

Toetsen van hypothesen: Breng statistiek tot leven in je klas!

Prof. dr. Steven Abrams

Dag van de Wiskunde en Wetenschappen
22 maart 2025



WWW.UHASSELT.BE/DSI

Inhoud



UHASSELT

KNOWLEDGE IN ACTION

Inhoud van de sessie

Inleiding: Wat is verklarende statistiek?

Deel 1: Hoe extreem mag je zijn?

Deel 2: Hypothesen

Deel 3: Toetsen van hypothesen

Deel 4: Proporties

Deel 5: Gemiddelden

(Deel 6: Toetsen voor twee properties of gemiddelden)



Voorkennis

We steunen op basisbegrippen rond:

- Kansexperimenten en kanstheorie
- Van steekproef naar populatie
- Beschrijvende statistiek
- Punt- en intervalschatten



Wat is verklarende statistiek?



UHASSELT

KNOWLEDGE IN ACTION

Verklarende statistiek

Kunnen we aan de hand van **gegevens** een **beslissing trekken** over een **specifieke onderzoeksvraag**, rekening houdend met de intrinsieke **onzekerheid** verbonden aan het uitvoeren van een (kans)experiment en/of het nemen van een **steekproef uit een populatie**?



Deel 1: Hoe extreem mag je zijn?



UHASSELT

KNOWLEDGE IN ACTION

Zeldzame gebeurtenissen



UHASSELT

KNOWLEDGE IN ACTION

Zeldzame gebeurtenissen

We spelen een (fictief) spel waarbij lukraak 0'en en 1'en gegenereerd worden door een computerprogramma dat ikzelf geschreven heb

Regels:

- Indien het een 1 geeft, verlies ik en geef ik jullie 10 euro (gelijk te verdelen over alle deelnemers)
- Indien het een 0 geeft, verliezen jullie en geven jullie mij 10 euro

Uitkomsten:

0, 0, 0, 0, ?

Na de vijfde keer een 0 te trekken, willen jullie mij (waarschijnlijk) niet meer betalen; er is mogelijk een **sterk vermoeden** dat mijn programma **niet eerlijk** is?!



UHASSELLT

KNOWLEDGE IN ACTION

Zeldzame gebeurtenissen

Soortgelijke experimenten werden uitgevoerd bij **verschillende personen (of groepen van individuen)**.

Telkens na 4 of 5 mislukkingen (d.i. een 0 trekken) hadden de meest mensen een sterk vermoeden dat er iets niet klopte. **Dit was een reactie op gevoel!**

Maar hoe waarschijnlijk is zo een gebeurtenis?



Zeldzame gebeurtenissen

In het geval van een eerlijk spel, is de kans op 0 even groot als de kans op 1 (beide uitkomsten treden op met een kans van 0.5 of 1/2).

Indien de gebeurtenissen onafhankelijk zijn, dan is de kans op 2 mislukkingen gelijk aan:

$$P(Y_1 = 0 \text{ en } Y_2 = 0) = P(Y_1 = 0, Y_2 = 0) = P(Y_1 = 0) \cdot P(Y_2 = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

Voor drie mislukkingen is de kans gelijk aan:

$$P(Y_1 = 0 \text{ en } Y_2 = 0 \text{ en } Y_3 = 0) = P(Y_1 = 0) \cdot P(Y_2 = 0) \cdot P(Y_3 = 0) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = 0.125$$

Voor vier of vijf 0'en daalt de kans verder tot 6.25% en 3.125%.



Zeldzame gebeurtenissen

De meeste mensen beschouwen een gebeurtenis als **onverwacht** wanneer ze een kleine kans heeft om op te treden.

Dit kadert in het geloof dat het spel *eerlijk* is.

Conclusie: Binnen die context is een **start met 4 à 5 mislukkingen** op een rij “uitzonderlijk”, met een kans dat zoiets gebeurt in de buurt van 5%.



Zeldzame gebeurtenissen

In de (verklarende) statistiek gebruikt men al lang een **5% criterium** (ook wel significantieniveau genoemd) om over **onverwachte** gebeurtenissen te spreken.

Je kan dit ook bekijken vanuit het perspectief van meerdere herhalingen van het experiment (of de studie):

- Indien je de proef 100 keer uitvoert tegen verschillende mensen en je gaat ervan uit dat het spel eerlijk is, dan gebeurt het slechts bij (ongeveer) 5 van de 100 personen dat ze starten met 4 à 5 mislukkingen op rij.
- Bij (ongeveer) 95 van de 100 personen zal het niet zo zijn dat de eerste 4 à 5 getallen gelijk zijn aan 0.



Toetsen van hypothesen

Bij **toetsen van hypothesen** gebruikt men een significantieniveau (of criterium) van 5% in de meeste studies om een uitkomst onverwacht, uitzonderlijk, zeldzaam of extreem te noemen.

Meer bepaald, indien de kans hoogstens 5% is dat je, binnen de context van de studie, een uitkomst ziet, dan is zulke uitkomst *uitzonderlijk*.



Waarden of gebieden?



UHASSELT

KNOWLEDGE IN ACTION

Waarden of gebieden?

Een bedrijf gebruikt een **grote machine in haar productieproces**. Die machine wordt elke maand gecontroleerd. Men meet dan gedurende meerdere dagen en op verschillende tijdstippen de kwaliteit van de output. Dat gebeurt volgens een vastgelegd schema waarbij **alle meetresultaten uiteindelijk tot één eindresultaat worden samengebracht**.

Men weet dat die manier van meten een **hoge eindscore** oplevert als de **machine dringend toe is aan onderhoud**. Als dit dan niet gebeurt, gaat ze stuk.

Waarden of gebieden?

Volgens de constructeur levert een **goed werkende machine niet altijd dezelfde eindscore** op.

De mogelijke eindscores en de **kansen dat je die scores opmeet** bij een goed werkende machine zien er als volgt uit:

Mogelijke eindscore	≤ 20	$(20, 24]$	$(24, 25]$	$(25, 30]$	$(30, 36]$
Kans bij een goed werkende machine	$\frac{30}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Waarden of gebieden?

Een machine die dringend aan onderhoud toe is levert meestal **een hoge score** (> 24). Een machine stilleggen heeft gevolgen:

- Een onderhoudsbeurt zorgt voor het stilleggen van het productieproces voor één dag stilleggen.
- Als de machine stuk gaat dan duurt het weken vooraleer ze kan hersteld worden.

Als je nu de klassieke 5% grens gebruikt om over een “extreme gebeurtenis” te spreken, **welke score moet je dan vinden om te beslissen dat de machine een onderhoudsbeurt nodig heeft?** Je weet dat je moet kijken naar hoge scores.



Waarden of gebieden?

Als de machine nog goed werkt dan is de kans om één van de uitkomsten voor de score te vinden gelijk aan:

$$P(Y \in (30, 36]) = 1/36 \approx 0.028 = 2.8\%$$

$$P(Y \in (25, 30]) = 2/36 \approx 0.056 = 5.6\%$$

$$P(Y \in (24, 25]) = 1/36 \approx 0.028 = 2.8\%$$



Waarden of gebieden?

Als de machine nog goed werkt dan is de kans om één van de uitkomsten voor de score te vinden gelijk aan:

$$P(Y \in (30, 36]) = 1/36 \approx 0.028 = 2.8\%$$

$$P(Y \in (25, 30]) = 2/36 \approx 0.056 = 5.6\%$$

$$P(Y \in (24, 25]) = 1/36 \approx 0.028 = 2.8\%$$

Dus de machine heeft een onderhoud nodig voor een score in het interval (30, 36] en (24, 25], maar niet in het interval (25, 30]



Waarden of gebieden?

Conclusie: de machine heeft een onderhoud nodig voor een score in het interval $(30, 36]$ en $(24, 25]$, maar niet in het interval $(25, 30]$

Tegenstrijdigheid:

Een hogere score zou moeten impliceren dat de machine een dringender onderhoud nodig heeft

Echter, meer extreme uitkomsten gaan niet noodzakelijk samen met kleinere kansen.

Oplossing: werken met gebieden



UHASSETT

KNOWLEDGE IN ACTION

Waarden of gebieden?

Oplossing: werken met gebieden

$$P(Y > 24) = \frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{4}{36} \approx 0.111 = 11.1\%$$

$$P(Y > 25) = \frac{2}{36} + \frac{1}{36} = \frac{3}{36} \approx 0.083 = 8.3\%$$

$$P(Y > 30) = 1/36 \approx 0.028 = 2.8\%$$

Dus, we vragen een onderhoud van zodra de **score groter is dan 30**



Toetsen van hypothesen

Bij **toetsen van hypothesen** kijk je niet naar extreme uitkomsten (extreme getalwaarden), maar naar extreme gebieden. Die gebieden zijn dan **rechterstaarten** (waarin extreem grote uitkomsten terechtkomen) of **linkerstaarten** (waarin extreem kleine uitkomsten in terechtkomen).

Relevante kansen zijn dus kansen om in staarten van verdelingen terecht te komen.



Deel 2: Hypothesen



UHASSELT

KNOWLEDGE IN ACTION

Een hypothese is ...

Een hypothese is **een bewering over een populatie**.

Meestal gaat zo een bewering over een populatieparameter.

Voorbeeld (1):

Indien je ervan overtuigd bent dat in 2024 even veel jongens als meisjes geboren werden in Vlaanderen, dan zeg je dat

$$\pi = \frac{1}{2} = 0.5$$

met π de **populatieproportie**, in dit voorbeeld de proportie jongens bij alle baby's geboren in Vlaanderen in 2024.



Een hypothese is ...

Een hypothese is **een bewering over een populatie**.

Meestal gaat zo een bewering over een populatieparameter.

Voorbeeld (2):

Ergens las je dat meisjes gemiddeld 172 cm groot zijn. Zelf denk je dat Vlaamse meisjes van 17 kleiner zijn. Dus de hypothese is

$$\mu < 172$$

met μ het **populatiegemiddelde**, in dit voorbeeld de gemiddelde lengte is van 17-jarige Vlaamse meisjes.



Een hypothese is ...

Een hypothese is **een bewering over een populatie**.

Meestal gaat zo een bewering over een populatieparameter.

Een hypothese formuleer je vooraleer je de steekproef trekt uit de populatie waarvoor je de **bewering** wenst te **toetsen**.



Nul- en alternatieve hypothese



UHASSELT

KNOWLEDGE IN ACTION

Nul- en alternatieve hypothese

Een **nulhypothese** H_0 is een bewering dat een populatieparameter een welbepaalde waarde aanneemt.

Meestal is de welbepaalde waarde de gangbare mening of de klassieke standaard. De bedoeling is immers om op basis van de steekproef de nulhypothese te kunnen verwerpen ten gunste van de **alternatieve hypothese**.

Voorbeeld (1):

$$H_0: \pi = \frac{1}{2}$$



Nul- en alternatieve hypothese

Een **alternatieve hypothese** H_1 is een bewering over dezelfde populatieparameter waarbij je iets *anders* zegt dan wat er in de nulhypothese staat.

In de meeste studies fixeert men zich hier niet op een welbepaalde waarde maar zegt de alternatieve hypothese dat de populatieparameter in **een bepaald gebied** ligt.

Voorbeeld (2):

$$H_0: \mu = 172$$

$$H_1: \mu < 172$$



Nul- en alternatieve hypothese

Andere alternatieven? Wat wens je te toetsen?

Voorbeeld (2):

$$H_1: \mu < 172 \text{ (1)}$$

$$H_1: \mu > 172 \text{ (2)}$$

$$H_1: \mu \neq 172 \text{ (3)}$$

(1)-(2) **éénzijdige** alternatieve hypothese

(3) **Tweezijdige** alternatieve hypothese



Oefening

In 2023 haalden de leerlingen van de derde graad ASO in Vlaanderen een **gemiddelde score van 57 op 100** op het **eindexamen wiskunde**. Jij twijfelt eraan of dat in 2024 ook het geval was. Dat wil je nu, met behulp van toetsen van hypothesen, onderzoeken.

- Wat is hier de **populatie** en welke **populatieparameter** (met juiste notatie) bestudeer je?
- Formuleer de **nulhypothese** (in woorden en met de juiste notatie).
- Formuleer de **alternatieve hypothese** (in woorden en met de juiste notatie). Is zij één- of tweezijdig?



UHASSELT

KNOWLEDGE IN ACTION

Extra oefening

Bij een klassieke behandeling bij de ziekte van Alzheimer is de gemiddelde tijd tot progressie 24 maanden. Men wil weten of de tijd tot progressie bij een nieuwe, experimentele behandeling, langer is. Daartoe meten onderzoekers de tijd tot progressie bij 100 patiënten die de experimentele behandeling krijgen.

Stel het bijbehorende toetsingsprobleem op.



Extra oefening

Een gynaecologe beweert dat meisjes bij de geboorte gemiddeld kleiner zijn dan de aangegeven 51 cm die als norm wordt gehanteerd in een klinisch handboek. Ze neemt een steekproef van 41 pasgeboren meisjes en noteert hun lengte.

Stel het bijbehorende toetsingsprobleem op.



Extra oefening

Het gemiddelde gewicht bij mannen tussen 55 en 70 jaar is 77 kg. Beschouw de bevolkingsgroep van mannen tussen 55 en 70 jaar, waarbij voor het eerst een hartprobleem is vastgesteld. Men wil weten of in deze groep het gemiddelde gewicht verschillend is van de algemene populatie van mannen tussen de 55 en 70 jaar.

Stel het bijbehorende toetsingsprobleem op.



Belangrijke opmerkingen

Opmerking 1:

- Merk op dat de hypothesen worden opgesteld op basis van het vraagstuk **zonder naar het resultaat van een steekproef te kijken**.
- In praktijk dienen de hypothesen steeds opgesteld te worden alvorens men de steekproef gaat trekken!

Opmerking 2:

Aangezien in de alternatieve hypothese staat wat we willen aantonen, zullen we bij het toetsen van de hypothesen nagaan wanneer de nulhypothese verworpen kan worden. In dat geval kan de alternatieve hypothese als correct beschouwd worden (zie *toetsen van hypothesen*).

Deel 3: Toetsen van hypothesen

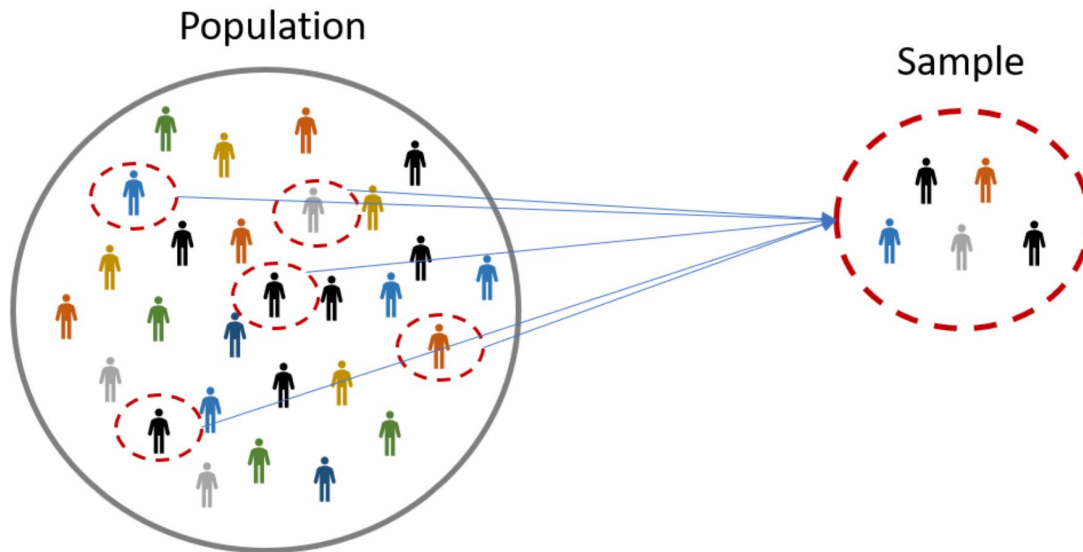


UHASSELT

KNOWLEDGE IN ACTION

Toetsen van hypothesen

- In de verklarende (of inferentiële) statistiek doe je, op basis van steekproefresultaten, een uitspraak over een populatieparameter.



Toetsen van hypothesen

- Bij **betrouwbaarheidsintervallen** gebruik je een steekproef om een betrouwbaarheidsinterval op te stellen voor een ongekende populatieparameter.
- Bij **toetsen van hypothesen** worden beweringen gemaakt over een populatieparameter. Nu gebruik je steekproefresultaten om die beweringen te beoordelen.



Algemene toetsingsprocedure

1. Formuleer het toetsingsprobleem door middel van de statistische nulhypothese H_0 en alternatieve hypothese H_1
2. Bereken de toetsingsgrootte voor het probleem
3. Bekijk of je H_0 kan verwerpen of niet op basis van de steekproef:
 - 3a. Bereken het kritisch punt (k.p.) van de verdeling van de toetsingsgrootte onder H_0 en vergelijk de waarde met het k.p.
 - 3b. Equivalent: bereken de p-waarde en vergelijk met het significantieniveau α
4. Formuleer een conclusie voor het probleem



Wat je (niet) bewijst



UHASSELT

KNOWLEDGE IN ACTION

Wat je (niet) bewijst

- Bij toetsen van hypothesen start je met de **nulhypothese** H_0 d.i. je gaat ervan uit dat de nulhypothese **waar** is
- Dat blijf je nu geloven, tenzij je in je steekproef resultaten vindt die de bewering van de nulhypothese zeer sterk **ongeloofwaardig** maken.
- In dat geval verwerp je H_0 ten gunste van de alternatieve hypothese



Wat je (niet) bewijst

- Laat ons terugkijken naar ons initieel voorbeeld van het computerprogramma (0-1 experiment). Indien jullie deelnemen aan de proef, dan moeten jullie mij eerst geloven dat $H_0: \pi = 0.5$.
- π wordt gedefinieerd als de kans op het genereren van een 0-waarde (herinner je dat **ik** dan win!)
- Echter, mogelijks zijn jullie op jullie hoede en geloven jullie dat de kans op winst voor mij (d.i. een 0 waarde) groter is, dus $H_1: \pi > 0.5$.
- Nu spelen we het spel **20 keer** (*steekproefgrootte*)



Wat je (niet) bewijst

- Volgende redenering wordt nu gevolgd:
 - Indien de **nulhypothese waar** is, is het dan uitzonderlijk dat jullie **hoogstens 9** keer winnen (of ik win minstens 11 keer)?
 - We definiëren Y als het **totaal aantal keer dat ik win** (op basis van hypothesen)
 - Kans om hoogstens 9 keer te winnen voor jullie is dus gelijk aan

$$P(Y \geq 11) = 0.4119$$

althans als de nulhypothese waar is.

Zo een uitkomst is dus helemaal niet uitzonderlijk onder de nulhypothese

Hoe worden deze kansen berekend (zie ook extra informatie)?



Wat je (niet) bewijst

- Volgende redenering wordt nu gevolgd:
 - Indien de **nulhypothese waar** is, is het dan uitzonderlijk dat jullie **hoogstens 5** keer winnen (of ik win meer dan 15 keer)?
 - We definiëren Y als het **totaal aantal keer dat ik win** (op basis van hypothesen)
 - Kans om hoogstens 5 keer te winnen voor jullie is dus gelijk aan

$$P(Y \geq 15) = 0.0207$$

althans als de nulhypothese waar is.

Hoogstens 5 keer winnen voor jullie is een uitzonderlijke gebeurtenis indien de nulhypothese waar is



Wat je (niet) bewijst

- Waar het hier om gaat, is dat je, **vanaf een bepaald ogenblik**, dingen begint te zien die je helemaal niet had verwacht als de nulhypothese waar zou zijn. En dan besluit je dat je de **nulhypothese verwierpt** en voor de alternatieve hypothese kiest.
- Bemerk dat je dit pas doet als je terechtkomt in “**extreme gebieden**” waarin je **slechts met 5% kans** valt als de **nulhypothese waar** is.
- In alle andere gevallen blijf je bij de nulhypothese (of verwierp je ze niet op basis van de steekproef).



Wat je (niet) bewijst

- Toetsen van hypothesen is een asymmetrische procedure waarbij je **heel lang aan de nulhypothese vasthoudt**. Pas bij extreme uitkomsten stap je over op de alternatieve hypothese.
- Je hebt dan zolang gewacht dat je mag zeggen dat er nu een statistisch bewijs is dat de alternatieve hypothese waar is, op het **gekozen significantieniveau** (5% criterium).



Wat je (niet) bewijst

- Over de **nulhypothese kan je zo'n uitspraak niet doen**. Het is niet omdat "ongeveer" 10 van de 20 cijfers 0 zijn dat je daarom te maken hebt met een eerlijk programma. Zo'n resultaat krijg je immers ook gemakkelijk indien de kans op 0 gelijk is 0.6.
- De **procedure die je gebruikt bij het toetsen van hypothesen levert geen statistisch bewijs dat de nulhypothese waar is**. Je start met een nulhypothese en je blijft er mee zitten zolang je ze niet kan verwerpen. Maar of ze juist is, dat vertellen de steekproefresultaten je niet.
- Toetsen van hypothesen gaat dus om **bewijsvoering tegen de nulhypothese** op basis van de steekproef



Soorten fouten



UHASSELT

KNOWLEDGE IN ACTION

Soorten fouten

- Indien we een toets van hypothese uitvoeren, kan het volgende gebeuren:
 - Wat de **populatie** betreft, gedraagt zij zich zoals beweerd wordt in de nulhypothese of niet
 - Wat **jouw beslissing** betreft, verwerp je de nulhypothese of niet



Type I en Type II fouten

- Bij het nemen van een beslissing: 2 soorten fouten

		Werkelijkheid	
		H_0 is waar	H_0 is niet waar
Statistisch besluit op basis van de steekproef	verwerp H_0	foute beslissing type I fout	correcte beslissing
	verwerp H_0 niet	correcte beslissing	foute beslissing type II fout

- Als $P(\text{Type I fout})$ daalt, dan stijgt $P(\text{Type II fout})$ en vice versa



Soorten fouten

- Ofwel is de nulhypothese **waar**, d.i. bijv. de kans op het genereren van een 0-waarde is gelijk aan 0.5
- Een **Type I fout** maak je als je de **nulhypothese verwerpt terwijl ze in feite waar is**
- Ofwel is de nulhypothese **niet waar**, d.i. bijv. de kans op het genereren van een 0-waarde is niet gelijk aan 0.5
- Een **Type II fout** maak je als je de **nulhypothese niet verwerpt terwijl ze in feite niet waar is**



Significantieniveau

- Bij het toetsen van hypothesen zorgt men ervoor dat de kans op een Type I fout te maken **klein** is.
- Gewoonlijk neemt men 5% zodat

$$P(\text{Type I fout}) = 0.05 = 5\%$$

Deze kans wordt het **significantieniveau** genoemd (zie ook 5% criterium hierboven), in het algemeen genoteerd door α



Oefening

Een Type I fout maken kan zeer ernstige gevolgen hebben. Waarom werkt men eigenlijk met een procedure waar de kans op een fout maken niet gelijk is aan 0? Wat zou dat betekenen in analogie met rechtspraak?



Het verhaal van signaal en ruis

- Om een toetsingsprobleem op te lossen en een toets van hypothesen uit te voeren, steunen we dus op een gekozen **significantieniveau** α om een beslissing te nemen.
- Het opstellen van een beslisregel voor het probleem is een kwestie van het bestuderen van de verhouding tussen **signaal en ruis**. We komen hier straks nog op terug.



Deel 4: Toetsen voor proporties



UHASSELT

KNOWLEDGE IN ACTION

Een toets opstellen

- Beschouw **een 0-1 populatie**, met een uitkomst die categorisch is en 2 niveau's heeft (success en mislukking), met de kans op succes genoteerd door π (*populatieproportie*)
- Als je uit die populatie een steekproef trekt dan heb je een aantal mislukkingen (0'en) en een aantal successen (1'en)
- De **steekproefproportie** $\hat{p}$ geeft de succesproportie in de steekproef als schatting voor de populatieproportie π
- Voor een andere steekproef uit dezelfde populatie, krijg je een andere steekproefproportie (*sampling variatie*)



Een toets opstellen

- Als de steekproef groot genoeg is, zodat voldaan is aan

$$\begin{aligned}n\pi &\geq 15 \\ n(1 - \pi) &\geq 15\end{aligned}$$

Dan kan je de kansverdeling van de steekproefproportie $\hat{P}$ benaderen door middel van een normale verdeling, onder de nulhypothese (d.i. als H_0 waar is)

- Meer bepaald, we hebben dan dat, onder H_0 ,

$$\frac{\hat{P} - E(\hat{P})}{se(\hat{P})} = \frac{\hat{P} - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} \sim N(0,1)$$



Een toets opstellen

- Meer bepaald, we hebben dan dat, onder H_0 ,

$$\frac{\hat{P} - E(\hat{P})}{se(\hat{P})} = \frac{\hat{P} - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} \sim N(0,1)$$

- π_0 is de **populatieproportie onder de nulhypothese**
- **Beslisregel:** indien de waarde van de toetsingsgrootte hierboven op basis van de steekproef(waarden) te *extreem* is, dan verwerpen we de nulhypothese en is er bewijs tegen de nulhypothese



Een toets opstellen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar).
- **Toetsingsprobleem:**

$$H_0: \pi = \frac{1}{2}$$

$$H_1: \pi > \frac{1}{2}$$



Een toets opstellen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar).
- **Toetsingsprobleem:**

$$H_0: \pi = \frac{1}{2}$$

$$H_1: \pi > \frac{1}{2}$$

Je start nu met de veronderstelling dat de nulhypothese waar is



Een toets opstellen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- **Toetsingsgrootte** (onder H_0 want $n0.5 \geq 15, n(1 - 0.5) \geq 15$):

$$\frac{\hat{P} - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} = \frac{\hat{P} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} \sim N(0,1)$$



Een toets opstellen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- Je hebt **een kansmodel** dat observeerbaar is want er is geen enkele onbekende parameter meer in de toetsingsgrootte.
- Het enige dat nog moet gebeuren is een *extreem gebied* bepalen, waarvoor een waarde voor de toetsingsgrootte in dat gebied bewijs geeft tegen de nulhypothese.

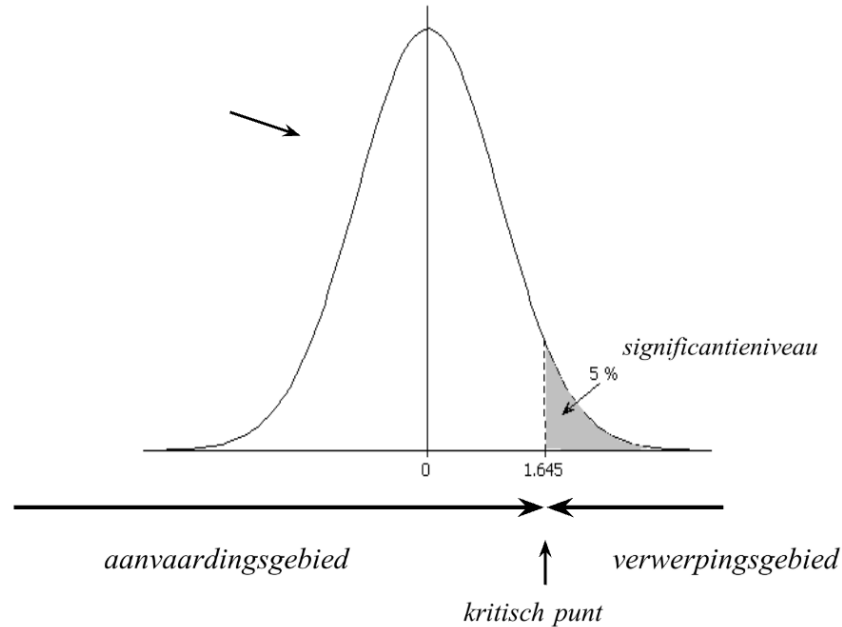


Een toets opstellen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- Het enige dat nog moet gebeuren is een *extreem gebied* bepalen, waarvoor een waarde voor de toetsingsgrootte in dat gebied bewijs geeft tegen de nulhypothese.
- Linker-, rechter- of twee staarten? **Kijk naar de alternatieve hypothese**



Beslisregel – het *extreme gebied* bepalen



Beslisregel – het *extreme gebied* bepalen

- Het kritisch punt is een ééNZijdig kritisch punt met betrekking tot de rechterstaart, waarvoor

$$P(Z \geq z) = 0.05$$

Indien we dit oplossen naar z , dan vinden we $z = 1.645$ als kritisch punt

Merk op: hoofdletter Z is de gangbare notatie voor een kansmodel of stochastische veranderlijke die een standaard normale verdeling volgt (d.i. onze toetsingsgrootheid).



Beslisregel – het *extreme gebied* bepalen

Twee gebieden geïdentificeerd:

- **Verwerpingsgebied:** $[1.645, \infty[$
- **Aanvaardingsgebied (of beter niet-verwerpingsgebied):**
 $] - \infty, 1.645[$



Beslisregel – waarde van de toetsingsgrootheid

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.* **Op basis van de steekproef zijn er 45 vrouwen met een kind in 2022 jonger dan 30 jaar.**
- **De waarde van de toetsingsgrootheid is**

$$\frac{\hat{p} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} = \frac{0.643 - 0.5}{0.06} \approx 2.38$$

Een toets uitvoeren

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- De waarde van de toetsingsgrootte ligt in het verwerpingsgebied (voor $\alpha = 0.05$). Als je in het verwerpingsgebied terechtkomt dan zeg je dat de geziene afwijking (ten opzichte van de bewering van de nulhypothese) geen “toevallig verschil” meer is, te wijten aan schommelingen van steekproeven.



Een toets uitvoeren

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- **Conclusie:** Er is een significant grotere kans op een jongere moeder dan op een oudere moeder op significantieniveau 5%.



De p-waarde

De **p-waarde** is de kans dat het kansmodel, als de nulhypothese waar is, even extreme of nog extremere waarden aanneemt, vergeleken met de waarde die jij hebt gevonden op basis van je steekproef.

“Extremiteit” betekent hier in de richting van de alternatieve hypothese.

Indien de **p-waarde kleiner** is dan het **significantieniveau** α dan verwerpen we de nulhypothese, anders niet.



De p-waarde

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- De **p-waarde** voor deze oefening is gelijk aan:

$$P(Z \geq z) = P\left(\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} \geq \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}}\right) = P\left(\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} \geq 2.38\right)$$



De p-waarde

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- De **p-waarde** voor deze oefening is gelijk aan:

$$P(Z \geq z) = P\left(\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} \geq 2.38\right) = 0.0087$$



De p-waarde

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. Jij bent er echter van overtuigd dat er meer jonge moeders zijn (< 30 jaar). *Testen met een steekproef van grootte $n = 70$.*
- De **p-waarde** voor deze oefening is gelijk aan:

$$P(Z \geq z) = P\left(\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} \geq 2.38\right) = 0.0087$$



De p-waarde

- De **p-waarde** voor deze oefening is gelijk aan:

$$P(Z \geq z) = P\left(\frac{\hat{P} - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{70}}} \geq 2.38\right) = 0.0087$$

- Conclusie:** $0.0087 < 0.05$, dus we verwerpen H_0 op significantieniveau 5%
- Conclusies** op basis van het kritisch punt en p-waarde moeten steeds hetzelfde zijn!



Tweezijdig toetsen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. **Jij trekt dit echter eenvoudigweg in twijfel.** *Testen met een steekproef van grootte $n = 70$.*
- **Toetsingsprobleem:**

$$H_0: \pi = \frac{1}{2}$$

$$H_1: \pi \neq \frac{1}{2}$$

Je start nu met de veronderstelling dat de nulhypothese waar is



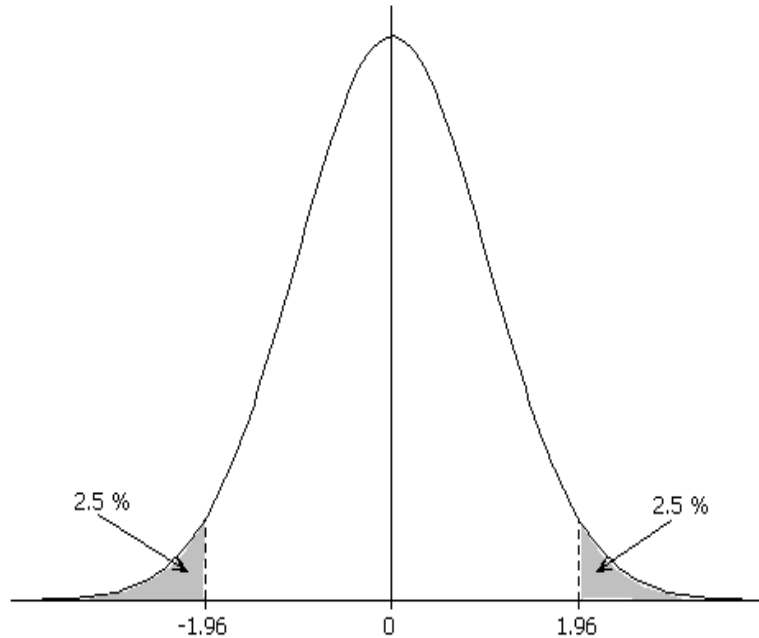
Tweezijdig toetsen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. **Jij trekt dit echter eenvoudigweg in twijfel.** *Testen met een steekproef van grootte $n = 70$.*
- **Verwerpingsgebied** bestaat nu uit twee staarten (extreem links en extreem rechts). Het significantieniveau α moet dan opgedeeld worden in twee stukken van $\alpha/2$



Tweezijdig toetsen

Waar zijn de verwerpings- en
aanvaardingsgebieden?



Tweezijdig toetsen

- **Voorbeeld:** Je vriendin beweert dat bij bevallingen van een kind er ongeveer evenveel moeders jonger zijn dan 30 jaar als moeders van minstens 30 jaar zijn. **Jij trekt dit echter eenvoudigweg in twijfel.** *Testen met een steekproef van grootte $n = 70$.*
- **Stel dat je de keuze voor een tweezijdige toets gemaakt hebt voor het verzamelen van de gegevens (trekken van de steekproef), wat zou dan je conclusie geweest zijn op basis van dezelfde steekproefwaarden als voorheen?**



Extra oefening

In de Verenigde Staten werd bij een goed getrokken steekproef van 116 astrologen de studie uitgevoerd om te bepalen of een testprocedure werkt. Er waren 40 astrologen die een goed antwoord gaven. Bevestigt dit dat astrologie (bij Amerikaanse astrologen) werkt?

Toets op het 5% significantieniveau.



Deel 5: Toetsen voor gemiddelden



UHASSELT

KNOWLEDGE IN ACTION

Toetsen voor gemiddelden

- Wanneer je te maken hebt met populatiekarakteristieken die je als continu kan behandelen (zoals leeftijd, gewicht, lengte, ...) dan kan je een hypothese toetsen over het **populatiegemiddelde μ**
- We maken hierbij gebruik van wat je weet over **toetsen van hypothesen voor proporties**



Een toets voor μ

- Het populatiegemiddelde μ kan geschat worden aan de hand van het steekproefgemiddelde $\bar{X}$ (kansmodel)
- Je weet dat $E(\bar{X}) = \mu$ aangezien $\bar{X}$ een overtekende schatter is voor μ
- Bovendien geldt dat $se(\bar{X}) = \sigma/\sqrt{n}$ is de standaardfout voor $\bar{X}$
- Volgens het principe van signaal versus ruis, vinden we nu de gestandaardiseerde toetsingsgrootte (kansmodel):

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Een toets voor μ

- Volgens het principe van signaal versus ruis, vinden we nu de gestandaardiseerde toetsingsgrootte (kansmodel):

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

- In geen enkele realistische studie is σ gekend dus moeten we deze grootte vervangen door de **steekproefstandaardafwijking**:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$



Een toets voor μ

- In geen enkele realistische studie is σ gekend dus moeten we deze grootte vervangen door de **steekproefstandaardafwijking** S :

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

met

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$



Een toets voor μ

- Onder de nulhypothese geldt nu dat:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

met

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

op voorwaarde dat de populatiekarakteristiek normaal verdeeld is.



Een toets voor μ

- **Voorbeeld:** Je vriend beweert dat het gemiddelde geboortegewicht van meisjes 3kg is. Je hebt geen idee ofdat waar is, maar je gelooft niet dat het 3kg is. *Je komt overeen om dit na te gaan op basis van geboortes van Vlaamse meisjes in 2000, op basis van een steekproef van grootte $n = 70$. **Geboortegewicht werd uitgedrukt in gram.***
- **Toetsingsprobleem:**

$$H_0: \mu = 3000$$

$$H_0: \mu \neq 3000$$



Een toets voor μ

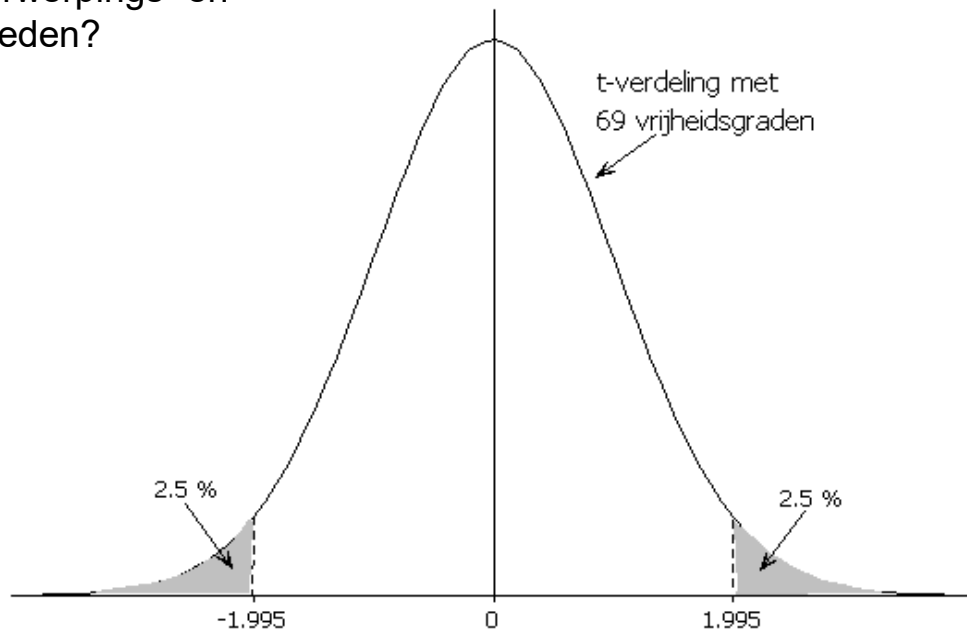
- **Voorbeeld:** Je vriend beweert dat het gemiddelde geboortegewicht van meisjes 3kg is. Je hebt geen idee ofdat waar is, maar je gelooft niet dat het 3kg is. *Je komt overeen om dit na te gaan op basis van geboortes van Vlaamse meisjes in 2000, op basis van een steekproef van grootte $n = 70$. **Geboortegewicht werd uitgedrukt in gram.***
- **Toetsingsgrootheid** (met verdeling onder H_0):

$$T = \frac{\bar{X} - 3000}{S/\sqrt{n}} \sim t(69)$$



Een toets voor μ

Waar liggen de verwerpings- en
aanvaardingsgebieden?



Een toets voor μ

- Veronderstel dat $\bar{x} = 3204.3$ en $s = 564.53$. Wat is nu de waarde van de toetsingsgrootte?

$$t = \frac{3204.3 - 3000}{564.53/\sqrt{70}} \approx 3$$

- Wat is de **beslissing** op basis van het geïdentificeerde verwerpingsgebied op een 5% significantieniveau?



Een toets voor μ

- Veronderstel dat $\bar{x} = 3204.3$ en $s = 564.53$. Wat is nu de waarde van de toetsingsgrootte?

$$t = \frac{3204.3 - 3000}{564.53/\sqrt{70}} \approx 3$$

- Als je in het verwerpingsgebied terechtkomt dan zeg je dat het gevonden verschil niet meer te wijten is aan toevallige schommelingen van steekproeven. **Je spreekt dan over een significant verschil en je stapt over van de nulhypothese op de alternatieve hypothese.**



Opmerking

- Indien n voldoende groot is (d.i. $n \geq 30$), dan geldt, onder de nulhypothese, dat:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$$

met

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

- In dat geval is geen veronderstelling van normaliteit nodig
- Bovendien, als $n \rightarrow \infty$, dan zal $t(n) \rightarrow N(0,1)$



Toetsen voor twee properties of gemiddelden



UHASSELT

KNOWLEDGE IN ACTION

Toetsen voor twee properties of gemiddelden

- Indien we twee proporties of gemiddelden willen vergelijken, moeten we andere toetsingsgrootheden beschouwen.
- Een **voorbeeld** voor continue gegevens in twee onafhankelijke groepen (voor een **tweezijdig toetsingsprobleem**)

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

OF na herschrijven van de hypothesen:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$



Toetsen voor twee eigenschappen of gemiddelden

- Indien we twee eigenschappen of gemiddelden willen vergelijken, moeten we andere toetsingsgrootheden beschouwen.
- Een **voorbeeld** voor continue gegevens in twee onafhankelijke groepen is de **twee steekproeven t-test** met

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{S_P^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2)$$

als de nulhypothese waar is. $(\mu_1 - \mu_2)_0$ is het **verschil in gemiddelden** verwacht **onder de nulhypothese**.



Toetsen voor twee properties of gemiddelden

Een **voorbeeld** voor continue gegevens in twee onafhankelijke groepen is de **twee steekproeven t-test** met

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{S_P^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2)$$

als de nulhypothese waar is. $(\mu_1 - \mu_2)_0$ is het **verschil in gemiddelden** verwacht **onder de nulhypothese**.

Voorwaarden: $n_1 \geq 30$ en $n_2 \geq 30$ **of** populatiekarakteristiek is normaal verdeeld in beide groepen! Bovendien veronderstellen we hier dat de **varianties** in beide groepen gelijk zijn.



Een voorbeeld

Men vermoedt dat een nieuwe behandeling tegen de ziekte van Alzheimer de **tijd tot progressie** verlengt. Om dit na te gaan, worden **122 Alzheimer patiënten** in 2 groepen gedeeld: groep 1 ($n_1 = 59$) krijgt de klassieke behandeling, terwijl groep 2 ($n_2 = 63$) de nieuwe behandeling krijgt.

Toetsingsprobleem:

$$H_0: \mu_{klassieke} = \mu_{nieuw}$$

$$H_0: \mu_{klassieke} < \mu_{nieuw}$$



Een voorbeeld

Men vermoedt dat een nieuwe behandeling tegen de ziekte van Alzheimer de **tijd tot progressie** verlengt. Om dit na te gaan, worden **122 Alzheimer patiënten** in 2 groepen gedeeld: groep 1 ($n_1 = 59$) krijgt de klassieke behandeling, terwijl groep 2 ($n_2 = 63$) de nieuwe behandeling krijgt.

Waarde van de toetsingsgrootte is gelijk aan 4.606.

Zijn de voorwaarden voldaan?

Een voorbeeld

Men vermoedt dat een nieuwe behandeling tegen de ziekte van Alzheimer de **tijd tot progressie** verlengt. Om dit na te gaan, worden **122 Alzheimer patiënten** in 2 groepen gedeeld: groep 1 ($n_1 = 59$) krijgt de klassieke behandeling, terwijl groep 2 ($n_2 = 63$) de nieuwe behandeling krijgt.

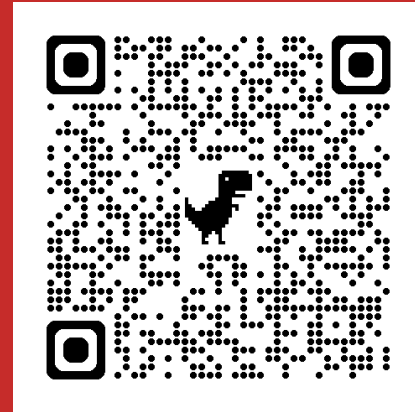
Waarde van de toetsingsgrootte is gelijk aan 4.606.

De bijbehorende **p-waarde** is gelijk aan < 0.0001 . **Wat is dan de conclusie?**

Wat vonden jullie van de sessie?

Laat het ons weten!

dsi@uhasselt.be



<https://forms.gle/kjuEjmJBQjVLpev16>



UHASSELT

KNOWLEDGE IN ACTION

Dankwoord

Dit materiaal kwam tot stand in samenwerking met de DSI werkgroep Training voor Secundair Onderwijs. De leden van de werkgroep danken Prof. em. dr. Herman Callaert, Hans Bekaert, Cecile Goethals, Lies Provoost en Marc Vancaudenberg voor het eerder ontwikkeld materiaal waarop deze slides werden gebaseerd.



UHASSELT

KNOWLEDGE IN ACTION

Extra informatie



UHASSELT

KNOWLEDGE IN ACTION

Bernoulli verdeling

- Onderstel dat de kans op succes gekend is en gelijk is aan p
- Dan is ook de kans op mislukking gelijk aan $1-p = q$

$$P(S) = p$$

$$P(M) = q$$



Binomiaal experiment

- Beschouw een experiment dat zodanig is dat het een zeker aantal keer na elkaar kan herhaald worden onder dezelfde omstandigheden
- Onderstel dat elke uitvoering van het experiment aanleiding geeft tot twee mogelijke uitslagen die we algemeen aanduiden met succes (S) en mislukking (M) (**Bernoulli experiment**)



Binomiale verdeling

De kans op juist x successen bij n onafhankelijke herhalingen van hetzelfde Bernoulli experiment (d.i. dezelfde kans op succes):

$$P(X = x) = \binom{n}{x} \times p^x \times q^{n-x}$$

$$\binom{n}{x} = \frac{n!}{x! \times (n-x)!}$$



Binomiale verdeling

Een discrete s.v. X heeft een binomiale verdeling met parameters n en p indien

$$P(X = x) = \binom{n}{x} \times p^x \times q^{n-x}$$

Notatie:

$$X \sim B(n, p)$$



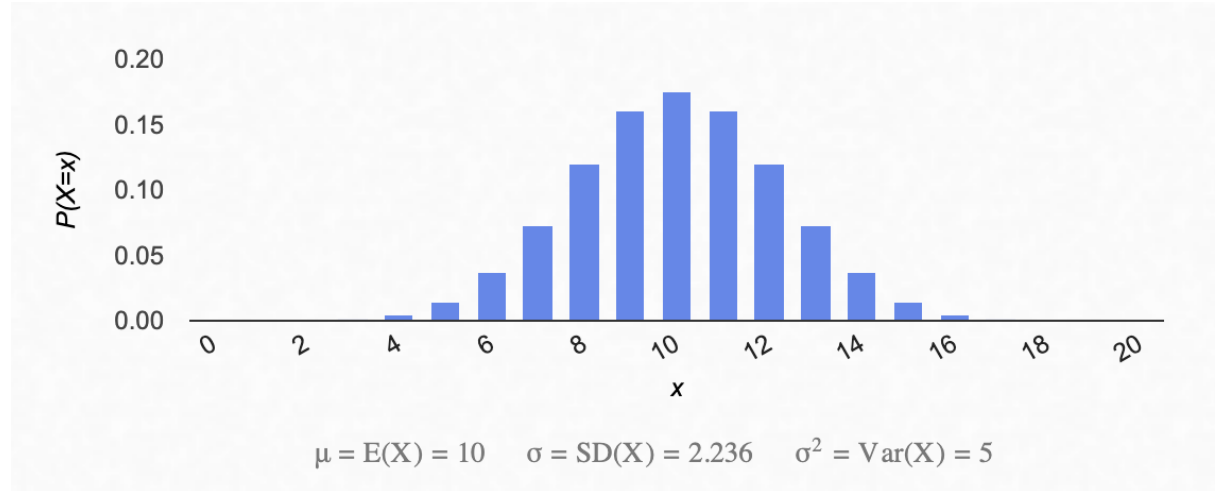
Voorbeeld 1: Het gooien van een muntstuk

- Definieer de stochastische veranderlijke
 X = het totaal aantal keren 'kop' gegooid
- $X \sim B(n,p)$ met $p = \textit{kans op kop gooien} = P(K) = 0.5$



Voorbeeld 1: Het gooien van een muntstuk

- Stel dat we het **Bernoulli experiment** (het computerprogramma) 20 maal uitvoeren (met gelijke kans op 0 gooien van 0.5), dan krijgen we volgende **discrete kansverdeling** (met gemiddelde ?, variantie ?):



<https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html>