Statistiek II

Sessie 1

Verzamelde vragen en feedback

Deel 1

VPPK

Universiteit Gent

2017-2018

Feedback Oefensessie 1

1 Staafdiagram

1. Wat is de steekproefgrootte?
   Op de horizontale as vinden we de respectievelijke waarden van X. Op de verticale as vinden we de frequentie van elke waarde aanwezig in de steekproef. We hebben dus een steekproef van variabele X met volgende waarden en respectievelijke frequentie:

   - 1 : 1
   - 2 : 2
   - 3 : 4
   - 4 : 2
   - 5 : 1
   - 6 : 0
   - 7 : 0
   - 8 : 0
   - 9 : 1

   Als we alle frequenties optellen komen we aan de steekproefgrootte:

   $$1 + 2 + 4 + 2 + 1 + 0 + 0 + 0 + 1 = 11$$

2. We maken gebruik van een staafdiagram, dus de variabele is discreet. Meer bepaald is deze variabele van absoluut meetniveau.

3. We zien een outlier aan de rechterkant. Dit zal ervoor zorgen dat het gemiddelde groter wordt, want het gemiddelde is gevoelig aan outliers. De mediaan is niet gevoelig aan outliers, dus deze zal kleiner zijn dan het gemiddelde.

   - De mediaan is de middelste observatie als we alles rangschikken van klein naar groot. In dit geval is dit de zesde observatie. Dus $med_x = 3$
   - De modus is de meest voorkomende waarde. Hier zien we dat de waarde 3, 4 keer voorkomt. Dus $mod_x = 3$
   - Voor het gemiddelde te berekenen maken we de som van elke observatie in de steekproef en delen deze door de steekproefgrootte:

     $$\frac{1 + 2 + 2 + 3 + 3 + 3 + 3 + 4 + 4 + 5 + 9}{11} = 3.55$$

     Let op, dit is enkel toegestaan indien onze variabele van interval, ratio of absoluut niveau is. Wanneer we van ordinaal of nominaal spreken is een gemiddelde onzinnig, en mogen we dit dus niet berekenen.

2 Staafdiagram 2

Bij deze oefening is het de bedoeling dat we niets gaan uitrekenen, maar een inzicht krijgen in wat de variantie, IQR en d is.

- Variantie: deze is gevoelig aan alle observaties. Om even technisch te gaan, de gekwadrateerde afstand tussen elke observatie en het steekproefgemiddelde, bepalen de variantie. Dus hoe extremer de waarden, hoe groter de variantie zal worden. Dus met deze info in het achterhoofd, kunnen we zien dat de linkse staafdiagram extremere waarden heeft, maar de middleste waarden hetzelfde zijn als de rechtse. Dus we gaan een hogere variantie observeren bij de linkse staafdiagram.

- IQR: Dit is het verschil tussen het 3de kwartiel en het 1ste kwartiel. Dit wil zeggen dat het verschil van de waarde van de observatie (of indien tussen 2 observaties, een iets complexere berekening) die het bovenste kwart van de rest scheidt en de waarde van de observatie die het onderste kwart van de rest scheidt. Als we gaan zien, we hebben hier 13 observaties, dus als we zouden zien, het 3de kwartiel is $13 \times 0.75 = 9.75$ - m.a.w. tussen de 9de en de 10de observatie. Het eerste kwartiel kunnen we gelijkaardig benaderen, en zal zich tussen de 3de en de 4de observatie bevinden. Als we nu kijken naar beide staafdiagrammen, dan zien we dat de 3de, 4de, 9de en 10de observatie voor beide hetzelfde zijn. Dus we krijgen een gelijke IQR.
- d: als we de definitie bekijken op pagina 36, dan zien we dat deze maat enkel afhankelijk is van de frequentie van de modus, de steekproefgrootte en het aantal verschillende waarden. We zien dat deze 3 zaken bij beide staafdiagrammen hetzelfde zijn. Dus deze zullen ook een gelijke d geven.

3 Bloedsuikerspiegel bij vrouwen

3.1 Puntschatting

(a) De puntschatting voor de verwachting in de populatie is $\hat{\mu} = 8.2$

(b) Ja, cf. p. 76

(c) Verwachting van $X = \mathbb{E}(X) = \mu$
Gemiddelde van $X = \bar{x}$
Schatter van verwachting van $X = \bar{X}$
Schatting van verwachting van $X = \hat{\mu} = \bar{x}$
Concreet: wanneer we over een schatter spreken dan is dit nog niet gerealiseerd. Dit is geen berekening, dit is gewoon een benaming. Daarentegen een schatting heeft een waarde, dit is een gerealiseerde berekening. De schatter schrijven we met een grote letter, omdat deze nog niet gerealiseerd is, de schatting met een kleine letter omdat deze wel gerealiseerd is. In dit geval is de schatting van de verwachting hetzelfde als het gemiddelde.

(d) Grotere steekproef nemen. Want hoe groter onze steekproef hoe meer extreme waarden elkaar gaan opheffen. We gaan meer data krijgen rond de populatieparameter.

(e) Ja, ongeacht de steekproefgrootte is mijn schatter zuiver

3.2 Intervalschatting

(a) Betrouwbaarheidsinterval: $\bar{x} \pm t_{n-1;\alpha/2} \frac{s_X}{\sqrt{n}}$
De eerste stap is dat we sn_X gaan omzetten in s_X omdat de formule voor beide verschillend is. De eerste is geen zuivere schatting, de tweede wel. We weten dat de formule voor sn_X^2:

$$sn_X^2 = \frac{SSR}{n}$$

Dit kunnen we dus herschrijven als:

$$SSR = sn_X^2 \times n$$

Gelijkaardig krijgen we door het hervormen van de formule van s_X^2 volgende vergelijking:

$$SSR = s_X^2 \times (n-1)$$

Met andere woorden, we hebben twee verschillende benaderingen om SSR te gaan berekenen. Als we deze twee nu met elkaar vergelijken, en de $n-1$ van de s_X^2 naar de andere kant verplaatsen krijgen we:

$$s_X^2 = sn_X^2 \frac{n}{n-1}$$

Om dan s_X te bekomen nemen we gewoon de vierkantswortel:

$$\begin{aligned} s_X &= sn_X\sqrt{\frac{n}{n-1}} \\ &= 2.170976 \end{aligned}$$

In de fomrule van ons betrouwbaarheidsinterval krijgen we ook een t-waarde met $n-1$ vrijheidsgraden en een kans van $\mathcal{P} = \alpha/2$ die zich aan de rechterkant van de waarde bevindt. M.a.w. we hebben $\alpha/2$ kans dat we die t-waarde of groter observeren. We zitten dus in de rechterstaart van onze verdeling. In de cursus wordt dit uitgedrukt als:

$$\mathcal{P}(T > t_{n-1;\alpha/2}) = \alpha/2$$

Nu, in de tips staan deze zaken anders omschreven:

$$t_{n-1;\alpha/2} = T$$

Deze twee betekenen hetzelfde, en hier hebben we dan ook dezelfde symbolen gebruikt om de link duidelijk te maken.

Nu voor een 98% betrouwbaarheidsinterval, welke tweezijdig wordt opgesteld, moeten we dus de waarde vinden waar dan in de rechterstaart nog 1% boven zit. Laten we de gekende gegevens eens invullen in de formule:

$$t_{51;0.01} = T$$

Dus we gaan op zoek naar iets dat daar op lijkt, enkel de waarde T kennen we niet. Wat vinden we in de tips?

$$t_{51;0.01} = 2.40$$

Dus als we al deze zaken nu invullen in de formule voor het betrouwbaarheidsinterval krijgen we:

$$8.2 \pm 2.40\frac{2.170976}{\sqrt{52}} = [7.48; 8.92]$$

(b) Neen, om s_X^2 te mogen berekenen moeten we ofwel een normaalverdeelde variabele hebben, ofwel een steekproefgrootte van minstens 30, of beide. Als geen van de twee voldaan is, dan mogen we niet verder. We weten niets over de verdeling, dus gaan we ervan uit dat het niet normaalverdeeld is. Onze steekproefgrootte is te klein. Dus voor hier verder te gaan zouden we technieken moeten toepassen die niet in de cursus gezien worden.

(c) Grotere steekproef nemen. Wanneer we een grotere steekproef gaan nemen, dan gaan we een kleinere variantie krijgen. Dit omdat de formule van de schatter van de variantie deelt door $n-1$. Hoe kleiner deze waarde, hoe kleiner dan ook de waarde die we gaan optellen bij ons gemiddelde om het betrouwbaarheidsinterval te berekenen. Dus bij een grotere steekproef gaan we een smaller betrouwbaarheidsinterval krijgen, zonder dat de betrouwbaarheid wijzigt.

(d) Smaller: $[7.59; 8.81]$. Hoe kunnen we dit intuïtief vatten? Stel dat je de getallen van 1 tot en met 100 naast elkaar opschrijft. Bij 98% laat je de 2 buitenste vallen. Maar om 95% over te houden laat je er meer vallen. Hoe lager onze betrouwbaarheid, hoe meer extreme waarden we buiten het betrouwbaarheidsinterval laten vallen, en dus hoe smaller ons interval gaat worden.

3.3 Toetsing

(a) $H_0 : \mu_X = 7.5$ versus $H_a : \mu_X \neq 7.5$

(b) Aangezien 7.5 in het interval $[7.48; 8.92]$ ligt aanvaarden we de nulhypothese dat de populatieverwachting 7.5 is. We hebben hier een interval opgemaakt symmetrisch rond het gemiddelde. Dit gemiddelde wijkt af van de verwachting. Nu hebben we een foutmarge ingesteld op ons gemiddelde, met name het 98% betrouwbaarheidsinterval. Wat wil dit nu zeggen? Als we oneindig veel steekproeven gaan nemen van dezelfde grootte, en we berekenen van elk het 98% betrouwbaarheidsinterval, zal in 98% van de gevallen het betrouwbaarheidsinterval de verwachting bevatten. Dit wil dus zeggen dat de waarde 8.2 niet significant verschilt van de waarde 7.5 op 98% betrouwbaarheid.

(c) Wat is een p-waarde? De p-waarde is de kans dat we bij een nieuwe steekproef een waarde minstens even extreem gaan observeren als in de huidige steekproef, gegeven dat de nulhypothese waar is. Dus als de nulhypothese waar is, en de verwachting is inderdaad 7.5, wat is dan de kans dat we minstens even extreem als 8.2 gaan observeren? We hebben hier een tweezijdige toets. Dus dit betekent dat we gaan zien wat de kans is dat we minstens $8.2 - 7.5 = 0.7$ waarden hoger dan 7.5 gaan observeren, maar ook dat we minstens 0.7 waarden lager dan 7.5 gaan observeren. Deze kans kunnen we noteren als:

$$\mathcal{P}\left(\bar{X} \leq 6.8 | \mu_X = 7.5\right) + \mathcal{P}\left(\bar{X} \geq 8.2 | \mu_X = 7.50\right)$$

Omdat we de kansen willen weten, conditioneel dat de nulhypothese waar is, hebben we een symmetrische verdeling rond 7.5. Dus de kans dat we kleiner gaan zitten dan 6.8 en de kans dat we groter gaan zitten dan 8.2 is dezelfde. Dit zorgt ervoor dat we maar 1 kans moeten berekenen, en deze moeten verdubbelen.

$$\mathcal{P}\left(\bar{X} \geq 8.2 | \mu_X = 7.5\right) \times 2$$

Maar omdat het onmogelijk zou zijn om voor elke bestaande waarde met elke bestaande standaardafwijking die kansen te gaan opschrijven en registreren gaan we onze waarden transformeren. Deze getransformeerde waarde heet de toetsingsgrootheid G. Voor de t-verdeling is dit:

$$G = T = \frac{\bar{x} - \mu_X}{s_X/\sqrt{n}} = \frac{8.2 - 7.5}{2.170976/\sqrt{52}} = 2.33$$

Dee transformatie zorgt ervoor dat de verwachting (7.5) equivalent is met 0 in de t-verdeling. Nu kunnen we onze p-waarde gaan herschrijven als:

(d)

$$\begin{aligned}\mathcal{P}\left(\bar{X} \geq 8.2 | \mu_X = 7.5\right) \times 2 &= \mathcal{P}\left(\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \geq \frac{8.2 - 7.5}{2.170976/\sqrt{52}}\right) \times 2 \\ &= \mathcal{P}\left(\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \geq 2.33\right) \times 2\end{aligned}$$

Cocnreet zoeken we dus de kans dat we een t-waarde bekomen van minstens 2.33. Dit vinden we in de tips. Daarstraks wisten we de vrijheidsgraden en de kans, en zochten we de t-waarde. Nu kennen we een t-waarde, maar geen kans:

$$t_{51;P} = 2.33$$

Nu zoeken we hetgeen dat hierop lijkt. We vinden dan voor de P de waarde 0.012. Maar we moeten vermenigvuldigen met 2, dus krijgen we 0.024. Onze p-waarde is dus 0.024. Ook hierbij gaan we de nulhypothese aanvaarden. Omdat we werken op betrouwbaarheid 98% gaan we zien of de p-waarde groter of kleiner is dan $\alpha = 1 - 0.98 = 0.02$. Als de p-waarde groter is aanvaarden we de nulhypothese, indien kleiner dan verwerpen we de nulhypothese.

(e) Dit kunnen we niet zeker weten zonder berekening. We weten dat het betrouwbaarheidsinterval smaller gaat worden met een toenemende steekproef. Dus er zal een moment komen dat ons gemiddelde uit het betrouwbaarheidsinterval valt, en we dus de nulhypothese gaan verwerpen. Maar we weten niet wanneer dit zal gebeuren.

Om nu de $t_{n-1;p} = T$ te interpreteren zien we dat de $n-1$ de vrijheidsgraden aanduiden. De p is de kans dat we een waarde minstens even groot als T gaan observeren. Dus we zitten weer in de rechterstaart. De T is onze T-waarde. Omdat onze verdeling symmetrisch is kunnen we dit ook gebruiken voor negatieve waarden van T. De kans p kunnen we dus ook lezen als de kans dat we een waarde van $-T$ of kleiner gaan observeren. Dan zitten we in de linkerstaart.
Bij het betrouwbaarheidsinterval kennen we de kans, dus vervangen we p door α of $\alpha/2$, afhankelijk of we een- of tweezijdig werken. Bij een p-waarde kennen we de T, en dan gaan we de bijhorende p zoeken.

4 Studietijd

1. $H_0 : \mu_X = 42$ versus $H_a : \mu_X < 42$
   Let op, we veronderstellen dat een student 3 uur per dag moet studeren, en er wordt gemeten over 14 dagen. Dus is onze nulhypothese $3 \times 14 = 42$ uren. De onderzoeker verwacht dat de studenten te weinig studeren. Dit wil zeggen dat we enkel interesse hebben in waarden lager dan 42. Al wat hoger ligt is in strijd met de verwachtingen van de onderzoeker, en dus evidentie tegen de alternatieve hypothese. Daarom gaan we eenzijdig testen.

2. Toetsingsgrootheid:

$$\begin{aligned} G &= \frac{38.4 - 42}{16.20383/\sqrt{40}} \\ &= -1.405 \end{aligned}$$

   p-waarde:

$$\mathcal{P}\left(\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \leq -1.405\right) = \mathcal{P}\left(\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \geq 1.405\right) = 0.08$$

   Aangezien onze p-waarde groter is dan $\alpha = 0.05$ aanvaarden we de nulhypothese

3. Zonder berekeningen is dit niet mogelijk om te zeggen. Zelfde redenering als de vorige oefening.

5 Dagelijkse energiebehoefte

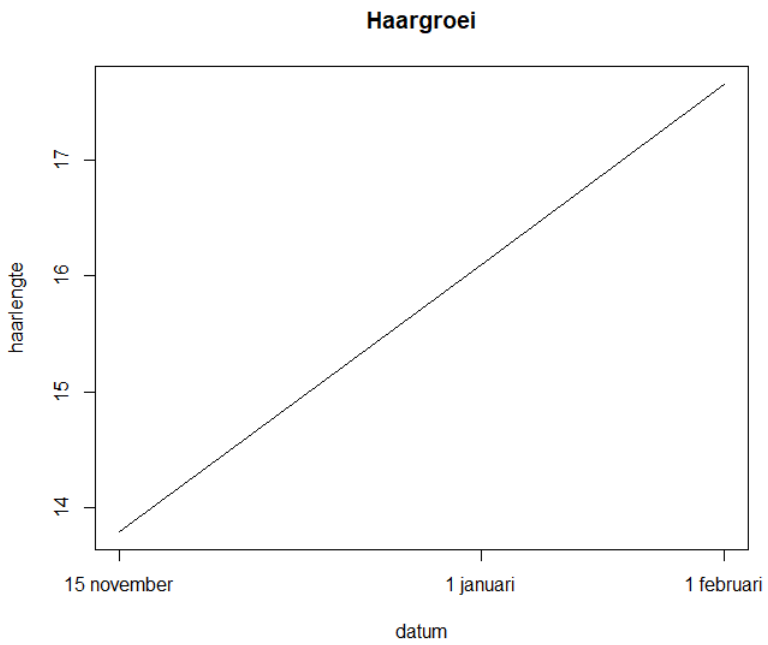
1. $H_0 : \mu_X = 8300$ versus $H_a : \mu_X < 8300$

2. We zien al dat $\bar{x} > \mu_X$, en dus de alternatieve hypothese niet kan kloppen. We aanvaarden de nulhypothese.

6 Haargroei

1. Per dag groeit die haar: $\frac{1.6}{31} = 0.0513129$. Want we weten dat het op 31 dagen 1.6cm groeit. Dit is onze b_1, of de richtingscoëfficiënt.
   Van 15 november tot 1 januari: -46 dagen
   Dus op 15 november was die haar $16.2 - (46 \times 0.0516129) = 13.8$cm. Dit is onze b_0, of het intercept.

2. $Y = 13.8 + 0.05X$



7 Facebook gebruik

1. $H_0 : \mu_{♂} = \mu_{♀}$ of $\mu_{♂} - \mu_{♀} = 0$ versus $H_a : \mu_{♂} \neq \mu_{♀}$ of $\mu_{♂} - \mu_{♀} \neq 0$
2. Tweezijdig
3. Afhankelijke steekproeven. Want we hebben data van 160 hetero-koppels. Dit wil dus zeggen data van 160 mannen en 160 vrouwen. Maar elk datapunt van de mannen hangt samen met een datapunt van de vrouwen, omdat het een koppel is. We gaan dus een t-toets voor one sample gebruiken. (verschillen van paren, cf. p. 99)
4. $\mu_{♂} = 187$, $\mu_{♀} = 204$, $\mu_{totaal} = 391$
5. $s^2_{♀} = sn^2_{♀} \frac{n}{n-1} = 31^2 \frac{160}{159} = 967.044$
   $s^2_{♂} = sn^2_{♂} \frac{n}{n-1} = 16^2 \frac{160}{159} = 257.6101$
6. $\rho = \frac{COV}{s_X s_Y} \Rightarrow COV = \rho s_X s_Y$

$$\begin{aligned} VAR(A+B) &= VAR(A) + VAR(B) + 2COV(A,B) \\ &= 967.044 + 257.6101 + (2 \times 0.35 \times \sqrt{967.044 \times 257.6101}) \\ &= 1574.05 \end{aligned}$$

8 Studietijd 2

1. $H_0 : \mu_\psi = \mu_\pi$ versus $H_a : \mu_\psi \neq \mu_\pi$
2. Tweezijdig
3. Onafhankelijk

9 Klanttevredenheid

1. $H_0 : med_{♀} = med_{♂}$ versus $H_a : med_{♀} \neq med_{♂}$
2. Tweezijdig
3. Onafhankelijk
4. Geen van de geziene toetsen, omdat we data hebben op ordinaal meetniveau. Daarom dat we ook de mediaan gebruiken in de hypotheses, omdat het gemiddelde zinloos is.

10 Klanttevredenheid 2

1. Voor de kans op een man tellen we alle kansen op voor een man, ongeacht de waarde van tevredenheid: $0.03 + 0.06 + 0.07 + 0.04 + 0.05 = 0.25$
2. De kans op een waarde groter dan 5 berekenen we door alle kansen voor de waarde 5, waarde 6 en waarde 7 op te tellen voor beide geslachten: $0.21 + 0.15 + 0.12 + 0.07 + 0.04 + 0.05 = 0.64$
3. Om te zien of deze twee variabelen afhankelijk of onafhankelijk zijn gaan we eerst voor elk de marginale kansen berekenen:
   - Man: 0.25
   - Vrouw: 0.75
   - 3: 0.12
   - 4: 0.24
   - 5: 0.28
   - 6: 0.19]item 7: 0.17

   Hierna gaan we telkens de gecombineerde kansen berekenen. Dus bijvoorbeeld de kans voor een Man met waarde voor tevredenheid 3: $0.25 \times 0.12 = 0.03$. Dit doen we voor elke combinatie. Als we exact dezelfde tabel terug uitkomen spreken we van onafhankelijke variabelen. We krijgen volgende tabel:

tevredenheid	1	2	3	4	5	6	7
vrouw	0	0	0.09	0.18	0.21	0.1425	0.1275
man	0	0	0.03	0.06	0.07	0.0475	0.0425

   We zien in de laatste 2 kolommen dat de waarden niet overeenkomen. Dus we spreken van afhankelijke variabelen.

11 Afhankelijkheid

1. Afhankelijk
2. Onafhankelijk (hoewel men ook kan opperen dat er volwassenen avondonderwijs of tweedekansonderwijs doen, en sommigen op latere leeftijd toch nog aan een universitaire opleiding beginnen - wat dan zorgt voor afhankelijkheid)
3. Afhankelijk. BMI wordt bepaald door gewicht en lengte. Dus we moeten lengte altijd mee in rekening brengen.

12 Zuivere munt

1. $H_0 : \pi = 1/2$ versus $H_a : \pi \neq 1/2$. Een tweezijdige alternatieve hypothese

2. We zitten hier bij de binomiaalverdeling. We observeren 4 keer munt. De p-waarde die we gaan berekenen is dat we minstens even extreem krijgen bij nieuwe steekproeven. Om dit even wat concreter te maken: als we de nulhypothese even herschrijven als 5/10 en het geobserveerde als 8/10, dan wil dit zeggen dat we de kans gaan berekenen dat we minstens 3/10 hoger zitten en minstens 3/10 lager. Als we dus 5 worpen doen wil dit zeggen 4 keer munt en 5 keer munt, maar ook 4 keer kop en 5 keer kop. Dit laatste herschrijven we naar 1 keer munt en 0 keer munt. Aangezien de binomiaalverdeling ook symmetrisch is, kunnen we dezelfde redenering gebruiken als de vorige p-waarde. Namelijk we berekenen de kans aan 1 kant, en verdubbelen deze.

$$\begin{aligned}\mathcal{P}(B(5,0.5) \geq 4) &= \mathcal{P}(B(5,0.5) = 4) + \mathcal{P}(B(5,0.5) = 5) \\ &= \frac{5}{32} + \frac{1}{32} = \frac{6}{32}\end{aligned}$$

   Aangezien we tweezijdig werken:

$$2\frac{6}{32} = \frac{3}{8} = 0.375$$

   We krijgen een p-waarde van 0.375 en deze is groter dan $\alpha = 0.05$, en dus aanvaarden we de nulhypothese

3. Neen, het kan aan toeval te wijten zijn. In een andere steekproef zouden we misschien meer kop krijgen. De hypotheses dienen a priori opgesteld te worden. Indien we dit niet zouden doen dan krijgen we een stijging van onze Type I fout.

4. p-waarde wordt kleiner naarmate de steekproefgrootte toeneemt. Dit is exact dezelfde redenering als bij de eerste oefeningen.

$$\begin{aligned}\mathcal{P}(B(10,0.5) \geq 8) &= \mathcal{P}(B(10,0.5) = 8) + \mathcal{P}(B(10,0.5) = 9) + \mathcal{P}(B(10,0.5) = 10) \\ &= \frac{56}{1024} = 0.0546875\end{aligned}$$

   Maar aangezien dit weer een eenzijdige p-waarde is, vermenigvuldigen we deze met 2 en krijgen we $p = 0.109375$. Deze is kleiner dan de vorige, toch aanvaarden we nog steeds de nulhypothese.

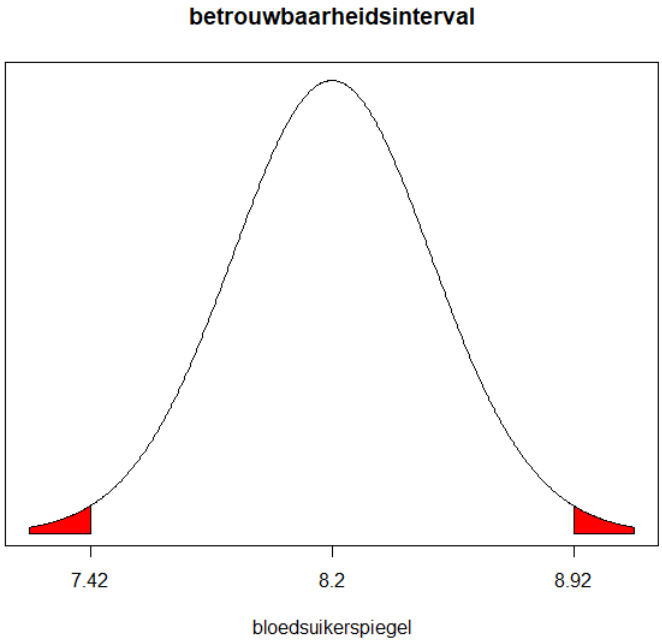
Stel dat we 1000 keer gaan opwerpen, dan gaan we de vorm beginnen krijgen van een normaalverdeling. Hoewel we dus discreet werken kunnen we wel sommige zaken extrapoleren. De binomiaalverdeling heeft ook een verwachting en een variantie. Net zoals bij een normaalverdeling en een t-verdeling zal de variantie kleiner worden naarmate we de steekproef laten toenemen. Dit zorgt ervoor dat ons betrouwbaarheidsinterval gaat versmallen, en dus de kans groter gaat worden dat we de nulhypothese verwerpen. Als de kans groter wordt dat we de nulhypothese verwerpen wil dit zeggen dat onze p-waarde daalt. We trekken namelijk ALTIJD dezelfde conclusie met een p-waarde als met een betrouwbaarheidsinterval als we met dezelfde α werken.

Betrouwbaarheidsinterval versus p-waarde

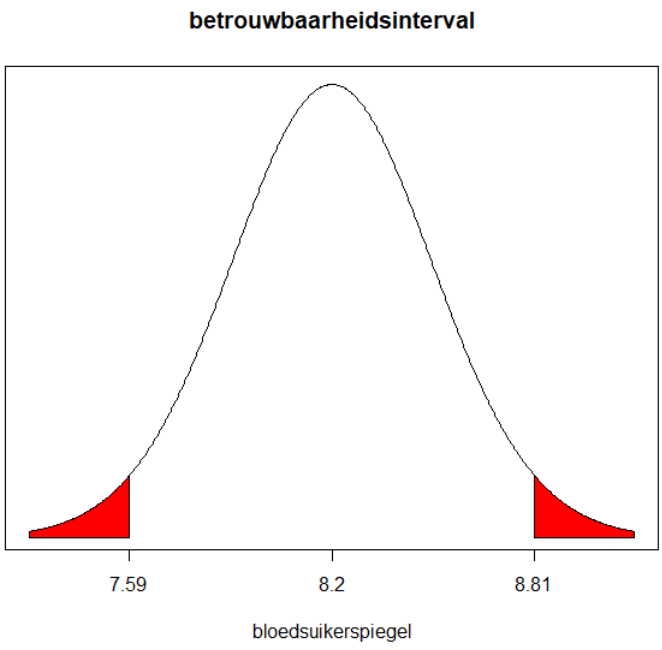
In de oefeningen is dit al grotendeels uitgelegd. Maar hier zal het ook grafisch verduidelijkt worden. We gebruiken als voorbeeld de oefening van de bloedsuikerspiegel. Dit wil zeggen dat we 51 vrijheidsgraden hebben.

Betrouwbaarheidsinterval

Bij een betrouwbaarheidsinterval gaan we een interval symmetrisch rond het geobserveerde gemiddelde opstellen. Bij een betrouwbaarheidsinterval bepalen we vooraf de kansen die we gebruiken om de nulhypothese te verwerpen. Dit centreren we rond het gemiddelde. In onze casus is dit dus rond 8.2:

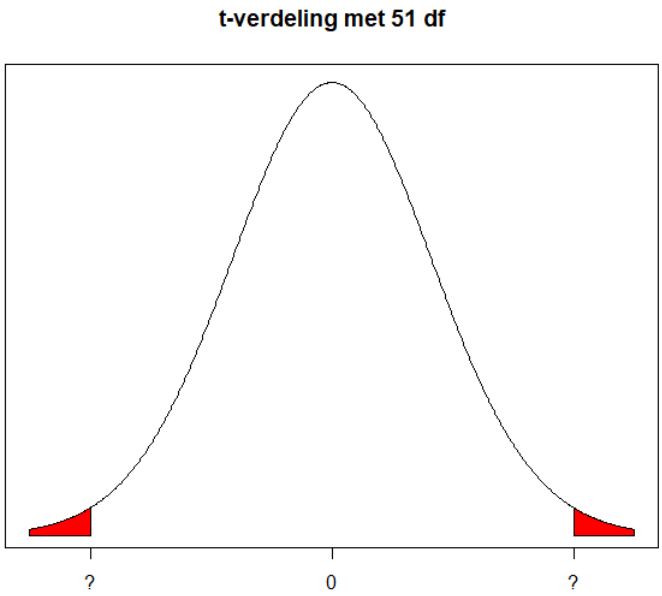


In de grafiek is het witte gedeelte ons aanvaardingsgebied. We hebben 98% van onze grafiek die wit is, en dit gaat van 7.42 tot 8.92. We gaan dan kijken of 7.5 - de verwachting - in het witte gedeelte valt. We zien dat het rode stuk links pas vanaf 7.42 begint. Dus we aanvaarden de nulhypothese.
We kunnen ook eens kijken naar het 95$-betrouwbaarheidsinterval:



Nu valt 7.5 in het rode gebied. Dus verwerpen we de nulhypothese.

Maar hoe komen we nu aan deze waarden? We gaan hier enkel voor 98% betrouwbaarheidsinterval overlopen. Het andere is dezelfde redenering. Als we kijken naar de t-verdeling, dan weten we dat deze symmetrisch is rond 0. We hebben in dit geval 51 vrijheidsgraden. We gaan in de t-verdeling de t-waarden zoeken die horen bij de kans dat er 1% boven ligt, en 1% onder:



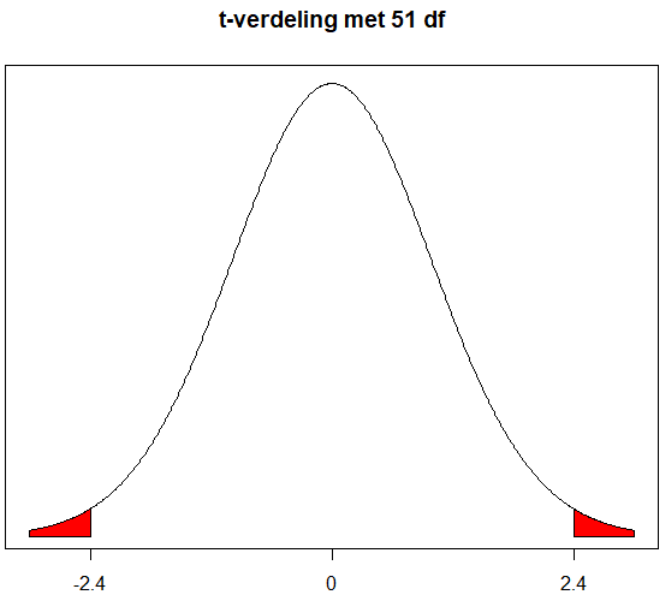
Aangezien de kans aan de linkerkant gelijk is aan de kans aan de rechterkant, en de verdeling symmetrisch is, kunnen we ook concluderen dat de gezochte waarden enkel verschillen qua teken. Dus we moeten maar 1 waarde kennen om de andere waarde te kennen. Aangezien de vroegere tabellen werden opgesteld voor de rechterkant van de grafiek worden de tips ook op deze manier uitgedrukt. Als we kijken naar de formulering:

$$t_{df,p} = T$$

dan kunnen we elk symbool wel plaatsen bij de gegevens. De *df* is het aantal vrijheidsgraden, in dit geval $n-1$. De p is de kans dat we groter observeren dan T, dus de oppervlakte van de rechterkant t.o.v. de waarde. Tenslotte T is de t-waarde die hoort bij de kans. Als we 2 van de 3 zaken weten, dan kunnen we de derde altijd afleiden van de rest. In ons betrouwbaarheidsinterval kennen we het aantal vrijheidsgraden, $df = n-1 = 51$ en de kans, de oppervlakte aan de rechterkant, $p = \alpha/2 = 0.01$. Dus als we deze zaken invullen krijgen we

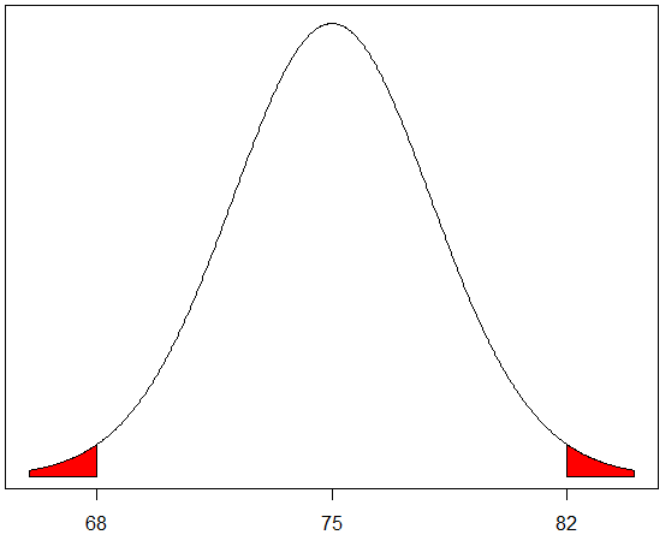
$$t_{51;0.01} = T$$

We zien bij de tips dat er 1 tip bij past. Op de plaats van de T in deze tip staat 2.40. Dit wil dus zeggen dat onze grafiek er zo uitziet:

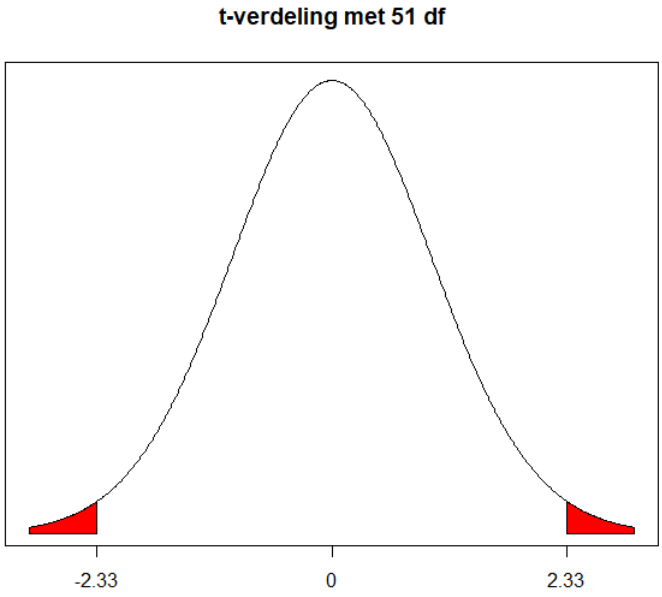


p-waarde

De p-waarde is de kans dat we een waarde minstens even extreem gaan observeren in een nieuwe steekproef als degene gevonden in de huidige steekproef, gegeven dat onze nulhypothese waar is. Omdat we ervan uitgaan dat onze nulhypothese waar is, hebben we een t-verdeling met $n-1$ vrijheidsgraden. Waar zijn we nu in geïnteresseerd?



Nu kennen we de waarden wel waar we de grens stellen, maar de kansen zijn onbekend. Om deze kansen te weten te komen gaan we onze waarden herschalen naar de t-verdeling. Deze herschaling zorgt ervoor dat onze verwachting 75 nu gelijk wordt aan 0 (voor de berekening, kijk in de oefening). De grenswaarden zijn nu bekend als de toetsingsgrootheid - en we weten door de symmetrie dat enkel het teken verschilt:



Nu moeten we de oppervlakte van de rode gebieden zien te weten te komen. Zoals al meermaals gezegd is de verdeling symmetrisch. Als we 1 kant weten, weten we de andere ook. Zoals ook eerder vermeld, de tabellen van vroeger werkten met de rechterzijde, en de tips zijn daarop gebaseerd. Dus we gaan nu de rode oppervlakte langs rechts zoeken. Dit is de kans dat we een t-waarde van 2.33 of groter observeren. Dus we vullen de formulering in:

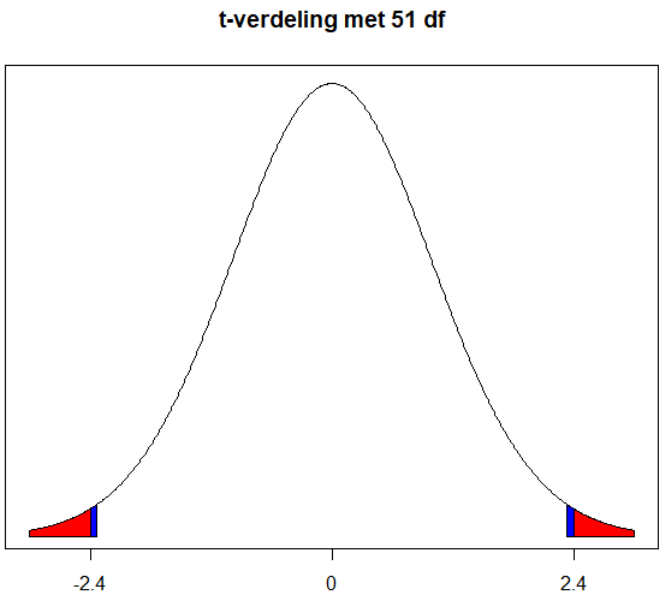
$$t_{51;p} = 2.33$$

Bij de tips vinden we nu een tip die overeenkomt met deze gegevens. De waarde p is hier 0.012. Dit wil dus zeggen dat het rode gebied aan de rechterkant 1.2% is. We hebben dus 1.2% kans om een waarde van 2.33 of groter te observeren, gegeven dat de nulhypothese juist is. Maar omdat we tweezijdig gaan, moeten we de rode oppervlakte langs links er nog bij tellen en deze is exact hetzelfde, dus ook 0.012. Dit zorgt ervoor dat we een p-waarde krijgen van 0.024. Dus de witte oppervlakte tussen onze grensgebieden is 0.976.

Als we met een betrouwbaarheid van 98% werken, dan is onze $\alpha = 0.02$. Onze p-waarde is groter dan 0.02. Wat wil dit nu concreet zeggen? Als we een oneindig aantal steekproeven zouden trekken, en we weten dat de nulhypothese waar is, dan willen we dat onze nulhypothese verworpen wordt in 0.02 gevallen. Dit is onze Type I fout. Maar we observeren een waarde die ons vertelt dat we nog 0.8% boven die 2% zitten. Dus als we oneindig veel steekproeven zouden trekken gaan we in 2.8% van de gevallen een waarde minstens even extreem uitkomen. Maar we willen slechts 2% verwerpen. Dus omdat we nog 0.8% ruimte hebben, zeggen we dat dit de nulhypothese ondersteunt.

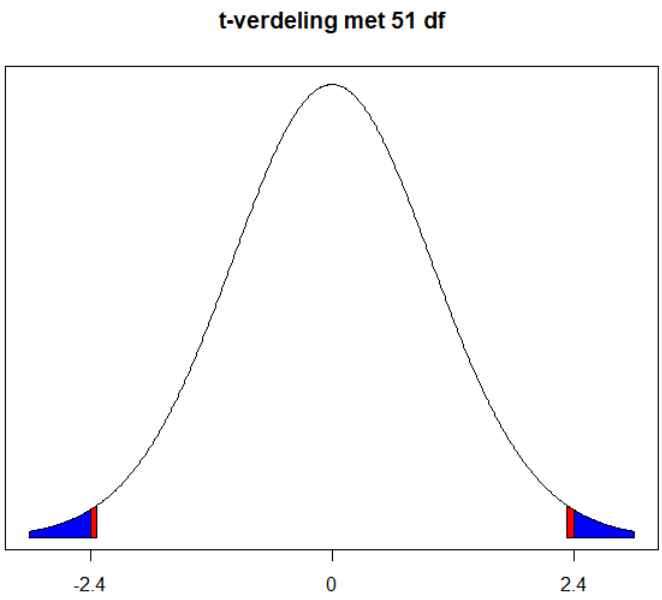
Daarentegen met een betrouwbaarheid van 95%, dan is onze $\alpha = 0.05$. Nu willen we dat in 5% van de gevallen onze nulhypothese verworpen wordt. We weten nu ook dat we 2.8% kans hebben om deze waarde of extremer te observeren. Dus nu hebben we nog zelfs ruimte voor de nulhypothese te verwerpen.

Laten we deze 2 casussen even grafisch voorstellen. Bij betrouwbaarheid van 98%:



De rode oppervlakte hier stelt de 2% voor in welke gevallen we onze nulhypothese gaan verwerpen. De extra blauwe oppervlakte is de extra kans in onze p-waarde. Omdat de oppervlakte van de p-waarde (rood + blauw) groter is dan enkel de 2% aanvaarden we de nulhypothese. Er is nog teveel kans dat we ten onrechte onze nulhypothese zouden verwerpen.

Bij een betrouwbaarheid van 95% zien we een ander verhaal:



Nu is de blauwe oppervlakte onze p-waarde. Dus de kans dat we minstens even extreem gaan observeren. De som van de rode en de blauwe oppervlakte is onze 5% kans dat we de nulhypothese gaan verwerpen. Nu merken we dat we veel meer kans hebben om de nulhypothese te verwerpen dan dat we de kans hebben om minstens even extreem te gaan observeren. Dit wil dus zeggen dat het al vrij uniek is dat we deze waarde observeren als de nulhypothese waar zou zijn. Daarom verwerpen we de nulhypothese.

Conclusie

Dus zoals we zien, we komen op dezelfde conclusies uit bij de p-waarde als bij het betrouwbaarheidsinterval. Dit is ook logisch omdat we telkens vergelijken met dezelfde waarden. Zoals je al zal gemerkt hebben, dezelfde t-waarden komen steeds terug bij de beslissingen.