

5 – Toetsingsprocedures

De Zz-toets

De **zz**-toets probeert een vermoeden over het *populatiegemiddelde* te weerleggen of bevestigen aan de hand van een voldoende grote steekproef.

De rechtszijdige Zz-toets

Algemeen wordt aangenomen dat superhelden gemiddeld $\mu=3.3$ mensen per dag redden. Onderzoekers willen dit controleren en voeren een steekproef uit bij 30 superhelden. In deze steekproef is het gemiddelde $\bar{x}=3.483$. We kunnen dus vermoeden dat superhelden gemiddeld *meer* redden uitvoeren.

De standaardafwijking in de populatie is verondersteld gekend en is $\sigma=0.55$.

De toetsingsprocedure loopt dan als volgt:

1. Formuleer de hypothesen:
 - $H_0: \mu=3.3$
 - $H_1: \mu>3.3$
2. Kies significantieniveau, bv. $\alpha=0.05$

```
3. # We hebben een steekproef met
4. n <- 30      # steekproefgrootte
5. sm <- 3.483  # steekproefgemiddelde
6. s <- 0.55    # standaardafwijking (verondersteld gekend)
7. a <- 0.05    # significantieniveau (gekozen door de onderzoeker)
```

```
m0 <- 3.3      # hypothetisch populatiegemiddelde (H0)
```

8. Bepaal de waarde van de toetsingsgrootte, in dit geval $\bar{x}=3.483$
9. Bepaal de overschrijdingskans p en verwerp H_0 als $p<\alpha$.

```
10. p <- 1 - pnorm(sm, m0, s/sqrt(n))
```

```
p
## [1] 0.03419546
if(p < a) {
  print("H0 verwerpen")
} else {
  print("H0 niet verwerpen")
}
## [1] "H0 verwerpen"
```

Een alternatieve berekening is het bepalen van het kritieke gebied, d.w.z. het gebied waarbinnen H_0 verworpen mag worden. De grens van dat gebied wordt de kritieke

grenswaarde **g** genoemd. Links ervan kan je **H₀** niet verwerpen (aanvaardingsgebied), rechts ervan wel (kritieke gebied). De oppervlakte van het aanvaardingsgebied is **1- α** , de oppervlakte van het kritieke gebied **α** .

```
g <- m0 + qnorm(1-a) * s / sqrt(n)
g
## [1] 3.465169

# Als het gevonden steekproefgemiddelde onder g ligt, kan je H0 niet verwerpen
if (sm < g) {
  print("H0 niet verwerpen")
} else {
  print("H0 verwerpen")
}
## [1] "H0 verwerpen"
```

We besluiten dus dat als we uitgaan van **H₀**, de kans dat we uit de populatie een steekproef trekken met het bekomen steekproefgemiddelde zeer klein is. We kunnen met het gekozen significantieniveau dan ook de nulhypothese verwerpen. Merk op dat beide methoden, de overschrijdingskans en de kritieke grenswaarde equivalent zijn en dus altijd identieke resultaten uitkomen!

Hieronder volgt een plot van deze casus.

```
# grenzen van de plot (x-waarden)
x <- seq(m0-4*s/sqrt(n), m0+4*s/sqrt(n), length=200)
# y-waarden (volgen de Gauss-curve)
dist <- dnorm(x, m0, s/sqrt(n))
plot (x, dist, type = 'l', xlab = '', ylab = '')

# Toon het gevonden steekproefgemiddelde ahv rode verticale lijn
abline(v=sm, col='red')
text(sm, 2, sm)

# Het aanvaardingsgebied plotten
i <- x <= g # Waarden van x links van g
polygon( # Plot deze waarden op de grafiek
  c(x[i], g, g),
  c(dist[i], dnorm(g, m0, s/sqrt(n)), 0),
  col = 'lightgreen')
text(g, .5, signif(g, digits=4)) # Toon grenswaarde
```

```
text(m0, 0.1, m0)           # Hypothetisch populatiegemiddelde
abline(v=m0)                # Trek daar een verticale lijn

text(m0, 1.5, 'aanvaardingsgebied (H0)')
```

De linkszijdige Zz-toets

Stel dat het resultaat van de steekproef $\bar{x} = 3.117$ was. Dan hebben we redenen om aan te nemen dat het populatiegemiddelde *kleiner* is dan 3.3. Dit kunnen we verifiëren met een linksszijdige zz-toets.

- Formuleer de hypothesen:
 - $H_0: \mu = 3.3$
 - $H_1: \mu < 3.3$
- Kies significantieniveau, bv. $\alpha = 0.05$

```
3. # We hebben een steekproef met
4. n <- 30      # steekproefgrootte
5. sm <- 3.117  # steekproefgemiddelde
6. s <- 0.55    # standaardafwijking (verondersteld gekend)
7. a <- 0.05    # significantieniveau (gekozen door de onderzoeker)
```

```
m0 <- 3.3      # hypothetisch populatiegemiddelde (H0)
```

- Bepaal de waarde van de toetsingsgrootte, in dit geval $\bar{x} = 3.117$
- Bepaal de overschrijdingskans p en verwerp H_0 als $p < \alpha$.

```
10. p <- pnorm(sm, m0, s/sqrt(n)) # Let op! geen 1 - pnorm(...). Waaro
    m?
```

```
p
## [1] 0.03419546
if(p < a) {
  print("H0 verwerpen")
} else {
  print("H0 niet verwerpen")
}
## [1] "H0 verwerpen"
```

De kritieke grenswaarde g ligt nu links van het gemiddelde. Rechts ervan kan je H_0 niet verwerpen (aanvaardingsgebied), links ervan wel (kritieke gebied).

```
g <- m0 - qnorm(1-a) * s / sqrt(n)
g
```

```
## [1] 3.134831

# Als het gevonden steekproefgemiddelde boven g ligt, kan je H0 niet
verwerpen

if (sm < g) {
  print("H0 verwerpen")
} else {
  print("H0 niet verwerpen")
}

## [1] "H0 verwerpen"
```

We besluiten dus dat als we uitgaan van H_0 , de kans dat we uit de populatie een steekproef trekken met het bekomen steekproefgemiddelde zeer klein is. We kunnen met het gekozen significantieniveau dan ook de nulhypothese verwerpen.
De plot van dit geval.

```
# grenzen van de plot (x-waarden)
x <- seq(m0-4*s/sqrt(n), m0+4*s/sqrt(n), length=200)
# y-waarden (volgen de Gauss-curve)
dist <- dnorm(x, m0, s/sqrt(n))
plot (x, dist, type = 'l', xlab = '', ylab = '')

# Toon het gevonden steekproefgemiddelde ahv rode verticale lijn
abline(v=sm, col='red')
text(sm, 2, sm)

# Het aanvaardingsgebied plotten
i <- x >= g                                # Waarden van x rechts van g
polygon(                                     # Plot deze waarden op de grafiek
  c(x[i],      g),
  c(dist[i], 0),
  col = 'lightgreen')
text(g, .5, signif(g, digits=4)) # Toon grenswaarde

text(m0, 0.1, m0)                          # Hypothetisch populatiegemiddelde
abline(v=m0)                               # Trek daar een verticale lijn

text(m0, 1.5, 'aanvaardingsgebied (H0)')
```

De tweezijdige Zz-toets

Hier willen we niet meteen testen of het populatiegemiddelde kleiner dan wel groter is dan een veronderstelde waarde, maar wel of een bekomen steekproefgemiddelde er mee in lijn ligt.

1. Formuleer de hypothesen:
 - $H_0: \mu = 3.3$
 - $H_1: \mu \neq 3.3$
2. Kies significantieniveau, bv. $\alpha = 0.05$

```
3. # We hebben een steekproef met
4. n <- 30      # steekproefgrootte
5. sm <- 3.483  # steekproefgemiddelde
6. s <- 0.55    # standaardafwijking (verondersteld gekend)
7. a <- 0.05    # significantieniveau (gekozen door de onderzoeker)
```

```
m0 <- 3.3      # hypothetisch populatiegemiddelde (H0)
```

8. Bepaal de waarde van de toetsingsgrootte, in dit geval $x^- = 3.483$
9. Bepaal de overschrijdingskans p en verwerp H_0 als $p < \alpha/2$ (waarom delen door 2?).

```
10. p <- pnorm(sm, m0, s/sqrt(n))
```

```
p
## [1] 0.9658045
if(p < a/2) {
  print("H0 verwerpen")
} else {
  print("H0 niet verwerpen")
}
## [1] "H0 niet verwerpen"
```

Nu zijn er twee kritieke grenswaarden: g_1 ligt nu links van het gemiddelde en g_2 rechts. Het kritieke gebied vertegenwoordigt nog altijd een oppervlakte van α .

```
g1 <- m0 - qnorm(1-a/2) * s / sqrt(n) # ondergrens aanvaardingsgebied
g2 <- m0 + qnorm(1-a/2) * s / sqrt(n) # bovengrens aanvaardingsgebied
c(g1, g2)
## [1] 3.103189 3.496811
# Als het gevonden steekproefgemiddelde tussen g1 en g2 ligt, kan je
# H0 niet verwerpen
if (g1 < sm & sm < g2) {
  print("H0 niet verwerpen")
}
```

```

} else {
  print("H0 verwerpen")
}
## [1] "H0 niet verwerpen"

```

De plot van deze situatie ziet er zo uit:

```

# grenzen van de plot (x-waarden)
x <- seq(m0-4*s/sqrt(n), m0+4*s/sqrt(n), length=200)
# y-waarden (volgen de Gauss-curve)
dist <- dnorm(x, m0, s/sqrt(n))
plot (x, dist, type = 'l', xlab = '', ylab = '')

# Het aanvaardingsgebied plotten
i <- g1 <= x & x <= g2          # Waarden van x tussen g1 en g2
dg1 <- dnorm(g1, m0, s/sqrt(n)) # kansdichtheid op g1 en g2
dg2 <- dnorm(g2, m0, s/sqrt(n))
polygon(                        # Plot deze waarden op de grafiek
  c(g1, g1, x[i], g2, g2),
  c(0, dg1, dist[i], dg2, 0),
  col = 'lightgreen')
text(g1, .5, signif(g1, digits=4)) # Toon grenswaarde
text(g2, .5, signif(g2, digits=4)) # Toon grenswaarde

text(m0, 0.1, m0)                # Hypothetisch populatiegemiddelde
abline(v=m0)                     # Trek daar een verticale lijn

text(m0, 1.5, 'aanvaardingsgebied (H0)')

# Toon het gevonden steekproefgemiddelde ahv rode verticale lijn
abline(v=sm, col='red')
text(sm, 2, sm)

```

Als we dus a priori geen uitspraak doen of het werkelijke populatiegemiddelde hetzij kleiner, hetzij groter is, dan blijkt het bekomen steekproefgemiddelde wel voldoende waarschijnlijk te zijn. We kunnen een niet-systematische steekproeffout niet uitsluiten. Of, met andere woorden, we kunnen hier de nulhypothese dus *niet* verwerpen.

De tt-toets

Als de veronderstellingen die gemaakt werden om de **zz**-toets te mogen toepassen *niet* gelden (bv. te kleine steekproef en/of de standaardafwijking van de populatie is onbekend), dan kan de **tt**-toets een alternatief zijn. Er zijn opnieuw drie varianten van deze toets: linkszijdig, rechtszijdig en tweezijdig. We werken hier enkel het geval voor de rechtszijdige toets uit.

Rechtszijdige tt-toets

Stel dat de onderzoekers van de superhelden door tijdsdruk niet in staat waren om een voldoende grote steekproef te nemen en slechts **n=20** $n=20$ observaties gedaan hebben, met hetzelfde steekproefgemiddelde **$\bar{x}=3.483$** $\bar{x}=3.483$. De standaardafwijking in deze steekproef bleek **s=0.55** $s=0.55$.

Kunnen we in deze omstandigheden, met eenzelfde significantieniveau **$\alpha=0.05$** $\alpha=0.05$, het besluit dat superhelden dagelijks *meer* dan 3.3 mensen redden aanhouden?

De toetsingsprocedure loopt dan als volgt:

1. Formuleer de hypothesen:
 - o **$H_0:\mu=3.3$** $H_0:\mu=3.3$
 - o **$H_1:\mu>3.3$** $H_1:\mu>3.3$
2. Kies significantieniveau, bv. **$\alpha=0.05$** $\alpha=0.05$

```
3. # We hebben een steekproef met
4. n <- 20      # steekproefgrootte
5. sm <- 3.483  # gemiddelde van de steekproef
6. ss <- 0.55   # standaardafwijking van de steekproef
7. a <- 0.05    # significantieniveau (gekozen door de onderzoeker)
```

```
m0 <- 3.3      # hypothetisch populatiegemiddelde (H0)
```

8. Bepaal de waarde van de toetsingsgrootte, in dit geval **$\bar{x}=3.483$** $\bar{x}=3.483$
9. Bepaal de overschrijdingskans **p** en verwerp **H_0** als **$p<\alpha$** $p<\alpha$. Merk op dat, omdat de functie `pt` niet toelaat gemiddelde en standaardafwijking mee te geven, we hier de waarden zelf moeten normaliseren.

```
10. p <- 1 - pt((sm - m0) / (s/sqrt(n)), df = n-1)
```

```
p
## [1] 0.07658003
if(p < a) {
  print("H0 verwerpen")
} else {
  print("H0 niet verwerpen")
}
## [1] "H0 niet verwerpen"
```

Bij de berekening van de kritieke grenswaarde gebruiken we de `qt`-functie met **$n-1$** $n-1$ vrijheidsgraden.

```
g <- m0 + qt(1-a, df = n-1) * s / sqrt(n)
```

```

g
## [1] 3.512655

# Als het gevonden steekproefgemiddelde onder g ligt, kan je H0 niet
verwerpen

if (sm < g) {
  print("H0 niet verwerpen")
} else {
  print("H0 verwerpen")
}

## [1] "H0 niet verwerpen"

```

De plot voor deze situatie:

```

# grenzen van de plot (x-waarden)
x <- seq(m0-4*ss/sqrt(n), m0+4*ss/sqrt(n), length=200)
# y-waarden (volgen de Gauss-curve van de t-verdeling)
dist <- dt((x-m0)/(ss/sqrt(n)), df = n-1) * ss/sqrt(n)
plot(x, dist, type = 'l', xlab = '', ylab = '')

# Het aanvaardingsgebied plotten
# Waarden van x links van g
i <- x < g
# waarde dichtheidsfunctie voor g
dg <- dt((g-m0)/(ss/sqrt(n)), df=n-1) * ss/sqrt(n)
# Plot deze waarden op de grafiek
polygon(
  c(x[i],      g, g),
  c(dist[i], dg, 0),
  col = 'lightgreen')
text(m0, 0.02, 'aanvaardingsgebied (H0)')

# Trek een verticale lijn bij het hypothetische populatiegemiddelde
text(m0, 0.01, m0)
abline(v=m0)

# Toon de kritieke grenswaarde
text(g+.025,.005, signif(g, digits=4))

# Toon het gevonden steekproefgemiddelde ahv rode verticale lijn

```



```
abline(v=sm, col='red')
text(sm-.025, .04, sm, col = 'red')
```

Met andere woorden, ook al krijgen we gelijkaardige resultaten in onze steekproef, kunnen we niet hetzelfde besluit trekken. Omdat onze steekproef te klein is, is er grotere onzekerheid of de waarde van het steekproefgemiddelde extreem genoeg is om de nulhypothese te verwerpen.

Gebruik van de functie `t.test`

In R bestaat er een functie `t.test` die je kan gebruiken om de toets uit te voeren op een verzameling van observaties.

```
test_sample <- c(                                # de steekproef
  3, 2, 3, 1, 10, 4, 2, 7, 3, 0,
  3, 1, 2, 3, 4, 0, 3, 8, 3, 7)
a <- 0.05                                         # significantieniveau
m0 <- 3.3                                         # hypothetisch populatiegemiddelde (H0)

t.test(test_sample,
  alternative = "greater", # rechtszijdige toets
  mu = m0,               # hypothese over gemiddelde
  conf.level = 1 - a)     # betrouwbaarheidsniveau

##
## One Sample t-test
##
## data:  test_sample
## t = 0.25359, df = 19, p-value = 0.4013
## alternative hypothesis: true mean is greater than 3.3
## 95 percent confidence interval:
##  2.427224      Inf
## sample estimates:
## mean of x
##      3.45
```

Het besluit hier is dat het steekproefgemiddelde van 3.45 hier nog geen aanleiding is voor het verwerpen van de nulhypothese omdat de overschrijdingskans **p=0.4013** $p=0.4013$ groter is dan het significantieniveau **$\alpha=0.05$** $\alpha=0.05$.

De `tt`-toets voor twee onafhankelijke steekproeven

De `tt`-toets kan ook gebruikt worden om twee steekproeven onderling te vergelijken. Eerst tonen we het geval van *onafhankelijke* steekproeven die afzonderlijk genomen zijn.

In een klinisch onderzoek wil men nagaan of een nieuw medicijn als bijwerking een verminderde reactiesnelheid heeft (Lindquist, g.d.). Zes deelnemers kregen een medicijn toegekend (interventiegroep) en zes anderen een placebo (controlegroep). Vervolgens werd hun reactietijd op een stimulus gemeten (in ms). We willen nagaan of er significante verschillen zijn tussen de interventie- en controlegroep.

- Controlegroep: 91, 87, 99, 77, 88, 91
- Interventiegroep: 101, 110, 103, 93, 99, 104

We noteren μ_1 voor het populatiegemiddelde van de patiënten die het medicijn nemen en μ_2 voor het gemiddelde van de niet behandelde populatie. Als het medicijn een impact heeft op de reactietijd, dan zal die van de controlegroep *lager* zijn van die van de interventiegroep.

```
control <- c( 91, 87, 99, 77, 88, 91)
intervention <- c(101, 110, 103, 93, 99, 104)
t.test(control, intervention, alternative="less")

##
##  Welch Two Sample t-test
##
## data:  control and intervention
## t = -3.4456, df = 9.4797, p-value = 0.003391
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -6.044949
## sample estimates:
## mean of x mean of y
##  88.83333 101.66667
```

De gemiddelde reactietijd van de controlegroep (88.83) blijkt inderdaad significant lager te zijn dan die van de interventiegroep (101.67).

Merk op dat in de oproep van `t.test` het significantieniveau (5%) en het veronderstelde populatiegemiddelde ($\mu_1 - \mu_2 = 0$) niet meegegeven werden. De waarden die we voor deze parameters willen kiezen zijn net de standaardwaarden van `t.test`.

In de functie `t.test` kan je ook de “group by” operator (`~`) gebruiken. Bijvoorbeeld, in de dataset `mtcars` kan je de vraag stellen “Zijn auto’s met een manuele versnellingsbak zuiniger dan die met een automaat?”

```
t.test(mpg ~ am, data = mtcars, alternative = 'less')

##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.0006868
## alternative hypothesis: true difference in means is less than 0
```

```
## 95 percent confidence interval:
##      -Inf -3.913256
## sample estimates:
## mean in group 0 mean in group 1
##      17.14737      24.39231
```

De **pp**-waarde is zeer klein, dus zelfs met een significantieniveau van 99% zou je de nulhypothese kunnen verwerpen.

De tt-toets voor gepaarde steekproeven

In deze variant van de **tt**-toets is een meting uitgevoerd op elk element van de steekproef, één keer vóór en één keer na een interventie. De bedoeling is om na te gaan of die interventie een significant effect gehad heeft op de meting.

In een studie werd nagegaan of auto's die rijden op benzine met additieven ook een lager verbruik hebben. Tien auto's werden eerst volgetankt met ofwel gewone benzine, ofwel benzine met additieven (bepaald door opgooien van een munt), waarna het verbruik werd gemeten (uitgedrukt in mijl per gallon). Vervolgens werden de auto's opnieuw volgetankt met de andere soort benzine en werd opnieuw het verbruik gemeten.

We gaan door middel van een gepaarde **tt**-toets na of auto's significant zuiniger rijden op benzine met additieven.

```
regular    <- c(16,20,21,22,23,22,27,25,27,28)
additives  <- c(19,22,24,24,25,25,26,26,28,32)
t.test(additives, regular, alternative="greater", paired=TRUE)

##
## Paired t-test
##
## data: additives and regular
## t = 4.4721, df = 9, p-value = 0.0007749
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.180207      Inf
## sample estimates:
## mean of the differences
##                        2
```

De optie `paired=TRUE` geeft aan dat het hier om een gepaarde t-toets gaat.

De **pp**-waarde, 0.0007749, ligt onder het significantieniveau 0.05, dus we kunnen de **nulhypothese verwerpen**. Volgens deze steekproef rijden auto's inderdaad zuiniger op benzine met additieven.