

OPLOSSINGEN BIJ HOOFDSTUK 5

- De onderzoekers van een preventiedienst vermoeden dat werknemers in een bedrijf zonder liften fitter zijn dan werknemers in een bedrijf mét liften. Ze nemen de proef op de som en selecteren willekeurig twaalf werknemers uit een bedrijf zonder en twaalf werknemers uit een bedrijf mét liften. Ze onderwerpen alle deelnemers aan een looptest waarbij men eenvoudigweg vraagt om zo lang mogelijk rondjes te lopen op een piste, zonder te stoppen. De geregistreeerde tijden van alle deelnemers vind je terug in het bestand *liftgebruik.sav*. Uit eerdere tests is gebleken dat de tijden van deze looptest normaal verdeeld zijn in de populatie.

Met lift: 30, 35, 45, 40, 50, 35, 55, 25, 30, 45, 40, 50

Zonder lift: 40, 35, 50, 55, 65, 55, 50, 35, 30, 50, 60, 39

Kan je beweren dat de deelnemers uit het bedrijf zonder liften fitter zijn dan de deelnemers uit het bedrijf mét liften? Waarom (niet)?

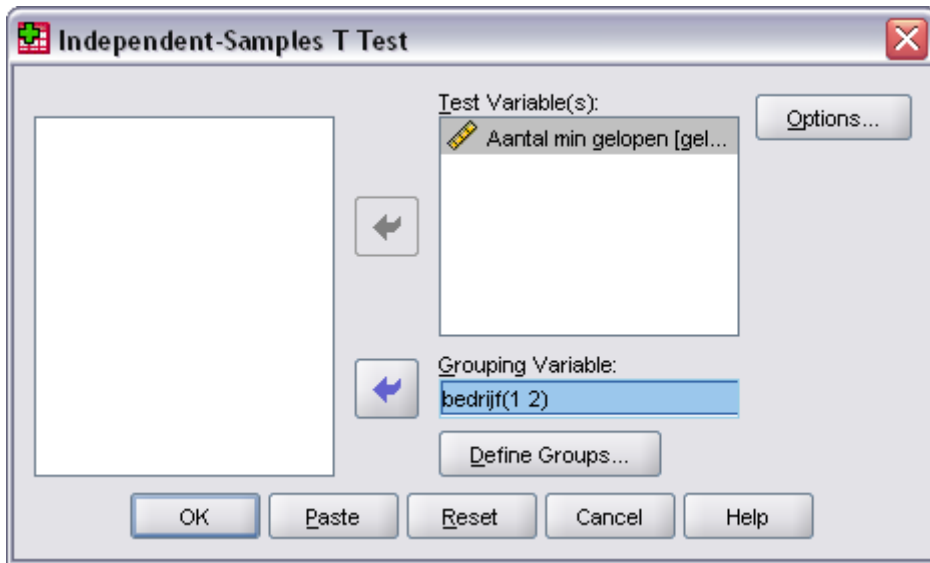
Het gaat om in totaal 24 deelnemers, die verdeeld zijn in 2 groepen. We willen weten of er een significant verschil is tussen de gemiddelden van de twee groepen op de tijd die ze gelopen hebben. Er zijn dus twee variabelen: de onafhankelijke variabele is het al dan niet beschikken over een lift; de afhankelijke variabele is de tijd die gelopen werd. De onafhankelijke variabele is gemeten op nominaal niveau, de afhankelijke variabele is gemeten op intervalniveau.

Beide groepen zijn duidelijk niet gelinkt aan elkaar en er vindt geen herhaalde meting plaats. We hebben dus te maken met een t-toets voor onafhankelijke steekproeven. Er is geen indicatie om éézijdig te toetsen, dus we toetsen tweezijdig. Voorstelling van de data in SPSS:

	bedrijf	gelopen
1	1	30
2	1	35
3	1	45
4	1	40
5	1	50
6	1	35
7	1	55
8	1	25
9	1	30
10	1	45
11	1	40
12	1	50
13	2	40
14	2	35
15	2	50
16	2	55
17	2	65
18	2	55
19	2	50
20	2	35
21	2	30
22	2	50
23	2	60
24	2	39

Het uitvoeren van de t-toets:

Analyse > compare means > Independent samples T-test



De verkregen output:

Group Statistics

		bedrijf	N	Mean	Std. Deviation	Std. Error Mean
Aantal min gelopen	met lift		12	40,00	9,293	2,683
	zonder lift		12	47,00	11,029	3,184

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
Aantal min gelopen	Equal variances assumed	,782	,386	-1,681	22	,107	-7,000	4,163	-15,634	1,634
	Equal variances not assumed			-1,681	21,385	,107	-7,000	4,163	-15,649	1,649

Voor deze toets:

H0: $\mu_1 = \mu_2$

H1: $\mu_1 \neq \mu_2$

De Levene's test toont aan dat de varianties niet significant verschillen van elkaar. We kunnen dus naar de rij "Equal variances assumed" kijken. Daarin vinden we een t-waarde van -1.681 en een p-waarde van .107. Dus:

$t(22) = -1.681, p > 0.05.$

Er blijkt geen significant verschil te zijn tussen beide groepen deelnemers.

Rapportering:

Gemiddeld werd er langer gelopen door de werknemers uit een bedrijf zonder lift ($M = 47$) dan door werknemers uit een bedrijf met lift ($M = 40$), maar dit verschil was niet significant, $t(22) = -1.681$, $p > 0.05$, $r = 0.34$.

We kunnen dus op basis van deze test niet concluderen dat de deelnemers uit het bedrijf zonder liften fitter zijn dan de deelnemers uit het bedrijf mét liften.

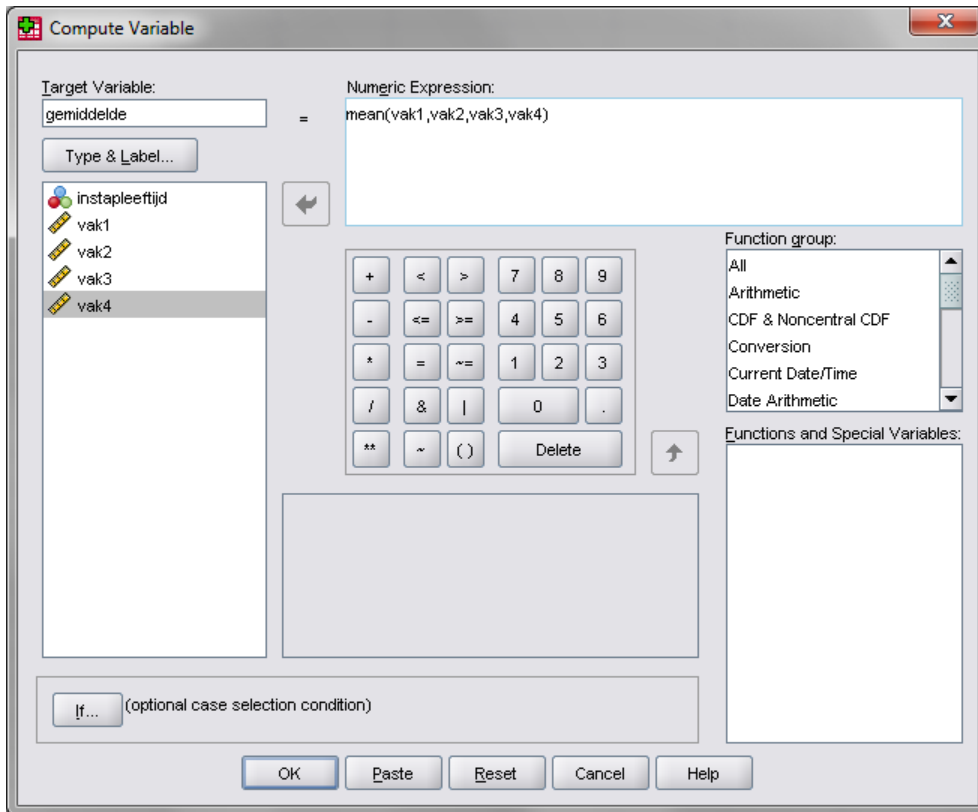
2. De directeur van een lagere school wil nagaan of de leerlingen uit het eerste leerjaar die in de kleuterklas zijn ingestapt vóór de leeftijd van 3 jaar betere resultaten halen dan leerlingen die na de leeftijd van 3 jaar zijn ingestapt. Hij verzamelt hiervoor de punten van de leerlingen op 4 vakken.
Je vindt de punten van de leerlingen in het bestand *instapleeftijd.sav*. Bereken voor elke leerling de gemiddelde score op de vier vakken (gebruik het *Compute*-commando!). Ga vervolgens na of er een verschil is tussen beide groepen leerlingen op gebied van deze scores. Je mag ervan uitgaan dat de scores normaal verdeeld zijn in de populatie.

Er is sprake van 2 variabelen: de afhankelijke variabele is de gemiddelde score op de vier vakken (die we eerst nog zullen moeten berekenen). De onafhankelijke variabele is de instapleeftijd in de kleuterklas. Aangezien deze OV 2 niveaus heeft, en de steekproeven onafhankelijk zijn van elkaar, komen we terecht bij de t-toets voor onafhankelijke steekproeven. Een Z-toets is niet mogelijk omdat de σ 's niet bekend zijn en de steekproeven te klein zijn om deze te schatten aan de hand van de steekproefstandaarddeviatie. We mogen wel parametrisch toetsen want we gaan ervan uit dat de scores normaal verdeeld zijn. De cases zijn de leerlingen.

Beide groepen zijn duidelijk niet gelinkt aan elkaar en er vindt geen herhaalde meting plaats. We hebben dus te maken met een t-toets voor onafhankelijke steekproeven.

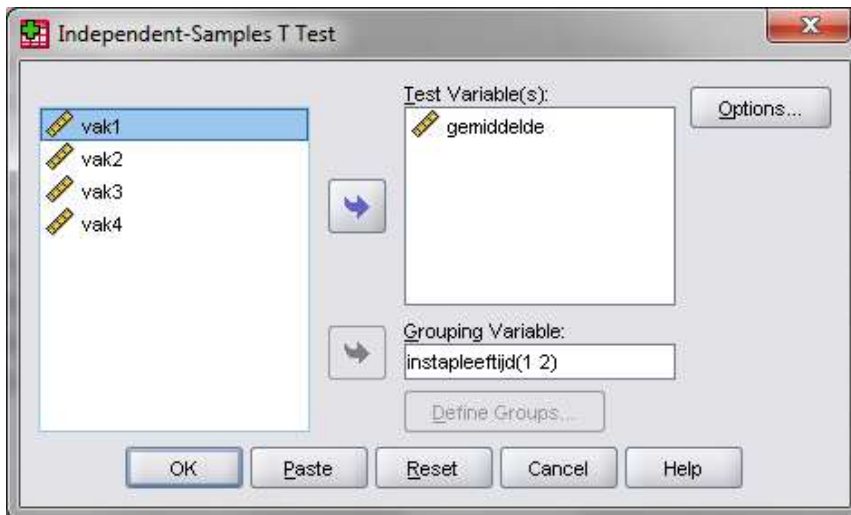
Berekening van de gemiddelde score:

Transform > Compute



Toetsen van het verschil:

Analyze > compare means > Independent samples T-test



Group Statistics

	instapleeftijd	N	Mean	Std. Deviation	Std. Error Mean
gemiddelde	voor 3 jaar	10	71,6750	11,27377	3,56508
	na 3 jaar	11	60,7500	7,56142	2,27985

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
gemiddelde	1,123	,303	2,631	19	,016	10,92500	4,15192	2,23494	19,61506
			2,582	15,529	,020	10,92500	4,23173	1,93197	19,91803

Voor deze toets:

H0: $\mu_1 = \mu_2$

H1: $\mu_1 \neq \mu_2$

De Levene's test toont aan dat de varianties niet significant verschillen van elkaar. We kunnen dus naar de rij "Equal variances assumed" kijken. Daarin vinden we een t-waarde van 2.63 en een p-waarde van .016.

$t(19) = 2.63, p = 0.016$.

Er blijkt dus een significant verschil te zijn tussen beide groepen deelnemers.

Rapportering:

Gemiddeld halen de leerlingen die in de kleuterklas instapten voor de leeftijd van 3 jaar hogere cijfers in het eerste leerjaar ($M = 71.68$) dan leerlingen die na de leeftijd van 3 jaar instapten ($M = 60.75$). Dit verschil is significant op .05-niveau ($t(19) = 2.63, p = 0.016, r = 0.52$).

- Een pedagoog wil nagaan of er een verschil is tussen twee manieren om bijscholingscursussen te geven aan volwassenen. In groep 1 zal methode A gebruikt worden terwijl groep 2 aan methode B wordt onderworpen. Alvorens de efficiëntie van beide methoden te kunnen meten, wil de onderzoeker nagaan of beide groepen evenwaardig zijn wat betreft opleidingsniveau, zodat dit geen storende variabele kan zijn in het interpreteren van de resultaten. In onderstaande tabel vind je de opleidingsniveaus van de deelnemers per groep.

Hoogst behaalde diploma	
Groep 1	Groep 2
middelbare school	master
master	master
bachelor	middelbare school
master	bachelor
master	bachelor
bachelor	doctor
middelbare school	master
doctor	master
master	bachelor

Is er een verschil in opleidingsniveau tussen beide groepen? Wat betekent dat voor de verdere analyse van de efficiëntie van de twee methoden?

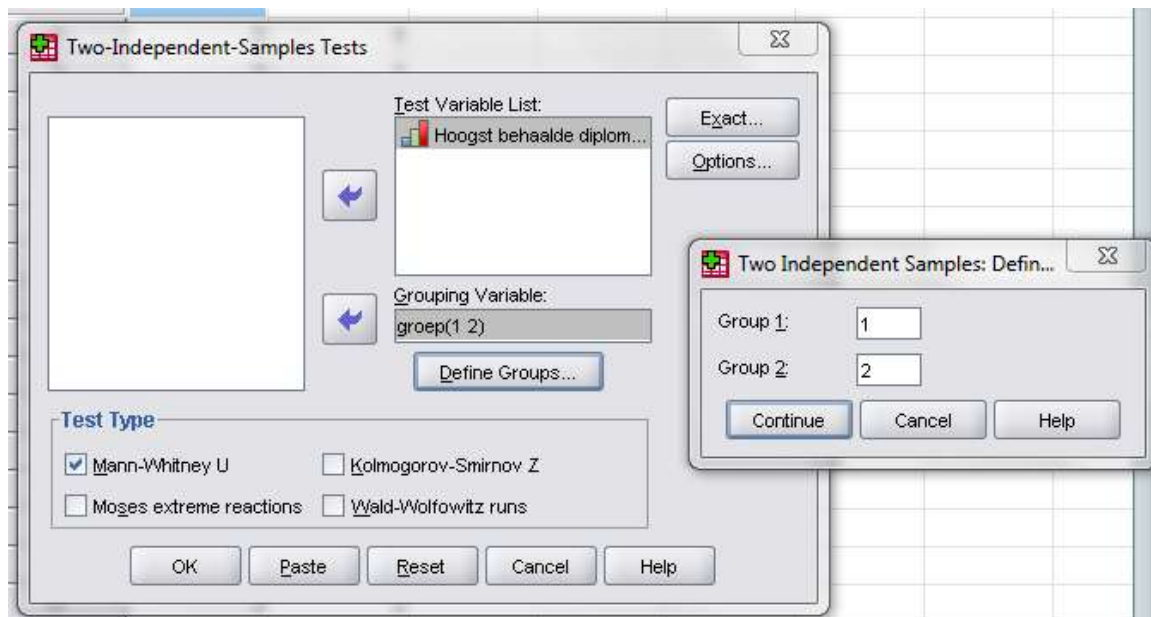
We hebben te maken met een nominale onafhankelijke variabele (groep). De afhankelijke variabele van dit deel van de analyse is het opleidingsniveau van de deelnemers, een ordinale variabele. We bestuderen twee populaties, waarbij de steeproeven onafhankelijk van elkaar zijn samengesteld. Aangezien de afhankelijke variabele ordinaal is, kunnen we niet parametrisch toetsen en zullen we een Mann-Whitney of Wilcoxon Rank Sum toets uitvoeren.

We geven de data in zoals hieronder voorgesteld. Merk op dat de opleidingsniveaus onder elkaar worden ingevoerd en een tweede variabele zorgt voor de indeling in groepen.

1 : oplniveau		1
	oplniveau	groep
1	1	1
2	3	1
3	2	1
4	3	1
5	3	1
6	2	1
7	1	1
8	4	1
9	3	1
10	3	2
11	3	2
12	1	2
13	2	2
14	2	2
15	4	2
16	3	2
17	3	2
18	2	2
19		

Toetsen van het verschil:

Analyse > nonparametric tests > 2 independent samples



De verkregen output:

Ranks

	groep	N	Mean Rank	Sum of Ranks
Hoogst behaalde diploma	1	9	9,28	83,50
	2	9	9,72	87,50
	Total	18		

Test Statistics^b

	Hoogst behaalde diploma
Mann-Whitney U	38,500
Wilcoxon W	83,500
Z	-,187
Asymp. Sig. (2-tailed)	,851
Exact Sig. [2*(1-tailed Sig.)]	,863 ^a

a. Not corrected for ties.

b. Grouping Variable: groep

Voor deze toets:

$$H_0: \theta_1 = \theta_2 \quad H_1: \theta_1 \neq \theta_2$$

We zien in de eerste tabel dat er weinig verschil is tussen de sum of ranks. De tweede tabel toont inderdaad ook dat er geen significant effect is. De nulhypothese wordt dus niet verworpen.

Er blijkt dus geen significant verschil te zijn tussen beide groepen deelnemers.

Rapportering:

Om na te gaan of het opleidingsniveau gelijk is in beide groepen van deelnemers werd een Mann-Whitney toets uitgevoerd. Het opleidingsniveau was iets hoger in groep 2 dan in groep 1, maar dit verschil was niet significant op $\alpha = .05$ -niveau, $U = 38.5$, $z = -0.187$, $p = .851$, $r = -.04$.