



Schatten en betrouwbaarheidsintervallen

Zie ook:
Cursusboek hoofdstuk 24

Inhoud

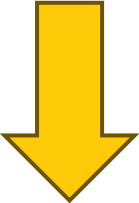
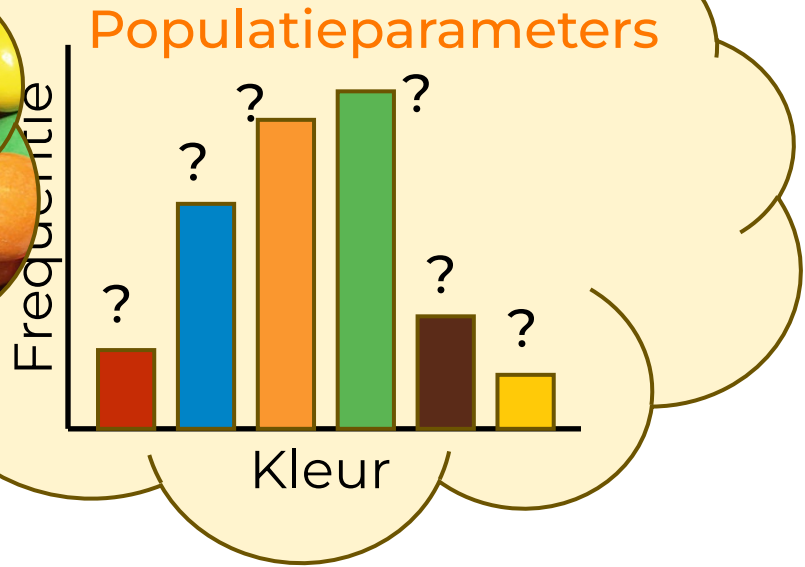
1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

Inhoud

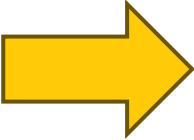
1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

Schattingen uit steekproeven

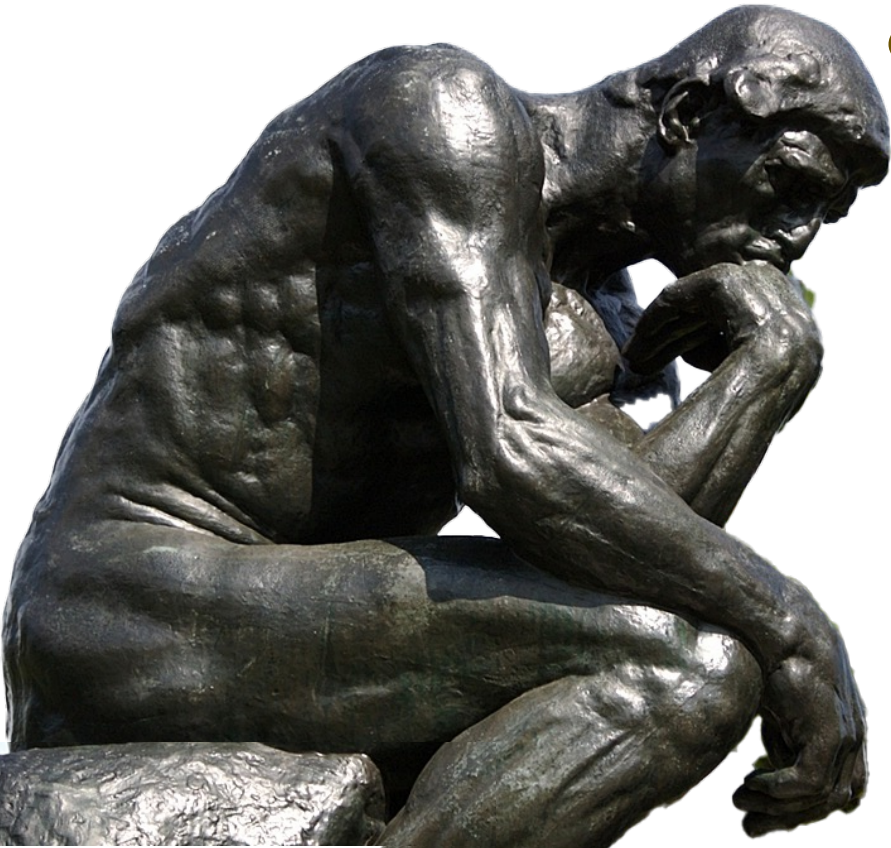
Populatie



Steekproef



Schatting



A decorative card with a blue floral border. The text is written in a blue, cursive script on a light blue rectangular background in the center of the card.

Een schatting
zonder een indicatie
van de precisie
is nutteloos

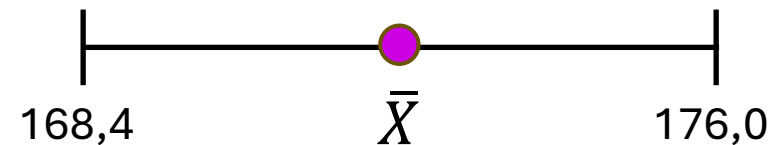
Betrouwbaarheidsintervallen

- Een betrouwbaarheidsinterval (BHI) is een interval rondom een schatter dat *waarschijnlijk* de populatieparameter bevat.
- Een 95%-BHI voor een populatieparameter bevat met 95% zekerheid de waarde van de parameter.

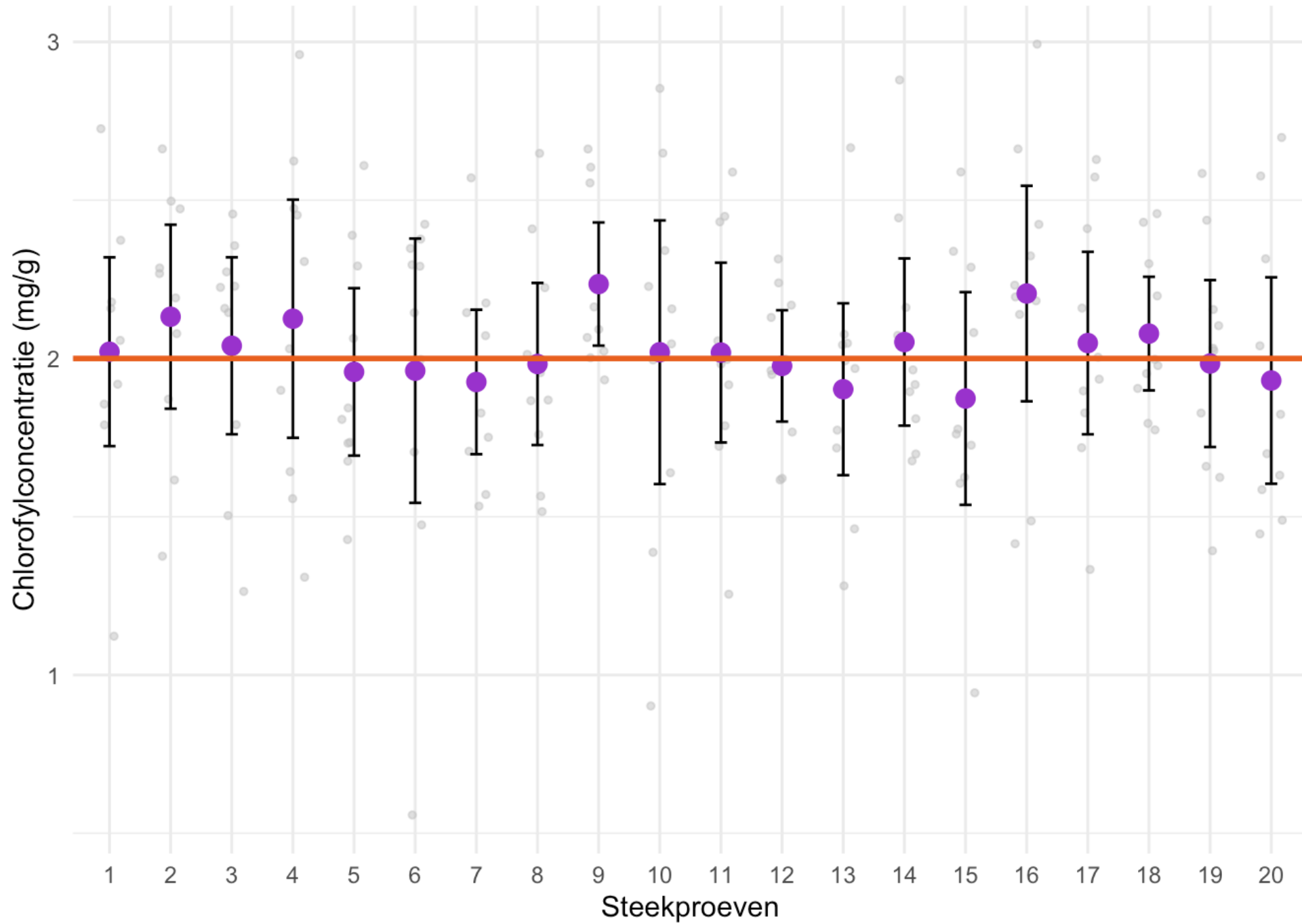
Voorbeeld:

“Op basis van de steekproef van 10 studenten is het 95%-BHI voor de gemiddelde lichaamslengte gelijk aan $(172,2 \pm 3,8)$ cm.”

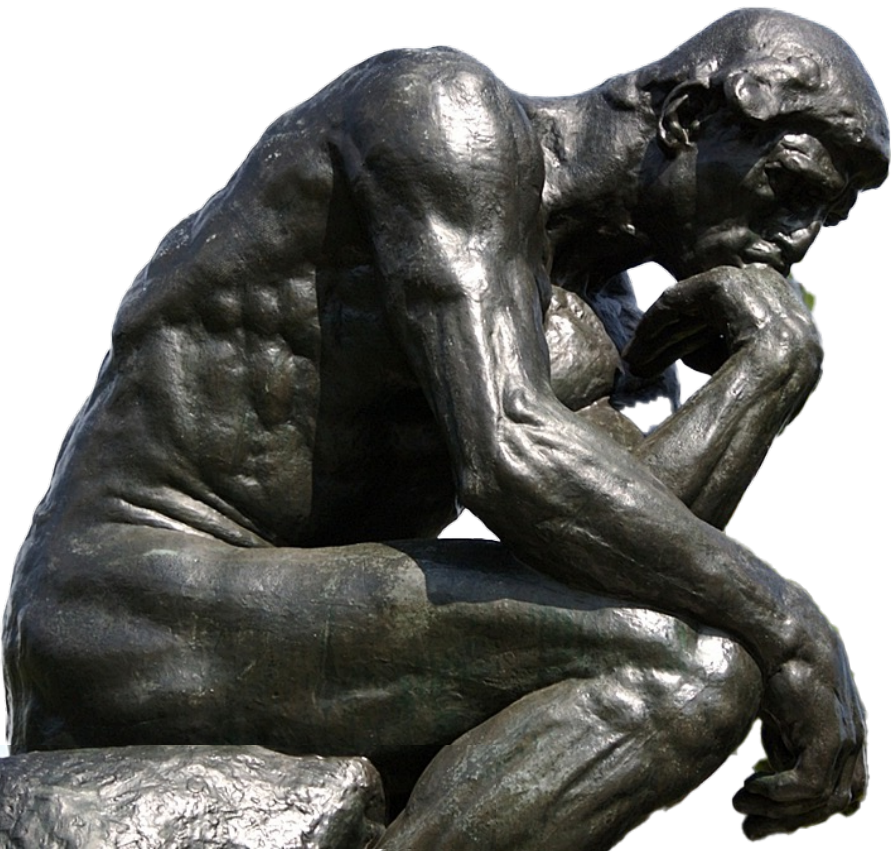
Interpretatie:



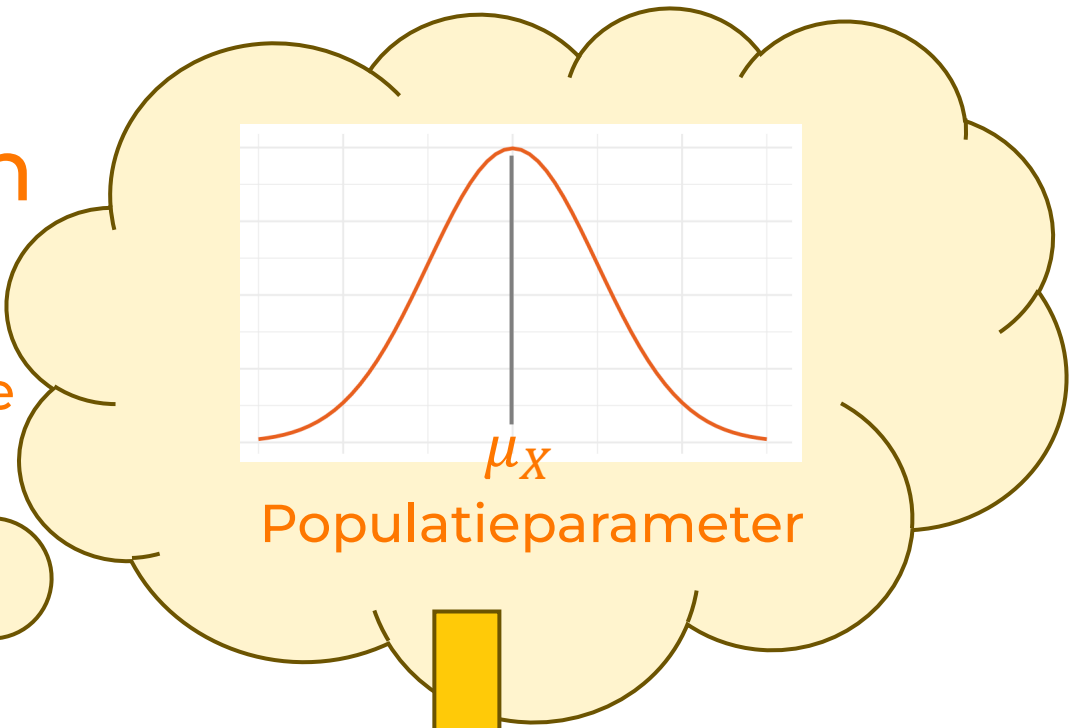
Waarschijnlijk (95% zeker) ligt de gemiddelde lengte μ_X tussen 168,4 cm en 176,0 cm.



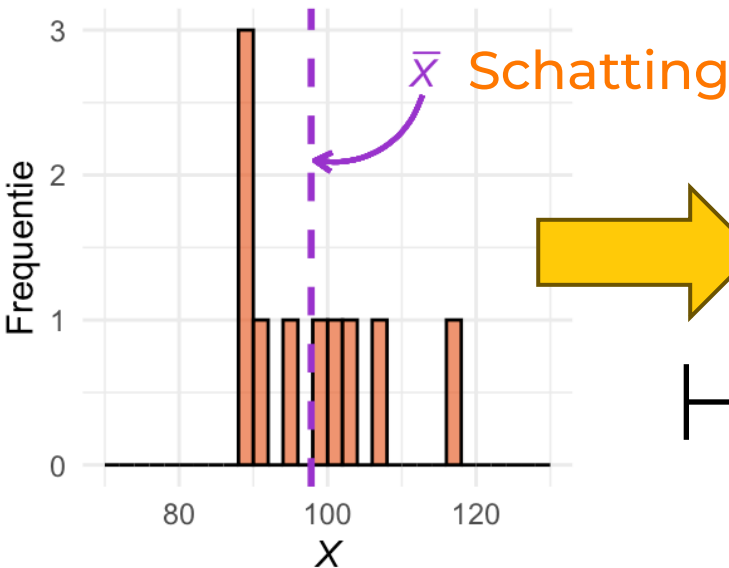
Schatten van het populatiegemiddelde van een continue variabele



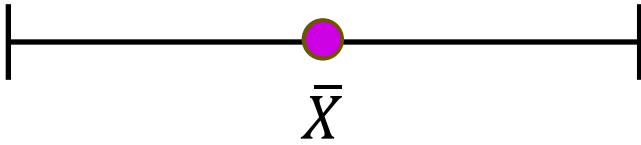
Populatie



Steekproef



Betrouwbaarheids-
interval



Inhoud

1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

Cruciale vraag:

Hoe ver liggen gemiddelden van steekproeven typisch af van het ware gemiddelde van de populatie?

Simuleren!

Het Normale Model

Veel statistische analyses gaan uit van een eenvoudig model:

1. De variabele is normaal verdeeld in de populatie.

Vaak redelijk omdat veel variabelen een klokvormige verdeling hebben.

2. Iedere waarde X_i van de steekproef is *onafhankelijk* uit deze normale verdeling getrokken

Redelijk als

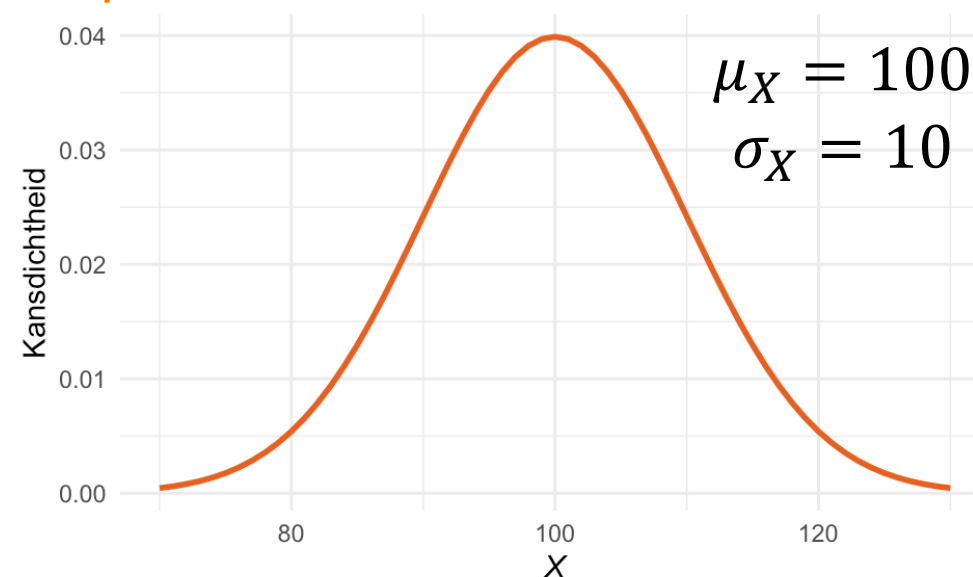
- de populatie veel groter is dan de steekproef
- de steekproef ongeveer eenvoudig aselekt is.

Cruciale vraag:

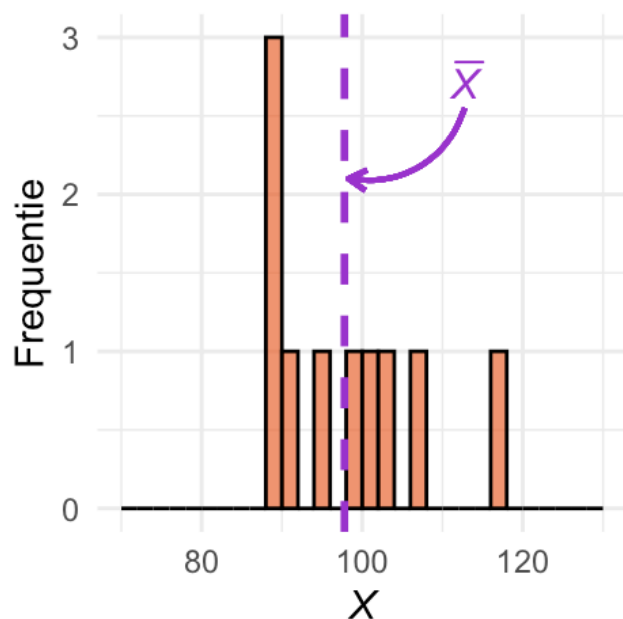
Hoe ver liggen gemiddelden van steekproeven typisch af van het ware gemiddelde van de populatie?

Simuleren!

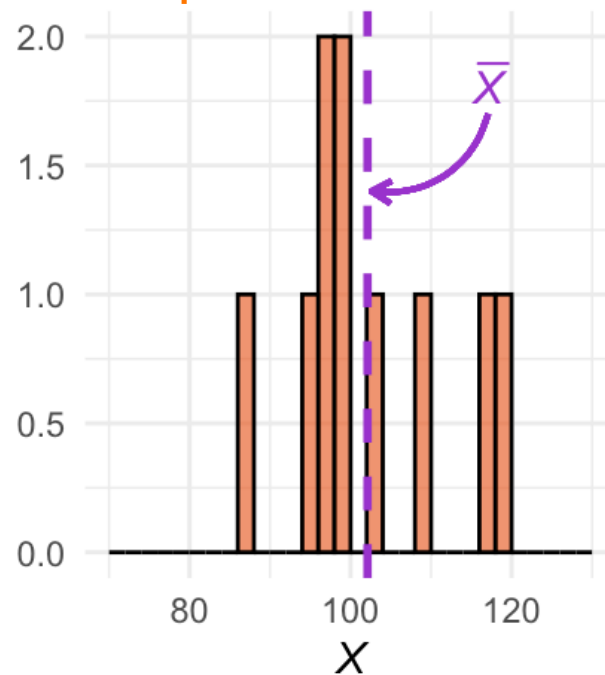
Populatie



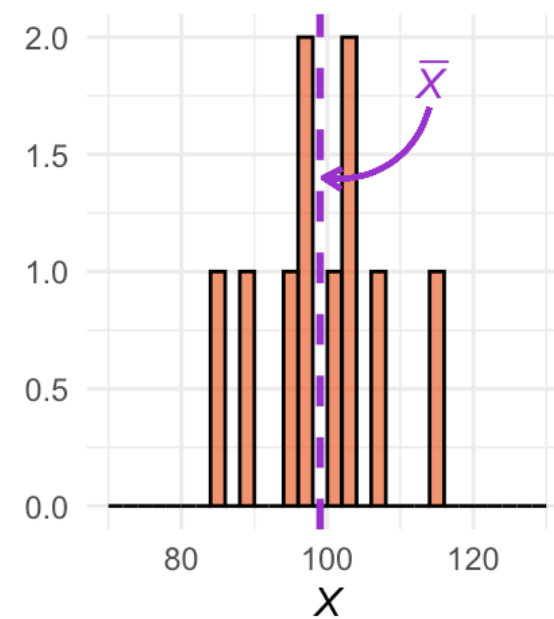
Steekproef 1



Steekproef 2



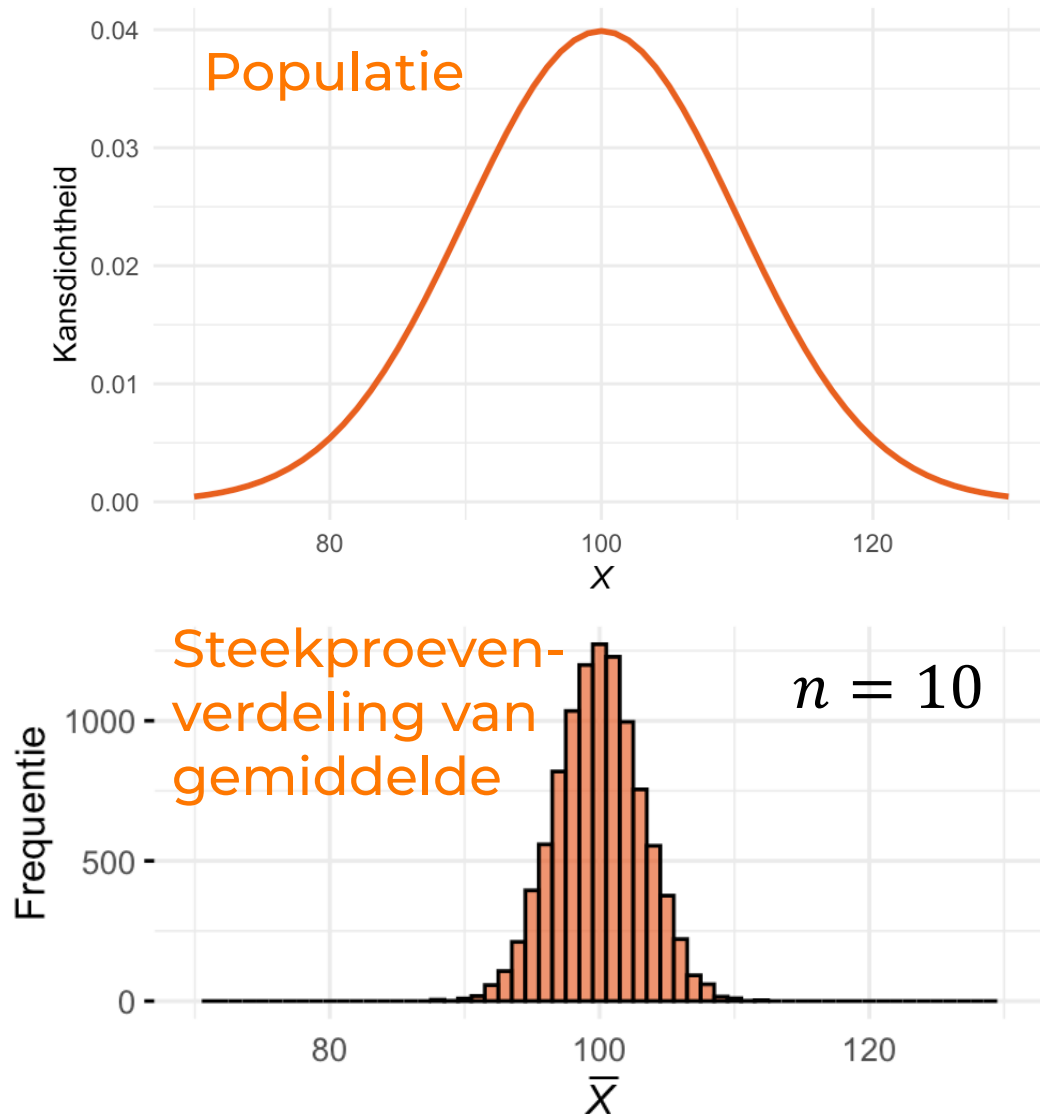
Steekproef 3



Steekproeven simuleren

https://onlinestatbook.com/stat_sim/sampling_dist/index.html

De steekproevenverdeling v.h. gemiddelde



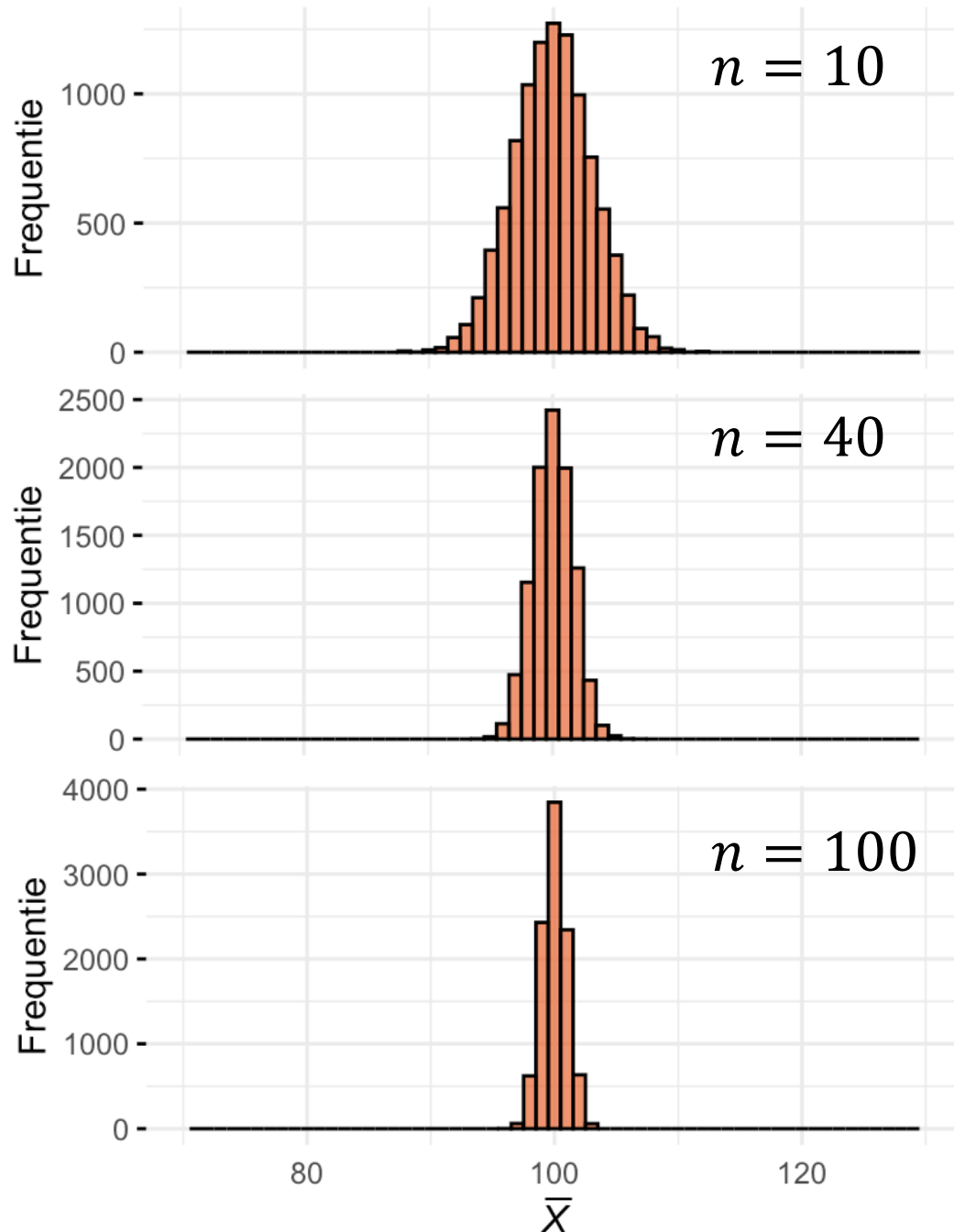
Opvallend:

- Gemiddeld zijn de schattingen juist:

$$\mathbb{E}[\bar{X}] = \mu_X$$

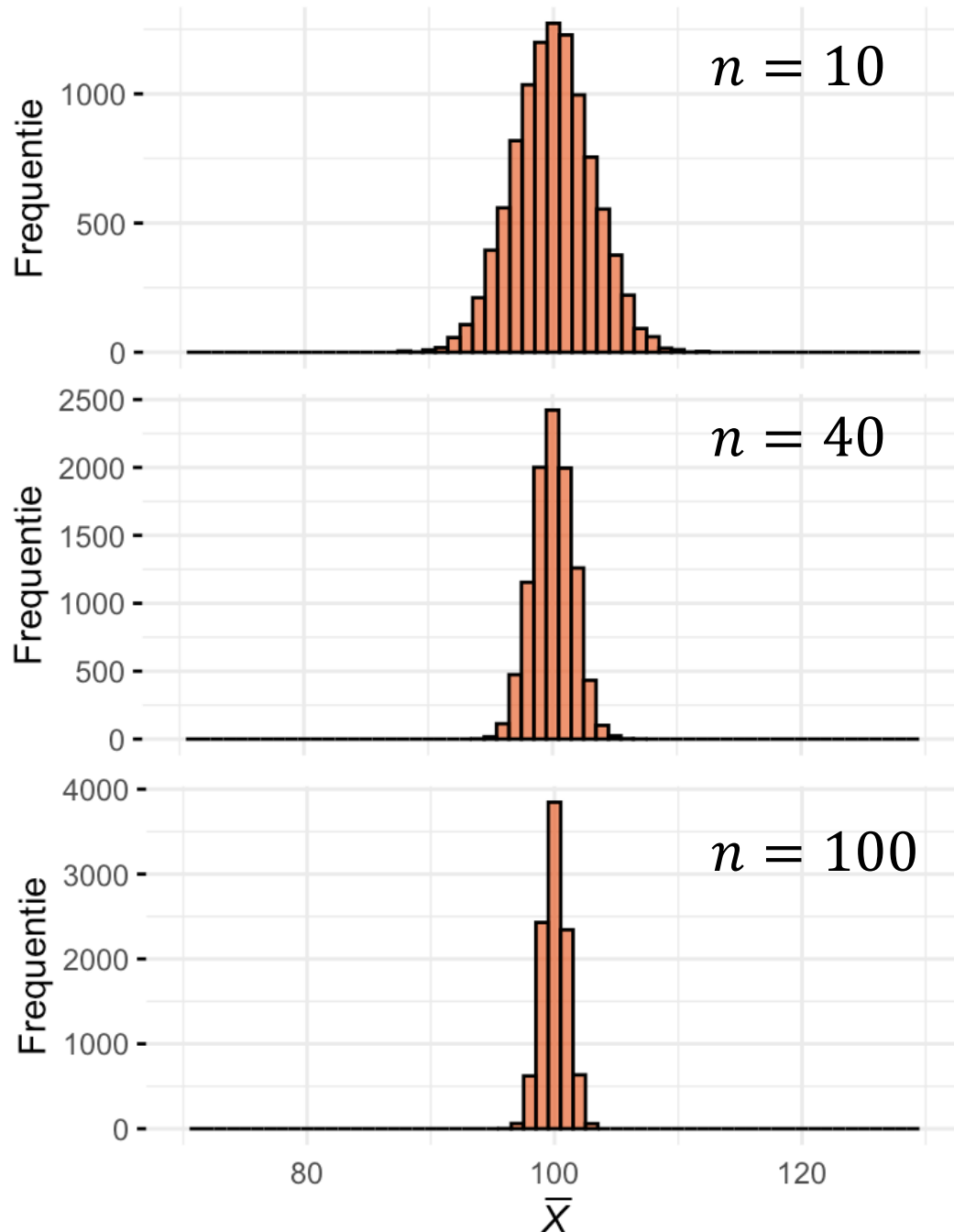
Dus: de schatter is zuiver.

- Als de populatie normaal verdeeld is, is $\bar{X}$ dat ook.
- Spreiding $\bar{X}$ is kleiner dan spreiding van X .



Spreading van de steekproevenverdeling neemt af met n

- Bij grotere steekproef, minder onderling verschil tussen de steekproefgemiddelden.
 - Typisch liggen de schattingen bij grotere steekproeven dus dichterbij de juiste waarde; zijn dus preciezer.
- Typische afwijking kwantificeren: standaarddeviatie van de steekproevenverdeling.



De standaarddeviatie van de steekproevenverdeling

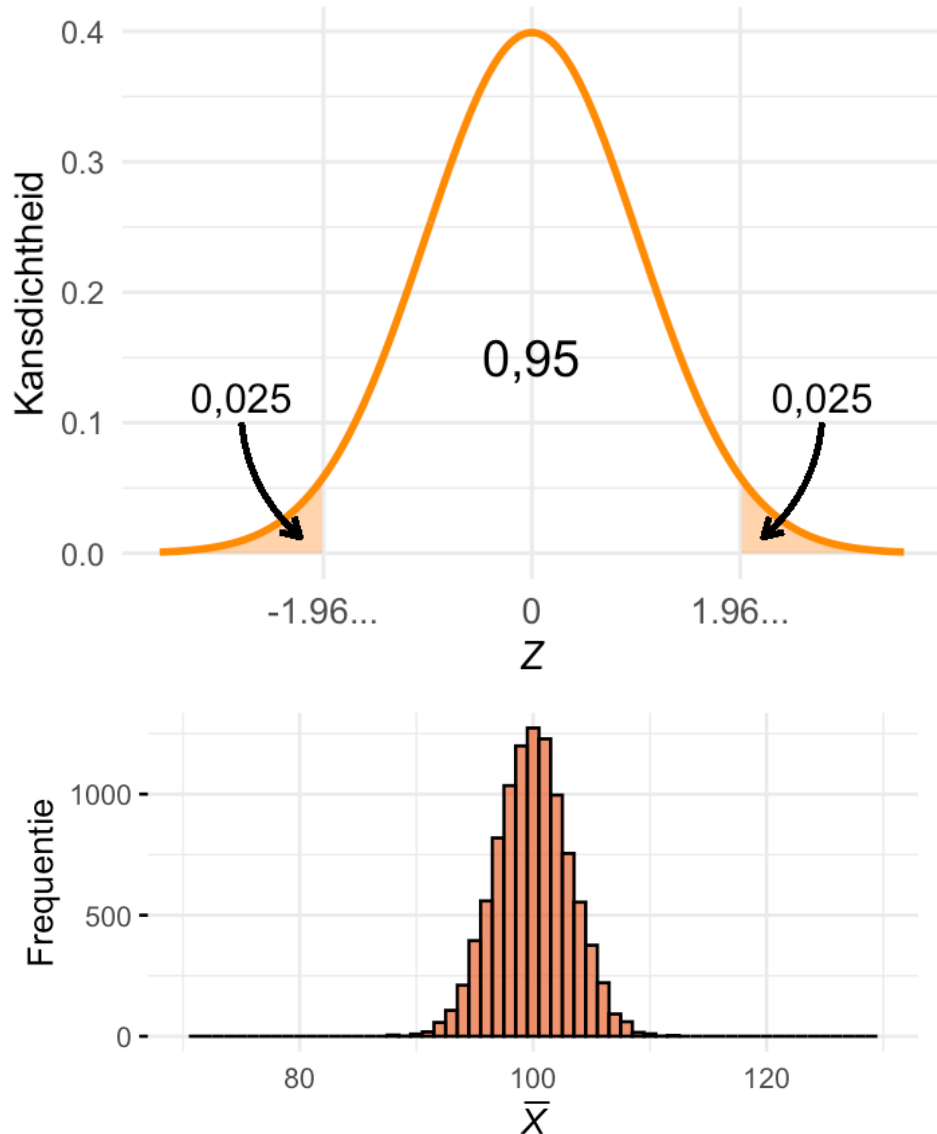
Wiskundig kan worden aangetoond:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}.$$

Dus we concluderen:

$\bar{X}$ is normaal verdeeld met gemiddelde μ_X en standaarddeviatie $\sigma_{\bar{X}} = \sigma_X / \sqrt{n}$.

De crux van een betrouwbaarheidsinterval



Voor iedere normale verdeling geldt:

95% van de waarnemingen ligt binnen 1,96 standaarddeviaties van het gemiddelde.

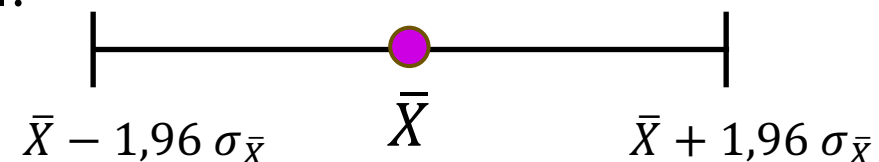
Voor de steekproevenverdeling geldt dus:

Bij 95% van de steekproeven ligt $\bar{X}$ binnen $1,96 \sigma_{\bar{X}}$ van de ware μ_X .

Andersom:

Bij 95% van de steekproeven ligt μ_X binnen $1,96 \sigma_{\bar{X}}$ van de gevonden $\bar{X}$.

Dus BHI:



Inhoud

1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

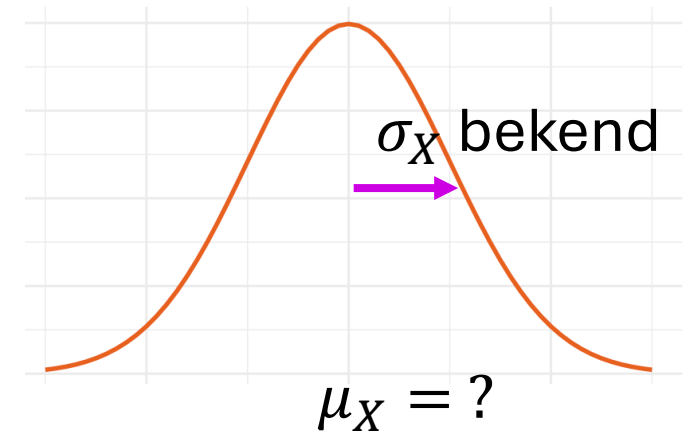
Situatie: μ_X onbekend, maar σ_X wel bekend

- We willen het populatiegemiddelde μ_X van variabele X schatten.
- We nemen een steekproef van grootte n .
- We nemen aan dat het Normale model van toepassing is
- Op een of andere manier is σ_X ons bekend.

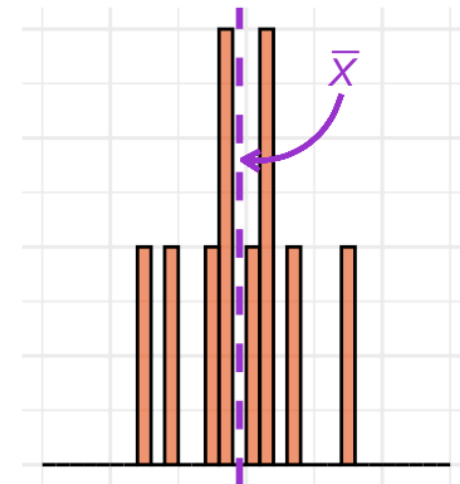
Opmerking:

- Deze situatie komt zelden voor.
- Is vooral een nuttige tussenstap.

Populatie



Steekproef, grootte n



Afleiding BHI:

- Onder het Normale Model geldt: $\bar{X}$ is normaal verdeeld met gemiddelde μ_X en standaarddeviatie $\sigma_{\bar{X}} = \sigma_X / \sqrt{n}$.
- Dan is $Z = \frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}}$ dus standaardnormaal verdeeld.

- Eigenschap standaard-normale verdeling:

$$\Pr[-1,96 < Z < 1,96] = 0,95$$

- Dus:

$$\Pr\left[-1,96 < \frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} < 1,96\right] = 0,95$$

- Omschrijven:

$$\Pr[\bar{X} - 1,96 \sigma_{\bar{X}} < \mu_X < \bar{X} + 1,96 \sigma_{\bar{X}}] = 0,95$$

- Dus 95%-BHI is:

$$[\bar{X} - 1,96 \sigma_{\bar{X}}, \bar{X} + 1,96 \sigma_{\bar{X}}]$$

- Andere notatie:

$$\mu_X = \bar{X} \pm 1,96 \sigma_{\bar{X}}$$

95%-BHI berekenen als σ bekend is



Je hebt in de kas van een bepaald aardappelras 9 planten opgekweekt en de aardappelopbrengst gemeten.

De gemiddelde opbrengst is $\bar{X} = 300\text{g}$. Je weet dat de standaarddeviatie van de opbrengst bij dit ras σ_X gelijk is aan 60 g. Bereken het 95%-BHI.

95%-BHI berekenen als σ bekend is



Je hebt in de kas van een bepaald aardappelras 9 planten opgekweekt en de aardappelopbrengst gemeten.

De gemiddelde opbrengst is $\bar{X} = 300\text{g}$. Je weet dat de standaarddeviatie van de opbrengst bij dit ras σ_X gelijk is aan 60 g. Bereken het 95%-BHI.

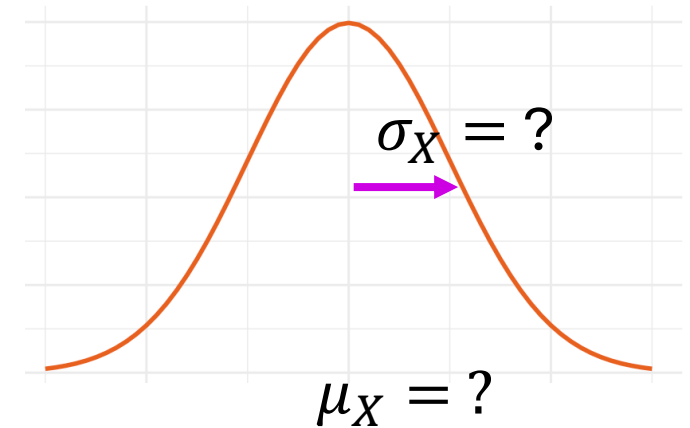
Inhoud

1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

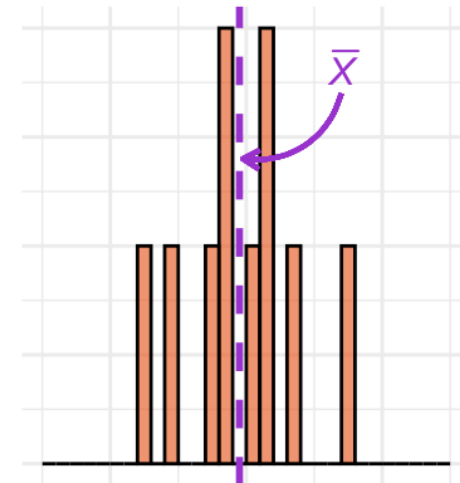
Situatie: μ_X en σ_X onbekend

- We willen het populatiegemiddelde μ_X van variabele X schatten.
- We nemen een steekproef van grootte n .
- We nemen aan dat het Normale model van toepassing is.
- σ_X is *niet* bekend.

Populatie



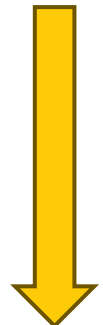
Steekproef, grootte n



Probleem: σ_X onbekend, dus $\sigma_{\bar{X}}$ onbekend

$$\mu_X = \bar{X} \pm 1,96 \sigma_{\bar{X}} = \bar{X} \pm 1,96 \cdot \frac{\sigma_X}{\sqrt{n}}$$


Oplossing:
gebruik s_X als schatter voor σ_X !

$$\mu_X = \bar{X} \pm (\text{andere kritieke waarde}) \cdot \frac{s_X}{\sqrt{n}}$$


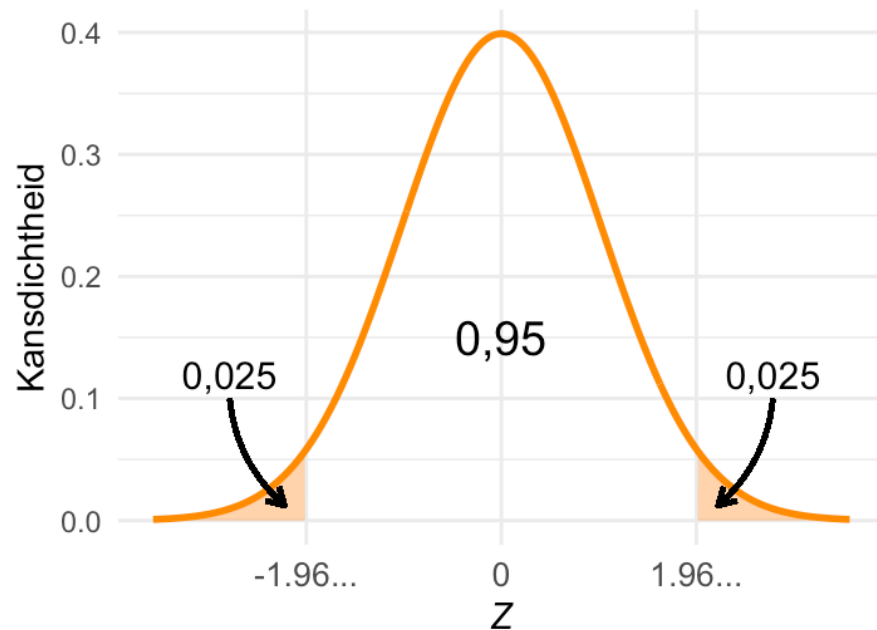
$SE_{\bar{X}}$



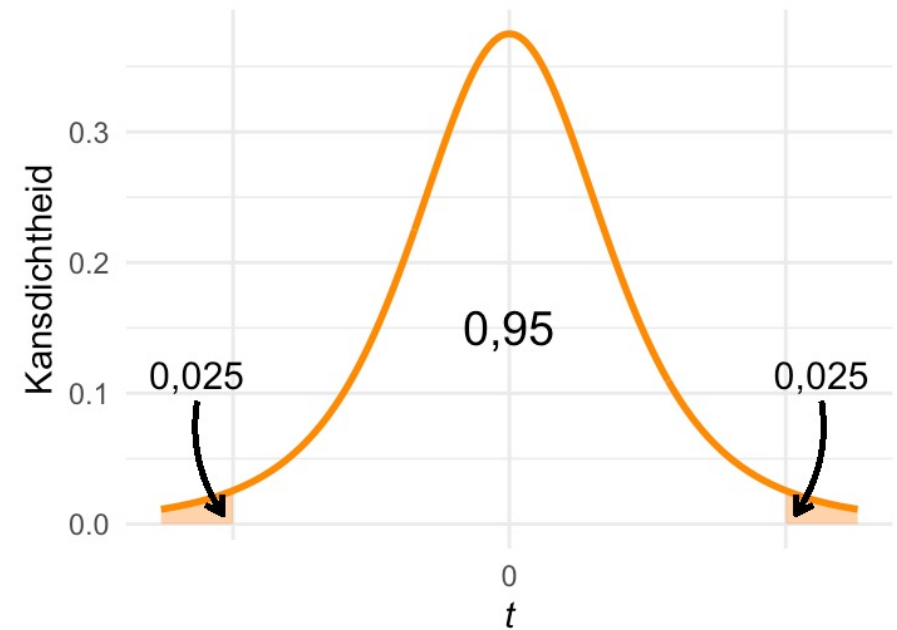
standaardfout van
het gemiddelde

Transformeren: niet Z maar t

$$Z = \frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} \longrightarrow \frac{\bar{X} - \mu_X}{SE_{\bar{X}}} = t$$



standaard-normaal



t -verdeling met $df = n - 1$.

Afleiding BHI:

- Onder het Normale Model geldt: $\bar{X}$ is normaal verdeeld met gemiddelde μ_X en standaarddeviatie $\sigma_{\bar{X}} = \sigma_X/\sqrt{n}$.
- Dan is de kansverdeling voor $t = \frac{\bar{X} - \mu_X}{SE_{\bar{X}}}$ een t -verdeling met $df = n - 1$.

- Eigenschap t -verdeling:

$$\Pr[-t_{0,05(2)df} < t < t_{0,05(2)df}] = 0,95$$

- Dus:

$$\Pr\left[-t_{0,05(2)n-1} < \frac{\bar{X} - \mu_X}{SE_{\bar{X}}} < t_{0,05(2)n-1}\right] = 0,95$$

- Omschrijven:

$$\Pr[\bar{X} - t_{0,05(2)n-1}SE_{\bar{X}} < \mu_X < \bar{X} + t_{0,05(2)n-1}SE_{\bar{X}}] = 0,95$$

- Dus 95%-BHI is:

$$[\bar{X} - t_{0,05(2)n-1}SE_{\bar{X}}, \bar{X} + t_{0,05(2)n-1}SE_{\bar{X}}]$$

- Andere notatie:

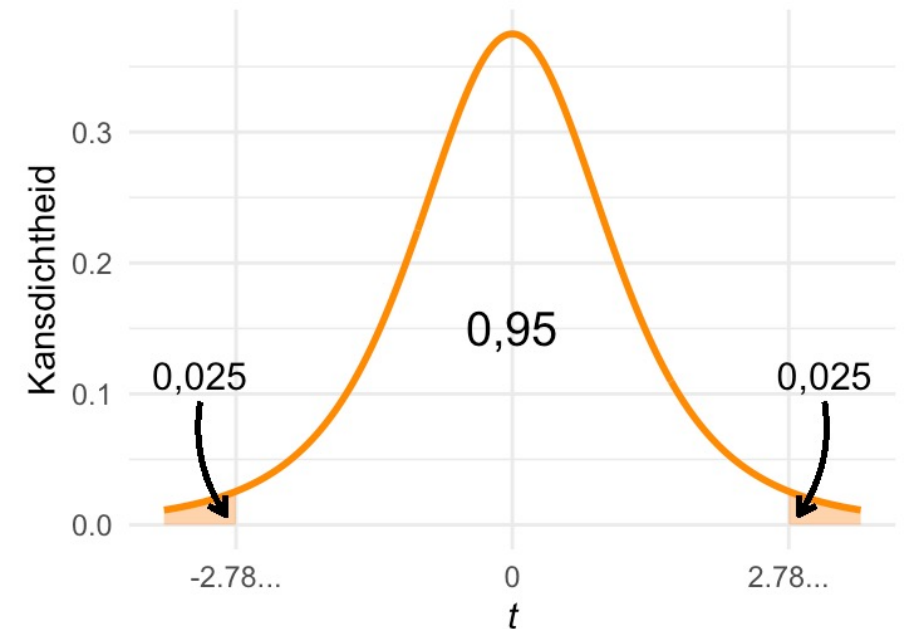
$$\mu_X = \bar{X} \pm t_{0,05(2)n-1}SE_{\bar{X}}$$

Dus hoe gaat de berekening?

1. Bereken gemiddelde $\bar{X}$ en standaarddeviatie s_X van de steekproef.
2. Bereken $SE_{\bar{X}} = \frac{s}{\sqrt{n}}$.
3. Bereken aantal vrijheidsgraden: $df = n - 1$.
4. Vraag de kritieke waarde $t_{0,05(2)n-1}$ aan R:

```
qt(0.975, df = n - 1)
```

5. Vul in: $[\bar{X} - t_{0,05(2)n-1}SE_{\bar{X}}, \bar{X} + t_{0,05(2)n-1}SE_{\bar{X}}]$

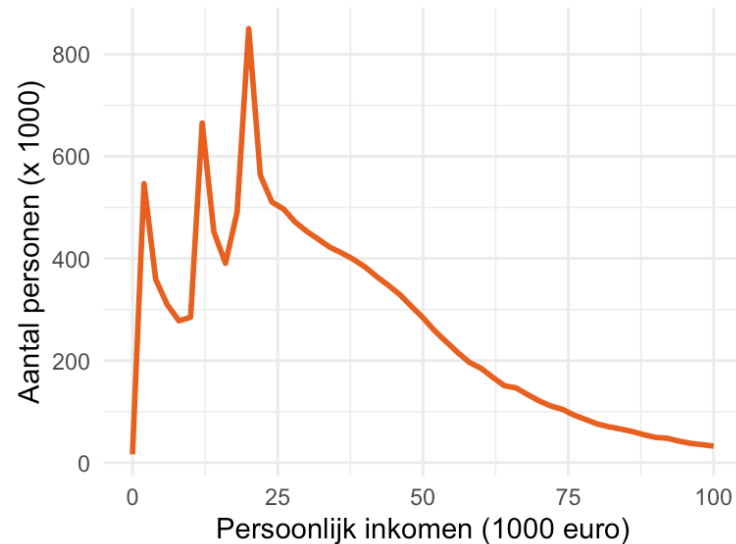


Inhoud

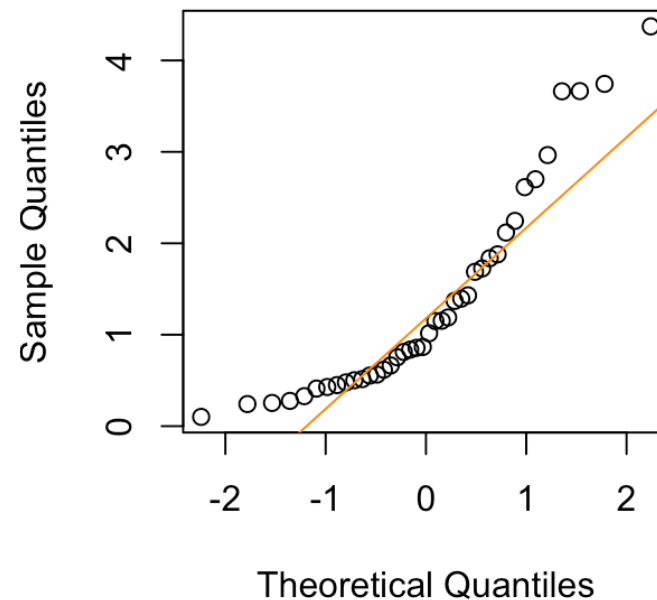
1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

Kunnen we het Normale Model gebruiken?

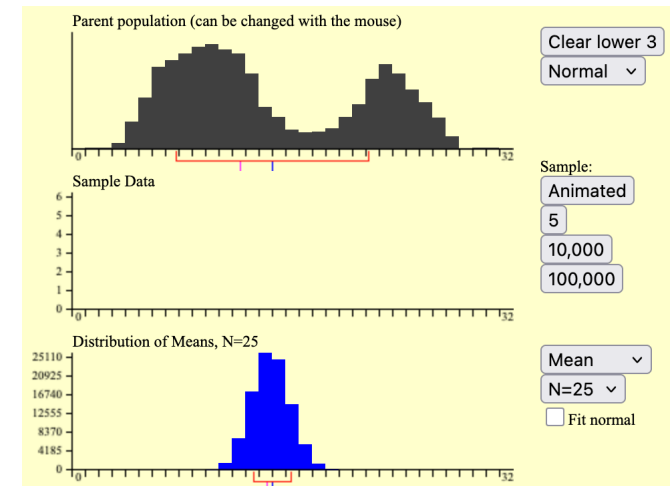
Denk na:
is het bij voorbaat
Normale Model redelijk
voor deze situatie?



Bekijk de gegevens:
sluiten die aan bij het
normale model?



Als je afwijkingen ziet:
bedenk of het
Normale model
robuust zou moeten
zijn voor zulke
afwijkingen.



Inhoud

1. Betrouwbaarheidsintervallen
2. Steekproeven en de steekproevenverdeling van het gemiddelde
3. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling weten
4. Het 95%-BHI voor als we de standaarddeviatie van de populatieverdeling niet weten
5. Controleren of het Normale Model van toepassing is
6. Robuustheid van het Normale Model

Demonstratie van de Centrale Limietstelling

https://onlinestatbook.com/stat_sim/sampling_dist/index.html

Richtlijnen:

Bij steekproeven van $n = 30$ of meer de formule voor het 95%-BHI ook van toepassing is als de populatie niet (helemaal) normaal verdeeld is.

Bij dramatische afwijkingen of sterke uitbijters:
zoek een andere methode.