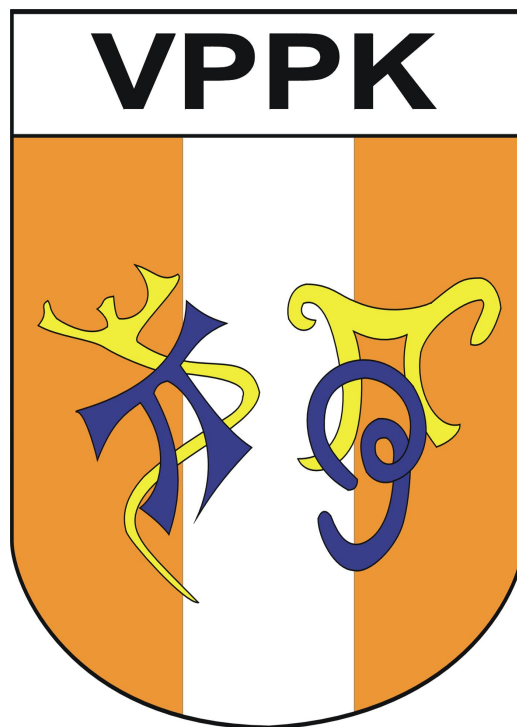


Statistiek II

Sessie 2

Verzamelde vragen en feedback

Deel 2



VPPK
Universiteit Gent
2017-2018

Feedback Oefensessie 2

We hebben besloten de bekomen grafieken in R niet in het document in te voegen, dit omdat het document met de code al vrij lang is, en de grafieken nog veel meer plaats zouden in beginnen nemen dat het onoverzichtelijk wordt, en ook niet meer echt milieuvriendelijk als er mensen dit willen printen. We gaan ervan uit dat jullie met het volgen van de code zelf de grafieken kunnen maken (aangezien ook de code van de grafieken in deze bundel staan).

1 Data manipulatie in R

- Een vector **naam** aanmaken:

```
> naam <- c("Ann", "Piet", "Bart")
> naam
[1] "Ann" "Piet" "Bart"
```

- Een vector **leeftijd** aanmaken:

```
> leeftijd <- c(8,11,7)
> leeftijd
[1] 8 11 7
```

- Een vector **geslacht** aanmaken:

```
> geslacht <- c("v", "m", "m")
> geslacht
[1] "v" "m" "m"
```

- Een vector **leerjaar** aanmaken - ordinaal meetniveau:

```
> leerjaar <- factor(c(3,5,2), ordered=TRUE)
> leerjaar
[1] 3 5 2
Levels: 2 < 3 < 5
```

- Een dataframe **kinderen** aanmaken met de 4 vorige vectoren:

```
> kinderen <- data.frame(naam, leeftijd, geslacht, leerjaar)
> kinderen
  naam leeftijd geslacht leerjaar
1 Ann         8        v         3
2 Piet        11        m         5
3 Bart         7        m         2
```

2 Data manipulatie in SPSS en Excel

Dit staat quasi letterlijk in de cursus.

3 Beschrijvende statistiek

```
> load("bloeddruk.RData")
```

1. Wanneer we de histogram maken via R dan zien we dat er een outlier is die onmogelijk is, aangezien de maximumwaarde 30 minuten is.

```
> hist(bloeddruk$slaperigheid2)
```

2. We zien dat de 38ste observatie de waarde 199 heeft. We veranderen deze dus naar 19. We zien dan op de histogram dat we nu allemaal plausible waarden hebben.

```
> bloeddruk$slaperigheid2[38]
[1] 199
> bloeddruk$slaperigheid2[38] <- 19
> hist(bloeddruk$slaperigheid2)
```

3. Met het commando `table` vragen we de frequentieverdeling op. Als we daarbij nog het commando `prop.table` bij voegen, dan krijgen we de relatieve frequentieverdeling.

```
> table(bloeddruk$dosis)

1  2  3  4  5
18 19 21 21 21
> prop.table(table(bloeddruk$dosis))

1    2    3    4    5
0.18 0.19 0.21 0.21 0.21
```

4. De bivariate frequentieverdeling vragen we ook op met het commando `table`, maar nu zetten we er 2 variabelen in. De eerste variabele wordt de rijen, de tweede de kolommen.

```
> table(bloeddruk$dosis, bloeddruk$slaperigheid, dnn=c("dosis", "slaperigheid"))
      slaperigheid
dosis  0  1  2  3  4  5
  1    5  6  5  1  1  0
  2    3  1  5  4  4  2
  3    5  3  2  3  3  5
  4    1  4  1  5  2  8
  5    1  0  0  4  3 13
```

5. Het spreidingsdiagram:

```
> plot(bloeddruk$dosis, bloeddruk$BD)
```

6. Op basis van de grafiek zouden we VERKEERDELIJK kunnen denken dat beide variabelen discreet zijn. Maar de onderliggende variabele van dosis is gewicht. We ronden deze gewoon telkens af naar de dichtsbijzijnde eenheid. In werkelijkheid is voor elke observatie de waarde van dosis een getal met oneindig veel cijfers na de komma. Dus we hebben hier een continue variabele. Meer bepaald van ratio-niveau, want we hebben een vast nulpunt. De bloeddruk kunnen we op dezelfde wijze gaan benaderen. Wanneer we bij de dokter onze bloeddruk laten meten krijgen we te horen dat we bijvoorbeeld een bloeddruk van 12 over 8 hebben. Dit is een afronding. De bloeddruk wordt gemeten in millimeters kwikdruk. Wanneer er geen kwikdruk is krijgen we 0. Dus we hebben terug een vast nulpunt. We hebben hier dus ook een variabele van ratio meetniveau.

7. Het staafdiagram van dosis geeft een grafische voorstelling weer van de bekomen frequentieverdeling uit oefening 3.

```
> barplot(table(bloeddruk$dosis))
```

8. Om het gemiddelde te bepalen van een variabele gebruiken we het commando `mean`. R kan dit enkel gebruiken bij numerieke variabelen. Aangezien `slaperigheid` als een factor staat geregistreerd in R krijgen we een foutmelding. Omdat het zinloos is om een gemiddelde van een ordinale variabele te berekenen

gaan we hier ook geen moeite doen om toch een gemiddelde te berekenen. Enkel voor `slaperigheid2` berekenen we dus het gemiddelde.

```
> mean(bloeddruk$slaperigheid2)
[1] 9.08
> mean(bloeddruk$slaperigheid)
[1] NA
Warning message:
In mean.default(bloeddruk$slaperigheid) :
  argument is not numeric or logical: returning NA
```

9. Voor `slaperigheid2` zien we dat we een plausibel antwoord krijgen:

```
> median(bloeddruk$slaperigheid2)
[1] 8
> barplot(table(bloeddruk$slaperigheid2))
```

Voor `slaperigheid` moeten we een andere redenering gaan toepassen.

```
> median(as.numeric(bloeddruk$slaperigheid))
[1] 4
> barplot(table(bloeddruk$slaperigheid))
```

Wanneer we een gefactoriseerde variabele, zoals `slaperigheid` gebruiken om een mediaan te berekenen gaat R deze omvormen. Het laagste niveau krijgt de waarde 1, het tweede laagste niveau de waarde 2,... Dit zorgt ervoor dat R niet de waarde geeft die overeenkomt met de dataset, maar met het niveau. Stel dat we onze niveaus gaan hernoemen als letters. Dan krijgen we als niveaus a, b, c, d, e, f. Ook hier zal R als mediaan dan 4 geven. Dit wil zeggen dat we het 4de niveau zoeken. In onze dataset geeft dit voor `slaperigheid` de waarde 3. Want 3 is ons 4de niveau. In een dataset met letters zou dit dan 'd' zijn.

```
> bloeddruk$slaperigheid
[1] 0 5 1 4 4 3 5 1 3 2 2 0 3 0 1 5 5 2 3 3 0 5 3 5 5 2 0 3 1 2 5 1 0 5 4 0 1 1 5 5
[71] 1 4 3 1 4 1 5 2 2 5 0 5 2 2 3 4 0 4 3 0 1 5 4 4 3 5 5 3 5 3
Levels: 0 1 2 3 4 5
> as.numeric(bloeddruk$slaperigheid)
[1] 1 6 2 5 5 4 6 2 4 3 3 1 4 1 2 6 6 3 4 4 1 6 4 6 6 3 1 4 2 3 6 2 1 6 5 1 2 2 6 6
[71] 2 5 4 2 5 2 6 3 3 6 1 6 3 3 4 5 1 5 4 1 2 6 5 5 4 6 6 4 6 4
```

10. Als we de formules volgen uit de cursus moeten we dus eerst het verschil nemen tussen elke observatie en het gemiddelde van de variabele. Deze verkregen waarden moeten we allemaal kwadrateren. De gekwadrateerde waarden gaan we dan allemaal sommeren, om uiteindelijk te delen door n . In R krijgen we dus volgende commando's (indien je dit op 1 regel wil zetten, let dan zeker op de haakjes, want een verkeerd haakje geeft een verkeerd resultaat):

```
> xbar <- mean(bloeddruk$dosis)
> n <- dim(bloeddruk)[1]
> sum((bloeddruk$dosis-xbar)^2)/n
[1] 1.9536
```

Als we de functie `var` gebruiken, dan moeten we rekening houden dat dit met de formule gedaan wordt waar gedeeld wordt door $n - 1$ en we dus moeten corrigeren:

```
> var(bloeddruk$dosis) * (n-1)/n
[1] 1.9536
```

11. De correlatiecoëfficiënt tussen `BD` en `dosis`:

```
> cor(bloeddruk$dosis, bloeddruk$BD)
[1] 0.07647263
```

12. De correlatiecoëfficiënt tussen 2 variabelen waar minstens 1 van de 2 variabelen ten hoogste van ordinaal niveau is, maakt gebruik van Kendall's Tau. Tussen 2 variabelen van minstens intervalniveau kunnen we lineaire trends detecteren. Maar zodra er 1 van de 2 niet meer minstens op intervalniveau zit moeten we monotone trends gaan onderzoeken. Daarom wordt voor de correlatie tussen `dosis` en `slaperigheid` gebruik gemaakt van Kendall's Tau

```
> cor(bloeddruk$dosis, as.numeric(bloeddruk$slaperigheid), method="kendall")
[1] 0.4195793
```

13. Aangezien zowel de variabele `dosis` als de variabele `slaperigheid2` van minstens intervalniveau zijn gaan we hier gebruik maken van de Pearson correlatiecoëfficiënt. Hiervoor moeten we geen extra informatie geven aan R.

```
> cor(bloeddruk$dosis, bloeddruk$slaperigheid2)
[1] 0.374037
```

4 Kansrekenen

1. Normaalverdeling

- $\mathcal{P}(N(100, 400) = 90) = 0$: Bij een continue verdeling is de kans dat we een exacte waarde observeren gelijk aan 0. Dit is niet met R te berekenen.
- $\mathcal{P}(N(100, 400) \leq 90) = 0.3085375$. In R wordt automatisch de rechterkant berekend van de grafiek. M.a.w. "kleiner of gelijk aan". Ook krijgen we in de geschreven opgave de variantie, maar in R moeten we de standaarddeviatie ingeven.

```
> pnorm(q=90, mean=100, sd=sqrt(400))
[1] 0.3085375
```

- $\mathcal{P}(N(100, 400) \geq 90) = 0.6914625$. Nu zoeken we de andere kant van de grafiek. Dit kunnen we in R op twee manieren bereiken. Ofwel geven we het argument `lower.tail = FALSE`, waardoor de rechterkant van de grafiek berekend wordt. Ofwel gaan we de vorige kans aftrekken van 1. Beide komen op hetzelfde neer.

```
> pnorm(q=90, mean=100, sd=sqrt(400), lower.tail = FALSE)
[1] 0.6914625
> 1-pnorm(q=90, mean=100, sd=sqrt(400))
[1] 0.6914625
```

- $\mathcal{P}(N(100, 400) \geq 86.5102) = 0.75$. Om de waarde te vinden waar 75% boven ligt in de grafiek gebruiken we het commando `qnorm`. Ook hier moeten we rekening houden met het feit dat R automatisch naar de linkerkant van de grafiek gaat. Dus ofwel vragen we de waarde op waar 25% onder ligt, ofwel geven we mee `lower.tail=FALSE`.

```
> qnorm(p=0.25, mean=100, sd=sqrt(400))
[1] 86.5102
> qnorm(p=0.75, mean=100, sd=sqrt(400), lower.tail = FALSE)
[1] 86.5102
```

- $\mathcal{P}(N(100, 400) \leq 113.4898) = 0.75$. Dezelfde redenering gebruiken we om de waarde te vinden waar 75% onder ligt.

```
> qnorm(p=0.75, mean=100, sd=sqrt(400))
[1] 113.4898
> qnorm(p=0.25, mean=100, sd=sqrt(400), lower.tail = FALSE)
[1] 113.4898
```

2. *t*-verdeling. Aangezien de redeneringen exact dezelfde zijn als bij de normaalverdeling geven we hier enkel de code mee, indien van toepassing. Namelijk dat R standaard de linkerkant van de figuur in rekening brengt.

- $\mathcal{P}(t_{70} = 1.2) = 0$
- $\mathcal{P}(t_{70} \leq 1.2) = 0.8829075$

```
> pt(q=1.2, df=70)
[1] 0.8829075
```
- $\mathcal{P}(t_{70} \geq 1.2) = 0.1170925$

```
> 1-pt(q=1.2, df=70)
[1] 0.1170925
> pt(q=1.2, df=70, lower.tail = FALSE)
[1] 0.1170925
```
- $\mathcal{P}(t_{70} \geq 0.526797) = 0.3$

```
> qt(p=0.7, df=70)
[1] 0.526797
```
- $\mathcal{P}(t_{70} \leq -0.526797) = 0.3$. In dit geval zien we de symmetrie van de *t*-verdeling. De waarde waar 30% onder ligt is de negatieve waarde van de waarde waar 30% boven ligt.

```
> qt(p=0.3, df=70)
[1] -0.526797
```

3. *F*-verdeling. Hier maken we gebruik van 2 vrijheidsgraden. De volgorde van deze vrijheidsgraden is van belang.

- $\mathcal{P}(F_{30,4} \geq 2) = 0.2641135$

```
> 1-pf(q=2, df1=30, df2=4)
[1] 0.2641135
> pf(q=2, df1=30, df2=4, lower.tail = FALSE)
[1] 0.2641135
```
- $\mathcal{P}(F_{30,4} \leq 0.3717987) = 0.05$

```
> qf(p=0.05, df1=30, df2=4)
[1] 0.3717987
```

5 Zuivere munt

1. Stel dat het muntstuk zuiver is, dan is de kans dat we munt gooien $1/2$. Dit is onze nulhypothese. Maar we vermoeden dat we meer munt krijgen. Dus de kans die we vermoeden is groter dan $1/2$. Dit is onze alternatieve hypothese. Deze is eenzijdig.

$$H_0 : \pi = 0.5 \text{ versus } H_a : \pi > 0.5$$

2. Om de power te berekenen in dit geval gaan we 3 stappen doorlopen.

- (a) Stel dat we vier keer opwerpen en drie keer munt uitkomen, dan is onze p -waarde de kans dat we 3 of 4 observeren. In de vorige bundel werd uitgelegd hoe we dit doen via de formule van de binomiaalverdeling. In R gebruiken we de functie `binom.test`. Voor deze functie moeten we verschillende zaken ingeven. Bij `n` geven we het aantal worpen in, dus in dit geval 4. Bij `x` geven we het aantal keer dat we munt observeren in, dus hier 3. Bij `p` geven we de kans van de nulhypothese in, dus hier 0.5. Bij `alternative` geven we in wat onze alternatieve hypothese zegt, hier verwachten we groter. Als laatste kunnen we de betrouwbaarheid specificeren bij `conf.level` - hier gebruiken we 0.95:

```
> binom.test(x=3, n=4, p=0.5, conf.level=0.95, alternative="greater")
```

```
Exact binomial test
```

```
data: 3 and 4
number of successes = 3, number of trials = 4, p-value = 0.3125
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.2486046 1.0000000
sample estimates:
probability of success
                0.75
```

Bij p -value krijgen we de bijhorende p -waarde dat we 3 of 4 keer munt zouden observeren. Deze is 0.3125, en aangezien deze groter is dan $\alpha = 0.05$ aanvaarden we de nulhypothese. Dezelfde conclusie kunnen we maken met het betrouwbaarheidsinterval. Dit interval kunnen we aflezen en is $[0.2486046 - 1]$. De waarde uit de nulhypothese, met name 0.5 bevindt zich in dit interval, en dus aanvaarden we de nulhypothese.

- (b) Wanneer we nu 4 keer munt observeren volgen we dezelfde redenering. Nu gebruiken we `x=4`:

```
Exact binomial test
```

```
data: 4 and 4
number of successes = 4, number of trials = 4, p-value = 0.0625
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.4728708 1.0000000
sample estimates:
probability of success
                1
```

Hier krijgen we een p -waarde van 0.0625 en dit is terug groter dan $\alpha = 0.05$. Dus we aanvaarden de nulhypothese.

- (c) Als we onze munt 4 keer opwerpen kunnen we nooit onze nulhypothese verwerpen. Onze p -waarde zal altijd groter zijn dan $\alpha = 0.05$.
- (d) Aangezien de power de kans is dat we de nulhypothese verwerpen, gegeven dat deze verkeerd is, is onze power 0. Want als we de nulhypothese altijd aanvaarden, dan kunnen we deze nooit verwerpen.

6 Zuivere munt 2

Het begin is hetzelfde als de vorige oefening, dus we gaan niet overal in detail terug op ingaan.

1. De p -waarde die we uitkomen als we 5 keer opgooien en 4 keer munt uitkomen is 0.1875. Aangezien deze groter is dan $\alpha = 0.05$ aanvaarden we de nulhypothese.


```
> binom.test(x=4, n=5, p=0.5, conf.level=0.95, alternative="greater")
```

```
Exact binomial test
```

```
data: 4 and 5
number of successes = 4, number of trials = 5, p-value = 0.1875
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.3425917 1.0000000
sample estimates:
probability of success
              0.8
```

2. Wanneer we daarentegen 5 keer munt observeren verwerpen we de nulhypothese. We krijgen een p -waarde van 0.03125 en deze is groter dan $\alpha = 0.05$.

```
> binom.test(x=5, n=5, p=0.5, conf.level=0.95, alternative="greater")
```

```
Exact binomial test
```

```
data: 5 and 5
number of successes = 5, number of trials = 5, p-value = 0.03125
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5492803 1.0000000
sample estimates:
probability of success
              1
```

3. Aangezien we nu de nulhypothese kunnen verwerpen, gaan we een power krijgen die groter is dan 0. We verwerpen de nulhypothese wanneer we 5 keer munt uitkomen. Dus gaan we op zoek naar de kans dat we 5 keer munt observeren onder de alternatieve hypothese.
4. We nemen als specifieke alternatieve hypothese $\pi = 0.55$. Dus voor de power onder deze assumptie zoeken we de kans dat we 5 keer munt observeren wanneer we 5 keer opgooien, gegeven dat onze nulhypothese verkeerd is (en dus de alternatieve correct). Dit kunnen we herschrijven als:

$$\mathcal{P}(B(5, 0.55) = 5)$$

In de tips zien we dat deze kans 0.05032844 is, dit is ineens onze power.

5. We kunnen dit ook berekenen via de functie `powerBinom`. Hier moeten we volgende argumenten ingeven: `n` = het aantal observaties (in dit geval 5), `p0` = de proportie onder de nulhypothese (hier 0.5), `p1` = de proportie onder de alternatieve hypothese (hier 0.55), en wat ons oorspronkelijke alternatieve hypothese was bij `alternative` (in dit geval `one.sided`). (Voor deze functie hebben we het package `exactci` nodig, vergeet dit dus niet te laden.)

```
> library(exactci)
Loading required package: ssanv
> powerBinom(n=5, p0=0.5, p1=0.55, sig.level=0.05, alternative="one.sided")
```

```
power and sample size for single binomial response
```

```
n = 5
```

```

p0 = 0.5
p1 = 0.55
power = 0.05032844
alternative = one.sided
sig.level = 0.05

```

NOTE: use rejections in correct direction only

We merken dat we dezelfde power verkrijgen als in het vorige puntje. Dus we hebben ongeveer 5% kans om de nulhypothese te verwerpen als deze verkeerd is. Dit is een zeer lage power. Onderzoek in de psychologie en de pedagogische wetenschappen proberen we altijd een betrouwbaarheid van ongeveer 80% te benaderen (dit is afhankelijk van het doel van het onderzoek - in het algemeen tussen de 70% en de 90%).

7 Slaperigheid2

We willen weten of het geobserveerde gemiddelde van onze data significant verschilt van de verwachting in de populatie, met name 7.4. Dus we stellen volgende hypothese op, met tweezijdige alternatieve hypothese:

$$H_0 : \mu = 7.4 \text{ versus } H_a : \mu \neq 7.4$$

We hebben hier 1 steekproef met onbekende verwachting. Dus we gaan een t -test voor 1 verwachting uitvoeren:

```

> t.test(bloeddruk$slaperigheid2, mu=7.4)

One Sample t-test

t = 2.2677, df = 99, p-value = 0.02552
alternative hypothesis: true mean is not equal to 7.4
95 percent confidence interval:
 7.610039 10.549961
sample estimates:
mean of x
 9.08

```

Wat kunnen we uit deze output halen?

- Op de eerste regel vinden we de toetsingsgrootte $t = 2.2677$, het aantal vrijheidsgraden $(n - 1) \text{ df} = 99$ en de p -waarde $p\text{-value} = 0.02552$. Dus op basis hiervan kunnen we de nulhypothese al verwerpen, aangezien onze p -waarde groter is dan $\alpha = 0.05$.
- Op de tweede regel vinden we onze alternatieve hypothese terug, met name $\mu \neq 7.4$
- De derde en vierde regel geven het 95% betrouwbaarheidsinterval. We zien dat 7.4 van onze nulhypothese zich niet in het interval bevindt, dus ook op basis hiervan verwerpen we de nulhypothese.
- Op de laatste regel zien we het geobserveerde gemiddelde, 9.08

We stellen nu als specifieke alternatieve hypothese dat $\mu_X = 10$. Dit wil dus zeggen dat we een verschil van $10 - 7.4 = 2.6$ willen detecteren - dit vullen we in onze functie in bij `delta`. Bij `sd` moeten we de geschatte standaardafwijking invullen. Daarnaast geven we ook nog mee dat onze $\alpha = 0.05$ bij `sig.level`, dat onze oorspronkelijke alternatieve hypothese tweezijdig was bij `alternative`, bij `type` dat het om 1 steekproef gaat en bij `n` de steekproefgrootte. Dit zorgt ervoor dat de power zal berekend worden.

```
> sx <- sd(bloeddruk$slaperigheid2)
> power.t.test(n = length(bloeddruk$slaperigheid2), delta = 2.6, sd = sx,
+             sig.level=0.05, alternative="two.sided", type="one.sample")
```

One-sample t test power calculation

```
      n = 100
  delta = 2.6
     sd = 7.408267
sig.level = 0.05
  power = 0.9351623
alternative = two.sided
```

We observeren dus een power van 0.9351623. Dit wil zeggen dat we 93.5% kans hebben om een verschil van 2.6 te detecteren gegeven dat onze populatieparameter niet 7.4 is.

Als we nu onze steekproefgrootte niet invullen, maar wel de power die we wensen, dan kunnen we de nodige steekproefgrootte schatten die we nodig hebben. Dus nu vullen we `power = 0.90` in bij de argumenten.

```
> power.t.test(power=0.90, delta = 2.6, sd = sx,
+             sig.level=0.05, alternative="two.sided", type="one.sample")
```

One-sample t test power calculation

```
      n = 87.25005
  delta = 2.6
     sd = 7.408267
sig.level = 0.05
  power = 0.9
alternative = two.sided
```

We moeten de steekproefgrootte naar boven afronden, dus we vinden dat we een steekproefgrootte van $n = 88$ nodig hebben om een power van 90% te bereiken.

8 Gewichtsverlies

1. Om het BMI te berekenen geven we de formule in R in, het gemiddelde berekenen we dan met de functie `mean`:

```
> load("obesitas.RData")
> BMI <- obesitas$gewicht1/obesitas$lengte^2
> mean(BMI)
[1] 30.07559
```

2. We kunnen dit op 2 manieren benaderen. We kunnen een verschillscore maken van het gewicht of we kunnen het argument `paired = TRUE` gebruiken. We verwachten dat het gewicht afneemt, dus we hebben een eenzijdige alternatieve hypothese

```
> t.test(obesitas$gewicht1, obesitas$gewicht2, paired=TRUE, alternative="greater")
```

Paired t-test

```

data:  obesitas$gewicht1 and obesitas$gewicht2
t = 6.4524, df = 63, p-value = 8.904e-09
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 3.632992      Inf
sample estimates:
mean of the differences
          4.901001

```

Als je nu `t.test(obesitas$gewicht1 - obesitas$gewicht2, mu = 0, alternative = "greater")` zou gebruiken, dan kom je op hetzelfde uit.

We observeren een zeer kleine p -waarde, dus we verwerpen de nulhypothese.

- In de vorige oefening hebben we gezien welke gegevens we nodig hebben in onze functie, dus we gaan dit hier niet herhalen. Het enige waar we de aandacht op willen vestigen is dat we nu invullen `alternative = "one.sided"` en `type = "paired"`. Alsook berekenen we nu de schatting van de standaardafwijking van de verschilscore.

```

> sx <-sd(obesitas$gewicht1-obesitas$gewicht2)
> power.t.test(n=dim(obesitas)[1], delta=2, sd=sx, sig.level=0.05, alternative="one.sided")

```

```

Paired t test power calculation

```

```

      n = 64
  delta = 2
    sd = 6.076469
sig.level = 0.05
  power = 0.8314236
alternative = one.sided

```

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs

We observeren dus een power van 83.14%.

- Nu vullen we de power in om te berekenen hoe groot onze steekproefgrootte moet zijn:

```

> power.t.test(power=0.90, delta=2, sd=sx, sig.level=0.05, alternative="one.sided",

```

```

Paired t test power calculation

```

```

      n = 80.42406
  delta = 2
    sd = 6.076469
sig.level = 0.05
  power = 0.9
alternative = one.sided

```

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs

We hebben dus minstens een steekproefgrootte van 81 nodig.

- Voor het 90% betrouwbaarheidsinterval ($1 - \alpha$) gaan we terug een t -test uitvoeren, maar nu geven we als argument `conf.level = 0.90`. (Als we niets ingeven wordt er in R vanuit gegaan dat dit 0.95 is)

```
> t.test(obesitas$gewicht1, obesitas$gewicht2, paired=TRUE, alternative="two.sided",
```

```
Paired t-test
```

```
data: obesitas$gewicht1 and obesitas$gewicht2
t = 6.4524, df = 63, p-value = 1.781e-08
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 3.632992 6.169010
sample estimates:
mean of the differences
          4.901001
```

We kunnen zien dat het betrouwbaarheidsinterval het enige is wat echt veranderd is. Het is nu tweezijdig. De p -waarde is niet echt veranderd, aangezien we hier over een miniem klein getal spreken (pas de 8ste of de 9de plaats na de komma).

9 Statismex: bloeddruk en geslacht

1. We zien dat er 47 vrouwen en 53 mannen in de steekproef zijn:

```
> BDV <- bloeddruk$BD[bloeddruk$geslacht=="v"]
> BDM <- bloeddruk$BD[bloeddruk$geslacht=="m"]
> nv <- length(BDV)
> nm <- length(BDM)
> nv
[1] 47
> nm
[1] 53
```

2. Beide QQ-plots tonen een lichte afwijking van de normaliteit. Aangezien er geen formele tests zijn gezien in de cursus, en we ons hier moeten baseren op een QQ-plot geven we het de voordeel van de twijfel. Ook hebben we een steekproef die groot genoeg is, dus mogen we verder met de test.

```
> qqnorm(BDV)
> qqnorm(BDM)
```

3. Nu zitten we met 2 onafhankelijke steekproeven met beide een onbekende verwachting. Dus gaan we de Welch-toets gebruiken. We hebben een eenzijdige alternatieve hypothese, aangezien we verwachten dat bij mannen een lagere bloeddruk is dan bij vrouwen.

```
> t.test(BDM, BDV, alternative="less", var.equal=FALSE)
```

```
Welch Two Sample t-test
```

```
data: BDM and BDV
t = -0.75959, df = 90.663, p-value = 0.2247
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 0.2484332
sample estimates:
mean of x mean of y
 11.47170  11.68085
```

We krijgen een p -waarde van 0.2247, welke groter is dan $\alpha = 0.05$ dus aanvaarden we de nulhypothese.

4. Voor de power te berekenen moeten we nu d berekenen met de formule

$$\frac{\mu_1 - \mu_2}{s_{\text{pooled}}}$$

We gebruiken de formule voor s_{pooled} zoals deze op het formuleblad staat, en vullen deze in R in:

```
> sp <- sqrt((52*var(BDM) + 46* var(BDV)) / (nm+nv-2))
> d <- 1/sp
```

Om de power te berekenen moeten we eerst het package `pwr` opladen, om vervolgens de functie `pwr.t2n.test` te gebruiken. Hier vullen we bij `n1` de steekproefgrootte van de mannen in en bij `n2` de steekproefgrootte van de vrouwen. We moeten rekening houden met het feit dat onze d positief is. We willen een effectgrootte van d observeren tussen mannen en vrouwen, waar mannen lager scoren dan vrouwen. Dus moeten we hier de negatieve versie van nemen, om vervolgens bij `alternative` aan te geven dat onze oorspronkelijke alternatieve hypothese `less` was. (We kunnen hetzelfde resultaat bekomen als we d niet negatief maken, maar aangeven dat we dan in de andere richting willen werken, namelijk `greater`)

```
> library(pwr)
> pwr.t2n.test(n1=nm, n2=nv, d=-d, sig.level=0.05, alternative="less")
```

```
t test power calculation

      n1 = 53
      n2 = 47
       d = 0.7349708
sig.level = 0.05
  power = 0.9771354
alternative = greater
```

We observeren hier een zeer hoge power, namelijk 97.71%. We zouden dus gerust een kleinere steekproef kunnen nemen.

5. Om nu de nodige steekproefgrootte bij mannen te gaan berekenen om een power van 90% te bekomen, met voor de rest dezelfde gegevens, gaan we bij de argumenten `n1` laten vallen en in de plaats `power = 0.90` zetten.

```
> pwr.t2n.test(power=0.90, n2=nv, d=d, sig.level=0.05, alternative="greater")
```

```
t test power calculation

      n1 = 24.64708
      n2 = 47
       d = 0.7349708
sig.level = 0.05
  power = 0.9
alternative = greater
```

Zoals in het vorige punt gezegd kunnen we een kleinere steekproef nemen. Voor een power van 90% met een steekproefgrootte van 47 vrouwen hebben we slechts 25 mannen nodig.

10 Shooter-games en stress

1. We hebben hier terug een t -test met 2 onafhankelijke steekproeven. Dus we krijgen als hypotheses:

$$H_0 : \mu_{\text{exp}} = \mu_{\text{cond}} \text{ versus } H_a : \mu_{\text{exp}} > \mu_{\text{cond}}$$

2. Als we de QQ-plots maken zien we een kleine afwijking van normaliteit, maar we geven hier weer het voordeel van de twijfel dat ze normaalverdeeld zijn. Aangezien onze steekproefgroottes minstens 30 zijn kunnen we sowieso een t -test uitvoeren.

```
> load("shooter.RData")
> cort1 <- shooter$cortisol[shooter$cond=="con"]
> cort2 <- shooter$cortisol[shooter$cond=="exp"]
> qqnorm(cort1)
> qqnorm(cort2)
> t.test(cort1, cort2, alternative="less")
```

Welch Two Sample t-test

```
data: cort1 and cort2
t = 1.5079, df = 65.823, p-value = 0.9318
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 3.061538
sample estimates:
mean of x mean of y
 13.41282  11.95937
```

We aanvaarden de nulhypothese, aangezien onze p -waarde groter is dan 0.05, met name 0.9318. Dit hadden we eigenlijk al kunnen voorspellen als we naar de gemiddeldes gingen kijken. We zien dat het gemiddelde van de controlegroep hoger is dan deze van de experimentele groep, wat tegen onze alternatieve hypothese ingaat.

11 Uitslagen en geslacht

1. Als we gaan kijken naar de QQ-plots van beide zien we dat er toch wel een lichte afwijking is van normaliteit, of deze afwijking significant is kunnen we niet weten zonder formele test. Maar omdat onze steekproefgrootte groot genoeg is maakt dit voor de rest niet uit.
2. We gebruiken terug een Welch-test, omdat het twee onafhankelijke steekproeven zijn met onbekende varianties. We hebben een eenzijdige alternatieve hypothese, dat mannen hoger scoren dan vrouwen:

```
> load("uitslagen.RData")
> man <- uitslagen$uitslag[uitslagen$geslacht=="m"]
> vrouw <- uitslagen$uitslag[uitslagen$geslacht=="v"]
> nm <- length(man)
> nv <- length(vrouw)
>
> qqnorm(man)
> qqnorm(vrouw)
> t.test(man, vrouw, alternative="greater")
```

Welch Two Sample t-test

```
data: man and vrouw
t = 3.3174, df = 140.04, p-value = 0.0005789
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.440077      Inf
sample estimates:
mean of x mean of y
 54.16202  51.28686
```

We krijgen een zeer kleine p -waarde, namelijk 0.05%, wat kleiner is dan onze vooropgestelde α van 5%. Dus we verwerpen de nulhypothese.

3. Gelijkaardig met oefening 9.4 gaan we d berekenen, om deze in te vullen in R. Aangezien dit volledig gelijkaardig is met oefening 9.4 gaan we ineens over naar R:

```
> sp <- sqrt((74*var(man) + 67*var(vrouw))/(nm+nv-2))
> d <- 3/sp
> pwr.t2n.test(n1=nm, n2=nv, d=d, sig.level=0.05, alternative="greater")
```

```
t test power calculation
```

```
      n1 = 75
      n2 = 68
      d  = 0.5791563
sig.level = 0.05
power    = 0.9638465
alternative = greater
```

We observeren een power van 96.38%, wat een zeer hoge power is.

12 Reactietijd

Hier zitten we terug met een t -test met 1 steekproef. De redenering is exact dezelfde als we al gezien hebben, het enige “addertje” hier is het formaat van het databestand. We zien hier nu wel een zware afwijking van normaliteit in de QQ-plot. Maar aangezien we een steekproefgrootte hebben die groter is dan 30 gaan we toch verder met de toetsingsprocedure.

$$H_0 : \mu = 810ms \text{ versus } H_a : \mu > 810ms$$

```
> data <- read.delim(file = "Stroop.txt", dec=',')
> qqnorm(data$reactietijd)
> t.test(x=data$reactietijd, mu = 0.810, alternative = "greater")
```

```
One Sample t-test
```

```
t = 1.7935, df = 56, p-value = 0.03914
alternative hypothesis: true mean is greater than 0.81
95 percent confidence interval:
 0.832989      Inf
```



```
sample estimates:  
mean of x  
1.150702
```

We hebben een p -waarde van 0.03914 en is groter dan 0.05, dus we verwerpen de nulhypothese, en concluderen dat bij de volwassen Vlaamse populatie met depressie de reactietijd op een Stroop-taak hoger ligt dan bij de volwassen Vlaamse populatie.

Vragen & Extra Informatie

Vraag over Zuivere Munt uit Oefensessie 1

Hoe bereken je deze kansen?

We gebruiken hiervoor de formule die ook terug te vinden is op het formuleblad dat jullie mogen gebruiken tijdens het examen. Op het eerste blad in het stukje Kansrekenen zien we:

$$\mathcal{P}(B(n, \pi) = k) = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$$

Dus voor $\mathcal{P}(B(5, 0.5) = 4)$ krijgen we (we vullen nu de n , π en k in de formule in):

$$\begin{aligned} \mathcal{P}(B(5, 0.5) = 4) &= \frac{5!}{4!(5-4)!} 0.5^4 (1-0.5)^{5-4} \\ &= \frac{5 \times 4 \times 3 \times 2 \times 1}{(4 \times 3 \times 2 \times 1)(1)} \times \frac{2}{32} \times \frac{1}{2} \\ &= \frac{5}{1} \times \frac{2}{32} \times \frac{1}{2} = \frac{5}{32} \end{aligned}$$

Gelijkaardig krijgen we:

$$\begin{aligned} \mathcal{P}(B(5, 0.5) = 5) &= \frac{5!}{5!0!} 0.5^5 (1-0.5)^0 \\ &= \frac{5 \times 4 \times 3 \times 2 \times 1}{(5 \times 4 \times 3 \times 2 \times 1)(1)} \times \frac{1}{32} \times 1 \\ &= \frac{1}{1} \times \frac{1}{32} = \frac{1}{32} \end{aligned}$$

Noot: $0! = 1$ en $x^0 = 1$ (met uitzondering van $x = 0$)

Type I fout, Type II fout, Betrouwbaarheid en Power

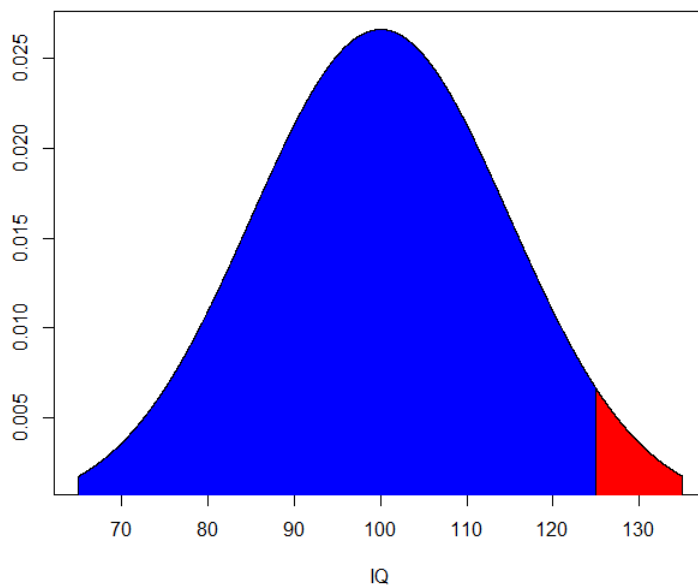
In de vorige bundel hebben we getracht een concreter beeld te geven over betrouwbaarheidsinterval, verwerpingsgebied en p -waarde. Nu gaan we met dat in het achterhoofd proberen een concreter beeld te geven over de power. We beginnen met een eenzijdige alternatieve hypothese, de stap naar het tweezijdige is dan zeer triviaal.

Type I fout en Betrouwbaarheid

Wanneer we over betrouwbaarheid spreken, dan spreken we impliciet ook over de Type I fout. Standaard gebruiken we $\alpha = 0.05$. Uit dat cijfer alleen kunnen we al afleiden dat onze betrouwbaarheid 95% is, want onze betrouwbaarheid kunnen we herschrijven als $1 - \alpha$. Dit is de kans dat we onze nulhypothese aanvaarden, gegeven dat de nulhypothese juist is. Het is altijd van belang de conditie erbij te nemen. Omdat we nu zeggen “gegeven dat onze nulhypothese juist is” (of in andere woorden, conditioneel dat de nulhypothese juist is), betekent dit dat we onder de nulverdeling werken - dit is de verdeling conditioneel dat de nulhypothese juist is. Als we 95% kans hebben om hier de juiste beslissing te nemen, conditioneel op onze nulhypothese, betekent dit dat we $\alpha = 5\%$ kans hebben een foute beslissing te nemen - dus de kans dat we de nulhypothese verwerpen, gegeven dat de nulhypothese juist is.

Om dit met een voorbeeldje wat duidelijker te maken. We weten dat het gemiddelde IQ in de populatie 100 is met een standaardafwijking van 15. Stel nu dat we in de faculteit een steekproef nemen en we daar het IQ van meten. We kunnen er dan met een grote veiligheid vanuit gaan dat we een eenzijdige alternatieve hypothese kunnen testen dat het gemiddelde IQ op de faculteit groter is dan 100.

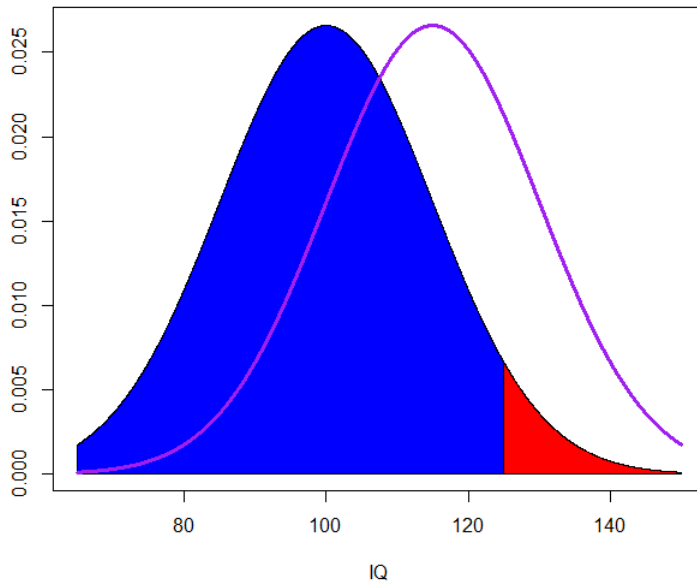
Grafisch krijgen we dit:



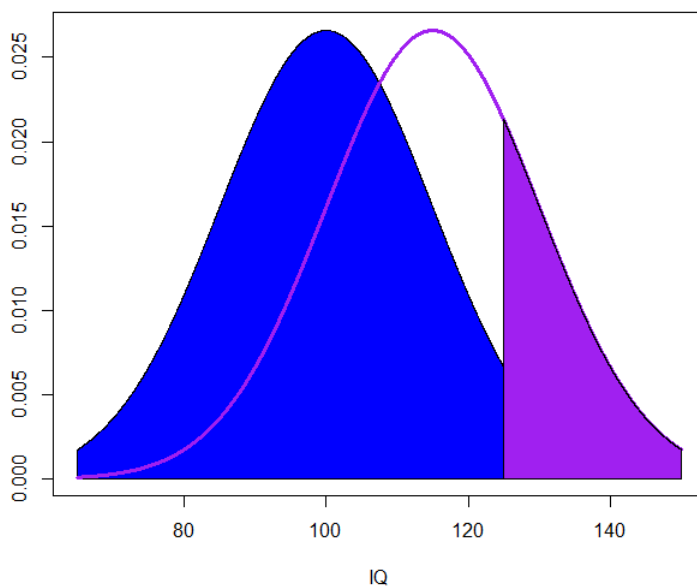
Het rode gebied geeft de α weer, en is dus de Type I fout. De kans dat we onze nulhypothese gaan verwerpen, gegeven dat deze juist is - in dit geval 5%. Het blauwe gebied is de betrouwbaarheid of $1 - \alpha = 0.95$. De kans dat we onze nulhypothese aanvaarden gegeven dat onze nulhypothese juist is. Dus stel dat het gemiddelde IQ in de faculteit ook 100 zou zijn, dan is er 5% kans dat we een gemiddelde observeren dewelke groter is dan 124.67, wat tot de conclusie zou leiden dat het gemiddelde niet gelijk is aan 100. Dus 5% van de tijd gaan we verkeerdelijk onze nulhypothese verwerpen.

Type II fout en Power

Maar stel nu dat het gemiddelde IQ op de faculteit 115 is. Wat is dan de kans dat we onze nulhypothese gaan verwerpen? Als het gemiddelde IQ op de faculteit 115 is dan is de nulhypothese foutief. Maar we weten dat we de nulhypothese pas verwerpen wanneer we een gemiddeld IQ observeren dat groter is dan 124.67. Dus we gaan de nulhypothese soms verkeerdelijk aanvaarden. Nu is de vraag, hoe vaak? Om dit te weten moeten we naast onze nulverdeling ook onze alternatieve verdeling opstellen. Dit is de verdeling, gegeven dat onze alternatieve hypothese correct is. Dit zou er dan als volgt uitzien:



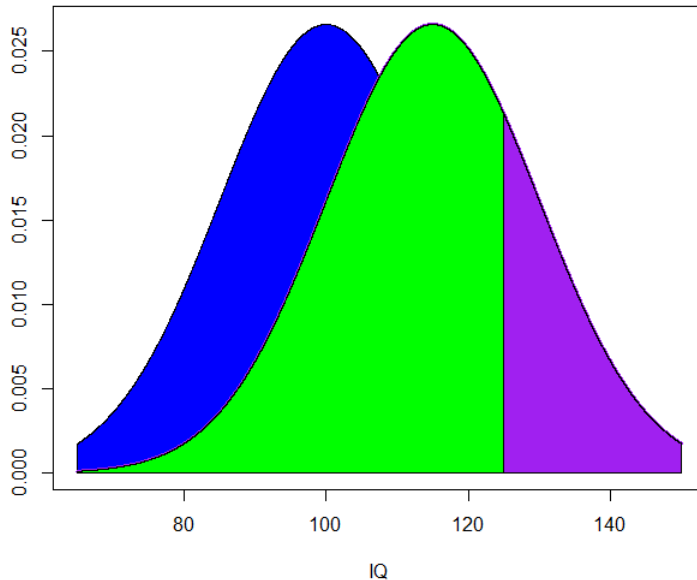
Wanneer gaan we nu de alternatieve aanvaarden als deze juist is? Pas wanneer we de nulhypothese gaan verwerpen. Want onze toetsingsprocedure gaat er altijd van uit dat onze nulhypothese correct is. Dus gaan we vanaf 124.67 pas onze alternatieve hypothese aanvaarden, gegeven dat deze juist is. Dit is in de volgende grafiek het paarse gebied.



We zien dat dit voor een deel overlapt met onze Type I fout. Met name, van onze nulverdeling is enkel

het rode gebied nu bedekt door de power. Deze kans is in dit geval 25.25%. We zien dat het maar een klein gebied is. Dus de kans dat we onze alternatieve hypothese aanvaarden gegeven dat deze correct is - of in andere woorden dat we onze nulhypothese verwerpen, gegeven dat deze verkeerd is - is $1 - \beta = 0.2525$.

Aangezien we onze alternatieve hypothese soms niet kunnen aanvaarden, ook al is deze correct, zorgt ervoor dat we een foute beslissing nemen. Dit is onze Type II fout. Dit geven we weer met β en is in dit geval 74.75%. In de volgende grafiek stellen we dit voor met een groene kleur.



Samengevat

We hebben dus 4 kansen. Elk gaat uit van een conditie, namelijk welke hypothese er correct is, en de beslissing die we maken:

- De nulhypothese is correct:
 - CORRECTE BESLISSING:
We aanvaarden de nulhypothese: Betrouwbaarheid: $1 - \alpha$ - het blauwe gebied
 - FOUTE BESLISSING:
We verwerpen de nulhypothese: Type I fout: α - het rode gebied
- De alternatieve hypothese is correct:
 - CORRECTE BESLISSING:
We verwerpen de nulhypothese: Power: $1 - \beta$ - het paarse gebied
 - FOUTE BESLISSING:
We aanvaarden de nulhypothese: β - het groene gebied

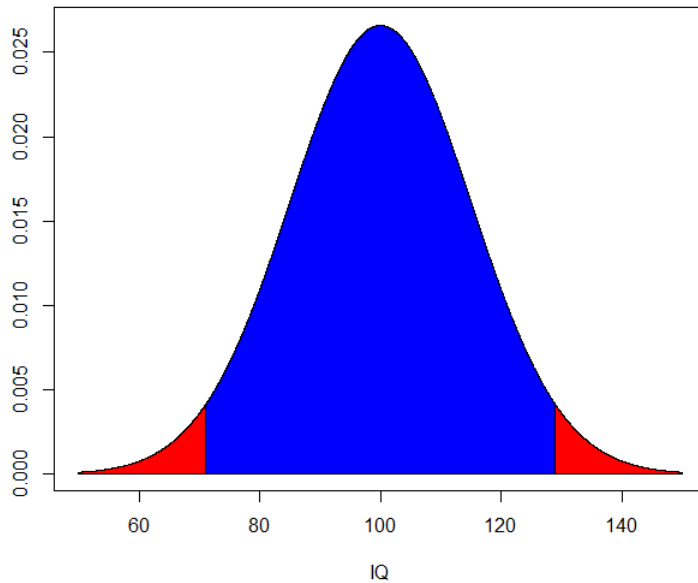
We kunnen hier ook zien dat als we de betrouwbaarheid verkleinen, bijvoorbeeld naar 90%, dan gaat ons rood gebied groter worden, en ons blauw gebied kleiner. Dit zorgt er ook voor dat de grens tussen paars en groen naar links gaat verschuiven. Dus we krijgen een grotere power, en een kleinere Type II fout.

Maar we kunnen ook onze power vergroten zonder onze Type I fout te vergroten. Dit doen we door een grotere steekproef te nemen. Als onze steekproefgrootte toeneemt dan worden onze curves smaller, dit wil zeggen dat we een grotere massa gaan krijgen rond de 100 voor de nulverdeling, en rond de 115 voor de alternatieve

verdeling. Dit zorgt ervoor dat er minder overlap is tussen de grafieken, en we dus een groter paars gebied krijgen.

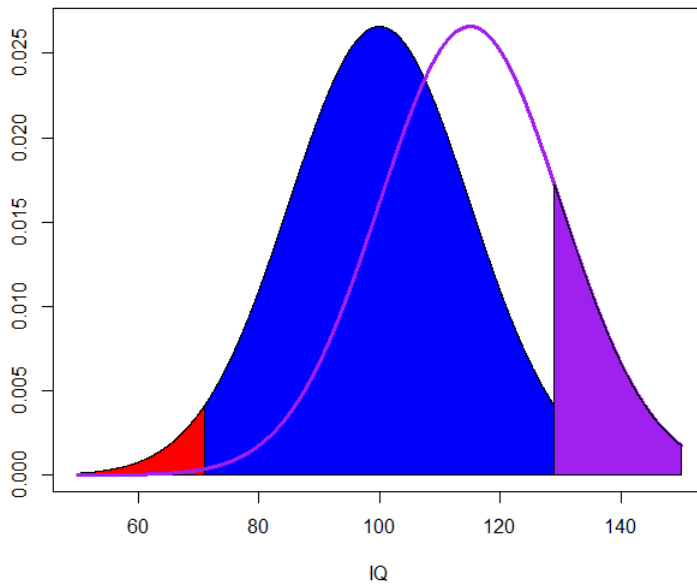
Wat nu met het tweezijdige?

Soms hebben we een alternatieve hypothese die tweezijdig is. Bijvoorbeeld dat we het gemiddelde IQ in een provincie gaan schatten. Aangezien dit zowel lager als hoger kan liggen dan de verwachting in de populatie, kunnen we best een alternatieve hypothese opstellen die tweezijdig is. Dit geeft volgende grafiek:

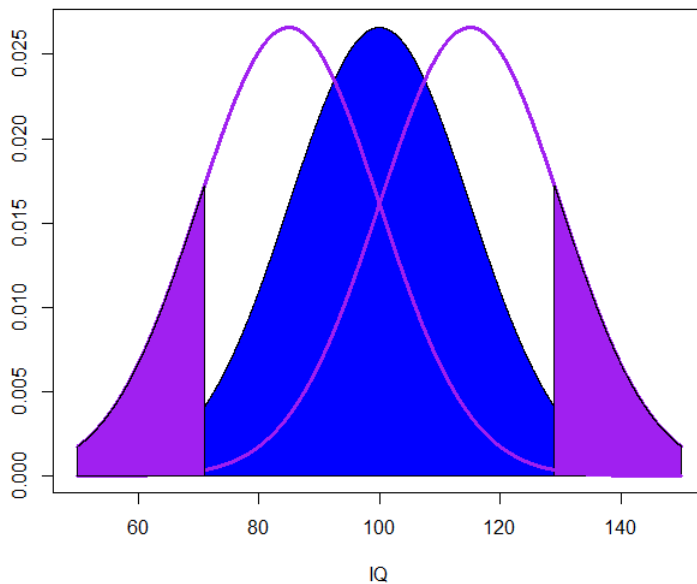


Nu is het rode gebied nog steeds 5%, maar is het verdeeld aan de twee kanten. Dus we hebben 2.5% aan de ene kant alsook aan de andere kant. Onze Type I fout blijft hetzelfde. Als de verwachting in de provincie ook 100 is hebben we 5% kans om dit te verwerpen, 2.5% kans in beide richtingen. We hebben dus ook 95% kans om de nulhypothese te aanvaarden.

Maar stel dat we een verschil willen kunnen detecteren van 15 IQ punten, net zoals in het vorige stuk. Dan moeten we nu in beide richtingen redeneren. We gaan eerst naar de rechterkant kijken, omdat dit volledig gelijkaardig is met het vorige deel:



Dus we gaan de alternatieve aanvaarden, gegeven dat deze juist is in het paarse gebied. Kijkende naar deze verdeling, met een verwachting van 115, geeft dit een kans van 17.5%. Dus als we enkel naar de rechterkant kijken hebben we een power van $1 - \beta = 0.175$. Dus we hebben een Type II fout van $\beta = 0.825$. Maar we werken tweezijdig. Dus we moeten ook aan de linkerkant hetzelfde doen. Als we dit grafisch voorstellen zien we dat dit het spiegelbeeld is van onze linkerkant:



Omdat dit het spiegelbeeld is, is het voldoende dat we dit eenzijdig uitwerken. Dus de power voor de tweezijdige toets is $1 - \beta = 0.175$.

Noot: dit zijn fictieve gegevens slechts gebruikt ter illustratie.