Statistiek deel 2

Week 1: Schatters

1. Inleiding

Statistiek 2: Schatters

In de statistiek wil je vaak uitspraken doen terwijl je niet alle informatie hebt, bijvoorbeeld:

- Wat vindt Nederland van invoering van een Coronapaspoort?
  Het is ondoenlijk om iedereen te bevragen.
- Over hoeveel materieel beschikt de vijand? Ze vertellen het je vaak niet.
- Hoeveel onontplofte granaten/boobytraps liggen in een bepaald gebied?
- Hoe bepaal je de inzet van Marechaussees op Schiphol?

Er kan dan gebruik worden gemaakt van steekproeven en schatters, Bijvoorbeeld:

- Enquêtes maken gebruik van een willekeurige groep, een vast panel, een representatieve deelgroep. Je kunt in de geënquêteerde groep het percentage uitrekenen. Als de groep groot genoeg is geeft dat een redelijke schatting over wat Nederland denkt.

- Metingen van wachttijden, aantallen, hoeveelheden, kosten, performance, etc.

Schatters voor parameters van standaardverdelingen

In de statistiek maak je vaak aannames voor kansverdelingen.

Voorbeeld: Het gewicht van Nederlandse mannen is normaal verdeeld (aanname). Dat wil zeggen dat de gewichtsverdeling ongeveer een normale verdeling volgt voor een bepaalde waarde van μ en σ.

Je weet alleen μ en σ niet.

Die kun je uitrekenen als je van **alle** Nederlandse mannen het gewicht weet, maar het is praktisch onmogelijk om het gewicht van alle mannen op een bepaald moment te meten. Gewichten veranderen, mannen overlijden of worden vrouw, het correct meten zou een enorme klus zijn om te organiseren en te verwerken, etc.

Daarom moet je wel met een steekproef werken. Maar hoe kun je daaruit een schatting geven voor μ en σ, en hoe groot moet die steekproef zijn?

Schatters voor parameters

In de praktijk zijn er twee soorten schatters:

1. Een **puntschatter** (*estimator*) is een formule waarmee op grond van een **steekproef** $x_1, \ldots x_n$ een goede schatting $\hat{\varphi}(x_1, \ldots x_n)$ kan worden uitgerekend voor een parameter φ die onbekend is.
   Voorbeeld: Je meet het gewicht van 68 volwassen mannen. Het gemiddelde gewicht is 84,6 kg en je gebruikt deze waarde als schatting voor het gemiddeld gewicht van **alle** volwassen Nedermannen.
   De schatter is dan $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g_i$. Schatters worden vaak genoteerd met ^ (circumflex, dakje).

   Een schatter $\hat{\varphi}$ heet **zuiver** (*unbiased*) als de verwachtingswaarde (het gemiddelde als je het heel vaak zou toepassen) ervan gelijk is aan de werkelijke waarde: $E(\hat{\varphi}) = \varphi$.
   Een schatter $\hat{\varphi}$ heet **efficiënt** als de standaarddeviatie ervan klein is.

2. Een stap verder is om niet één waarde te geven als schatting, maar een **betrouwbaarheidsinterval** (*confidence interval, CI*), een interval waar de werkelijke waarde die je wilt schatten met een bepaalde zekerheid in zit, bijvoorbeeld een 95% betrouwbaarheidsinterval.
   Voorbeeld: Je bepaalt met een steekproef dat het **gemiddelde gewicht** van alle volwassen Nedermannen zich met 95% zekerheid in het interval [81,87] kg bevindt.
   Let op: Dat is wat anders dan dat het gewicht van 95% van de Nedermannen tussen 81 en 89 kg ligt!

Schatters voor μ en σ voor een willekeurige kansverdeling

Herhaling: De verwachtingswaarde $E(\underline{x})$ van een kansvariabele $\underline{x}$ is als volgt gedefinieerd:
Voor een discrete kansvariabele $\underline{k}$ met een kansfunctie $f(k) = P(\underline{k} = k)$ is $\boldsymbol{E(\underline{k}) = \sum_{\text{alle } k} kf(k)}$.
Voor een continue kansvariabele $\underline{x}$ met een kansdichtheidsfunctie $f(x) =$ is $\boldsymbol{E(\underline{x}) = \int xf(x)dx}$.
Hiervoor moet je alle mogelijke uitkomsten kennen en alle bijbehorende kansen.

In plaats van $\underline{k}$ zelf kun je ook de verwachtingswaarde van een functie $h(\underline{k})$ van $\underline{k}$ bekijken: $\boldsymbol{E\left(h(\underline{k})\right)}$.

Bijvoorbeeld: $\text{Var}(\underline{k}) = E\left((\underline{k} - \mu)^2\right) = E(\underline{k}^2) - \mu^2$

Bijvoorbeeld: Hierin zijn de fabricagekosten, marge en transportkosten per product samen €21 en de vaste bestelkosten € 45 per bestelling. Door $\underline{l} = h(\underline{k})$ als nieuwe kansvariabele te nemen kun je laten zien dat de verwachtingswaarde van de kosten $E\left(h(\underline{k})\right) = \sum_{\text{alle } k} h(k)f(k)$ is en analoog voor continue kansvariabelen. Je vervangt in de formule voor de verwachtingswaarde dus gewoon $kf(k)$ door $h(k)f(k)$.

Rekenregel 1: Als $\underline{x}$ en $\underline{y}$ kansvariabelen zijn en a en b zijn vaste getallen dan geldt:

$$E\left(a\underline{x} + b\underline{y}\right) = aE(\underline{x}) + bE\left(\underline{y}\right)$$

Rekenregel 2: Als $\underline{x}$ een kansvariabele is met $\mu = E(\underline{x})$, dan geldt $\text{Var}(\underline{x}) = E\left((\underline{x} - \mu)^2\right) = E(\underline{x}^2) - \mu^2$

Schatter voor de gemiddelde waarde μ

We bekijken eerst het voor de hand liggende idee om de gemiddelde waarde van een kansverdeling te schatten door een steekproef uit die kansverdeling te nemen: $x_1, \ldots, x_n$ en daarvan het gemiddelde te nemen $\hat{\mu} = \frac{1}{n}(x_1 + \cdots + x_n)$.

Je hoopt dat dit een goede benadering is voor de gemiddelde waarde μ van de kansverdeling. D.w.z., als je dit heel vaak doet, dan wil je dat de gemiddelden van al die steekproeven μ is.

Preciezer gezegd, je hoopt dat $E(\hat{\mu}) = E\left(\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n)\right) = \mu$.

Dit is eenvoudig aan te tonen door Rekenregel 1 een paar keer toe te passen:

$$E(\hat{\mu}) = E\left(\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n)\right) = \frac{1}{n}E(\underline{x}_1 + \cdots + \underline{x}_n) = \frac{1}{n}\left(E(\underline{x}_1) + \cdots + E(\underline{x}_n)\right)$$

Alle $\underline{x}_i$ zijn getrokken uit dezelfde kansverdeling met gemiddelde waarde μ, dus er geldt voor elke i:

$$E(\underline{x}_i) = \mu$$

Hierdoor is

$$E(\hat{\mu}) = E\left(\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n)\right) = \frac{1}{n}(\mu + \cdots + \mu) = \frac{n\mu}{n} = \mu$$

Dit is niet verrassend, maar logisch, het klopt met je intuïtie, en het klopt ook nog eens wiskundig.

Conclusie: **$\hat{\mu}$, het steekproefgemiddelde, is een zuivere schatter voor μ. Voor elke kansverdeling!**

Schatter voor de standaardvariatie σ

We hebben de gemiddelde waarde van een kansverdeling geschat door het gemiddelde te nemen van een steekproef van een aantal waarden uit die kansverdeling $(x_1, \ldots, x_n)$:

$$\hat{\mu} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$$

Het ligt voor de hand om hetzelfde voor de standaarddeviatie te doen: Bereken van de steekproef de standaarddeviatie en gebruik die als schatter $\hat{\sigma}$ voor de standaarddeviatie σ van de kansverdeling, dus

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

Helaas blijkt dat niet zo te werken en dat is een technisch verhaal.

Het is niet direct noodzakelijk dat je de details snapt, zolang je het maar weet dat het zo is en weet hoe je steekproefstandaarddeviaties moet uitrekenen.

Technische uitleg schatter voor σ (1) (Hoef je niet te snappen)

De uitleg in Buijs Appendix A bevat foutjes en is te kort door de bocht en daardoor onbegrijpelijk.

Bekijk de schatter $u = \sum_{i=1}^{n}(x_i - \hat{\mu})^2$. We gaan proberen daarvan de verwachtingswaarde uit te rekenen, in de hoop dat daar uitkomt: $n\,\sigma^2$. Allereerst: Zowel alle $\underline{x_i}$ als $\underline{\hat{\mu}}$ zijn kansvariabelen, omdat we de steekproeven vaker willen nemen. Werk eerst de haakjes uit:

$$\underline{u} = \sum_{i=1}^{n}\left(\underline{x_i} - \underline{\hat{\mu}}\right)^2 = \sum_{i=1}^{n}\left(\underline{x_i}^2 - 2\underline{x_i}\underline{\hat{\mu}} + \underline{\hat{\mu}}^2\right) = \sum_{i=1}^{n}\underline{x_i}^2 - 2\underline{\hat{\mu}}\sum_{i=1}^{n}\underline{x_i} + \underline{\hat{\mu}}^2\sum_{i=1}^{n}1 =$$

$$= \left(\sum_{i=1}^{n}\underline{x_i}^2\right) - 2\underline{\hat{\mu}}\cdot n\underline{\hat{\mu}} + n\underline{\hat{\mu}}^2 = \left(\sum_{i=1}^{n}\underline{x_i}^2\right) - n\underline{\hat{\mu}}^2.$$

We gebruiken hier o.a. dat $\sum_{i=1}^{n}\underline{x_i} = n\mu$. Gevolg (Rekenregel 1 en alle $\underline{x_i}$ komen uit dezelfde verdeling):

$$E\left(\underline{u}\right) = \sum_{i=1}^{n} E\left(\underline{x_i}^2\right) - nE\left(\underline{\hat{\mu}}^2\right) = nE\left(\underline{x}^2\right) - nE\left(\underline{\hat{\mu}}^2\right)$$

Gebruik twee keer Rekenregel 2:

$$E\left(\underline{x}^2\right) = \mathrm{Var}\left(\underline{x}^2\right) + E\left(\underline{x}\right)^2 = \sigma^2 + \mu^2$$

en

$$E\left(\underline{\hat{\mu}}^2\right) = \mathrm{Var}\left(\underline{\hat{\mu}}^2\right) + E\left(\underline{\hat{\mu}}\right)^2 = \frac{\sigma^2}{n} + \mu^2$$

want voor het gemiddelde geldt: de verwachtingswaarde is de oorspronkelijke μ, maar de standaarddeviatie is: de oorspronkelijke σ gedeeld door $\sqrt{n}$.

Technische uitleg schatter voor σ (2). Het resultaat is belangrijk!

Samenvoegen levert

$$E(\underline{u}) = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\,\sigma^2.$$

Dit is een verrassend resultaat, omdat je zou verwachten dat $E(\underline{u}) = n\,\sigma^2$.
Dit betekent dat een zuivere schatter van de standaarddeviatie niet is:

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}\,,$$

maar:

$$\hat{\sigma} = s = \sqrt{\boldsymbol{\frac{1}{n-1}}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}\,.$$

De eerste formule geeft de standaarddeviatie als je het voor **alle** waarden van de kansverdeling uitrekent, maar is een onzuivere schatter voor een steekproef(klopt gemiddeld gesproken niet). De tweede is zuiver.

Verschil tussen STDEV.P en STDEV.S in Excel

STDEV.**P** (formule met $\frac{1}{n}$) pas je toe als je de hele **P**opulatie gebruikt. Je hebt het dan over een **discrete** kansverdeling met eindig veel mogelijkheden en je **berekent** met STDEV.P de standaarddeviatie door **alle mogelijkheden** mee te nemen. Deze waarde is echt de standaarddeviatie en wordt met σ aangeduid.

STDEV.**S** (formule met $\frac{1}{n-1}$) pas je toe op een **S**teekproef die je neemt uit een grotere populatie. Dat kan een continue verdeling zijn zoals de normale verdeling, maar ook een discrete verdeling waarbij je een deel van alle mogelijke waarden in je berekening gebruikt. Je **schat** daarmee de standaarddeviatie m.b.v. een kleine steekproef. Deze waarde is de **steekproefstandaarddeviatie** en **wordt vaak met s aangeduid**.

De waarde van STDEV.S is altijd groter dan die van STDEV.P, een iets ruimere schatting, maar het verschil tussen de twee waarden is alleen relevant voor kleine waarden van de steekproefgrootte n. Vuistregel:

Voor $n > 5$ is STDEV.S minder dan 10% groter dan STDEV.P.
Voor $n \geq 50$ is STDEV.S minder dan 1% groter dan STDEV.P.
Voor $n \geq 500$ is STDEV.S minder dan 0,1% groter dan STDEV.P.

Voorbeeld: Verschillende schatters voor een uniforme verdeling

Let op. Dit voorbeeld laat zien hoe je soms verschillende schatters kunt verzinnen met verschillende resultaten. Dit verzinnen hoeft je niet zelf te kunnen.

De schurkenstaat Borea heeft zich op unfaire wijze toegang verschaft tot nucleaire technologie en voert geregeld ondergrondse kernproeven uit om de rest van de wereld te imponeren. Op grond van seismische informatie en satellietfoto's van kraters op Boreaals gebied kan een inschatting worden gemaakt van de kracht van de bommen.

Men wil aan de hand van gegevens van eerdere proeven een schatting maken over de grootte van de bom die Borea in staat is te maken. Het volgende is zeer vereenvoudigde versie:
Probleem: Je hebt drie waarnemingen van kernexplosies en voor elke explosie een schatting van de explosieve kracht. Je wilt een schatting geven voor de kracht van de sterkste bom die Borea kan maken.

Aanname: Borea kan een kernbom maken met een kracht b, en de kracht van de bommen tot nu toe zijn willekeurig gekozen, d.w.z. een steekproef van 3 waarden genomen uit een uniforme verdeling op $[0, b]$, waarbij de waarde van de bovengrens b onbekend is. Hoe kun je hieruit de waarde van b schatten?

Voorbeeld: Schatters voor een uniforme verdeling

Statistisch probleem:
Je hebt een steekproef van n waarden genomen uit een uniforme verdeling op $[0, b]$, de waarde van de bovengrens b is onbekend. Hoe kun je de waarde van b schatten uit deze steekproef?

Poging 1:
Neem de maximale waarde uit de steekproef als schatter: $\hat{b}_1 = \max(x_1, \dots, x_n)$

Voorbeeld: Verschillende schatters voor een uniforme verdeling

Met Excel is 20 keer (regels) het volgende gedaan:
Neem een steekproef van drie random getallen (1, 2, 3) uit een uniforme verdeling op [0,10], dus $n = 3$ en $b = 10$.
We gaan met vier verschillende schatters (laatste 4 kolommen) de bovengrens 10 proberen te schatten uit de drie steekproefwaarden.
Voor de vier schatters zijn gemiddelde en standaarddeviatie over de 20 pogingen rechtsonder uitgerekend.

Bij Poging 1 (Maximum) valt op dat de geschatte waarde gemiddeld veel te laag is (7,09 < 10)

Dit komt omdat het maximum **altijd** onder 10 zit.
Dit is geen zuivere schatter, hij klopt gemiddeld niet.

Nr	1	2	3	max			
1	0,35	0,64	1,05	1,05			
2	4,73	9,85	9,67	9,85			
3	3,33	3,57	2,17	3,57			
4	4,89	2,56	5,61	5,61			
5	2,31	0,84	2,29	2,31			
6	4,42	9,54	5,81	9,54			
7	9,65	0,18	6,13	9,65			
8	1,43	3,26	5,75	5,75			
9	2,19	9,19	5,83	9,19			
10	4,78	8,20	1,49	8,20			
11	6,12	9,54	5,02	9,54			
12	7,59	9,39	3,15	9,39			
13	0,23	8,48	7,83	8,48			
14	3,02	0,05	2,83	3,02			
15	5,41	0,05	8,53	8,53			
16	5,55	1,10	1,12	5,55			
17	1,78	7,79	8,78	8,78			
18	8,73	3,82	4,35	8,73			
19	7,20	5,99	4,83	7,20			
20	5,95	7,80	6,84	7,80			
			mu=	7,09			
			sigma=	2,73			

Voorbeeld: Schatters voor een uniforme verdeling

Poging 2:

We weten dat voor de gemiddelde waarde van de uniforme verdeling op $[0, b]$ geldt : $\mu = \frac{b}{2}$.

Dat betekent dat $b = 2\mu$. De bovengrens is tweemaal de gemiddelde waarde van de verdeling.

We zouden dus als schatter kunnen tweemaal het gemiddelde van de **steekproef** kunnen nemen:

$$\hat{b}_2 = 2\hat{\mu} = 2\frac{x_1 + \cdots + x_n}{n}$$

Voorbeeld: Schatters voor een uniforme verdeling

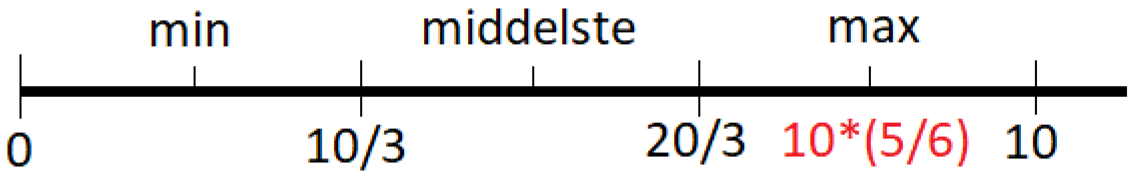
Bij Poging 2 is de geschatte waarde gemiddeld best goed (9,69).

Dit is wel een zuivere schatter (omdat $\widehat{\mu}$ een zuivere schatter is voor μ, dus gemiddeld klopt de schatting).

Nr	1	2	3	max	2*mean
1	0,35	0,64	1,05	1,05	1,36
2	4,73	9,85	9,67	9,85	16,17
3	3,33	3,57	2,17	3,57	6,05
4	4,89	2,56	5,61	5,61	8,71
5	2,31	0,84	2,29	2,31	3,62
6	4,42	9,54	5,81	9,54	13,19
7	9,65	0,18	6,13	9,65	10,64
8	1,43	3,26	5,75	5,75	6,96
9	2,19	9,19	5,83	9,19	11,47
10	4,78	8,20	1,49	8,20	9,65
11	6,12	9,54	5,02	9,54	13,79
12	7,59	9,39	3,15	9,39	13,42
13	0,23	8,48	7,83	8,48	11,02
14	3,02	0,05	2,83	3,02	3,94
15	5,41	0,05	8,53	8,53	9,32
16	5,55	1,10	1,12	5,55	5,18
17	1,78	7,79	8,78	8,78	12,23
18	8,73	3,82	4,35	8,73	11,27
19	7,20	5,99	4,83	7,20	12,02
20	5,95	7,80	6,84	7,80	13,73
			mu=	7,09	9,69
			sigma=	2,73	3,99

Voorbeeld: Schatters voor een uniforme verdeling

Poging 3: Je hebt een steekproef van n waarden en je ordent ze van klein naar groot. Je verdeelt het interval $[0\ ,\ b]$ in n gelijke stukken dan zal de maximale waarde waarschijnlijk in het laatste intervalletje komen en gemiddeld ongeveer in het midden van het laatste interval, dus op de waarde $\frac{2n-1}{2n}b$.



We zouden dus als schatter kunnen nemen

$$\hat{b}_3 = \frac{2n}{2n-1}\max(x_1, \dots, x_n).$$

In ons voorbeeld is $n = 3$, dus het maximum wordt vermenigvuldigd met $6/5$. Dit corrigeert voor dat het maximum van de steekproefwaarden altijd structureel onder de bovengrens zit.

Voorbeeld: Schatters voor een uniforme verdeling

Poging 3 laat weer een te lage waarde zien (8,50), maar wel beter dan Poging 1.

Dit is geen zuivere schatter (want de maximale waarde blijkt gemiddelde niet precies halverwege het laatste interval te liggen, dus die aanname klopt niet).

Nr	1	2	3	max	2*mean	**6/5*max**
1	0,35	0,64	1,05	1,05	1,36	**1,26**
2	4,73	9,85	9,67	9,85	16,17	**11,82**
3	3,33	3,57	2,17	3,57	6,05	**4,29**
4	4,89	2,56	5,61	5,61	8,71	**6,73**
5	2,31	0,84	2,29	2,31	3,62	**2,77**
6	4,42	9,54	5,81	9,54	13,19	**11,45**
7	9,65	0,18	6,13	9,65	10,64	**11,58**
8	1,43	3,26	5,75	5,75	6,96	**6,90**
9	2,19	9,19	5,83	9,19	11,47	**11,03**
10	4,78	8,20	1,49	8,20	9,65	**9,84**
11	6,12	9,54	5,02	9,54	13,79	**11,45**
12	7,59	9,39	3,15	9,39	13,42	**11,27**
13	0,23	8,48	7,83	8,48	11,02	**10,18**
14	3,02	0,05	2,83	3,02	3,94	**3,63**
15	5,41	0,05	8,53	8,53	9,32	**10,23**
16	5,55	1,10	1,12	5,55	5,18	**6,66**
17	1,78	7,79	8,78	8,78	12,23	**10,53**
18	8,73	3,82	4,35	8,73	11,27	**10,48**
19	7,20	5,99	4,83	7,20	12,02	**8,65**
20	5,95	7,80	6,84	7,80	13,73	**9,36**
			mu=	7,09	9,69	**8,50**
			sigma=	2,73	3,99	**3,28**

Voorbeeld: Schatters voor een uniforme verdeling

Poging 4: In de literatuur vind je dat de maximale waarde van een steekproef van *n* waarden uit de uniforme verdeling op $[0, b]$ een verwachtingswaarde heeft van $\frac{n}{n+1}b$.
Dat is dus niet $\frac{2n-1}{2n}b$, zoals we in de vorige poging hadden beredeneerd, die aanname klopte niet.

We zouden als verbeterde schatter kunnen nemen

$$\hat{b}_4 = \frac{n+1}{n}\max(x_1, \ldots, x_n).$$

Google

maximum of uniform distribution estimator

altervista.org
http://lucaballan.altervista.org › pdfs › Max_of... PDF

Maximum of a set of IID Random Variables - Luca Ballan

Given a set of n i.i.d. **uniform random variables** Xi,. Xi ~ U (a, b) let Y be the random variable representing the **maximum** of this set.

27 pagina's

Voorbeeld: Schatters voor een uniforme verdeling

Poging 4 lijkt weer een goede waarde te hebben (9,45), de standaarddeviatie lijkt kleiner dan in Poging 2.

Dit is inderdaad een zuivere schatter.

De variantie van Poging 2 is gemiddeld $\frac{b^2}{3n}$

De variantie van Poging 4 is gemiddeld $\frac{b^2}{n(n+2)}$

De laatste is altijd kleiner (want $n + 2 \geq 3$).

Dat is in ons voorbeeld ook zo, want 3,99 > 3,64

Nr	1	2	3	max	2*mean	6/5*max	**4/3*max**
1	0,35	0,64	1,05	1,05	1,36	1,26	**1,40**
2	4,73	9,85	9,67	9,85	16,17	11,82	**13,14**
3	3,33	3,57	2,17	3,57	6,05	4,29	**4,77**
4	4,89	2,56	5,61	5,61	8,71	6,73	**7,48**
5	2,31	0,84	2,29	2,31	3,62	2,77	**3,07**
6	4,42	9,54	5,81	9,54	13,19	11,45	**12,72**
7	9,65	0,18	6,13	9,65	10,64	11,58	**12,87**
8	1,43	3,26	5,75	5,75	6,96	6,90	**7,67**
9	2,19	9,19	5,83	9,19	11,47	11,03	**12,25**
10	4,78	8,20	1,49	8,20	9,65	9,84	**10,94**
11	6,12	9,54	5,02	9,54	13,79	11,45	**12,73**
12	7,59	9,39	3,15	9,39	13,42	11,27	**12,52**
13	0,23	8,48	7,83	8,48	11,02	10,18	**11,31**
14	3,02	0,05	2,83	3,02	3,94	3,63	**4,03**
15	5,41	0,05	8,53	8,53	9,32	10,23	**11,37**
16	5,55	1,10	1,12	5,55	5,18	6,66	**7,40**
17	1,78	7,79	8,78	8,78	12,23	10,53	**11,71**
18	8,73	3,82	4,35	8,73	11,27	10,48	**11,64**
19	7,20	5,99	4,83	7,20	12,02	8,65	**9,61**
20	5,95	7,80	6,84	7,80	13,73	9,36	**10,40**
			mu=	7,09	9,69	8,50	**9,45**
			sigma=	2,73	3,99	3,28	**3,64**

Voorbeeld: Schatters voor een uniforme verdeling

Wat leer je van dit voorbeeld?

- Het schatten van een parameter in de statistiek kan vaak op meer manieren.

- Je kunt “goede” (zuivere) schatters maken en minder goede (onzuivere), maar onzuivere schatters zijn niet per definitie verkeerd en kunnen ook redelijke schattingen geven.

- Bij zuivere schatters heb je ook nog goede en minder goede, qua spreiding van het resultaat (σ).

- In de praktijk zijn er bijna altijd goede standaardschatters te vinden.

- In kansrekening/statistiek kun je soms argumenten gebruiken die logisch klinken, maar niet kloppen.

Betrouwbaarheid

Onbetrouwbaarheid wordt vaak aangeduid met een (kleine) fractie α, een getal tussen 0 en 1.
Dicht bij 0 (lage onbetrouwbaarheid) is betrouwbaar, dicht bij 1 is onbetrouwbaar.
Betrouwbaarheid is dan de fractie $1-\alpha$, ofwel als percentage: $100(1-\alpha)\%$.

Een onbetrouwbaarheid van $\alpha = 0{,}05$ betekent dat het resultaat in gemiddeld 5% van de gevallen mogelijk niet klopt en dat de betrouwbaarheid 95% is, gemiddeld klopt het resultaat in 95% van de gevallen.

Wetenschappelijk gezien wordt een betrouwbaarheid van $\alpha = 0{,}95$ of groter vaak gezien “aangetoond”.
In wetenschappelijke artikelen wordt vaak gesproken over een **$\boldsymbol{p}$-waarde ($\boldsymbol{p = \alpha}$)**: $p < 0{,}05$ of $p < 0{,}01$ betekent dat de kans dat het aangetoonde resultaat niet klopt kleiner is dan 0,05, resp. 0,01.

Als je kunt aantonen dat een medicijn of behandeling in 40% van de gevallen werkt en de betrouwbaarheid van die 40% is minstens 95% $(p < 0{,}05)$ dan kun je daar een valide publicatie over schrijven. Hoe kleiner de p-waarde, hoe sterker het resultaat.

Er is wel een verschil tussen de betrouwbaarheid (eigenlijk effectiviteit) van de behandeling (40%) en de betrouwbaarheid (95%) waarmee je de effectiviteit van de behandeling (het getal 40%) aantoont.

Betrouwbaarheidsinterval voor μ voor normale verdeling bij bekende σ

Stel je hebt een normaal verdeelde kansvariabele $\underline{x}$ met **bekende $\boldsymbol{\mu}$ en $\boldsymbol{\sigma}$**.
Het steekproefgemiddelde $\bar{x}$ van een steekproef van n waarden $x_1, \ldots, x_n$ voldoet dan aan

$$\bar{x} \sim N\left(\mu?, \frac{\sigma}{\sqrt{n}}\right)$$

Voor een bepaalde betrouwbaarheid, bv. 0,95 kunnen we een **voorspellingsinterval voor $\bar{\boldsymbol{x}}$** vinden dat met kans 0,95 de waarde $\bar{x}$ bevat, nl.:

$$\mu - z\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z\frac{\sigma}{\sqrt{n}}$$

Hierin is de z-waarde bij een rechteroverschrijdingskans van 0,05/2 = 0,025.

$$z = \text{invNorm}(0{,}975) = 1{,}9600$$

De grenzen zijn wel uitgedrukt in de bekende μ en σ. Door omschrijven van de ongelijkheden ($\mu - z\frac{\sigma}{\sqrt{n}} \leq \bar{x} \;\rightarrow \mu \leq \bar{x} + z\frac{\sigma}{\sqrt{n}}$) zie je dat je feitelijk $\bar{x}$ en μ kunt verwisselen:

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z\frac{\sigma}{\sqrt{n}}$$

Dit is een (95%) **betrouwbaarheidsinterval voor** de onbekende $\boldsymbol{\mu}$ in termen van de bekende $\bar{x}$ en σ:

$$\left[\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}}\right]$$

Betrouwbaarheidsinterval versus Voorspellingsinterval

Stel je hebt waarden die voldoen aan een normale verdeling met (onbekend) gemiddelde μ en (bekende) standaarddeviatie σ. Het steekproefgemiddelde $\bar{x}$ van een steekproef van n waarden $x_1, \ldots, x_n$ voldoet aan

$$\bar{x} \sim N\left(\mu?, \frac{\sigma}{\sqrt{n}}\right)$$

Een 95% **voorspellingsinterval** is een interval voor $\bar{x}$ dat met kans 0,95 de waarde $\bar{x}$ bevat, bij gegeven μ en σ:

$$\mu - z\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z\frac{\sigma}{\sqrt{n}}\bar{x}$$

$z = \text{invNorm}(0.95)$ Dit kun je omschrijven als:

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z\frac{\sigma}{\sqrt{n}}$$

Een 95% **betrouwbaarheidsinterval** is een interval voor (de onbekende) μ, waar deze met kans 0,95 in zit, op basis van het gemeten steekproefgemiddelde:

$$\left[\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}}\right]$$

Voorbeeld Betrouwbaarheidsinterval en steekproefgrootte voor μ

Van een bepaald soort accu's is bekend dat de gebruiksduur van een geladen accu normaal verdeeld is met $\sigma = 5$ uur, maar de verwachtingswaarde μ van de gebruiksduur is niet bekend. Deze wordt geschat door de gemiddelde gebruiksduur van 10 accu's te bepalen. Dit levert op: $\bar{x} = 35$ uur.

We kunnen met $\bar{x} = 35$ een 95% betrouwbaarheidsinterval voor de gemiddelde gebruiksduur μ berekenen:

$$\left[\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}}\right] = \left[35 - 1{,}96\frac{5}{\sqrt{10}}, 35 + 1{,}96\frac{5}{\sqrt{10}}\right] = [31{,}90\,, 38{,}10]$$

Een **99%** betrouwbaarheidsinterval is ($z = \text{invNorm}(0{,}995) = 2{,}5758$)

$$\left[\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}}\right] = \left[35 - 2{,}5758\frac{5}{\sqrt{10}}, 35 + 2{,}5758\frac{5}{\sqrt{10}}\right] = [30{,}93\,, 39{,}07]$$

Een hogere betrouwbaarheid leidt tot een breder interval, dus een lagere nauwkeurigheid.

De **onnauwkeurigheid** (de halve lengte van het betrouwbaarheidsinterval) is $z\frac{\sigma}{\sqrt{n}}$.

De onnauwkeurigheid kan verkleind worden door n te vergroten.

Bij 95% betrouwbaarheid is de onnauwkeurigheid (38,1 – 31,9)/2 = 3,10 uur.

Als je daar 1 uur van wilt maken, dan moet $z\frac{\sigma}{\sqrt{n}} = 1$ zijn, dus $n = 96$.

Let op: Meer metingen doen kan ook de waarde van het steekproefgemiddelde wat veranderen.

Dit is een voorbeeld van een **steekproefomvang** berekening.

Voorbeeld Betrouwbaarheidsinterval en steekproefgrootte voor μ bij bekende σ

Let op: We hebben net een **95% betrouwbaarheidsinterval voor de gemiddelde gebruiksduur μ** bepaald:

$$\left[\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}}\right] = \left[35 - 1{,}96\frac{5}{\sqrt{10}}, 35 + 1{,}96\frac{5}{\sqrt{10}}\right] = [31{,}90\,,38{,}10]$$

Dit betekent dat de gemiddelde gebruiksduur van de accu's is geschat op grond van een steekproef van 10 accu's en dat de werkelijke waarde van de gemiddelde gebruiksduur met 95% waarschijnlijkheid ligt tussen 31,90 en 38,10.

Dit betekent **niet** dat 95% van de accu's een gebruiksduur heeft tussen 31,90 en 38,10!

De kans dat een accu een gebruiksduur heeft tussen deze twee waarden is namelijk (even aangenomen dat de gemiddelde waarde 35 uur is)

$$\text{normalcdf}(31.90\,,38.10\,,35\,,5) = 0{,}4647$$

Dit geldt dus maar voor 46% van de accu's.