



Statistiek in de alfa en gamma studies

Aansluiting wiskunde VWO-WO
16 april 2018



Universiteit Utrecht

Wie ben ik?

Marieke Westeneng

Docent bij afdeling Methoden en
Statistiek

Faculteit Sociale Wetenschappen
Universiteit Utrecht

M.A.A.Westeneng@uu.nl



Universiteit Utrecht



Curriculum

- Curriculum afgestemd op studenten en de rol die zij later krijgen
- Onderzoeker
- Gebruiker van onderzoek
- Sloppy science; kritisch kijken naar onderzoek



Kritisch lezen van artikelen

British Educational Research Journal
Vol. 37, No. 1, February 2011, pp. 105–129

 Routledge
Taylor & Francis Group

Three-year results of the Friendly Schools whole-of-school intervention on children's bullying behaviour

Donna Cross^{a*}, Helen Monks^a, Marg Hall^a, Thérèse Shaw^a,
Yolanda Pintabona^a, Erin Erceg^a, Greg Hamilton^a, Clare
Roberts^b, Stacey Waters^a and Leanne Lester^a

^a*Edith Cowan University, Child Health Promotion Research Centre, Perth, Australia;*

^b*Curtin University of Technology, Perth, Australia*



Universiteit Utrecht



Kritisch lezen van artikelen

Introduction

Bullying between students at school can seriously affect their social, physical and psychological well-being as well as their academic achievement. Students who are bullied, compared to those who are not, tend to experience poorer health, more somatic complaints and greater risk of injury (Williams *et al.*, 1996; Rigby, 1998b; Wolke *et al.*, 2001; Strabstein & Piazza, 2008; Gini & Pozzoli, 2009); poorer self-esteem (Rigby & Slee, 1991; Slee & Rigby 1993; Salmivalli *et al.*, 1999; Jankauskiene *et al.*, 2008); more interpersonal difficulties (Kumpulainen *et al.*, 1998); higher levels of loneliness (Kochenderfer & Ladd, 1996a; Forero *et al.*, 1999; Nansel *et al.*, 2001);



Kritisch lezen van artikelen

Table 1. Friendly Schools study design

Study condition	Baseline Grade 4 (Mar 2000)	Intervention (year 1)	Posttest 1 Grade 4 (Nov 2000)	Intervention (year 2)	Posttest 2 Grade 5 (Nov 2001)	Maintenance	Posttest 3 Grade 6 (Nov 2002)
Intervention	O ₁	X ₁	O ₂	X ₂	O ₃	X ₃	O ₄
Comparison	O ₁	X ₄	O ₂	X ₅	O ₃	X ₆	O ₄

Notes.

O = Student cohort data collection.

X_{1,2} = Whole-of-school bullying intervention (active school support and training).

X₃ = Maintenance of whole-of-school bullying intervention (passive school support).

X_{4,5} = Regular school bullying prevention program and road safety curriculum.

X₆ = Regular school bullying prevention program and road safety curriculum *plus* release of X_{1,2} to schools.

Table 2. Response rates of the longitudinal cohort of students from baseline to post-test 3

	Baseline (March 2000) <i>n</i> (%)	Posttest 1 (Nov 2000) <i>n</i> (%)	Posttest 2 (Nov 2001) <i>n</i> (%)	Posttest 3 (Nov 2002) <i>n</i> (%)
Intervention	1046 (100)	984 (94.1)	861 (82.3)	688 (65.8)
Comparison	922 (100)	863 (93.6)	775 (84.1)	688 (74.6)
Total	1968 (100)	1847 (93.9)	1636 (83.1)	1376 (69.9)



Kritisch lezen van artikelen

Table 4. Student baseline report of being bullied, bullying others, telling if bullied and seeing another Grade 4 or younger student bullied during last term

Bullying outcomes	Frequency	Intervention		Comparison		Total		<i>p</i> value
		<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	
Was bullied	Every few weeks or more often	168	(16.1)	152	(16.6)	320	(16.3)	.841
	1–2 times a term	262	(25.1)	220	(24.0)	482	(24.6)	
	Not at all	615	(58.9)	546	(59.5)	1161	(59.1)	
	Total	1045	(100)	918	(100)	1963	(100)	
Bullied others	Every few weeks or more often	27	(2.6)	28	(3.0)	55	(2.8)	.403
	1–2 times a term	108	(10.4)	111	(12.1)	219	(11.2)	
	Not at all	903	(87.0)	780	(84.9)	1683	(86.0)	
	Total	1038	(100)	919	(100)	1957	(100)	
Told if bullied	Told someone/ Was not bullied	889	(85.7)	778	(84.7)	1667	(85.2)	.505
	Told no one	148	(14.3)	141	(15.3)	289	(14.8)	
	Total	1037	(100)	919	(100)	1956	(100)	
Saw another bullied	Did not see	(531)	51.2	(516)	56.8	(1047)	53.8	.013
	Saw	(506)	48.8	(392)	43.2	(898)	46.2	
	Total	(1037)	100	(908)	100	(1945)	100	



Kritisch lezen van artikelen

when the students were still in Grade 4, students in the comparison group had an increased likelihood of being bullied vs. not bullied ($p = .004$). The comparison group students were 1.5 times more likely to be bullied than those in the intervention group with an effect size of -16% representing a 16% lower rate in the intervention compared to the comparison group, relative to the baseline rate. However, when comparing the probabilities of being bullied regularly in Grade 4, no significant differences between groups were observed at this time point ($p = .323$). No significant differences were found in Grade 5 between the study conditions, either with regard to being bullied ($p = .117$) or being bullied regularly ($p = .996$). At posttest 3, when in Grade 6, the students in the comparison group were no more likely to be bullied overall ($p = .133$) but had 1.5 times higher odds (an effect size of -31%) than the intervention group students of being bullied regularly ($p = .010$) with an effect size of -31% indicating a relative lower rate of 31% in the intervention compared to the comparison group.

No statistically significant differences were observed between the intervention and comparison group students regarding the frequency of bullying other students (Table 5).



Kritisch lezen van artikelen

Table 5. Binary logistic regression results for student responses to how often *they were bullied* and *bullied others* last term

Time point	Outcome category	Odds ratio ^a	95% confidence interval	<i>p</i> value ^b	Effect size ^c %
Posttest 1 <i>n</i> = 1834	<i>Bullied</i> every few weeks vs. Less often/Not	1.16	(.87; 1.54)	.323	-4.3
	<i>Bullied</i> vs. Not at all	1.49	(1.14; 1.94)	.004	-16.2
Posttest 2 <i>n</i> = 1624	<i>Bullied</i> every few weeks vs. Less often/Not	1.00	(.70; 1.43)	.996	0.6
	<i>Bullied</i> vs. Not at all	1.25	(.95; 1.64)	.117	-8.3
Posttest 3 <i>n</i> = 1359	<i>Bullied</i> every few weeks vs. Less often/Not	1.50	(1.10; 2.05)	.010	-30.6
	<i>Bullied</i> vs. Not at all	1.26	(.93; 1.71)	.133	-38.5
Posttest 1 <i>n</i> = 1827	<i>Bullied others</i> every few weeks vs. Less often/Not	.66	(.39; 1.13)	.134	39.3
	<i>Bullied others</i> vs. Not at all	.81	(.54; 1.20)	.295	9.3
Posttest 2 <i>n</i> = 1613	<i>Bullied others</i> every few weeks vs. Less often/Not	1.14	(.71; 1.83)	.580	-17.9
	<i>Bullied others</i> vs. Not at all	1.15	(.81; 1.63)	.432	-9.3
Posttest 3 <i>n</i> = 1357	<i>Bullied others</i> every few weeks vs. Less often/Not	.87	(.48; 1.59)	.657	39.3
	<i>Bullied others</i> vs. Not at all	1.02	(.75; 1.40)	.890	3.6

Notes. ^aOdds for comparison vs. intervention group adjusted for values for dependent variable at baseline, gender and school size. ^bBonferroni adjusted level of significance for group comparisons $\alpha = .025$. ^cEffect size calculated as difference in intervention and comparison rates at posttest expressed as a percentage of the average baseline rate across the two groups.



Kritisch lezen van artikelen

One possible explanation for the lack of program effects on reports of bullying others is that it may not be possible to change perpetrators' behaviour using only universal whole-school activities. This group of students may require greater support for behaviour change through selective and targeted approaches. Rigby and Slee (2008) propose that different treatments may be required depending on the severity of bullying, and the age, social and psychological characteristics of the children involved. While not widespread, some intervention schools in the current study reported successfully using the Method of Shared Concern (Pikas, 2002) in the second and third study years, as recommended in the FS intervention, to support students involved in bullying. Process data suggest intervention schools required more comprehensive training and support to implement this and other strategies to



Beschrijvende statistieken

- Populatie, steekproef
- Frequentietabel, proporties, percentages
- Grafieken
- Vorm van de verdeling (symmetrisch, scheve verdeling, bimodaal, uitschieters)
- Percentielen
- Meetniveaus
- Standaardnormaalverdeling, z-scores



Centrum- en spreidingsmaten

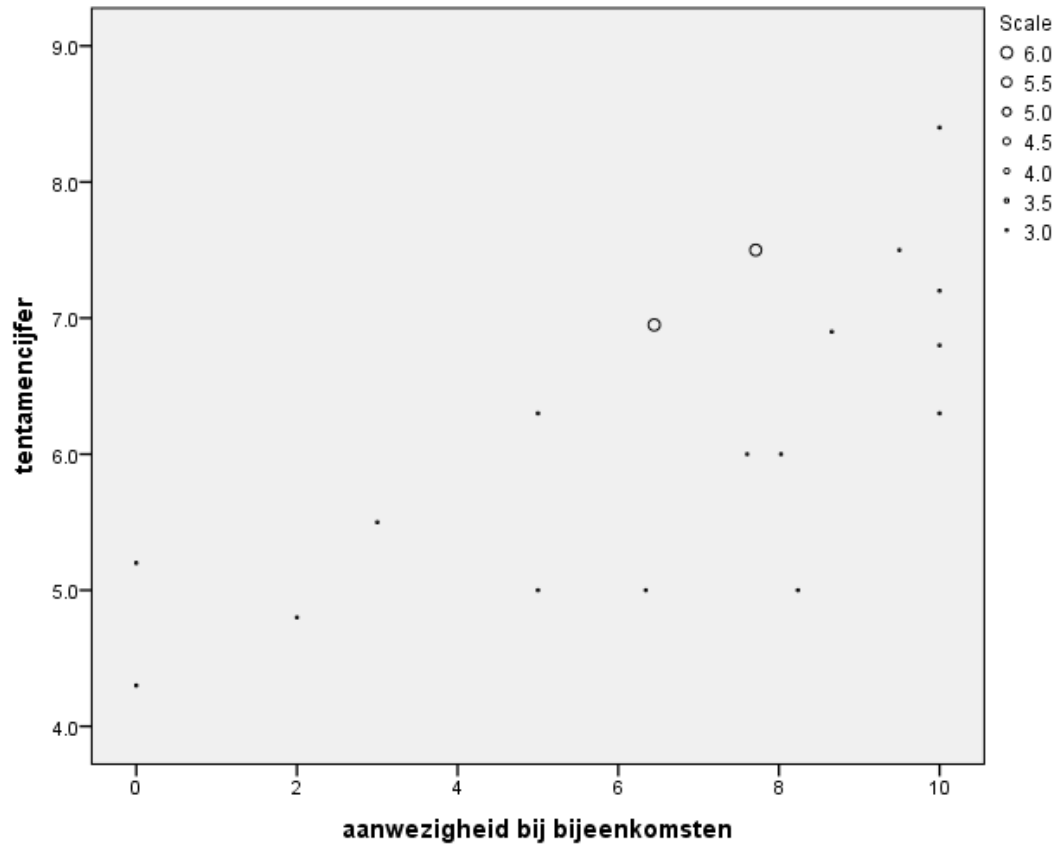
- Gemiddelde en standaardafwijking/ variantie

$$s^2 = \frac{\sum (X - M)^2}{n - 1}$$

- Mediaan en interkwartielafstand
- Modus en bereik



Correlatie



- Samenhang?
- Vorm?
- Richting?
- Sterkte?
- Uitschieters?



Voorbeeld

Een onderzoeker wil weten of kinderen die soms naar Sesamstraat kijken en kinderen die altijd naar Sesamstraat kijken van elkaar verschillen in hun gemiddelde kennis van getallen.

De onderzoeker heeft een onderzoeksgroep met 60 kinderen die soms naar Sesamstraat kijken en 62 kinderen die altijd naar Sesamstraat kijken. Alle kinderen maken een toets om hun kennis van getallen te meten.

De onderzoeker verwacht dat kinderen die altijd naar Sesamstraat kijken meer kennis hebben.



Hypothesetoetsen Stappen

- Stap 1: hypothesen bepalen en significantieniveau kiezen
- Stap 2: kritieke waarden bepalen
- Stap 3: H_0 uitgangspunt en de toetsingsgrootte uitrekenen
- Stap 4: beslissing nemen over H_0



Hypothesen

- Hypothesen

- $H_0: \mu_{\text{altijd}} = \mu_{\text{soms}}$

- $H_1: \mu_{\text{altijd}} > \mu_{\text{soms}}$ of

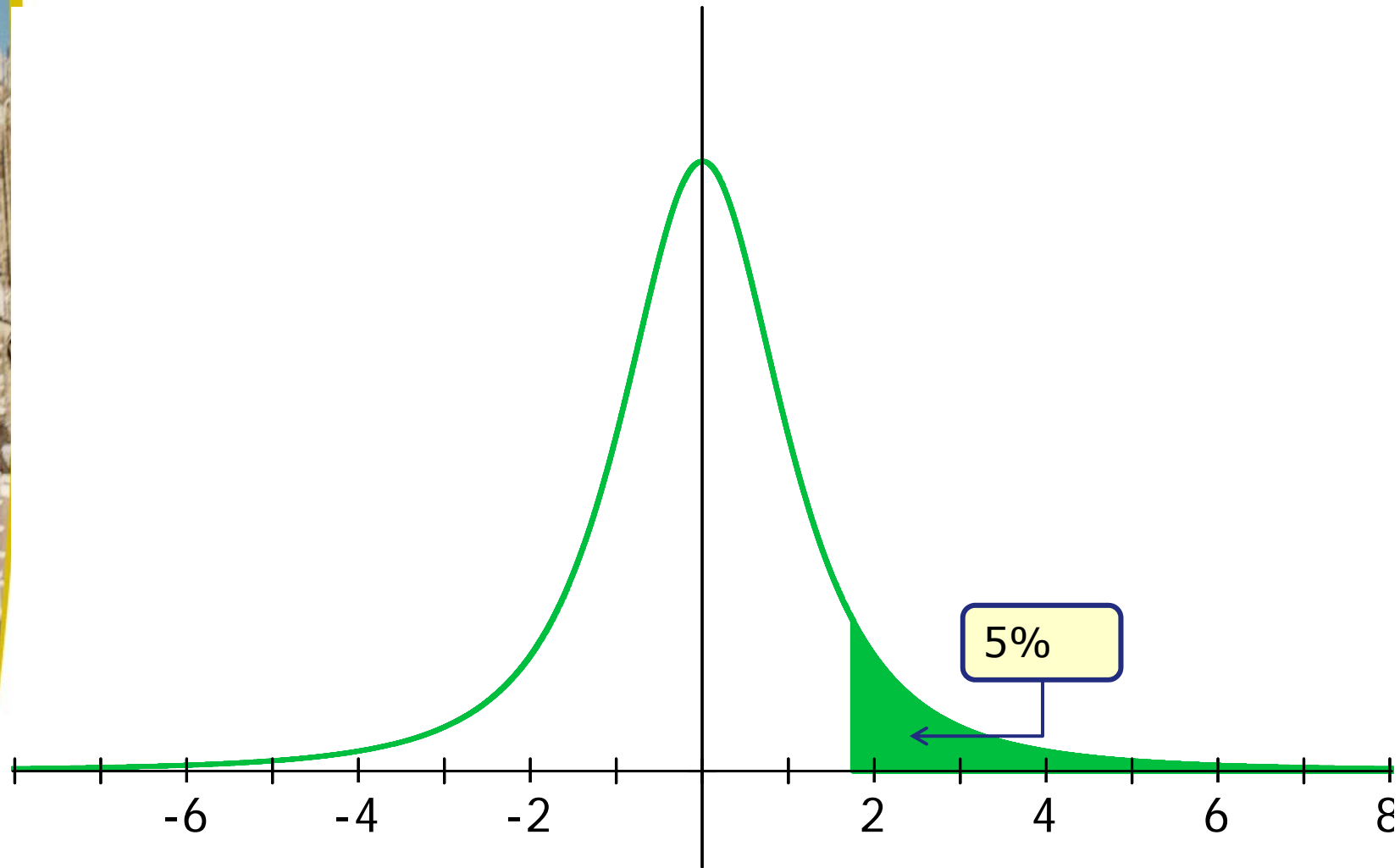
- $H_0: \mu_{\text{altijd}} - \mu_{\text{soms}} = 0$

- $H_1: \mu_{\text{altijd}} - \mu_{\text{soms}} > 0$

- Significantieniveau: $\alpha = .05$.



Steekproevenverdeling



Kritiek gebied -- voorbeeld

- o Aantal vrijheidsgraden = $(n_1 - 1) + (n_2 - 1) = 59 + 61 = 120$
- o $\alpha = .05$; éénzijdig

$$t_{\text{kritiek}} = 1.658$$

df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

Toetsingsgrootheid

Group Statistics

	Kijkgedrag	N	Mean	Std. Deviation	Std. Error Mean
Nameting kennis van getallen (0-54)	soms	60	27.50	11.527	1.488
	altijd	62	36.77	11.076	1.407

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Nameting kennis van getallen (0-54)	Equal variances assumed	.034	.854	-4.532	120	.000	-9.274	2.046	-13.326	-5.223
	Equal variances not assumed			-4.529	119.365	.000	-9.274	2.048	-13.329	-5.220



Toetsingsgrootheid

De toetsingsgrootheid is een t -waarde.

- t voor 1 gemiddelde:

$$t = \frac{M - \mu_0}{S_M}$$

- t voor gekoppelde metingen:

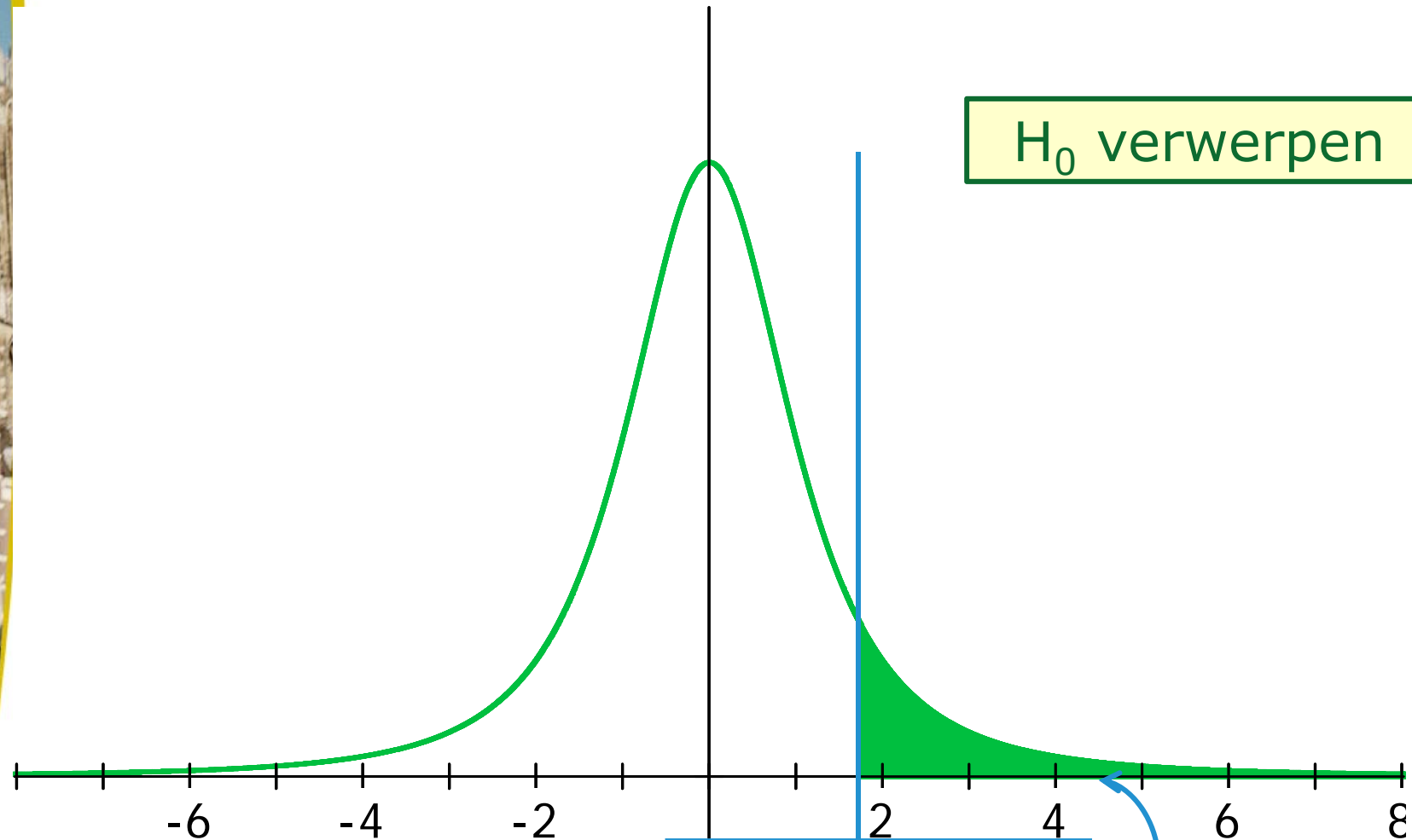
$$t = \frac{M_D - \mu_D}{S_{M_D}}$$

- t voor twee onafh. groepen:

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S_{(M_1 - M_2)}}$$



Beslissing over H_0



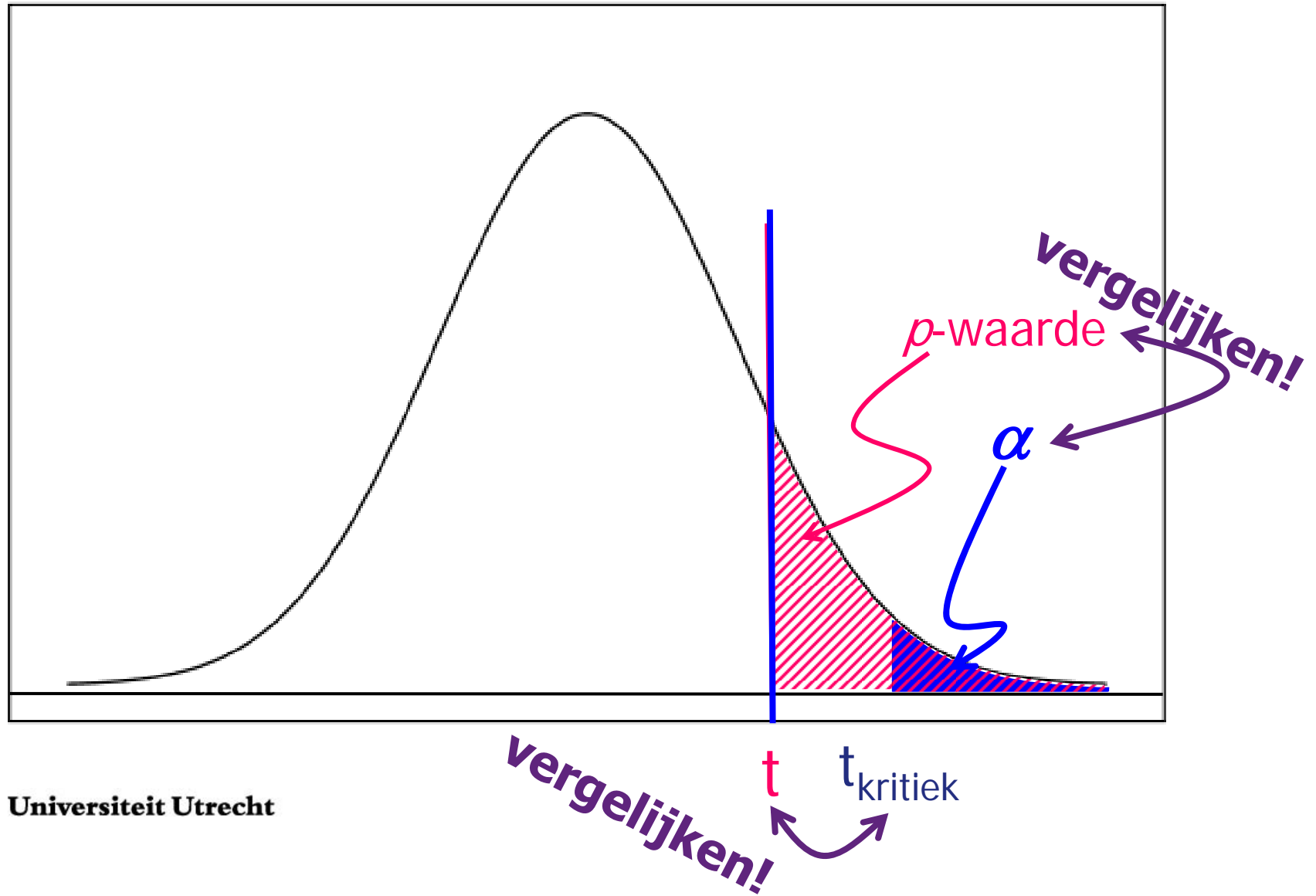
H_0 verwerpen

$$t_{\text{kritiek}} = 1.658$$

$$t = 4.532$$



p -waarde? kritieke waarde?



Conclusie

Kinderen die altijd naar Sesamstraat kijken hebben gemiddeld meer kennis van getallen dan kinderen die soms naar Sesamstraat kijken.

Geschatte Cohen's d

$$d = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2}} = \frac{(36.77 - 27.50) - 0}{\sqrt{127.6898}} = 0.82$$





Betrouwbaarheidsinterval

BI voor verschil tussen twee gemiddelden
(onafhankelijke steekproeven)

$$M_1 - M_2 \pm t \cdot s_{M_1 - M_2}$$

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Nameting kennis van getallen (0-54)	Equal variances assumed	.034	.854	-4.532	120	.000	-9.274	2.046	-13.326	-5.223
	Equal variances not assumed			-4.529	119.365	.000	-9.274	2.048	-13.329	-5.220





SPSS

- Werken met formules
- Werken met grote databestanden



Toetsing

Onderzoekers van de Faculteit Sociale Wetenschappen hebben een onderzoek uitgevoerd naar de invloed van het kijken naar Sesamstraat op de kennis die kinderen hebben van letters, lichaamsdelen, vormen en getallen.

Aan het onderzoek hebben 180 kinderen meegewerkt in de leeftijd van 2 t/m 5 jaar oud. Van deze kinderen keken 54 kinderen nooit naar Sesamstraat; 126 kinderen keken regelmatig naar Sesamstraat.

Maarten onderzoekt of er verschil is tussen de kinderen die nooit naar Sesamstraat kijken en de kinderen die regelmatig naar Sesamstraat kijken wat betreft hun kennis van letters. Een hogere score op de variabele *Kennis van Letters* (0 - 58) duidt erop dat kinderen meer kennis hebben van letters. Hij voert alle toetsen uit bij $\alpha = .05$. Hieronder zie je zijn SPSS Output:

Group Statistics

	Kijkgedrag	N	Mean	Std. Deviation	Std. Error Mean
Nameting kennis van letters (0-58)	nooit	54	16,91	9,481	1,290
	regelmatig	124	22,11	13,831	1,242

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Nameting kennis van letters (0-58)	Equal variances assumed	13,628	,000	-2,518	176	,013	-5,205	2,067	-9,285	-1,126
	Equal variances not assumed			-2,907	143,602	,004	-5,205	1,791	-8,745	-1,666



Toetsing

Group Statistics

	Kijkgedrag	N	Mean	Std. Deviation	Std. Error Mean
Nameting kennis van letters (0-58)	nooit	54	16,91	9,481	1,290
	regelmatig	124	22,11	13,831	1,242

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Nameting kennis van letters (0-58)	Equal variances assumed	13,628	,000	-2,518	176	,013	-5,205	2,067	-9,285	-1,126
	Equal variances not assumed			-2,907	143,602	,004	-5,205	1,791	-8,745	-1,666

De alternatieve hypothese is dat kinderen die regelmatig naar Sesamstraat kijken *meer* kennis hebben van letters dan kinderen die nooit naar Sesamstraat kijken. Deze hypothese wordt *eenzijdig* getoetst bij $\alpha = .05$.

Welke uitspraak is juist bij bovenstaande gegevens?

- Het resultaat is significant, maar Maarten kan een type I fout hebben gemaakt.
- Het resultaat is significant, maar Maarten kan een type II fout hebben gemaakt.
- Het resultaat is niet significant, maar Maarten kan een type I fout hebben gemaakt.
- Het resultaat is niet significant, maar Maarten kan een type II fout hebben gemaakt.



SPSS toets

Naam: _____

Studentnummer: _____ Datum: _____

1	Het aantal respondenten waarvan het geboortjaar bekend is, is...	
2	De gemiddelde score op de totaalscore ZIL op meetmoment 1 is...	
3	De maximaal behaalde score op <u>hyperarousal</u> is...	
4	De mediaan van uren slaap is...	
	Voer een <u>t-toets</u> voor twee afhankelijke steekproeven uit, ter beantwoording van de vraag of de gemiddelde totaalscore ZIL op <i>meetmoment 1</i> verschilt van de gemiddelde totaalscore ZIL op <i>meetmoment 2</i> , en vraag het 80%- betrouwbaarheidsinterval op.	
5	De <i>t</i> -waarde is...	
6	De bovengrens van het betrouwbaarheidsinterval is...	
	Voer een <u>Spearman</u> rangcorrelatietoets uit, ter beantwoording van de vraag of er een positief verband is tussen wakker worden met een droge mond en wakker worden met een zure smaak in de mond.	
7	De waarde van r_s is...	
8	De <i>p</i> -waarde is...	
	Voer een <u>regressie-analyse</u> uit waarbij je <i>uren slaap</i> voorspelt aan de hand van ZIL totaalscore meetmoment 2.	
9	De waarde van de standaardschattingfout van dit model is...	
10	De <u>regressiecoëfficiënt</u> van het effect van <i>ZIL totaalscore meetmoment 2</i> is ...	
	Voeg <i>sleepimpact</i> toe als voorspeller aan het regressiemodel.	
11	Wat is de VIF-waarde voor <i>ZIL totaalscore meetmoment 2</i> ?	
12	De <i>p</i> -waarde voor de toets of <i>sleepimpact</i> significant is in het model is gelijk aan...	



Universiteit Utrecht

Tools in ons onderwijs

- Online platform voor zelfstudie
- [Shiny-apps](https://utrecht-university.shinyapps.io/correlation/) (<https://utrecht-university.shinyapps.io/correlation/>)
 - Tools voor begrip van statistische maten
- [Kennisclips](https://www.youtube.com/channel/UCshOousN2WkuIps-JhhKgoA)
 - (<https://www.youtube.com/channel/UCshOousN2WkuIps-JhhKgoA>)



Verder in de bachelor

- Verschillen kinderen in hun gemiddelde score op de voormeting kennis van getallen en nameting kennis van getallen?
 - *t*-toets voor afhankelijke groepen
- Is er een verschil in gemiddelde kennis van letters tussen kinderen die nooit, soms, regelmatig of altijd naar Sesamstraat kijken?
 - ANOVA



Verder in de bachelor

- Verschillen jongens en meisjes in hun gemiddelde kennis van letters op de nameting, wanneer we corrigeren voor kennis van letters op de voormeting?
 - ANCOVA
- Is er een verschil tussen voor- en nameting kennis van vormen? En is dit verschil anders voor jongens dan voor meisjes?
 - Herhaalde metingen/ mixed design





Verder in de bachelor

- Is er samenhang tussen leeftijd en kennis van lichaamsdelen?
 - **Correlatie**
- Kunnen we kennis van lichaamsdelen voorspellen door middel van leeftijd?
 - **Enkelvoudige regressie**
 - **Uitbreiden naar multipele regressie, logistische regressie**



Verder in de bachelor

- Is de relatie tussen mate van stress en cijfer verschillend voor mannen en vrouwen?
 - Moderator-analyse
- Verloopt de relatie tussen voorlezen in de klas en leesvaardigheid, via de hoeveelheid uren die leerlingen zelf lezen?
 - Mediatie-analyse



Vragen?



Universiteit Utrecht