Ongepaarde t-toets

Table of Contents

<div class="navbar-header">  
 <button type="button" class="navbar-toggle collapsed" data-toggle="collapse" data-target="#navbar">  
 <span class="icon-bar"></span>  
 <span class="icon-bar"></span>  
 <span class="icon-bar"></span>  
 </button>  
 <a class="navbar-brand" href="Index.html">Statistisch Handboek Studiedata</a>  
</div>  
<div id="navbar" class="navbar-collapse collapse">  
 <ul class="nav navbar-nav navbar-start">  
 <li>

Toetsmatrix (R)

Toetsmatrix (Python)

Begrippenlijst

Over

Verantwoording

Feedback

Nieuwsbrief

</ul>  
 <ul class="nav navbar-nav navbar-right">  
  
 </ul>  
</div><!--/.nav-collapse -->

Disclaimer: Het peer review proces voor deze toets is nog niet afgerond; daarom is deze pagina nog in concept.

# Toepassing

Gebruik de *ongepaarde t-toets* om de gemiddelden van twee onafhankelijke groepen te vergelijken.[[1]](#footnote-21)

# Onderwijscasus

Vanaf 2011 heeft de opleiding Taalwetenschap een Bindend Studieadvies (BSA) die de selectiviteit van het eerste jaar moet vergroten. De studieadviseur vraagt zich af of het gemiddelde cijfer van de opleiding Taalwetenschap op 1 februari, na invoering van het BSA, veranderd is. De data is beschikbaar voor het cohort gestart in 2010 en voor het cohort gestart in 2011.

Dit onderzoek vertaalt zich in de volgende combinatie van hypothesen, waarbij de nulhypothese zo geformuleerd is dat er geen effect of verschil is en de alternatieve hypothese zo geformuleerd is dat er wel een effect of verschil is.

*H0*: Het gemiddelde tentamencijfer dat de studenten halen aan de opleiding Taalwetenschap is niet veranderd na de invoer van het BSA, µ0 = µ1

*HA*: Het gemiddelde tentamencijfer dat de studenten halen aan de opleiding Taalwetenschap is veranderd na de invoer van het BSA, µ0 ≠ µ1

# Assumpties

Voor een valide resultaat moeten de data aan een aantal voorwaarden voldoen voordat de toets uitgevoerd kan worden.

## Normaliteit

De *t-toets* gaat ervan uit dat de data van de steekproef normaal verdeeld is. Ga er bij een aantal observaties *n* > 100 vanuit dat de t-toets robuust genoeg is om uit te voeren zonder dat de data een normale verdeling volgt.[[2]](#footnote-27)

Controleer de assumptie van normaliteit met de volgende stappen:  
1. Controleer de data visueel met een histogram, een boxplot en een Q-Q plot.  
2. Toets of de data normaal verdeeld is met de *Kolmogorov-Smirnov test* of bij een kleinere steekproef (n < 50) met de *Shapiro-Wilk test*.[[3]](#footnote-29), [[4]](#footnote-31)

Als blijkt dat de data niet normaal verdeeld is, transformeer[[5]](#footnote-33) dan de data eventueel en bepaal daarna of deze wel normaal verdeeld is.

Als er geen sprake is van normaliteit gebruik de [Mann-Whitney U toets](08-Mann-Whitney-U-toets-R.html).[[6]](#footnote-35), [[7]](#footnote-37)

## Wel of geen gepoolde variantie

De *ongepaarde t-toets* kan met en zonder gepoolde variantie uitgevoerd worden. Bij een gepoolde variantie is de berekening van de variantie van het verschil in gemiddelden anders en wordt aangenomen dat de varianties van beide steekproeven even groot zijn. Deze aanname is te toetsen met de *Levene’s Test*, waarbij een significant resultaat aangeeft dat er een verschil is in de varianties van beide groepen. De hedendaagse consensus is echter om altijd deze aanname niet te toetsen en de *ongepaarde t-toets* zonder gepoolde variantie uit te voeren.[[8]](#footnote-39) Een gepoolde variantie zorgt ervoor dat het onderscheidend vermogen[[9]](#footnote-41) van de *ongepaarde t-toets* iets hoger is als de varianties van beide groepen ongeveer gelijk zijn, maar geeft verkeerde resultaten als de varianties van elkaar afwijken. Daarnaast heeft *Levene’s Test* een laag onderscheidend vermogen, wat betekent dat het lastig is om ongelijke varianties goed te toetsen. Gebruik daarom de *ongepaarde t-toets* zonder gepoolde variantie; deze staat ook wel bekend als *Welch’s t-toets*.[[10]](#footnote-42)

# Uitvoering

Er is een dataset ingeladen met gemiddelde cijfers van eerstejaarsstudenten bij de opleiding Taalwetenschap: Cijfers\_gemiddeld. De data bevatten cijfers van 180 studenten begonnen in 2010 en cijfers van 160 studenten begonnen in 2011.

## De data bekijken

Gebruik head() en tail() om de structuur van de data te bekijken.

## Eerste 6 observaties  
head(Cijfers\_gemiddeld)  
## Studentnummer Cohort Cijfers  
## 1 302256 2010 5.548  
## 2 344374 2010 6.520  
## 3 302078 2010 5.297  
## 4 325370 2010 8.214  
## 5 362912 2010 6.695  
## 6 335804 2010 5.315  
  
## Laatste 6 observaties  
tail(Cijfers\_gemiddeld)  
## Studentnummer Cohort Cijfers  
## 335 321518 2011 6.247  
## 336 317635 2011 7.185  
## 337 396683 2011 7.264  
## 338 393299 2011 7.132  
## 339 362510 2011 5.763  
## 340 335445 2011 4.814

Selecteer beide groepen en sla deze op in een vector om deze makkelijker aan te kunnen roepen.

Cijfers\_2010 <- Cijfers\_gemiddeld$Cijfers[Cijfers\_gemiddeld$Cohort == "2010"]  
Cijfers\_2011 <- Cijfers\_gemiddeld$Cijfers[Cijfers\_gemiddeld$Cohort == "2011"]

Inspecteer de data met length(), mean()en sd() om meer inzicht te krijgen in de data.

## Aantallen, gemiddelde en standaarddeviatie 2010  
length(Cijfers\_2010)  
## [1] 180  
mean(Cijfers\_2010)  
## [1] 6.369189  
sd(Cijfers\_2010)  
## [1] 1.117662

## Aantallen, gemiddelde en standaarddeviatie 2011  
length(Cijfers\_2011)  
## [1] 160  
mean(Cijfers\_2011)  
## [1] 6.462363  
sd(Cijfers\_2011)  
## [1] 1.199976

* Gemiddeld tentamencijfer 2010 (standaarddeviatie): 6,37 (1,12). *n* = 180.
* Gemiddeld tentamencijfer 2011 (standaarddeviatie): 6,46 (1,2). *n* = 160.

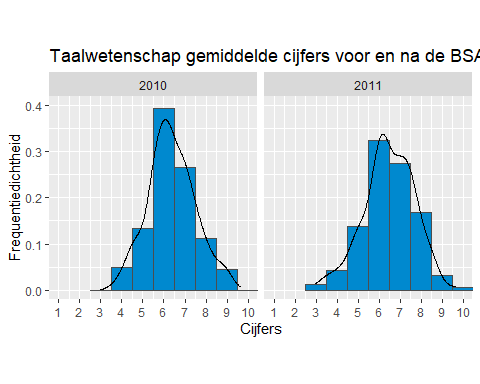
## Visuele inspectie van normaliteit

Geef de verdeling van de data visueel weer met een histogram, Q-Q plot en boxplot.

### Histogram

Focus bij het analyseren van een histogram op de symmetrie van de verdeling, de hoeveelheid toppen (modaliteit) en mogelijke uitbijters. Een normale verdeling is symmetrisch, heeft één top en geen uitbijters.[[11]](#footnote-47), [[12]](#footnote-49)

## Histogram met ggplot2  
library(ggplot2)  
  
ggplot(Cijfers\_gemiddeld,  
 aes(x = Cijfers)) +  
 geom\_histogram(aes(y = ..density..),  
 binwidth = 1,  
 color = "grey30",  
 fill = "#0089CF") +  
 facet\_wrap(~ Cohort) +  
 geom\_density(alpha = .2, adjust = 1) +  
 ylab("Frequentiedichtheid") +  
 scale\_x\_continuous(  
 labels = as.character(seq(1, 10)),  
 breaks = seq(1, 10)) +  
 coord\_fixed(ylim = c(0, 0.4),  
 xlim = c(1, 10),  
 ratio = 22) +  
 labs(title = "Taalwetenschap gemiddelde cijfers voor en na de BSA")



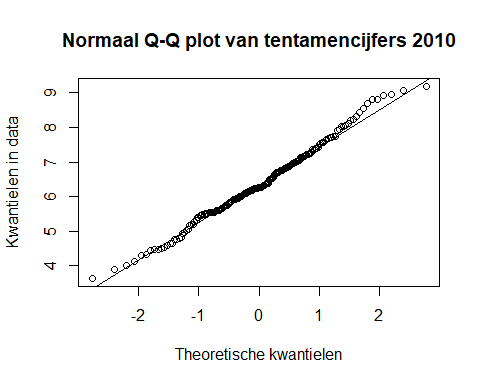
Beide histogrammen laten een verdeling zien die redelijk symmetrisch is, één top heeft en geen uitbijters. Daarom zijn beide verdeling bij benadering normaal verdeeld.

### Q-Q plot

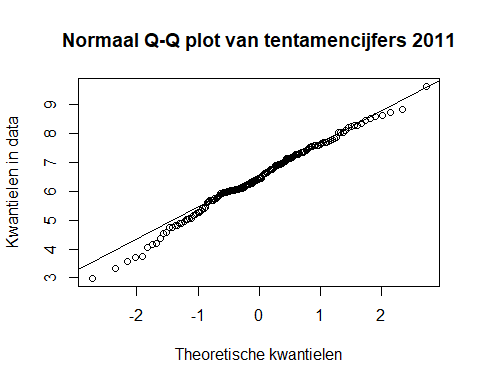
Gebruik qqnorm() en qqline() met pch = 1om een Q-Q plot te maken, met als datapunten kleine cirkels.

Als over het algemeen de meeste datapunten op de lijn liggen, kan aangenomen worden dat de data normaal verdeeld is.

qqnorm(Cijfers\_2010, pch = 1,  
 main = "Normaal Q-Q plot van tentamencijfers 2010",  
 ylab = "Kwantielen in data",  
 xlab = "Theoretische kwantielen")  
qqline(Cijfers\_2010)



qqnorm(Cijfers\_2011, pch = 1,  
 main = "Normaal Q-Q plot van tentamencijfers 2011",  
 ylab = "Kwantielen in data",  
 xlab = "Theoretische kwantielen")  
qqline(Cijfers\_2011)

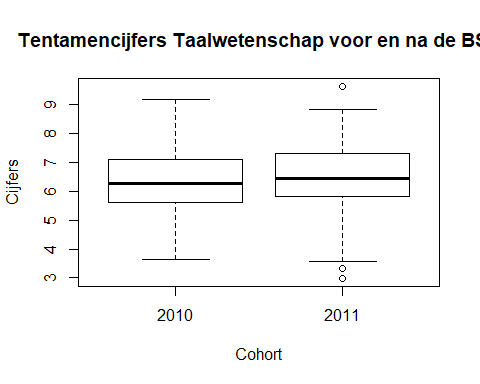


Voor beide Q-Q plots liggen de meeste datapunten op of vlakbij de lijn. Hoewel er bij de uiteinden van de verdeling wat afwijkingen zijn, duidt deze grafiek op een goede benadering van de normaalverdeling voor beide cohorten.

### Boxplot

De box geeft de middelste 50% van de tentamencijfers weer. De zwarte lijn binnen de box is de mediaan. In de staarten of snorreharen zitten de eerste 25% en de laatste 25%. Cirkels visualiseren mogelijke uitbijters.[[13]](#footnote-55) Hoe meer de boxen overlappen, hoe waarschijnlijker er geen significant verschil is tussen de groepen.

boxplot(Cijfers ~ Cohort, Cijfers\_gemiddeld,  
 main = "Tentamencijfers Taalwetenschap voor en na de BSA")



De boxplotten geven de spreiding weer van het gemiddelde tentamencijfer voor de BSA en na de BSA. De boxplotten en de staarten lijken symmetrisch, wat een teken is van bij benadering normaal verdeelde data. Het cohort van 2011 heeft een aantal mogelijke uitbijters.[[14]](#footnote-57)

## Toetsen van normaliteit

Om te controleren of de data normaal verdeeld is, kan de normaliteit getoetst worden. Twee veelgebruikte toetsen zijn: de *Kolmogorov-Smirnov test* en de *Shapiro-Wilk test*.

### Kolmogorov-Smirnov

De *Kolmogorov-Smirnov test* toetst het verschil tussen twee verdelingen. Standaard toetst deze test het verschil tussen een normale verdeling en de verdeling van de steekproef. De Lilliefors correctie is vereist als het gemiddelde en de standaardafwijking niet van tevoren bekend of bepaald zijn, wat meestal het geval is bij een steekproef. Als de p-waarde kleiner dan 0,05 is, is de verdeling van de data significant verschillend van de normale verdeling.

library(nortest)

lillie.test(Cijfers\_2010)

##   
## Lilliefors (Kolmogorov-Smirnov) normality test  
##   
## data: Cijfers\_2010  
## D = 0.05404, p-value = 0.2244

lillie.test(Cijfers\_2011)

##   
## Lilliefors (Kolmogorov-Smirnov) normality test  
##   
## data: Cijfers\_2011  
## D = 0.061106, p-value = 0.1526

Bij deze casus is van beide groepen de p-waarde groter dan 0,05, dus er zijn geen significante verschillen gevonden tussen de verdeling van de steekproef en de normale verdeling. De *ongepaarde t-toets* kan uitgevoerd worden.

### Shapiro-Wilk Test

De *Shapiro-Wilk test* is een soortgelijke test als de *Kolmogorov-Smirnov test* en vooral geschikt bij kleine steekproeven (*n* < 50). Als de p-waarde kleiner dan 0,05 is, is de verdeling van de data significant verschillend van de normale verdeling. Er zijn twee subsets van Cijfers\_gemiddeld ingeladen: Cijfers\_2010\_n30 en Cijfers\_2011\_n30. Beide subsets bevatten 30 studenten. Voor relatief kleine steekproeven als deze is de *Shapiro-Wilk Test* geschikt.

shapiro.test(Cijfers\_2010\_n30)

##   
## Shapiro-Wilk normality test  
##   
## data: Cijfers\_2010\_n30  
## W = 0.98796, p-value = 0.9765

shapiro.test(Cijfers\_2011\_n30)

##   
## Shapiro-Wilk normality test  
##   
## data: Cijfers\_2011\_n30  
## W = 0.97627, p-value = 0.7202

De p-waarde is groter dan 0,05 voor beide groepen, dus er zijn geen significante verschillen gevonden tussen de verdeling van de steekproef en de normale verdeling. De *ongepaarde t-toets* kan uitgevoerd worden.

## Ongepaarde t-toets

Voer een ongepaarde t.test() uit met paired = FALSE (vanwege de ongepaarde groepen) en var.equal = FALSE (omdat de varianties niet per se aan elkaar gelijk zijn). Het eerste argument bestaat uit de afhankelijke variabele Cijfers en de groepvariabele Cohort.

t.test(Cijfers ~ Cohort, Cijfers\_gemiddeld,   
 paired = FALSE,  
 alternative = "two.sided",  
 var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: Cijfers by Cohort  
## t = -0.738, df = 326.37, p-value = 0.461  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.3415432 0.1551960  
## sample estimates:  
## mean in group 2010 mean in group 2011   
## 6.369189 6.462363

* *t* = -0,74, *p* = 0,46
* p-waarde > 0,05, dus de H0 wordt niet verworpen [[15]](#footnote-62)
* Vrijheidsgraden: *df* = 326,37, niet gelijk aan aantal observaties min één bij een ongepaarde t-toets zonder gepoolde varianties
* 95%-betrouwbaarheidsinterval: bij het herhalen van het experiment met verschillende steekproeven van de populatie zal 95% van de betrouwbaarheidsintervallen de daadwerkelijke parameter bevatten, het verschil tussen de hoogte van de cijfers voor en na de BSA cursus, µverschil = µT2011 - µT2010. In deze casus is het interval tussen -0,34 en 0,16. Aangezien 0 in dit interval zit, is er geen significant verschil tussen beide gemiddelden in 2010 en 2011.
* Het gemiddelde van de steekproef is in 2010 6,37
* Het gemiddelde van de steekproef is in 2011 6,46

### Effectmaat: Cohen’s d

De p-waarde geeft aan of het verschil tussen twee groepen significant is. De grootte van het verschil of effect is echter ook relevant. Een effectmaat is een gestandaardiseerde maat die de grootte van een effect weergeeft, zodat effecten van verschillende onderzoeken met elkaar vergeleken kunnen worden.[[16]](#footnote-64) Een veel gebruikte effectmaat is Cohen’s *d*. Cohen’s *d* geeft een gestandaardiseerd verschil weer: het verschil in gemiddelden tussen twee groepen gecorrigeerd voor de gecombineerde standaardafwijking van de twee groepen. Een indicatie om *d* te interpreteren is: rond 0,3 is het een klein effect, rond 0,5 is het een gemiddeld effect en rond 0,8 is het een groot effect.[[17]](#footnote-65)

In dit voorbeeld is de p-waarde groter dan 0,05, dus is een effectmaat uitrekenen onnodig. Pas de volgende stappen toe bij een p-waarde kleiner dan 0,05.

Gebruik de functie cohensD() met de argumenten Cijfers ~ Cohort waarbij Cijfers de afhankelijke variabele is en Cohort de onafhankelijke variabele die de groepen aangeeft, het argument Cijfers\_gemiddeld dat de dataset aangeeft en het argument method = unequal omdat er niet aangenomen wordt dat de varianties aan elkaar gelijk zijn.

library(lsr)  
cohensD(Cijfers ~ Cohort, Cijfers\_gemiddeld, method = "unequal")

## [1] 0.08035328

*d* = 0.0803533. De sterkte van het effect van het BSA op het cijfer is verwaarloosbaar.

# Rapportage

Een *ongepaarde t-toets* is uitgevoerd om te toetsen of het gemiddelde tentamencijfer is veranderd na de invoer van het BSA. Het verschil tussen het gemiddelde tentamencijfer van cohort 2010 (*M2010* = 6,37, *SD2010* = 1,12) en het gemiddelde tentamencijfer van cohort 2011 (*M2011* = 6,46, *SD2011* = 1,2) is niet significant, *t* = -0,74, *p* = 0,46. Het 95% betrouwbaarheidsinterval voor het verschil tussen het gemiddelde van beide groepen loopt van -0,34 tot 0,16. Het gemiddelde tentamencijfer lijkt niet veranderd te zijn na de invoer van het BSA.

|  |  |  |  |
| --- | --- | --- | --- |
| Cohort | N | M | SD |
| 2010 | 180 | 6,37 | 1,12 |
| 2011 | 160 | 6,46 | 1,2 |

*Tabel 1. Groepsgrootte, gemiddeld tentamencijfer en standaarddeviatie per cohort*

Deze pagina maakt onderdeel uit van het Statistisch Handboek Studiedata, ontwikkeld binnen de zone Veilig en betrouwbaar benutten van studiedata van het Versnellingsplan. R code is uitgevoerd met R versie 3.6.3; Python code is uitgevoerd in Python 3.7. © 2020 Versnellingsplan - Statistisch Handboek Studiedata - Licentie Laatst gewijzigd op:17-07-2020

1. Van Geloven, N. (25 mei 2016). *T-toets* [Wiki Statistiek Academisch Medisch Centrum](https://wikistatistiek.amc.nl/index.php/T-toets#ongepaarde_t-toets). [↑](#footnote-ref-21)
2. Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annu Rev Public Health, 23*, 151-69. doi: 10.1146/annurev.publheath.23.100901.140546 <http://rctdesign.org/techreports/arphnonnormality.pdf> [↑](#footnote-ref-27)
3. Laerd statistics. (2018). [Testing for Normality using SPSS Statistics](https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php). [↑](#footnote-ref-29)
4. Normaliteit. (14 juli 2014). [UvA Wiki Methodologiewinkel](https://wiki.uva.nl/methodologiewinkel/index.php/Normaliteit). [↑](#footnote-ref-31)
5. Bereken de logaritme van de variabele en de inverse (1 gedeeld door de variabele), en bekijk of deze normaal verdeeld is. Als dat zo is, mag je die gebruiken voor de analyse. [↑](#footnote-ref-33)
6. Van Geloven, N. (13 maart 2018). *Mann-Whitney U toets* [Wiki Statistiek Academisch Medisch Centrum](https://wikistatistiek.amc.nl/index.php/Mann-Whitney_U_toets). [↑](#footnote-ref-35)
7. De [Mann-Whitney U toets](08-Mann-Whitney-U-toets-R.html) maakt een rangschikking van de data. Hierdoor is de test verdelingsvrij en is normaliteit geen assumptie. Ook zijn uitbijters minder van invloed op het eindresultaat. Toch wordt er voor deze test minder vaak gekozen, doordat bij het maken van een rankschikking de data informatie verliest. Als de data wel normaal verdeeld is, heeft de [Mann-Whitney U toets](08-Mann-Whitney-U-toets-R.html) minder onderscheidend vermogen dan wanneer de *ongepaarde t-toets* uitgevoerd zou worden. [↑](#footnote-ref-37)
8. Lakens, D. (26 januari 2015). *Always use Welch’s t-test instead of Student’s t-test*. [The 20% Statistician](http://daniellakens.blogspot.com/2015/01/always-use-welchs-t-test-instead-of.html). [↑](#footnote-ref-39)
9. Onderscheidend vermogen, in het Engels power genoemd, is de kans dat de nulhypothese verworpen wordt wanneer de alternatieve hypothese ‘waar’ is. [↑](#footnote-ref-41)
10. Lakens, D. (26 januari 2015). *Always use Welch’s t-test instead of Student’s t-test*. [The 20% Statistician](http://daniellakens.blogspot.com/2015/01/always-use-welchs-t-test-instead-of.html). [↑](#footnote-ref-42)
11. Outliers (13 augustus 2016). [UvA Wiki Methodologiewinkel](https://wiki.uva.nl/methodologiewinkel/index.php/Outliers). [↑](#footnote-ref-47)
12. Uitbijters kunnen bepalend zijn voor de uitkomst van toetsen. Bekijk of de uitbijters valide uitbijters zijn en niet een meetfout of op een andere manier incorrect verkregen data. Het weghalen van uitbijters kan de uitkomst ook vertekenen, daarom is het belangrijk om verwijderde uitbijters te melden in een rapport. [↑](#footnote-ref-49)
13. Outliers (13 augustus 2016). [UvA Wiki Methodologiewinkel](https://wiki.uva.nl/methodologiewinkel/index.php/Outliers). [↑](#footnote-ref-55)
14. Uitbijters kunnen bepalend zijn voor de uitkomst van toetsen. Bekijk of de uitbijters valide uitbijters zijn en niet een meetfout of op een andere manier incorrect verkregen data. Het weghalen van uitbijters kan de uitkomst ook vertekenen, daarom is het belangrijk om verwijderde uitbijters te melden in een rapport. [↑](#footnote-ref-57)
15. In dit voorbeeld wordt uitgegaan van een waarschijnlijkheid van 95% c.q. een p-waardegrens van 0,05. De grens is naar eigen inzicht aan te passen; houd hierbij rekening met type I en type II fouten. [↑](#footnote-ref-62)
16. Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage publications. [↑](#footnote-ref-64)
17. Marshall, E., & Boggis, E. (2016). *The statistics tutor’s quick guide to commonly used statistical tests*. <http://www.statstutor.ac.uk/resources/uploaded/tutorsquickguidetostatistics.pdf>. [↑](#footnote-ref-65)