een weegschaal om het gewicht te bepalen.

#### Stap 1. Vraagstelling formuleren.

Heeft muziek invloed op de groei van wietplantjes?

#### Stap 2. Hypothesen opstellen.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

#### Stap 3. Gegevens verzamelen.

X = het gewicht van een plant die is gegroeid in stilte.

Y = het gewicht van een plant die is gegroeid onder het genot van reggea.

X: 1.234, 1.295, 1.349, 1.358, 1.361, 1.377, 1.387, 1.424, 1.476, 1.576, 1.635, 1.658

Y: 1.287, 1.304, 1.311, 1.366, 1.421, 1.456, 1.463, 1.498, 1.523, 1.555, 1.589, 1.608, 1.621, 1.645

#### Stap 4. Gegevens verwerken.

Neem  $\alpha = 0,05$ .

Neem aan dat er n waarnemingen X zijn en m waarnemingen Y met  $n \leq m$ .

(Er zijn dus gelijk of meer waarnemingen Y dan er waarnemingen van X zijn.)

Maak een tabel waarin je X en Y rangschikt van klein naar groot. Het kleinste getal krijgt rangnummer 1, het één na kleinste getal rangnummer 2, enz. Bereken dan de som van de rangnummers van S ( $S_x$ ).

X gerangschikt van klein naar groot	Rangnummer	Y gerangschikt van klein naar groot	Rangnummer
.234	1	1.287	2
1.295	3	1.304	4
1.349	6	1.311	5
1.358	7	1.366	9
1.361	8	1.421	12
1.377	10	1.456	14
1.387	11	1.463	15
1.424	13	1.498	17
1.476	16	1.523	18
1.576	20	1.555	19
1.635	24	1.589	21
1.658	26	1.608	22
		1.621	23
		1.645	25

Het totaal van de rangnummers van S is in dit geval dus:

$1 + 3 + 6 + 7 + 8 + 10 + 11 + 13 + 16 + 20 + 24 + 26 = 145$ .

### Stap 5. Toetsen.

Als er geen verschil is tussen de werking van de muziek dan is:

$$E(Sx) = \frac{1}{2} \cdot n(n+m+1)$$

Verschilt  $Sx$  te veel van  $E(Sx)$ , dan is er een significant verschil tussen de rijen waarnemingsgetallen.

$E(Sx)$  is de verwachte waarde van  $Sx$ . Je kunt  $E(Sx)$  gelijk stellen aan  $\mu$ , want dat is de waarde die je verwacht als de nulhypothese klopt. Hierbij is de nulhypothese dat je verwacht dat de waarnemingen  $X$  niet significant verschillen van waarnemingen  $Y$ .

$$\mu = E(Sx) = \frac{1}{2} \cdot n(n+m+1)$$

Je kunt nu de significantie berekenen d.m.v. normalcdf. Hierbij moet je ook de standaardafwijking weten. Bij de Wilcoxon toets kun je de standaard afwijking berekenen met de volgende formule:

$$\sigma = \sqrt{\frac{1}{12} \cdot nm(n+m+1)}$$

We gaan nu weer met behulp van het voorbeeld laten zien hoe je dit dan moet berekenen.

We hebben eerder uitgerekend dat  $Sx=145$ .

Om de significantie te bepalen willen we nu  $\mu$  en  $\sigma$  berekenen.

$$\mu = E(Sx) = \frac{1}{2} \cdot n(n+m+1) = \frac{1}{2} \cdot 12(12+14+1) = 162$$

$$\sigma = \sqrt{\frac{1}{12} \cdot nm(n+m+1)} = \sqrt{\frac{1}{12} \cdot 12 \cdot 14(12+14+1)} = 19,4$$

We toetsen tweezijdig, dus de overschrijdingskans van 145 is:

$$P(Sx \leq 145) = \text{normalcdf}(-10^99, 145, 162, 19.40) \approx -0,190$$

### Stap 6. Gegevens analyseren

We hebben een significantie niveau van  $\alpha = 0,05$ .

Bij tweezijdig toetsen moet 0,190 dus minder zijn dan  $\frac{1}{2}\alpha = 0,025$ .

Dat is hier niet zo.

Er is dus geen aanleiding te veronderstellen dat muziek invloed heeft op de groei.

## 4.5 Mann-Whitneytoets

De Mann-Whitney toets is een toets die nagaat of twee onafhankelijke steekproeven uit dezelfde populatie komen of eenzelfde verdeling hebben. De onderlinge ligging van de gegevens worden vergeleken door te tellen hoeveel van steekproef A een lager rangnummer hebben dan B, of andersom.

Dit opgetelt heet de U-waarde, die wordt vergeleken met de U-waarde uit de U-tabel, die je kunt aflezen door middel van de steekproefomvang. Ligt de gevonden waarde boven die van de tabel, dan wordt de nulhypothese behouden. Ligt de gevonden waarde eronder, dan wordt de nulhypothese verworpen.

### Voorbeeld

Freek vraagt zich af of de verkoop Cup-a-Soup in de kantine afhangt van het seizoen.

Hij gaat de verkoop in twee seizoenen vergeleken: 10 dagen in de zomer (A) en 10 dagen in de winter (B).

#### Stap 1. Vraagstelling formuleren.

Hangt de verkoop van Cup-a-Soup af van het seizoen?

#### Stap 2. Hypothesen opstellen.

$H_0: \mu_A = \mu_B$

$H_1: \mu_A \neq \mu_B$

#### Stap 3 en 4. Gegevens verzamelen en verwerken.

	A	A	A	A	B	A	A	B	A	B
score	1	2	3	4	6	6	6	7	8	8
rang	1	2	3	4	6	6	6	8	10.5	10.5

	A	B	A	B	A	B	B	B	B	B
score	8	8	9	10	10	11	12	13	13	14
rang	10.5	10.5	13	14.5	14.5	16	17	18.5	18.5	20

#### Stap 5. Toetsen.

De toetsingsgrootheid U berekent voor elk van de elementen van de ene steekproef (zeg B) hoeveel elementen van de ander steekproef (A) kleiner zijn.

$$U = 0 + 0 + 0 + 0 + 1 + 1 + 2 + 3 + 4 + 5 = 16$$

Deze waarde vergelijken we met de waarde die we vinden in de U-tabel, die kun je vinden in de onderstaande link:

<http://math.usask.ca/~laverty/S245/Tables/wmw.pdf>

We hebben twee steekproeven van 10, zoeken we die op in de tabel dan vinden we de U-waarde van 23.

#### Stap 6. Gegevens analyseren.

Onze waarde van 16 ligt onder 23, dus onze nulhypothese wordt verworpen.

De verkoop van Cup-a-Soup is seizoensgebonden.

## 5. Nawoord

Het viel lang niet mee om een handleiding statistische toetsen te maken voor PWS-ers.

Er zijn bibliotheken vol over geschreven, maar waarom kan niemand het nou eens gewoon lekker simpel uitleggen? De sectie zelf had geen tijd om dat allemaal uit te gaan zoeken en dus werden wij met deze lastige, maar interessante opdracht opgezadeld.

Met een heleboel boeken en sites die elkaar tegenspraken in alle mogelijke manieren zijn we erin geslaagd (vinden we) een begrijpelijke handleiding te maken, die niet saai is om te lezen en toetsen toelicht met behulp van toepasselijke voorbeelden. We weten nu wel dat je niet in tien uurtjes alle statistische toetsen kunt leren toe te passen, maar we hebben geprobeerd zoveel mogelijk relevante dingen voor deze handleiding te selecteren en het overbodige weg te laten, zodat die tijd bespaard blijft die gebruikt moet worden voor het PWS, want dat kost al zoveel tijd.

Dus beste PWS-er, we hopen dat je nu zonder problemen een waterdicht statistisch bewijs kunt leveren!

## 6. Literatuurlijst

1. P., DE PELMACKER; P. VAN KENHOVE. Marktonderzoek. Amsterdam, Pearson Education, 2003.
2. Meetschalen. <http://nl.wikipedia.org/wiki/Meetniveau>.
3. F. VAN DER ZEE. Kennisverwerving in de Empirische Wetenschappen, de methodologie van - wetenschappelijk onderzoek. BMOOO, Groningen, 2004.
4. <http://www.moaweb.nl/bibliotheek/digitaal-woordenboek/s/scheefheid>
5. D. HOWITT; D. CRAMER. Statistiek in de sociale wetenschappen. Amsterdam, Pearson Education, 2007.

**PWS**