

## Hoofdstuk 13. Lineaire regressieanalyse – Oefenreeks – Oplossingen

### Oefening 13.1. Regressie.

**Antwoord:** 8.7 % van de variantie in depressieve gevoelens wordt voorspeld door het model met alleen sociodemografische predictoren. Voegen we echter twee individuele predictoren toe aan het model, angst en coronastress, dan stijgt de verklaarde variantie significant met 38.9 procentpunten naar 47.6 %.

**Stap 1.** Kiezen van de juiste analysemethode.

**1. Hoeveel variabelen bevat mijn onderzoeksraag?**

Zes variabelen, namelijk depressie, leeftijd, geslacht, relatiestatus, angst en coronastress.

**2. Welk meetniveau hebben de variabelen in mijn onderzoeksraag?**

‘W1\_Depressie1’ tot en met ‘W1\_Depressie3’: drie ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, ‘W1\_Depressie\_Gem’

‘W1\_Leeftijd’: metrische variabele (2020-W1\_Gebaar)

‘W1\_Geslacht’ en ‘W1\_Relatiestatus’: twee nominale variabelen met telkens twee categorieën

‘W1\_Angst1’ tot en met ‘W1\_Angst6’: zes ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, ‘W1\_Angst\_Gem’

‘W1\_Corstress1’ tot en met ‘W1\_Corstress3’: drie ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, ‘W1\_Corstress\_Gem’

**3. Welke rol vervullen de variabelen?**

De variabelen leeftijd, geslacht, relatiestatus, angst en coronastress zijn de onafhankelijke variabelen, waarvan we het effect willen kennen op de afhankelijke variabele depressie: we gebruiken de onafhankelijke variabelen dus als predictoren om de uitkomstvariabele ‘W1\_Depressie\_Gem’ te verklaren/voorspellen.

Flowchart:

Eén afhankelijke variabele → Metrisch → Twee of meer onafhankelijke variabelen → Beide

→ Meervoudige lineaire regressie of ANCOVA.

Kijk naar de verwoording van de onderzoeksraag: aangezien het hier gaat om het voorspellen van een variabele aan de hand van andere variabelen (predictoren) en niet om het onderzoeken van de verschillen tussen twee of meer groepen (gecontroleerd voor een derde variabele), is **meervoudige lineaire regressie** de juiste analysetechniek. Omdat de predictoren bloksgewijs ingebracht worden, gaat het hier meer specifiek om hiërarchische meervoudige regressie.

**Stap 2.** Datacontrole.

Inconsistentiecheck:

#### W1\_Depressie1 DEPRESSIE - ...was ik gelukkig

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Zelden of nooit	54	2,7	2,7
	2 Soms of weinig	457	22,6	22,6
	3 Regelmatig	1.006	49,8	75,1
	4 Meestal of altijd	503	24,9	100,0
	Total	2.020	100,0	100,0

#### W1\_Depressie2 DEPRESSIE - ...had ik plezier in het leven

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Zelden of nooit	55	2,7	2,7
	2 Soms of weinig	565	28,0	28,0
	3 Regelmatig	1.010	50,0	50,0
	4 Meestal of altijd	390	19,3	19,3
	Total	2.020	100,0	100,0

#### W1\_Depressie3 DEPRESSIE - ...voelde ik me droevig

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Zelden of nooit	493	24,4	24,4
	2 Soms of weinig	823	40,7	40,7
	3 Regelmatig	643	31,8	31,8
	4 Meestal of altijd	61	3,0	3,0
	Total	2.020	100,0	100,0

Gezien de grote hoeveelheid aan variabelen, tonen we hier slechts diegene waarbij iets opvalt. De twee eerste items over depressie moeten omgekeerd worden, zodat een hogere score meer depressiviteit betekent. Daarna kunnen we de nodige metrische variabelen aanmaken, namelijk leeftijd, angst, coronastress en depressie. Vooraf kun je het best nog nagaan of de schalen angst, coronastress en depressie intern consistent zijn aan de hand van een Cronbach's alpha toets, maar dit valt buiten de scope van deze oefening.

RECODE W1\_Depressie1 (1=4) (2=3) (3=2) (4=1) (**MISSING=SYSMIS**) INTO W1\_Depressie1\_Recode.

VARIABLE LABELS W1\_Depressie1\_Recode 'Omkering van W1\_Depressie1'.

EXECUTE.

RECODE W1\_Depressie2 (1=4) (2=3) (3=2) (4=1) (**MISSING=SYSMIS**) INTO W1\_Depressie2\_Recode.

VARIABLE LABELS W1\_Depressie2\_Recode 'Omkering van W1\_Depressie2'.

EXECUTE.

COMPUTE W1\_Depressie\_Gem=**MEAN**(W1\_Depressie1\_Recode,W1\_Depressie2\_Recode,W1\_Depressie3).

VARIABLE LABELS W1\_Depressie\_Gem 'Gemiddelde score op drie items rond depressie'.

EXECUTE.

COMPUTE W1\_Leeftijd=**MEAN**(2020-W1\_Gebjaar).

VARIABLE LABELS W1\_Leeftijd 'Leeftijd respondenten'.

EXECUTE.

COMPUTE W1\_Angst\_Gem=**MEAN**(W1\_Angst1,W1\_Angst2,W1\_Angst3,W1\_Angst4,W1\_Angst5,W1\_Angst6).

VARIABLE LABELS W1\_Angst\_Gem 'Gemiddelde score op zes items over angstgevoelens'.

EXECUTE.

COMPUTE W1\_Corstress\_Gem=**MEAN**(W1\_Corstress1,W1\_Corstress2,W1\_Corstress3).

VARIABLE LABELS W1\_Corstress\_Gem 'Gemiddelde score op drie items over coronastress'.

EXECUTE.

Assumptiecheck:

1. De afhankelijke variabele (of uitkomstvariabele) is van metrisch meetniveau. **IN ORDE**
2. Er zijn twee of meer onafhankelijke variabelen (of predictoren). **IN ORDE**
3. Onafhankelijkheid: de residuele termen zijn niet gecorreleerd met elkaar. **NAGAAN IN OUTPUT REGRESSIONANALYSE**
4. Lineariteit: de afhankelijke variabele heeft een collectief lineair verband met alle predictoren. **NAGAAN VIA RESIDUENANALYSE**
5. Normaliteit: de residuele termen zijn normaal verdeeld. **NAGAAN VIA RESIDUENANALYSE**

6. Homoscedasticiteit: de residuele termen hebben homogene (of anders gezegd: gelijke) varianties voor elke waarde van de afhankelijke variabele. **NAGAAN VIA RESIDUENANALYSE**
7. Er zijn geen uitschieters onder de residuele termen (de gestandaardiseerde residuele termen zijn niet groter dan drie standaardafwijkingen). **NAGAAN VIA RESIDUENANALYSE**
8. Geen multicollineariteit, waarbij twee of meer predictoren te sterk gecorreleerd zijn met elkaar.  
**NAGAAN IN OUTPUT REGRESSIEANALYSE**

**Stap 3.** Hypothesen formuleren. Dit moet hier voor (1) het model met de eerste blok predictoren, (2) elke predictor van model 1, (3) het model met de tweede blok predictoren, (4) elke predictor van model 2 en (5) het verschil in verklaringskracht tussen model 1 en model 2. Hier doen we dit bij wijze van voorbeeld enkel voor de beide modellen en het verschil tussen beide:

#### **Verklaringskracht model 1**

$H_0$ : *Leeftijd, geslacht en relatiestatus hebben geen effect op depressie: de verklaringskracht van dit lineair model is gelijk aan nul.*

$H_a$ : *Leeftijd, geslacht en relatiestatus hebben wel een effect op depressie: de verklaringskracht van dit lineair model is groter of kleiner dan nul.*

#### **Verklaringskracht model 2**

$H_0$ : *Leeftijd, geslacht, relatiestatus, angst en coronastress hebben geen effect op depressie: de verklaringskracht van dit lineair model is gelijk aan nul.*

$H_a$ : *Leeftijd, geslacht, relatiestatus, angst en coronastress hebben wel een effect op depressie: de verklaringskracht van dit lineair model is groter of kleiner dan nul.*

#### **Verschil in verklaringskracht tussen model 1 en 2**

$H_0$ : *Leeftijd, geslacht en relatiestatus (model 1) hebben evenveel effect op depressie als leeftijd, geslacht, relatiestatus, angst en coronastress (model 2): de verklaringskracht van beide modellen is gelijk aan elkaar.*

$H_a$ : *De predictoren uit model 1 hebben niet evenveel effect op depressie als de predictoren uit model 2: de verklaringskracht van model 2 is groter dan die van model 1.*

**Stap 4.** Significantieniveau bepalen.

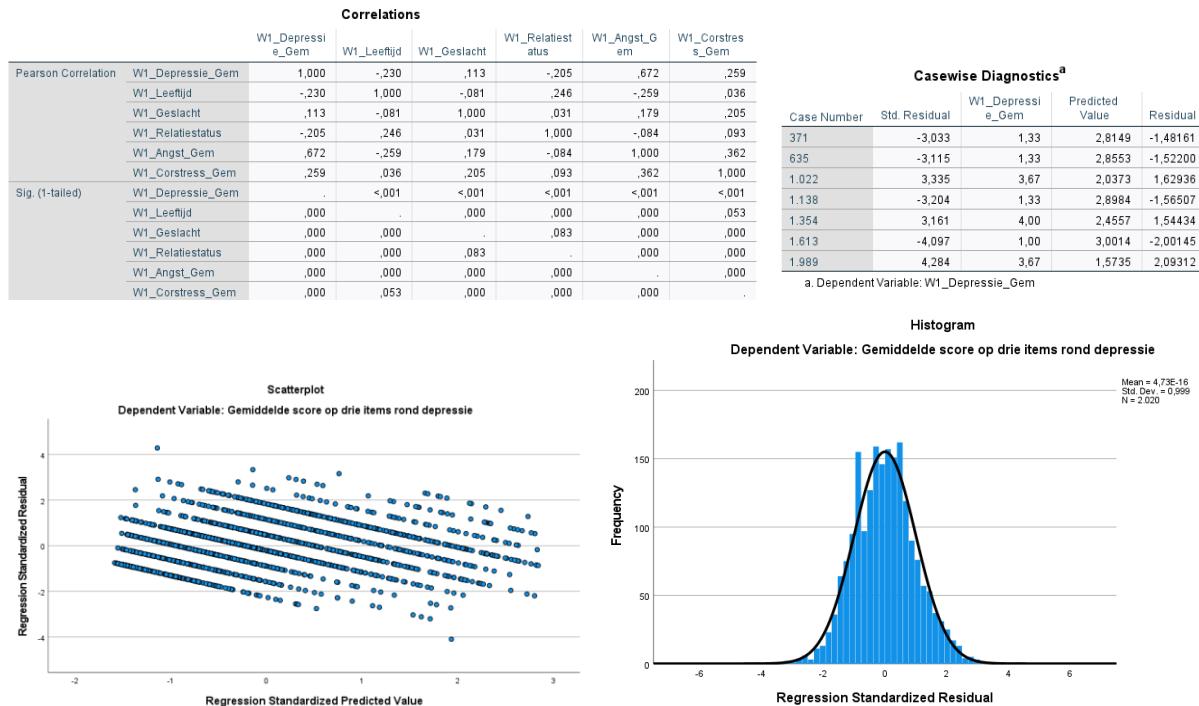
Het gaat om een steekproef met meer dan 400 onderzoekseenheden ( $n = 2020$ ), dus hanteer  $\alpha = .01$  of 1 %.

**Stap 5.** Toetsingsgrootte en p-waarde berekenen.

Doorloop de reeks SPSS-acties in het boek om het volgende commando te verkrijgen:

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT W1_Depressie_Gem
/METHOD=ENTER W1_Leeftijd W1_Geslacht W1_Relatiestatus
/METHOD=ENTER W1_Angst_Gem W1_Corstress_Gem
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESiduals DURBIN HISTOGRAM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

**Output.** Controleer eerst de overige zes assumpties aan de hand van de correlatietafel, VIF-waarden (assumptie 8) en de output voor residuenanalyse (assumptie 3-7). **Assumptie 8 is oké**, want geen enkele correlatie tussen twee predictoren is groter dan .60 en geen enkele VIF-waarde is groter dan 5. Verder zijn **assumptie 4, 5 en 6 ook oké**, want de scatterplot vertoont geen non-lineair verband (lineariteit in orde), de datapunten vormen een ovaal en geen kegel- of zandlopervorm (homoscedasticiteit in orde) en het histogram van de residuen vertoont een normale verdelingsvorm (normaliteit in orde). Bij *Casewise Diagnostics* zie je wel dat twee uitschieters in het bijzonder (nummer 1613 en 1989) voor problemen kunnen zorgen. Die zou je beter bestuderen om na te gaan of die uit de analyse verwijderd kunnen worden, maar net zoals in het boek zullen we om didactische redenen **assumptie 7 rond uitschieters als vervuld beschouwen**. Tot slot ligt de toetsingsgroothed van de Durbin-Watson toets (zie *Model Summary* tabel) dicht bij 2, waardoor **assumptie 3 rond onafhankelijkheid van residuele termen ook in orde is**.



Model Summary <sup>c</sup>										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics				Durbin-Watson
						F Change	df1	df2	Sig. F Change	
1	,295 <sup>a</sup>	,087	,085	,64426	,087	63,905	3	2.016	<.001	
2	,690 <sup>b</sup>	,475	,474	,48853	,389	746,075	2	2.014	<.001	1,970

a. Predictors: (Constant), W1\_Relatiestatus, W1\_Geslacht, W1\_Leeftijd

b. Predictors: (Constant), W1\_Relatiestatus, W1\_Geslacht, W1\_Leeftijd, W1\_Corstress\_Gem, W1\_Angst\_Gem

c. Dependent Variable: W1\_Depressie\_Gem

ANOVA <sup>a</sup>						Coefficients <sup>a</sup>										
Model	Sum of Squares	df	Mean Square	F	Sig.	Unstandardized Coefficients	Standardized Coefficients		95,0% Confidence Interval for B		Collinearity Statistics					
							B	Std. Error	t	Sig.						
1	Regression	79,576	3	26,525	63,905	<.001 <sup>b</sup>	(Constant)	.2486	.052	47,460	,000	2,383	2,589			
	Residual	836,793	2,016	,415			W1_Leeftijd	-.010	.001	-.181	<.001	-.013	-,008	,931	1,074	
	Total	916,368	2,019				W1_Geslacht	,164	.034	,104	,849	<.001	,098	,231	,991	1,009
	Regression	435,699	5	87,140	365,115	<.001 <sup>c</sup>	W1_Relatiestatus	-.243	.033	-,163	-,7433	<.001	-,307	-,179	,937	1,067
	Residual	480,670	2,014	,239												
2	Total	916,368	2,019													

a. Dependent Variable: W1\_Depressie\_Gem

b. Predictors: (Constant), W1\_Relatiestatus, W1\_Geslacht, W1\_Leeftijd

c. Predictors: (Constant), W1\_Relatiestatus, W1\_Geslacht, W1\_Leeftijd, W1\_Corstress\_Gem, W1\_Angst\_Gem

Model	Unstandardized Coefficients	Standardized Coefficients	95,0% Confidence Interval for B			Collinearity Statistics				
	B	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF		
1	(Constant)	1,088	,059	18,520	<.001	,973	1,203			
	W1_Leeftijd	-,002	,001	-,031	-,179	,073	-,004	,000	,866	1,154
	W1_Geslacht	-,012	,026	-,008	-,471	,638	-,064	,039	,940	1,063
	W1_Relatiestatus	-,220	,025	-,148	-,8,818	,<.001	-,269	-,171	,927	1,078
	W1_Angst_Gem	,556	,016	,636	34,943	<.001	,525	,587	,785	1,273
2	W1_Corstress_Gem	,034	,013	,045	2,523	,012	,007	,060	,822	1,216

a. Dependent Variable: W1\_Depressie\_Gem

## Stap 6. Conclusie en rapportering.

“Om de invloed van twee groepen onafhankelijke variabelen op depressie te toetsen, werd gebruik gemaakt van een hiërarchische meervoudige regressieanalyse. In een basismodel (model 1) werden drie sociodemografische predictoren (leeftijd, geslacht en relatiestatus) opgenomen, waarna in een tweede model twee bijkomende © 2022, Toegepaste Statistiek – Wim Hardyns, Koen Ponnet en Pelckmans Uitgevers nv

individuele predictoren toegevoegd werden (angst en coronastress). Hoewel model 1 ( $R^2 = .087$ ,  $R^2_{Adjusted} = .085$ ,  $F(3, 2016) = 63.91$ ,  $p < .001$ ) en model 2 ( $R^2 = .475$ ,  $R^2_{Adjusted} = .474$ ,  $F(5, 2014) = 365.12$ ,  $p < .001$ ) beide een significant deel van de variantie in depressie verklaren, is model 2 een stuk sterker. We zien dat model 2 door de toevoeging van het ervaren van angst en coronastress 38,9 procentpunten meer van de totale variantie in depressie verklaart ( $\Delta R^2 = .389$ ,  $F(2, 2014) = 746.08$ ,  $p < .001$ ) dan model 1.”

### Oefening 13.2. Regressie.

**Antwoord:** Indien we de assumpties van de regressieanalyse als vervuld beschouwen, dan kunnen we stellen dat zowel ouderschap als geslacht een significant effect heeft op het ervaren van angstgevoelens. De analyse toont aan dat ouders minder angstgevoelens ervaren dan respondenten zonder kinderen en dat vrouwen in hogere mate angst ervaren in vergelijking met mannen. Van beide predictoren heeft geslacht het grootste effect op angst.

**Stap 1.** Kiezen van de juiste analysemethode.

**1. Hoeveel variabelen bevat mijn onderzoeksraag?**

Drie variabelen, namelijk ouderschap, geslacht en angst.

**2. Welk meetniveau hebben de variabelen in mijn onderzoeksraag?**

‘W1\_Ouder’ en ‘W1\_Geslacht’: twee nominale variabelen met telkens twee categorieën

‘W1\_Angst1’ tot en met ‘W1\_Angst6’: zes ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, ‘W1\_Angst\_Gem’

**3. Welke rol vervullen de variabelen?**

De variabelen ouderschap en geslacht zijn de onafhankelijke variabelen, waarvan we het effect willen kennen op de afhankelijke variabele angst: we gebruiken de onafhankelijke variabelen dus als predictoren om de uitkomstvariabele ‘W1\_Angst\_Gem’ te verklaren/voorspellen.

Flowchart:

Eén afhankelijke variabele → Metrisch → Twee of meer onafhankelijke variabelen → Metrisch

→ Meervoudige lineaire regressie of partiële correlatietoets.

Kijk naar de verwoording van de onderzoeksraag: aangezien het hier gaat om het voorspellen van een variabele aan de hand van andere variabelen (predictoren) en niet om het onderzoeken van de samenhang tussen twee variabelen (gecontroleerd voor een derde variabele), is **meervoudige lineaire regressie** de juiste analysetechniek.

**Stap 2.** Datacontrole.

Inconsistentiecheck:

W1_Ouder						W1_Geslacht					
	Frequency	Percent	Valid Percent	Cumulative Percent		Frequency	Percent	Valid Percent	Cumulative Percent		
Valid	0 Nee	1.109	54,9	60,5	60,5						
	1 Ja	724	35,8	39,5	100,0						
Total		1.833	90,7	100,0							
Missing	999	187	9,3								
Total		2.020	100,0								

W1_Angst1					W1_Angst2				
	Frequency	Percent	Valid Percent	Cumulative Percent		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Helemaal niet	446	22,1	22,1	22,1	897	44,4	44,4	44,4
	2 Meerdere dagen, maar minder dan de helft	927	45,9	45,9	68,0	685	33,9	33,9	78,3
	3 Meer dan de helft van de dagen	386	19,1	19,1	87,1	273	13,5	13,5	91,8
	4 Bijna elke dag	261	12,9	12,9	100,0	165	8,2	8,2	100,0
	Total	2.020	100,0	100,0		2.020	100,0	100,0	

W1_Angst3					W1_Angst4				
	Frequency	Percent	Valid Percent	Cumulative Percent		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Helemaal niet	512	25,3	25,3	25,3	677	33,5	33,5	33,5
	2 Meerdere dagen, maar minder dan de helft	883	43,7	43,7	69,1	773	38,3	38,3	71,8
	3 Meer dan de helft van de dagen	389	19,3	19,3	88,3	349	17,3	17,3	89,1
	4 Bijna elke dag	236	11,7	11,7	100,0	221	10,9	10,9	100,0
	Total	2.020	100,0	100,0		2.020	100,0	100,0	

W1_Angst5					W1_Angst6				
	Frequency	Percent	Valid Percent	Cumulative Percent		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Helemaal niet	1.119	55,4	55,4	55,4	587	29,1	29,1	29,1
	2 Meerdere dagen, maar minder dan de helft	557	27,6	27,6	83,0	848	42,0	42,0	71,0
	3 Meer dan de helft van de dagen	219	10,8	10,8	93,8	369	18,3	18,3	89,3
	4 Bijna elke dag	125	6,2	6,2	100,0	216	10,7	10,7	100,0
	Total	2.020	100,0	100,0		2.020	100,0	100,0	

Geen inconsistenties, behalve de 187 ontbrekende waarden bij ouderschap: die zijn afkomstig van de respondenten die aangeduid hebben dat ze alleen wonen, waardoor ze de vraag van ouderschap niet moesten beantwoorden. Deze respondenten zullen door SPSS automatisch niet meegenomen worden in de analyse. Nu moet je alleen nog de metrische schaalvariabele 'W1\_Angst\_Gem' aanmaken. Vooraf kun je het best nog nagaan of de schaal intern consistent is aan de hand van een Cronbach's alpha toets, maar dit valt buiten de scope van deze oefening.

COMPUTE W1\_Angst\_Gem=MEAN(W1\_Angst1,W1\_Angst2,W1\_Angst3,W1\_Angst4,W1\_Angst5,W1\_Angst6).  
VARIABLE LABELS W1\_Angst\_Gem 'Gemiddelde score op zes items over angstgevoelens'.

EXECUTE.

Assumptiecheck:

1. De afhankelijke variabele (of uitkomstvariabele) is van metrisch meetniveau. **IN ORDE**
2. Er zijn twee of meer onafhankelijke variabelen (of predictoren). **IN ORDE**
3. Onafhankelijkheid: de residuele termen zijn niet gecorreleerd met elkaar. **NAGAAN IN OUTPUT REGRESSIONANALYSE**
4. Lineariteit: de afhankelijke variabele heeft een collectief lineair verband met alle predictoren. **NAGAAN VIA RESIDUENANALYSE**
5. Normaliteit: de residuele termen zijn normaal verdeeld. **NAGAAN VIA RESIDUENANALYSE**
6. Homoscedasticiteit: de residuele termen hebben homogene (of anders gezegd: gelijke) varianties voor elke waarde van de afhankelijke variabele. **NAGAAN VIA RESIDUENANALYSE**
7. Er zijn geen uitschieters onder de residuele termen (de gestandaardiseerde residuele termen zijn niet groter dan drie standaardafwijkingen). **NAGAAN VIA RESIDUENANALYSE**
8. Geen multicollineariteit, waarbij twee of meer predictoren te sterk gecorreleerd zijn met elkaar. **NAGAAN IN OUTPUT REGRESSIONANALYSE**

**Stap 3.** Hypothesen formuleren. Dit moet hier voor (1) het volledige model met alle predictoren en (2) elke predictor van het model apart.

#### **Verklaringskracht model**

$H_0$ : *Ouderschap en geslacht hebben geen effect op angst: de verklaringskracht van dit lineair model is gelijk aan nul.*

$H_a$ : *Ouderschap en geslacht hebben wel een effect op angst: de verklaringskracht van dit lineair model is groter of kleiner dan nul.*

#### **Verklaringskracht predictor ouderschap**

$H_0$ : *Ouderschap heeft geen effect op angst: de verklaringskracht van ouderschap is gelijk aan nul.*

$H_a$ : *Ouderschap heeft wel een effect op angst: de verklaringskracht van ouderschap is groter of kleiner dan nul.*

#### **Verklaringskracht predictor geslacht**

$H_0$ : *Geslacht heeft geen effect op angst: de verklaringskracht van geslacht is gelijk aan nul.*

$H_a$ : *Geslacht heeft wel een effect op angst: de verklaringskracht van geslacht is groter of kleiner dan nul.*

**Stap 4.** Significantieniveau bepalen.

Het gaat om een steekproef met meer dan 400 onderzoekseenheden ( $n = 1833$ ), dus hanteer  $\alpha = .01$  of 1 %.

**Stap 5.** Toetsingsgrootte en p-waarde berekenen.

Doorloop de reeks SPSS-acties in het boek om het volgende commando te verkrijgen:

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT W1_Angst_Gem  
/METHOD=ENTER W1_Ouder W1_Geslacht  
/SCATTERPLOT=(*ZRESID ,*ZPRED)  
/RESIDUALS DURBIN HISTOGRAM(ZRESID)  
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

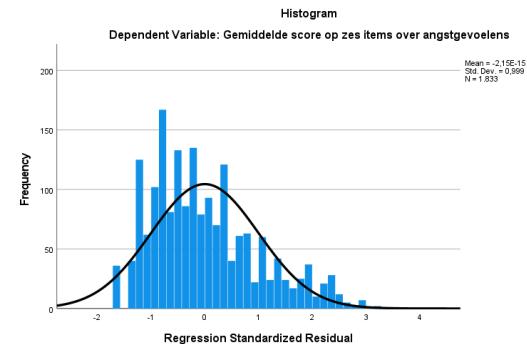
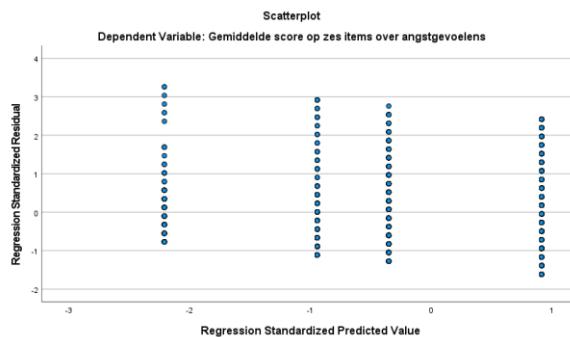
**Output.** Controleer eerst de overige zes assumpties aan de hand van de correlatietafel, VIF-waarden (assumptie 8) en de output voor residuenanalyse (assumptie 3-7). Als we **assumptie 5 (normaliteit van residuen)** visueel controleren, kun je twijfelen of er sprake is van een normale verdeling. Het histogram heeft namelijk links van het midden een paar extreme pieken en rechts een langere staart, waardoor die als rechts scheef bestempeld wordt en niet als normaal verdeeld beschouwd kan worden. We kunnen als tweede stap kijken naar de z-scores van scheefheid en kurtosis van de residuen om een besluit te maken inzake normaliteit (zie hieronder). Ook de scatterplot valt moeilijk te interpreteren omdat er enkel binaire variabelen als predictoren gebruikt werden, waardoor we ook geen zekerheid hebben over lineariteit en homoscedasticiteit.

Correlations					
	W1_Angst_Gem	W1_Ouder	W1_Geslacht		
Pearson Correlation	W1_Angst_Gem	1,000	-,165	,204	
	W1_Ouder	-,165	1,000	-,019	
	W1_Geslacht	,204	-,019	1,000	
Sig. (1-tailed)	W1_Angst_Gem		<,001	<,001	
	W1_Ouder	,000		,209	
	W1_Geslacht	,000	,209		
N	W1_Angst_Gem	1.833	1.833	1.833	
	W1_Ouder	1.833	1.833	1.833	
	W1_Geslacht	1.833	1.833	1.833	

### Casewise Diagnostics<sup>a</sup>

Case Number	Std. Residual	W1_Angst_Gem	Predicted Value	Residual
1.848	3,036	3,83	1,5746	2,25878
1.862	3,260	4,00	1,5746	2,42545
1.898	3,260	4,00	1,5746	2,42545

a. Dependent Variable: W1\_Angst\_Gem



### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,260 <sup>a</sup>	,068	,067	,74407	1,979

a. Predictors: (Constant), W1\_Geslacht, W1\_Ouder

b. Dependent Variable: W1\_Angst\_Gem

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	73,559	2	36,780	66,432	<,001 <sup>b</sup>
	Residual	1.013,172	1.830	,554		
	Total	1.086,731	1.832			

a. Dependent Variable: W1\_Angst\_Gem

b. Predictors: (Constant), W1\_Geslacht, W1\_Ouder

### Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error				Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	1,828	,040	46,208	,000	1,750	1,906		
	W1_Ouder	-,254	,036	-,7130	<,001	-,323	-,184	1,000	1,000
	W1_Geslacht	,372	,042	,8921	<,001	,290	,454	1,000	1,000

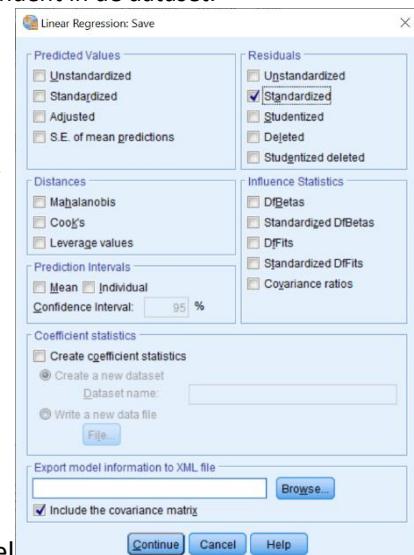
a. Dependent Variable: W1\_Angst\_Gem

### Assumptie 5. Normaliteit van residuen controleren a.d.h.v. z-scores scheefheid en kurtosis:

Actie 1. Vink tijdens het uitvoeren van het regressiecommando in de dialoogvensters onder de knop Save de optie Standardized aan in het kader Residuals. Je kunt ook eenvoudigweg in de Syntax het lijtje /SAVE ZRESID toevoegen bij het regressiecommando onderaan. Je vraagt hiermee aan SPSS om een nieuwe variabele aan te maken, waarin de gestandaardiseerde residuen van het regressiemodel opgeslagen worden van elke respondent in de dataset.

#### REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT W1_Angst_Gem
/METHOD=ENTER W1_Ouder W1_Geslacht
/SCATTERPLOT=*ZRESID ,*ZPRED
/RESIDUALS DURBIN HISTOGRAM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE ZRESID.
```



Actie 2. Voer het bovenstaande regressiecommando uit. Een nieuwe variabele met de gestandaardiseerde residuen zal nu verschijnen in je databestand, genaamd 'ZRE\_1' (of een ander nummer dan 1, afhankelijk van of je al eerder andere residuen opgeslagen hebt).

Actie 3. Vraag via het *Frequencies*-commando scheefheid, kurtosis en hun standaardfouten op van de variabele 'ZRE\_1'.

Statistics		
ZRE_1		
N	Valid	1.833
	Missing	187
Skewness	,764	
Std. Error of Skewness	,057	
Kurtosis	,022	
Std. Error of Kurtosis	,114	

Actie 4. Bereken de z-scores van scheefheid en kurtosis. Je ziet hier dat de z-score van scheefheid buiten de grenzen van +/- 1.96 valt, waardoor we kunnen besluiten dat de residuen niet normaal verdeeld zijn.

$$z_{\text{scheefheid}} = 0.76 / 0.06 = 12.67 \quad z_{\text{kurtosis}} = 0.02 / 0.11 = 0.18$$

De CLS geldt hier niet, dus moet je andere oplossingen hanteren om aan de vijfde assumptie te kunnen voldoen (bv. de variabele transformeren of het toepassen van bootstrapping, zie hoofdstuk 10). In principe kun je hier geen lineaire regressie toepassen, maar om didactische redenen zullen we in deze oefening de assumpties even als vervuld beschouwen en de resultaten interpreteren.

**Stap 6.** Conclusie en rapportering. De vrijheidsgraden van de t-toetsen van elke predictor vind je als volgt:  $df = n - k$ , met  $k$  = aantal variabelen in de regressieanalyse (hier 3). **LET OP, deze rapportering is bij wijze van voorbeeld. Aangezien o.a. de normaliteitsassumptie niet vervuld is van de lineaire regressieanalyse, is de analysetechniek niet helemaal betrouwbaar en interpreteer je de resultaten beter niet.**

"Het effect van ouderschap en geslacht als predictoren van angst werden gezamenlijk getoetst aan de hand van een meervoudig lineair regressiemodel. Dit lineair model verklaart een significant deel van de variantie in de scores van angst, met  $R^2 = .068$ ,  $R^2_{Adjusted} = .067$ ,  $F(2, 1830) = 66.43$ ,  $p < .001$ . Het effect van geslacht op angst was in dit model groter ( $\beta = 0.20$ ,  $t(1830) = 8.92$ ,  $p < .001$ ) dan het effect van ouderschap ( $\beta = -0.16$ ,  $t(1830) = -7.13$ ,  $p < .001$ )."

### Oefening 13.3. Regressie.

**Antwoord:** Indien we de assumpties van de regressieanalyse als vervuld beschouwen, dan kunnen we stellen dat het voorgestelde model (met financiële stress, het hebben van een gezondheidsprobleem en geslacht als predictoren) een significant deel van de variantie in coronastress (5 %) verklaart. Van alle predictoren heeft geslacht het grootste significante effect op coronastress, gevolgd door financiële stress, terwijl het hebben van een gezondheidsprobleem geen significante predictor blijkt te zijn.

#### Stap 1. Kiezen van de juiste analysemethode.

##### 1. Hoeveel variabelen bevat mijn onderzoeksfrage?

Vier variabelen, namelijk financiële stress, het hebben van een gezondheidsprobleem, geslacht en coronastress.

##### 2. Welk meetniveau hebben de variabelen in mijn onderzoeksfrage?

'W1\_FINSTRESS1' tot en met 'W1\_FINSTRESS3': drie ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W1\_FINSTRESS\_Gem'

'W1\_Handicap' en 'W1\_Geslacht': twee nominale variabelen met telkens twee categorieën

'W1\_Corstress1' tot en met 'W1\_Corstress3': drie ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W1\_Corstress\_Gem'

##### 3. Welke rol vervullen de variabelen?

De variabelen financiële stress, het hebben van een gezondheidsprobleem en geslacht zijn de onafhankelijke variabelen, waarvan we het effect willen kennen op de afhankelijke variabele coronastress: we gebruiken de onafhankelijke variabelen dus als predictoren om de uitkomstvariabele 'W1\_Corstress\_Gem' te verklaren/voorspellen.

Flowchart:

Eén afhankelijke variabele → Metrisch → Twee of meer onafhankelijke variabelen → Metrisch

→ Meervoudige lineaire regressie of partiële correlatietoets.

Kijk naar de verwoording van de onderzoeksfrage: aangezien het hier gaat om het voorspellen van een variabele aan de hand van andere variabelen (predictoren) en niet om het onderzoeken van de samenhang tussen twee variabelen (gecontroleerd voor een derde variabele), is **meervoudige lineaire regressie** de juiste analysetechniek.

#### Stap 2. Datacontrole.

Inconsistentiecheck:

W1_Handicap					W1_Geslacht				
	Frequency	Percent	Valid Percent	Cumulative Percent		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Nee	1.827	90,4	90,4	90,4	0 Man	479	23,7	23,7
	1 Ja	193	9,6	9,6	100,0	1 Vrouw	1.541	76,3	76,3
Total		2.020	100,0	100,0		Total	2.020	100,0	100,0

**W1\_FINSTRESS1 FINANCIËLE STRESS - Met mijn/ons huidig inkomen is het moeilijk om veel meer te veroorloven dan de basisbenodigdheden**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	772	38,2	59,6
	2 Eerder niet akkoord	266	13,2	80,1
	3 Nog niet akkoord/noch akkoord	134	6,6	90,4
	4 Eerder akkoord	89	4,4	97,3
	5 Akkoord	35	1,7	100,0
	Total	1.296	64,2	100,0
Missing	999	724	35,8	
Total	2.020	100,0		

**W1\_FINSTRESS2 FINANCIËLE STRESS - Ik heb het gevoel dat mijn/ons huidig inkomen toelaat een levensstandaard te behouden die ik wens**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	29	1,4	2,2
	2 Eerder niet akkoord	88	4,4	6,8
	3 Nog niet akkoord/noch akkoord	100	5,0	7,7
	4 Eerder akkoord	450	22,3	34,7
	5 Akkoord	629	31,1	48,5
	Total	1.296	64,2	100,0
Missing	999	724	35,8	
Total	2.020	100,0		

**W1\_FINSTRESS3 FINANCIËLE STRESS - Met mijn/ons huidig inkomen is het moeilijk om rond te komen (dit wil zeggen 'de eindjes aan elkaar te knopen')**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	880	43,6	67,9
	2 Eerder niet akkoord	247	12,2	79,0
	3 Nog niet akkoord/noch akkoord	101	5,0	79,8
	4 Eerder akkoord	55	2,7	99,0
	5 Akkoord	13	,6	100,0
	Total	1.296	64,2	100,0
Missing	999	724	35,8	
Total	2.020	100,0		

**W1\_Corstress1 CORONASTRESS - Ik maak mij zorgen dat ik met het coronavirus besmet zal raken**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	247	12,2	12,2
	2 Eerder niet akkoord	549	27,2	39,4
	3 Nog niet akkoord, noch akkoord	380	18,8	58,2
	4 Eerder akkoord	617	30,5	88,8
	5 Akkoord	227	11,2	100,0
	Total	2.020	100,0	100,0

**W1\_Corstress2 CORONASTRESS - Ik maak mij zorgen dat mijn naasten met het coronavirus besmet zullen raken**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	51	2,5	2,5
	2 Eerder niet akkoord	164	8,1	8,1
	3 Nog niet akkoord, noch akkoord	215	10,6	10,6
	4 Eerder akkoord	869	43,0	43,0
	5 Akkoord	721	35,7	35,7
	Total	2.020	100,0	100,0

**W1\_Corstress3 CORONASTRESS - Ik vind de huidige coronasituatie zeer stresserend**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	108	5,3	5,3
	2 Eerder niet akkoord	336	16,6	16,6
	3 Nog niet akkoord, noch akkoord	427	21,1	21,1
	4 Eerder akkoord	757	37,5	37,5
	5 Akkoord	392	19,4	19,4
	Total	2.020	100,0	100,0

Er ontbreken 724 respondenten bij de items over financiële stress, maar dat zijn de respondenten die nog student zijn en die deze vragen niet beantwoord hebben. Die groep zal door SPSS automatisch niet meegenomen worden in de regressieanalyse. Alvorens de schaalvariabele van financiële stress aan te maken, moet het tweede item omgekeerd worden zodat een hogere score meer financiële stress betekent. Vooraf kun je het best nog nagaan of de schaal intern consistent is aan de hand van een Cronbach's alpha toets, maar dit valt buiten de scope van deze oefening.

RECODE W1\_FINSTRESS2 (1=5) (2=4) (3=3) (4=2) (5=1) (**MISSING=SYSMIS**) INTO W1\_FINSTRESS2\_Recode.

VARIABLE LABELS W1\_FINSTRESS2\_Recode 'Omkering van W1\_FINSTRESS2'.

EXECUTE.

COMPUTE W1\_FINSTRESS\_Gem=**MEAN**(W1\_FINSTRESS1, W1\_FINSTRESS2\_Recode, W1\_FINSTRESS3).

VARIABLE LABELS W1\_FINSTRESS\_Gem 'Gemiddelde score op drie items over financiële stress'.

EXECUTE.

COMPUTE W1\_Corstress\_Gem=**MEAN**(W1\_Corstress1,W1\_Corstress2,W1\_Corstress3).

VARIABLE LABELS W1\_Corstress\_Gem 'Gemiddelde score op drie items over coronastress'.

EXECUTE.

Assumptiecheck:

1. De afhankelijke variabele (of uitkomstvariabele) is van metrisch meetniveau. **IN ORDE**
2. Er zijn twee of meer onafhankelijke variabelen (of predictoren). **IN ORDE**
3. Onafhankelijkheid: de residuele termen zijn niet gecorreleerd met elkaar. **NAGAAN IN OUTPUT REGRESSIONANALYSE**
4. Lineariteit: de afhankelijke variabele heeft een collectief lineair verband met alle predictoren. **NAGAAN VIA RESIDUENANALYSE**

5. Normaliteit: de residuele termen zijn normaal verdeeld. **NAGAAN VIA RESIDUENANALYSE**
  6. Homoscedasticiteit: de residuele termen hebben homogene (of anders gezegd: gelijke) varianties voor elke waarde van de afhankelijke variabele. **NAGAAN VIA RESIDUENANALYSE**
  7. Er zijn geen uitschieters onder de residuele termen (de gestandaardiseerde residuele termen zijn niet groter dan drie standaardafwijkingen). **NAGAAN VIA RESIDUENANALYSE**
  8. Geen multicollineariteit, waarbij twee of meer predictoren te sterk gecorreleerd zijn met elkaar.
- NAGAAN IN OUTPUT REGRESSIEANALYSE**

**Stap 3.** Hypothesen formuleren. Dit moet hier voor (1) het volledige model met alle predictoren en (2) elke predictor van het model apart (maar hier doen we dit bij wijze van voorbeeld alleen voor financiële stress):

#### **Verklaringskracht model**

$H_0$ : *Financiële stress, het hebben van een gezondheidsprobleem en geslacht hebben geen effect op coronastress: de verklaringskracht van dit lineair model is gelijk aan nul.*

$H_a$ : *Financiële stress, het hebben van een gezondheidsprobleem en geslacht hebben wel een effect op coronastress: de verklaringskracht van dit lineair model is groter of kleiner dan nul.*

#### **Verklaringskracht predictor financiële stress**

$H_0$ : *Financiële stress heeft geen effect op coronastress: de verklaringskracht van ouderschap is gelijk aan nul.*

$H_a$ : *Financiële stress heeft wel een effect op coronastress: de verklaringskracht van ouderschap is groter of kleiner dan nul.*

**Stap 4.** Significantieniveau bepalen.

Het gaat om een steekproef met meer dan 400 onderzoekseenheden ( $n = 1296$ ), dus hanteer  $\alpha = .01$  of 1 %.

**Stap 5.** Toetsingsgrootte en p-waarde berekenen.

Doorloop de reeks SPSS-acties in het boek om het volgende commando te verkrijgen:

#### REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT W1_Corstress_Gem
/METHOD=ENTER W1_Finstress_Gem W1_Handicap W1_Geslacht
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS DURBIN HISTOGRAM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

**Output.** Controleer eerst de overige zes assumpties aan de hand van de correlatietafel, VIF-waarden (assumptie 8) en de output voor residuenanalyse (assumptie 3-7). **Assumptie 8 is oké**, want geen enkele correlatie tussen twee predictoren is groter dan .60 en geen enkele VIF-waarde is groter dan 5. Verder zijn **assumptie 4 en 6 ook oké**, want de scatterplot vertoont geen non-lineair verband (lineariteit in orde) en de datapunten vormen een ovaal en geen kegel- of zandlopervorm (homoscedasticiteit in orde). Bij *Casewise Diagnostics* staan er twee uitschieters maar aangezien hun gestandaardiseerde residuele term telkens heel dicht bij 3 ligt, kun je ook **assumptie 7 rond uitschieters als vervuld** beschouwen. Verder ligt de toetsingsgrootte van de Durbin-Watson toets (zie *Model Summary* tabel) heel dicht bij 2, waardoor **assumptie 3 rond onafhankelijkheid van residuele**

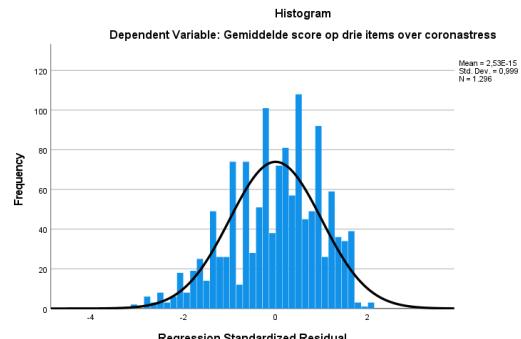
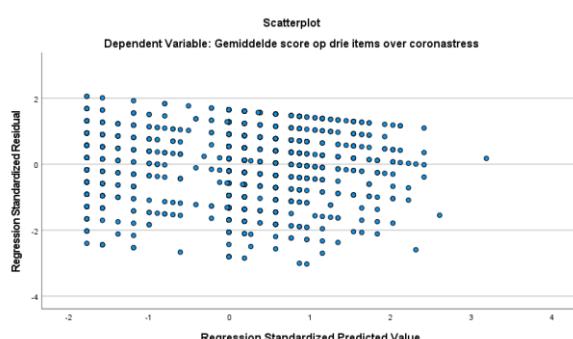
**termen ook in orde is. Assumptie 5 (normaliteit van residuen) lijkt echter niet vervuld.** Het histogram heeft verspreid een paar extreme pieken en links een dikkere staart, waardoor die als links scheef bestempeld wordt en niet als normaal verdeeld beschouwd kan worden. We kunnen als tweede stap kijken naar de z-scores van scheefheid en kurtosis van de residuen om een besluit te maken inzake normaliteit (zie hieronder).

Correlations					
	W1_Corstress_Gem	W1_FINSTRESS_Gem	W1_Handicap	W1_Geslacht	
Pearson Correlation	W1_Corstress_Gem	1,000	,120	,075	,185
	W1_FINSTRESS_Gem	,120	1,000	,024	,049
	W1_Handicap	,075	,024	1,000	,063
	W1_Geslacht	,185	,049	,063	1,000
Sig. (1-tailed)	W1_Corstress_Gem	.	<.001	,004	<.001
	W1_FINSTRESS_Gem	,000	.	,198	,039
	W1_Handicap	,004	,198	.	,012
	W1_Geslacht	,000	,039	,012	.
N	W1_Corstress_Gem	1,296	1,296	1,296	1,296
	W1_FINSTRESS_Gem	1,296	1,296	1,296	1,296
	W1_Handicap	1,296	1,296	1,296	1,296
	W1_Geslacht	1,296	1,296	1,296	1,296

#### Casewise Diagnostics<sup>a</sup>

Case Number	Std. Residual	W1_Corstress_Gem	Predicted Value	Residual
1,162	-3,026	1,00	3,7145	-2,71453
1,360	-3,005	1,00	3,6956	-2,69562

a. Dependent Variable: W1\_Corstress\_Gem



#### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,224 <sup>a</sup>	,050	,048	,89698	1,880

a. Predictors: (Constant), W1\_Geslacht, W1\_FINSTRESS\_Gem, W1\_Handicap

b. Dependent Variable: W1\_Corstress\_Gem

#### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	55,001	3	18,334	22,787	<,001 <sup>b</sup>
	Residual	1,039,503	1,292	,805		
	Total	1,094,504	1,295			

a. Dependent Variable: W1\_Corstress\_Gem

b. Predictors: (Constant), W1\_Geslacht, W1\_FINSTRESS\_Gem, W1\_Handicap

#### Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta	t			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	3,032	,068	44,761	<,001		2,899	3,165		
	W1_FINSTRESS_Gem	,120	,030	,110	4,037	<,001	,061	,178	,997	1,003
	W1_Handicap	,180	,080	,061	2,250	,025	,023	,338	,996	1,004
	W1_Geslacht	,363	,056	,176	6,468	<,001	,253	,474	,994	1,006

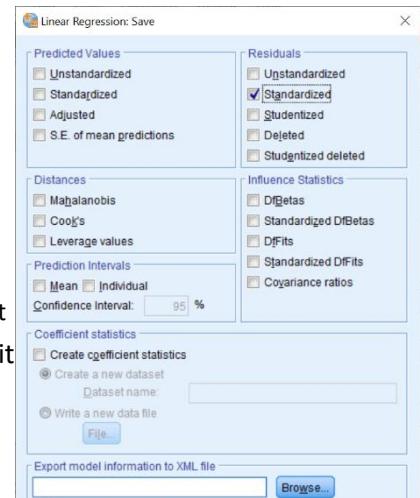
a. Dependent Variable: W1\_Corstress\_Gem

#### Assumptie 5. Normaliteit van residuen controleren a.d.h.v. z-scores scheefheid en kurtosis:

Actie 1. Vink tijdens het uitvoeren van het regressiecommando in de dialoogvensters onder de knop *Save* de optie *Standardized* aan in het kader *Residuals*. Je kunt ook eenvoudigweg in de *Syntax* het lijtje /SAVE ZRESID toevoegen bij het regressiecommando onderaan. Je vraagt hiermee aan SPSS om een nieuwe variabele aan te maken, waarin de gestandaardiseerde residuen van het regressiemodel opgeslagen worden van elke respondent in de dataset.

#### REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT W1_Corstress_Gem
/METHOD=ENTER W1_Finstress_Gem W1_Handicap W1_Geslacht
```



```

/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS DURBIN HISTOGRAM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE ZRESID.

```

Actie 2. Voer het bovenstaande regressiecommando uit. Een nieuwe variabele met de gestandaardiseerde residuen zal nu verschijnen in je databestand, genaamd 'ZRE\_1' (of een ander nummer dan 1, afhankelijk van of je al eerder andere residuen opgeslagen hebt).

Actie 3. Vraag via het *Frequencies*-commando scheefheid, kurtosis en hun standaardfouten op van de variabele 'ZRE\_1'.

Statistics		
ZRE_1		
	N	Valid
/FREQUENCIES VARIABLES=ZRE_1		1.296
/STATISTICS=SKEWNESS SESKEW KURTOSIS SEKURT		724
/ORDER=ANALYSIS.		
	Skewness	-.431
	Std. Error of Skewness	,068
	Kurtosis	-,336
	Std. Error of Kurtosis	,136

Actie 4. Bereken de z-scores van scheefheid en kurtosis. Je ziet hier dat beide z-scores buiten de grenzen van +/- 1.96 vallen, waardoor we kunnen besluiten dat de residuen niet normaal verdeeld zijn.

$$z_{\text{scheefheid}} = -0.43 / 0.07 = -6.14 \quad z_{\text{kurtosis}} = -0.34 / 0.14 = -2.42$$

In principe kun je hier geen lineaire regressie toepassen, maar om didactische redenen zullen we in deze oefening de assumpties even als vervuld beschouwen en de resultaten interpreteren.

**Stap 6.** Conclusie en rapportering. De vrijheidsgraden van de t-toetsen van elke predictor vind je als volgt:  $df = n - k$ , met  $k$  = aantal variabelen in de regressieanalyse (hier 4). **LET OP, deze rapportering is bij wijze van voorbeeld. Aangezien de normaliteitsassumptie niet vervuld is van de lineaire regressieanalyse, is de analysetechniek niet helemaal betrouwbaar en interpreteer je de resultaten beter niet.**

"Het effect van financiële stress, het hebben van een gezondheidsprobleem en geslacht als predictoren van coronastress werden gezamenlijk getoetst aan de hand van een meervoudig lineair regressiemodel. Dit lineair model verklaart een significant deel van de variantie in de scores van coronastress, met  $R^2 = .050$ ,  $R^2_{\text{Adjusted}} = .048$ ,  $F(3, 1292) = 22.79$ ,  $p < .001$ . Het effect van geslacht op coronastress was in dit model het grootst ( $\beta = 0.18$ ,  $t(1292) = 6.47$ ,  $p < .001$ ) en het effect van het hebben van een gezondheidsprobleem was niet significant ( $\beta = 0.06$ ,  $t(1292) = 2.25$ ,  $p = .025$ )."

### Oefening 13.4. Regressie.

**Antwoord:** Het voorgestelde model (met relatietevredenheid, steun van de partner, coronastress en geslacht als predictoren) verklaart een significant deel van de variantie in hoop (12.3 %). Van alle predictoren heeft relatietevredenheid het grootste significante effect op hoop, gevolgd door coronastress en geslacht, terwijl steun van de partner geen significante predictor van hoop blijkt te zijn.

**Stap 1.** Kiezen van de juiste analysemethode.

**1. Hoeveel variabelen bevat mijn onderzoeksraag?**

Vijf variabelen, namelijk relatietevredenheid, steun van de partner, coronastress, geslacht en hoop.

**2. Welk meetniveau hebben de variabelen in mijn onderzoeksraag?**

'W2\_QMI1' tot en met 'W2\_QMI5': vijf ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_QMI\_Gem'

'W2\_Steun\_partner1' tot en met 'W2\_Steun\_partner5': vijf ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_Steun\_partner\_Gem'

'W2\_Corstress1' tot en met 'W2\_Corstress4': vier ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_Corstress\_Gem'

'W2\_Geslacht': nominale variabele met twee categorieën

'W2\_HOPE1' tot en met 'W2\_HOPE6': zes ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_HOPE\_Gem'

**3. Welke rol vervullen de variabelen?**

De variabelen relatietevredenheid, steun van de partner, coronastress en geslacht zijn de onafhankelijke variabelen, waarvan we het effect willen kennen op de afhankelijke variabele hoop: we gebruiken de onafhankelijke variabelen dus als predictoren om de uitkomstvariabele 'W2\_HOPE\_Gem' te verklaren/voorspellen.

Flowchart:

Eén afhankelijke variabele → Metrisch → Twee of meer onafhankelijke variabelen → Metrisch

→ Meervoudige lineaire regressie of partiële correlatiotoets.

Kijk naar de verwoording van de onderzoeksraag: aangezien het hier gaat om het voorspellen van een variabele aan de hand van andere variabelen (predictoren) en niet om het onderzoeken van de samenhang tussen twee variabelen (gecontroleerd voor een derde variabele), is **meervoudige lineaire regressie** de juiste analysetechniek.

**Stap 2.** Datacontrole.

Deze analyse moet op een subset van de data uitgevoerd worden, namelijk mensen die een relatie hebben. Maak eerst die selectie, voor je de inconsistentiecheck en de assumptiecheck uitvoert:

USE ALL.

COMPUTE filter\_\$(W2\_Relatiestatus = 1).

VARIABLE LABELS filter\_\$ 'W2\_Relatiestatus = 1 (FILTER)'.

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_\$(f1.0).

FILTER BY filter\_\$.

EXECUTE.

## Inconsistentiecheck:

### W2\_QMI1 RELATIE TEVREDENHEID - We hebben een goede relatie

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	1	,3	,3
	2 Eerder niet akkoord	15	4,3	4,6
	3 Nog niet akkoord, noch akkoord	20	5,8	5,8
	4 Eerder akkoord	123	35,4	45,8
	5 Akkoord	188	54,2	100,0
	Total	347	100,0	100,0

### W2\_Steun\_partner1 STEUN PARTNER - Ik vertel mijn partner openlijk hoe ik me voel en dat ik zijn/haar steun zou waarderen

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 (Bijna) nooit	7	2,0	2,0
	2 Zelden	22	6,3	8,4
	3 Soms	75	21,6	21,6
	4 Vaak	157	45,2	75,2
	5 Zeer vaak	86	24,8	100,0
	Total	347	100,0	100,0

### W2\_Corstress1 CORONASTRESS - Ik maak mij zorgen dat ik met het coronavirus besmet zal raken

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Niet akkoord	59	17,0	17,0
	2 Eerder niet akkoord	92	26,5	43,5
	3 Nog niet akkoord, noch akkoord	74	21,3	64,8
	4 Eerder akkoord	97	28,0	92,8
	5 Akkoord	25	7,2	100,0
	Total	347	100,0	100,0

### W2\_Geslacht Wat is jouw geslacht?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 Man	85	24,5	24,5
	1 Vrouw	262	75,5	75,5
	Total	347	100,0	100,0

### W2\_HOPE1 HOOP - Als ik in moeilijkheden zit, kan ik verschillende manieren bedenken om er uit te geraken

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Helemaal oneens	3	,9	,9
	2 Oneens	9	2,6	2,6
	3 Eerder oneens	25	7,2	7,2
	4 Eerder eens	147	42,4	42,4
	5 Eens	113	32,6	32,6
	6 Helemaal eens	50	14,4	14,4
	Total	347	100,0	100,0

We tonen hier slechts een selectie van alle variabelen die gecontroleerd moeten worden op inconsistenties. Indien je ze allemaal opvraagt, zie je dat geen enkel item omgekeerd hoeft te worden voor we de metrische schaalvariabelen aanmaken. Er zijn ook geen andere inconsistenties op te merken. Vooraf kun je het best nog nagaan of de schalen intern consistent zijn aan de hand van een Cronbach's alpha toets, maar dit valt buiten de scope van deze oefening.

COMPUTE W2\_QMI\_Gem=MEAN(W2\_QMI1,W2\_QMI2,W2\_QMI3,W2\_QMI4,W2\_QMI5).

VARIABLE LABELS W2\_QMI\_Gem 'Gemiddelde score op vijf items over relatietevredenheid'.

EXECUTE.

COMPUTE W2\_Steun\_partner\_Gem=MEAN(W2\_Steun\_partner1, W2\_Steun\_partner2, W2\_Steun\_partner3, W2\_Steun\_partner4, W2\_Steun\_partner5).

VARIABLE LABELS W2\_Steun\_partner\_Gem 'Gemiddelde score op vijf items over steun van de partner'.

EXECUTE.

COMPUTE W2\_Corstress\_Gem=MEAN(W2\_Corstress1,W2\_Corstress2,W2\_Corstress3,W2\_Corstress4).

VARIABLE LABELS W2\_Corstress\_Gem 'Gemiddelde score op vier items over coronastress'.

EXECUTE.

COMPUTE W2\_HOPE\_Gem=MEAN(W2\_HOPE1,W2\_HOPE2,W2\_HOPE3,W2\_HOPE4,W2\_HOPE5,W2\_HOPE6).

VARIABLE LABELS W2\_HOPE\_Gem 'Gemiddelde score op zes items over hoop'.

EXECUTE.

## Assumptiecheck:

1. De afhankelijke variabele (of uitkomstvariabele) is van metrisch meetniveau. **IN ORDE**
2. Er zijn twee of meer onafhankelijke variabelen (of predictoren). **IN ORDE**
3. Onafhankelijkheid: de residuele termen zijn niet gecorreleerd met elkaar. **NAGAAN IN OUTPUT REGRESSIONS**

4. Lineariteit: de afhankelijke variabele heeft een collectief lineair verband met alle predictoren.

#### NAGAAN VIA RESIDUENANALYSE

5. Normaliteit: de residuele termen zijn normaal verdeeld. **NAGAAN VIA RESIDUENANALYSE**
6. Homoscedasticiteit: de residuele termen hebben homogene (of anders gezegd: gelijke) varianties voor elke waarde van de afhankelijke variabele. **NAGAAN VIA RESIDUENANALYSE**
7. Er zijn geen uitschieters onder de residuele termen (de gestandaardiseerde residuele termen zijn niet groter dan drie standaardafwijkingen). **NAGAAN VIA RESIDUENANALYSE**
8. Geen multicollineariteit, waarbij twee of meer predictoren te sterk gecorreleerd zijn met elkaar. **NAGAAN IN OUTPUT REGRESSIEANALYSE**

**Stap 3.** Hypothesen formuleren. Dit moet hier voor (1) het volledige model met alle predictoren en (2) elke predictor van het model apart (maar hier doen we dit bij wijze van voorbeeld alleen voor relatietevredenheid):

#### Verklaringskracht model

$H_0$ : Relatietevredenheid, steun van de partner, coronastress en geslacht hebben geen effect op hoop: de verklaringskracht van dit lineair model is gelijk aan nul.

$H_a$ : Relatietevredenheid, steun van de partner, coronastress en geslacht hebben wel een effect op hoop: de verklaringskracht van dit lineair model is groter of kleiner dan nul.

#### Verklaringskracht predictor relatietevredenheid

$H_0$ : Relatietevredenheid heeft geen effect op hoop: de verklaringskracht van relatietevredenheid is gelijk aan nul.

$H_a$ : Relatietevredenheid heeft wel een effect op hoop: de verklaringskracht van relatietevredenheid is groter of kleiner dan nul.

**Stap 4.** Significantieniveau bepalen.

Het gaat om een steekproef met minder dan 400 onderzoekseenheden ( $n = 347$ ), dus hanteer  $\alpha = .05$  of 5 %.

**Stap 5.** Toetsingsgrootte en p-waarde berekenen.

Doorloop de reeks SPSS-acties in het boek om het volgende commando te verkrijgen:

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT W2_HOPE_Gem  
/METHOD=ENTER W2_QMI_Gem W2_Steun_partner_Gem W2_Corstress_Gem W2_Geslacht  
/SCATTERPLOT>(*ZRESID,*ZPRED)  
/RESIDUALS DURBIN HISTOGRAM(ZRESID)  
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

**Output.** Controleer eerst de overige zes assumpties aan de hand van de correlatietafel, VIF-waarden (assumptie 8) en de output voor residuenanalyse (assumptie 3-7). **Assumptie 8 is oké**, ook al is de correlatie tussen relatietevredenheid en steun van de partner groter dan .60: geen enkele VIF-waarde is groter dan 5, dus er is geen sprake van multicollineariteit. Verder is **assumptie 4 ook oké**, want de scatterplot vertoont geen non-lineair verband (lineairiteit in orde). Bij *Casewise Diagnostics* staan er drie uitschieters waarvan de gestandaardiseerde

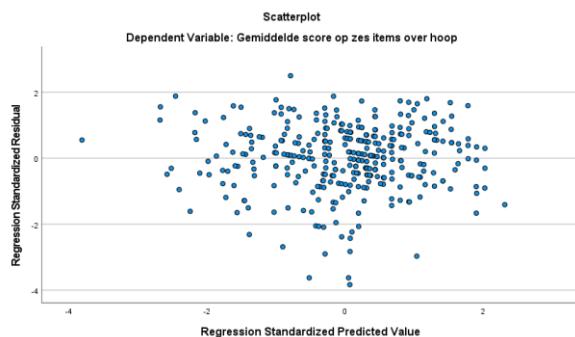
residuele term telkens ver voorbij 3 ligt. Om **assumptie 7 rond uitschieters als vervuld te beschouwen**, is het beter om die **uitschieters te verwijderen**. Laten we aannemen dat je om die te verwijderen ook de nodige theoretische of statistische gronden hebt. **Het zal ook helpen voor assumptie 6 rond homoscedasticiteit**, want als je de nullijn van links naar rechts afgaat, dan lijken de datapunten naar onderen toe in een punt te evolueren en gaan dan terug naar boven, wat eerder wijst op heteroscedasticiteit.

Correlations						
	W2_HOPE_Gem	W2_QMI_Gem	W2_Steun_partner_Gem	W2_Corstress_Gem	W2_Geslacht	
Pearson Correlation	1,000	,255	,180	-,151	-,159	
	W2_QMI_Gem	1,000	,720	,053	,013	
	W2_Steun_partner_Gem	,180	1,000	,007	,064	
	W2_Corstress_Gem	-,151	,053	1,000	,148	
	W2_Geslacht	-,159	,013	,064	,148	1,000
Sig. (1-tailed)						
	W2_HOPE_Gem	.	<.001	<.001	,002	,001
	W2_QMI_Gem	,000	.	,000	,163	,402
	W2_Steun_partner_Gem	,000	,000	.	,446	,118
	W2_Corstress_Gem	,002	,163	,446	.	,003
	W2_Geslacht	,001	,402	,118	,003	.
N						
	W2_HOPE_Gem	347	347	347	347	347
	W2_QMI_Gem	347	347	347	347	347
	W2_Steun_partner_Gem	347	347	347	347	347
	W2_Corstress_Gem	347	347	347	347	347
	W2_Geslacht	347	347	347	347	347

Casewise Diagnostics <sup>a</sup>						
Case Number	Std. Residual	W2_HOPE_Gem	Predicted Value	Residual		
68	-3,834	1,00	4,1772	-3,17718		
141	-3,626	1,17	4,1717	-3,00507		
186	-3,626	1,00	4,0052	-3,00516		

a. Dependent Variable: W2\_HOPE\_Gem



Uitschieters verwijderen via *Select Cases* (zorg hierbij dat de oude filter nog van kracht blijft):

- Zoek eerst op welke waarde rijnummer 68, 141 en 186 hebben bij de variabele *Number* in de *Data View*.

	Nummer	Nummer	Nummer
68	329	141	778
			186
			1.024

- Voer nu een selectie uit waarbij die drie cases uitgesloten worden van de analyse.

USE ALL.

COMPUTE filter\_\$(W2\_Relatiestatus = 1 & Nummer ~=~ 329 & Nummer ~=~ 778 & Nummer ~=~ 1024).  
VARIABLE LABELS filter\_\$ 'W2\_Relatiestatus = 1 & Nummer ~=~ 68 & Nummer ~=~ 141 & Nummer ~=~ 186 (FILTER)'.

VALUE LABELS filter\_\$(0 'Not Selected' 1 'Selected').

FORMATS filter\_\$(f1.0).

FILTER BY filter\_\$(

EXECUTE.

- Voer nu opnieuw het bovenstaande *Regression*-commando uit.

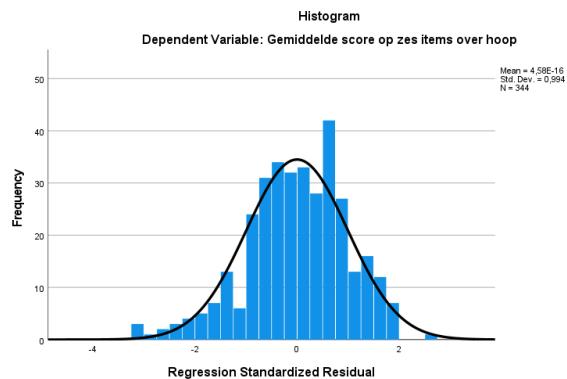
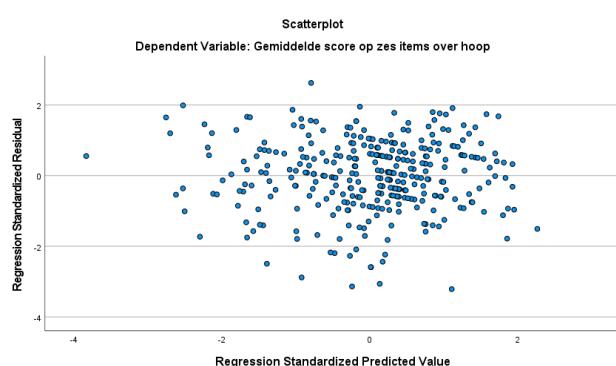
**Output.** Ga hier nu verder met het controleren van de assumpties. SPSS duidt nu enkele andere uitschieters aan, maar die zijn minder problematisch (de gestandaardiseerde residuele termen liggen dicht bij 3) dan de voorgaande en kun je dus laten staan. **Assumptie 7 is nu dus vervuld (geen uitschieters), alsook assumptie 6 rond homoscedasticiteit:** door de drie uitschieters te verwijderen vormt de scatterplot een meer gelijkmatig geheel. Verder kunnen we **assumptie 5 rond normaliteit** ook als vervuld beschouwen: het histogram ziet er normaal verdeeld uit. Tot slot ligt de toetsingsgroothed van de Durbin-Watson toets (zie *Model Summary* tabel) heel dicht bij 2, waardoor **assumptie 3 rond onafhankelijkheid van residuele termen ook in orde is**.

Correlations						
	W2_HOPE_Gem	W2_QMI_Gem	W2_Steun_partner_Gem	W2_Corstress_Gem	W2_Geslacht	
Pearson Correlation	1,000	,280	,207	-,153	-,151	
	W2_QMI_Gem	1,000	,719	,051	,012	
	W2_Steun_partner_Gem	,207	1,000	,004	,062	
	W2_Corstress_Gem	-,153	,051	1,000	,147	
	W2_Geslacht	-,151	,012	,062	,147	1,000
Sig. (1-tailed)	W2_HOPE_Gem	.	<.001	<.001	,002	,003
	W2_QMI_Gem	,000	.	,000	,172	,411
	W2_Steun_partner_Gem	,000	,000	.	,469	,127
	W2_Corstress_Gem	,002	,172	,469	.	,003
	W2_Geslacht	,003	,411	,127	,003	.
N	W2_HOPE_Gem	344	344	344	344	344
	W2_QMI_Gem	344	344	344	344	344
	W2_Steun_partner_Gem	344	344	344	344	344
	W2_Corstress_Gem	344	344	344	344	344
	W2_Geslacht	344	344	344	344	344

### Casewise Diagnostics<sup>a</sup>

Case Number	Std. Residual	W2_HOPE_Gem	Predicted Value	Residual
38	-3,063	1,83	4,2228	-2,38948
40	-3,137	1,67	4,1137	-2,44707
158	-3,212	2,00	4,5061	-2,50608

a. Dependent Variable: W2\_HOPE\_Gem



### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,351 <sup>a</sup>	,123	,113	,78013	1,886

a. Predictors: (Constant), W2\_Geslacht, W2\_QMI\_Gem, W2\_Corstress\_Gem, W2\_Steun\_partner\_Gem

b. Dependent Variable: W2\_HOPE\_Gem

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29,066	4	7,266	11,939	<.001 <sup>b</sup>
	Residual	206,318	339	,609		
	Total	235,384	343			

a. Dependent Variable: W2\_HOPE\_Gem

b. Predictors: (Constant), W2\_Geslacht, W2\_QMI\_Gem, W2\_Corstress\_Gem, W2\_Steun\_partner\_Gem

### Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error				Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	3,571	,275	12,985	<.001	3,030	4,112		
	W2_QMI_Gem	,277	,073	,3,769	<.001	,132	,421	,479	2,087
	W2_Steun_partner_Gem	,016	,071	,017	,228	,820	-,124	,157	,479
	W2_Corstress_Gem	-,137	,048	-,147	-,2,856	,005	-,231	-,043	,973
	W2_Geslacht	-,256	,099	-,133	-,2,583	,010	-,450	-,061	,971

a. Dependent Variable: W2\_HOPE\_Gem

**Stap 6. Conclusie en rapportering.** De vrijheidsgraden van de t-toetsen van elke predictor vind je als volgt:  $df = n - k$ , met  $k$  = aantal variabelen in de regressieanalyse (hier 5).

"Het effect van relatietevredenheid, steun van de partner, coronastress en geslacht als predictoren van hoop werden gezamenlijk getoetst aan de hand van een meervoudig lineair regressiemodel. Dit lineair model verklaart een significant deel van de variantie in hoop, met  $R^2 = .123$ ,  $R^2_{Adjusted} = .113$ ,  $F(4, 339) = 11.94$ ,  $p < .001$ . Het effect van relatietevredenheid op hoop was in dit model het grootst ( $\beta = 0.28$ ,  $t(339) = 3.77$ ,  $p < .001$ ), gevuld door het effect van coronastress ( $\beta = -0.15$ ,  $t(339) = -2.86$ ,  $p = .005$ ) en het effect van geslacht ( $\beta = -0.13$ ,  $t(339) = -2.58$ ,  $p = .010$ ). Het effect van steun van de partner was niet significant in dit model ( $p = .820$ )."

### Oefening 13.5. Regressie.

**Antwoord:** 4.5 % van de variantie in coronastress wordt voorspeld door het model met alleen sociodemografische predictoren. Voegen we twee relationele predictoren toe aan het model, namelijk relatietevredenheid en steun van de partner, dan stijgt de verklaarde variantie met 0.06 procentpunten naar 5.1 %, wat echter geen significante verbetering is van het model.

**Stap 1.** Kiezen van de juiste analysemethode.

**1. Hoeveel variabelen bevat mijn onderzoeksraag?**

Zes variabelen, namelijk opleidingsniveau, geslacht, het hebben van een gezondheidsprobleem, relatietevredenheid, steun van de partner en coronastress.

**2. Welk meetniveau hebben de variabelen in mijn onderzoeksraag?**

'W2\_Diploma': ordinale variabele met drie categorieën

'W2\_Geslacht' en 'W2\_Handicap': twee nominale variabelen met telkens twee categorieën

'W2\_QMI1' tot en met 'W2\_QMI5': vijf ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_QMI\_Gem'

'W2\_Steun\_partner1' tot en met 'W2\_Steun\_partner5': vijf ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_Steun\_partner\_Gem'

'W2\_Corstress1' tot en met 'W2\_Corstress4': vier ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_Corstress\_Gem'

**3. Welke rol vervullen de variabelen?**

De variabelen opleidingsniveau, geslacht, het hebben van een gezondheidsprobleem, relatietevredenheid en steun van de partner zijn de onafhankelijke variabelen, waarvan we het effect willen kennen op de afhankelijke variabele coronastress: we gebruiken de onafhankelijke variabelen dus als predictoren om de uitkomstvariabele 'W2\_Corstress\_Gem' te verklaren/voorspellen.

Flowchart:

Eén afhankelijke variabele → Metrisch → Twee of meer onafhankelijke variabelen → Metrisch

→ Meervoudige lineaire regressie of partiële correlatietoets.

Kijk naar de verwoording van de onderzoeksraag: aangezien het hier gaat om het voorspellen van een variabele aan de hand van andere variabelen (predictoren) en niet om het onderzoeken van de samenhang tussen twee variabelen (gecontroleerd voor een derde variabele), is **meervoudige lineaire regressie** de juiste analysetechniek. Omdat de predictoren bloksgewijs ingebracht worden, gaat het hier meer specifiek om hiërarchische meervoudige regressie.

**Stap 2.** Datacontrole.

Inconsistentiecheck:

Deze analyse moet op een subset van de data uitgevoerd worden, namelijk mensen die een relatie hebben. Maak eerst die selectie, voor je de inconsistentiecheck en de assumptiecheck uitvoert:

USE ALL.

COMPUTE filter\_\$(W2\_Relatiestatus = 1).

VARIABLE LABELS filter\_\$ 'W2\_Relatiestatus = 1 (FILTER)'.

VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.

© 2022, Toegepaste Statistiek – Wim Hardyns, Koen Ponnet en Pelckmans Uitgevers nv

FORMATS filter\_\$(f1.0).

FILTER BY filter\_\$(

EXECUTE.

**W2\_Geslacht**

	Frequency	Percent	Valid Percent	Cumulative
				Percent
Valid	0 Man	85	24,5	24,5
	1 Vrouw	262	75,5	75,5
	Total	347	100,0	100,0

**W2\_Handicap**

	Frequency	Percent	Valid Percent	Cumulative
				Percent
Valid	0 Nee	280	80,7	80,7
	1 Ja	67	19,3	19,3
	Total	347	100,0	100,0

**W2\_Diploma**

	Frequency	Percent	Valid Percent	Cumulative
				Percent
Valid	1 Geen diploma, een diploma lager onderwijs of een diploma lagere graad middelbaar onderwijs	6	1,7	1,7
	2 Een diploma middelbaar onderwijs	61	17,6	17,6
	3 Een diploma hoger onderwijs	280	80,7	80,7
	Total	347	100,0	100,0

**W2\_QMI1**

	Frequency	Percent	Valid Percent	Cumulative
				Percent
Valid	1 Niet akkoord	1	,3	,3
	2 Eerder niet akkoord	15	4,3	4,3
	3 Nog niet akkoord, noch akkoord	20	5,8	5,8
	4 Eerder akkoord	123	35,4	35,4
	5 Akkoord	188	54,2	54,2
	Total	347	100,0	100,0

**W2\_Steun\_partner1**

	Frequency	Percent	Valid Percent	Cumulative
				Percent
Valid	1 (Bijna) nooit	7	2,0	2,0
	2 Zelden	22	6,3	6,3
	3 Soms	75	21,6	21,6
	4 Vaak	157	45,2	45,2
	5 Zeer vaak	86	24,8	24,8
	Total	347	100,0	100,0

**W2\_Corstress1**

	Frequency	Percent	Valid Percent	Cumulative
				Percent
Valid	1 Niet akkoord	59	17,0	17,0
	2 Eerder niet akkoord	92	26,5	43,5
	3 Nog niet akkoord, noch akkoord	74	21,3	64,8
	4 Eerder akkoord	97	28,0	92,8
	5 Akkoord	25	7,2	100,0
	Total	347	100,0	100,0

Gezien de grote hoeveelheid aan variabelen, tonen we hier slechts een selectie. Geen enkele van de schaalitems moet omgekeerd worden voor we de schaalvariabelen berekenen. Er zijn verder ook geen inconsistenties op te merken. Vooraf kun je het best nog nagaan of de schalen intern consistent zijn aan de hand van een Cronbach's alpha toets, maar dit valt buiten de scope van deze oefening.

COMPUTE W2\_QMI\_Gem=MEAN(W2\_QMI1,W2\_QMI2,W2\_QMI3,W2\_QMI4,W2\_QMI5).

VARIABLE LABELS W2\_QMI\_Gem 'Gemiddelde score op vijf items over relatietevredenheid'.

EXECUTE.

COMPUTE W2\_Steun\_partner\_Gem=MEAN(W2\_Steun\_partner1, W2\_Steun\_partner2, W2\_Steun\_partner3, W2\_Steun\_partner4, W2\_Steun\_partner5).

VARIABLE LABELS W2\_Steun\_partner\_Gem 'Gemiddelde score op vijf items over steun van de partner'.

EXECUTE.

COMPUTE W2\_Corstress\_Gem=MEAN(W2\_Corstress1,W2\_Corstress2,W2\_Corstress3,W2\_Corstress4).

VARIABLE LABELS W2\_Corstress\_Gem 'Gemiddelde score op vier items over coronastress'.

EXECUTE.

Assumptiecheck:

1. De afhankelijke variabele (of uitkomstvariabele) is van metrisch meetniveau. **IN ORDE**

2. Er zijn twee of meer onafhankelijke variabelen (of predictoren). **NIET IN ORDE**
3. Onafhankelijkheid: de residuele termen zijn niet gecorreleerd met elkaar. **NAGAAN IN OUTPUT REGRESSIEANALYSE**
4. Lineariteit: de afhankelijke variabele heeft een collectief lineair verband met alle predictoren. **NAGAAN VIA RESIDUENANALYSE**
5. Normaliteit: de residuele termen zijn normaal verdeeld. **NAGAAN VIA RESIDUENANALYSE**
6. Homoscedasticiteit: de residuele termen hebben homogene (of anders gezegd: gelijke) varianties voor elke waarde van de afhankelijke variabele. **NAGAAN VIA RESIDUENANALYSE**
7. Er zijn geen uitschieters onder de residuele termen (de gestandaardiseerde residuele termen zijn niet groter dan drie standaardafwijkingen). **NAGAAN VIA RESIDUENANALYSE**
8. Geen multicollineariteit, waarbij twee of meer predictoren te sterk gecorreleerd zijn met elkaar. **NAGAAN IN OUTPUT REGRESSIEANALYSE**

**Assumptie 2.** Deze assumptie lijkt in orde, maar hier hebben we te maken met een uitzondering (zie handboek p. 502-505). De predictoren in een lineair regressiemodel moeten ofwel binair (categorische variabele met twee categorieën) ofwel metrisch zijn. ‘W2\_Diploma’ valt hier niet onder, aangezien dit een categorische variabele met drie categorieën is. Deze variabele moeten we dummificeren voor we ze kunnen toevoegen als predictor in het model. Het gaat hier om een variabele met drie categorieën, dus maken we twee nieuwe binaire variabelen aan ( $k - 1$ , met  $k = \text{aantal categorieën}$ ). Als referentiecategorie nemen we hier ‘3 = Een diploma hoger onderwijs’, omdat die het meest voorkomt en een logische vergelijking is met de twee lagere categorieën die veel minder voorkomen in deze dataset. Via de stappen in het boek kom je op het onderstaande commando uit:

```
RECODE W2_DIPLOMA (1=1) (MISSING=SYSMIS) (ELSE=0) INTO Diploma_Dummy_1.  
VARIABLE LABELS Diploma_Dummy_1 'Geen of laag diploma vs hoger onderwijs'.  
EXECUTE.
```

```
RECODE W2_DIPLOMA (2=1) (MISSING=SYSMIS) (ELSE=0) INTO Diploma_Dummy_2.  
VARIABLE LABELS Diploma_Dummy_2 'Middelbaar onderwijs vs hoger onderwijs'.  
EXECUTE.
```

Controleer ook altijd even de nieuwe variabelen door hun frequentietabel te vergelijken met die van de originele variabele:

```
FREQUENCIES VARIABLES= W2_DIPLOMA Diploma_Dummy_1 Diploma_Dummy_2  
/ORDER=ANALYSIS.
```

Nu is de tweede assumptie wel in orde, als we deze twee nieuwe binaire variabelen gebruiken als predictor voor opleidingsniveau.

**Stap 3.** Hypothesen formuleren. Dit moet hier voor (1) het model met de eerste blok predictoren, (2) elke predictor van model 1, (3) het model met de tweede blok predictoren, (4) elke predictor van model 2 en (5) het verschil in verklaringskracht tussen model 1 en model 2. Hier doen we dit bij wijze van voorbeeld enkel voor de beide modellen en het verschil tussen beide:

### Verklaringskracht model 1

$H_0$ : Opleidingsniveau, geslacht en het hebben van een gezondheidsprobleem hebben geen effect op coronastress: de verklaringskracht van dit lineair model is gelijk aan nul.

$H_a$ : Opleidingsniveau, geslacht en het hebben van een gezondheidsprobleem hebben wel een effect op coronastress: de verklaringskracht van dit lineair model is groter of kleiner dan nul.

### Verklaringskracht model 2

$H_0$ : Opleidingsniveau, geslacht, het hebben van een gezondheidsprobleem, relatietevredenheid en steun van de partner hebben geen effect op coronastress: de verklaringskracht van dit lineair model is gelijk aan nul.

$H_a$ : Opleidingsniveau, geslacht, het hebben van een gezondheidsprobleem, relatietevredenheid en steun van de partner hebben wel een effect op coronastress: de verklaringskracht van dit lineair model is groter of kleiner dan nul.

### Verschil in verklaringskracht tussen model 1 en 2

$H_0$ : Opleidingsniveau, geslacht en het hebben van een gezondheidsprobleem (model 1) hebben evenveel effect op coronastress als opleidingsniveau, geslacht, het hebben van een gezondheidsprobleem, relatietevredenheid en steun van de partner (model 2): de verklaringskracht van beide modellen is gelijk aan elkaar.

$H_a$ : De predictoren uit model 1 hebben niet evenveel effect op coronastress als de predictoren uit model 2: de verklaringskracht van model 2 is groter dan die van model 1.

### Stap 4. Significantieniveau bepalen.

Het gaat om een steekproef met minder dan 400 onderzoekseenheden ( $n = 347$ ), dus hanteer  $\alpha = .05$  of 5 %.

### Stap 5. Toetsingsgrootte en p-waarde berekenen.

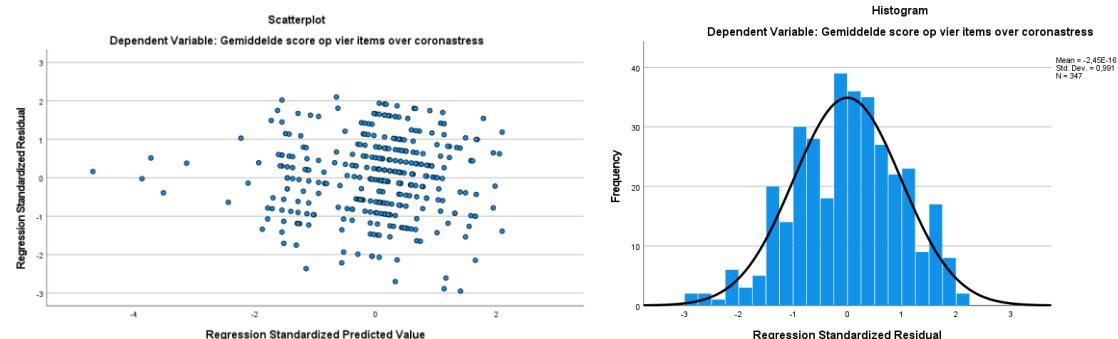
Doorloop de reeks SPSS-acties in het boek om het volgende commando te verkrijgen:

#### REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL CHANGE  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT W2_Corstress_Gem  
/METHOD=ENTER Diploma_Dummy_1 Diploma_Dummy_2 W2_Geslacht W2_Handicap  
/METHOD=ENTER W2_QMI_Gem W2_Steun_partner_Gem  
/SCATTERPLOT=(*ZRESID ,*ZPRED)  
/RESIDUALS DURBIN HISTOGRAM(ZRESID)  
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

**Output.** Assumptie 3, 4, 5, 7 en 8 lijken hier in orde te zijn, maar het valt meteen op dat **assumptie 6 rond homoscedasticiteit niet vervuld is**. De scatterplot ziet er kegelvormig uit, zoals in scenario c van figuur 12.11 van het handboek (p. 451). We hebben hier dus te maken met heteroscedasticiteit, waardoor de assumptie geschonden is. Om didactische redenen zullen we in deze oefening de assumpties toch als vervuld beschouwen en de resultaten alsnog interpreteren.

Correlations								
	W2_Corstress_Gem	Diploma_Dummy_1	Diploma_Dummy_2	W2_Geslacht	W2_Handicap	W2_QMI_Gem	W2_Steun_partner_Gem	
Pearson Correlation	W2_Corstress_Gem	1,000	-,106	,036	,148	,100	,053	,007
	Diploma_Dummy_1	-,106	1,000	-,061	,024	-,009	-,050	-,046
	Diploma_Dummy_2	,036	-,061	1,000	-,089	,023	-,005	,030
	W2_Geslacht	,148	,024	-,089	1,000	,024	,013	,064
	W2_Handicap	,100	-,009	,023	,024	1,000	-,031	-,029
	W2_QMI_Gem	,053	-,050	-,005	,013	-,031	1,000	,720
	W2_Steun_partner_Gem	,007	-,046	,030	,064	-,029	,720	1,000
Sig. (1-tailed)	W2_Corstress_Gem		,024	,251	,003	,031	,163	,446
	Diploma_Dummy_1	,024		,128	,327	,435	,177	,198
	Diploma_Dummy_2	,251	,128		,049	,332	,460	,291
	W2_Geslacht	,003	,327	,049		,328	,402	,118
	W2_Handicap	,031	,435	,332	,328		,285	,293
	W2_QMI_Gem	,163	,177	,460	,402	,285		,000
	W2_Steun_partner_Gem	,446	,198	,291	,118	,293	,000	



### Model Summary<sup>c</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics					Durbin-Watson
						F Change	df1	df2	Sig. F Change		
1	,212 <sup>a</sup>	,045	,034	,87394	,045	4,014	4	342	,003		
2	,225 <sup>b</sup>	,051	,034	,87383	,006	1,044	2	340	,353	2,079	

a. Predictors: (Constant), W2\_Handicap, Diploma\_Dummy\_1, W2\_Geslacht, Diploma\_Dummy\_2

b. Predictors: (Constant), W2\_Handicap, Diploma\_Dummy\_1, W2\_Geslacht, Diploma\_Dummy\_2, W2\_QMI\_Gem, W2\_Steun\_partner\_Gem

c. Dependent Variable: W2\_Corstress\_Gem

ANOVA <sup>a</sup>						Coefficients <sup>a</sup>										
Model	Sum of Squares	df	Mean Square	F	Sig.	Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B	Collinearity Statistics			
							B	Std. Error				Lower Bound	Upper Bound	Tolerance	VIF	
1	Regression	12,263	4	3,066	4,014	,003 <sup>b</sup>	1 (Constant)	,303	,101	29,718	<.001	2,814	3,213			
		Residual	261,209	342	,764		Diploma_Dummy_1	-,727	,361	-,107	-,2015	,045	-,1436	-,017	,996 1,004	
		Total	273,473	346			Diploma_Dummy_2	,095	,124	,041	,768	,443	-,149	,339	,988 1,012	
2	Regression	13,858	6	2,310	3,025	,007 <sup>c</sup>	1	W2_Geslacht	,314	,110	,152	,2,862	,004	,098	,529	,991 1,009
		Residual	259,615	340	,764		W2_Handicap	,213	,119	,095	,1,794	,074	-,021	,447	,999 1,001	
		Total	273,473	346			2 (Constant)	,287	,269	10,536	<.001	2,308	3,367			
a. Dependent Variable: W2_Corstress_Gem																
b. Predictors: (Constant), W2_Handicap, Diploma_Dummy_1, W2_Geslacht, Diploma_Dummy_2																
c. Predictors: (Constant), W2_Handicap, Diploma_Dummy_1, W2_Geslacht, Diploma_Dummy_2, W2_QMI_Gem, W2_Steun_partner_Gem																

a. Dependent Variable: W2\_Corstress\_Gem

**Stap 6. Conclusie en rapportering. LET OP, deze rapportering is bij wijze van voorbeeld. Aangezien de assumptie van homoscedasticiteit niet vervuld is van de lineaire regressieanalyse, is de analysetechniek niet helemaal betrouwbaar en interpreteer je de resultaten beter niet.**

"Om de invloed van twee groepen onafhankelijke variabelen op coronastress te toetsen, werd gebruik gemaakt van een hiërarchische meervoudige regressieanalyse. In een basismodel (model 1) werden drie sociodemografische predictoren (opleidingsniveau, geslacht en het hebben van een gezondheidsprobleem) opgenomen, waarna in een tweede model twee bijkomende relationele predictoren toegevoegd werden (relatietevredenheid en steun van de partner). Zowel model 1 ( $R^2 = .045$ ,  $R^2_{Adjusted} = .034$ ,  $F(4, 342) = 4.01$ ,  $p = .003$ ) als model 2 ( $R^2 = .051$ ,  $R^2_{Adjusted} = .034$ ,  $F(6, 340) = 3.02$ ,  $p = .007$ ) verklaren beide een significant deel van de variantie in coronastress. Het is echter niet zo dat model 2 door de toevoeging van relatietevredenheid en coronastress significant meer variantie verklaart dan model 1: model 2 verklaart slechts 0.6 procentpunten meer van de totale variantie, wat geen significante verandering is ( $\Delta R^2 = .006$ ,  $F(2, 340) = 1.04$ ,  $p = .353$ ).

### Oefening 13.6. Regressie.

**Antwoord:** Het voorgestelde model (met aantal huisgenoten, werk- of studiestress en angst als predictoren) verklaart een significant deel van de variantie in werk- of studieconcentratie (7.4 %). Geen enkele van de predictoren op zichzelf heeft voldoende verklaringskracht in de variantie van werk- of studieconcentratie: alle regressiecoëfficiënten zijn niet significant.

**Stap 1.** Kiezen van de juiste analysemethode.

**1. Hoeveel variabelen bevat mijn onderzoeksraag?**

Vier variabelen, namelijk aantal huisgenoten, werk- of studiestress, angst en werk- of studieconcentratie.

**2. Welk meetniveau hebben de variabelen in mijn onderzoeksraag?**

'W2\_RESID01': metrische variabele

'W2\_W\_STRESS1' tot en met 'W2\_W\_STRESS3': drie ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_W\_STRESS\_Gem'

'W2\_Angst1' tot en met 'W2\_Angst4': vier ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_Angst\_Gem'

'W2\_W\_CONC1' tot en met 'W2\_W\_CONC3': drie ordinale schaalitems die omgezet kunnen worden naar één metrische schaalvariabele, 'W2\_W\_CONC\_Gem'

**3. Welke rol vervullen de variabelen?**

De variabelen aantal huisgenoten, werk- of studiestress en angst zijn de onafhankelijke variabelen, waarvan we het effect willen kennen op de afhankelijke variabele werk- of studieconcentratie: we gebruiken de onafhankelijke variabelen dus als predictoren om de uitkomstvariabele 'W2\_W\_CONC\_Gem' te verklaren/voorspellen.

Flowchart:

Eén afhankelijke variabele → Metrisch → Twee of meer onafhankelijke variabelen → Metrisch

→ Meervoudige lineaire regressie of partiële correlatietoets.

Kijk naar de verwoording van de onderzoeksraag: aangezien het hier gaat om het voorspellen van een variabele aan de hand van andere variabelen (predictoren) en niet om het onderzoeken van de samenhang tussen twee variabelen (gecontroleerd voor een derde variabele), is **meervoudige lineaire regressie** de juiste analysetechniek.

**Stap 2.** Datacontrole.

Inconsistentiecheck:

**W2\_RESID01 Uit hoeveel personen bestaat jouw huishouden, jezelf meegerekend? Kinderen, partners of anderen die gedurende deze periode deeltijds bij je wonen, worden ook meegerekend als huisgenoten.**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	85	16,6	16,6
	2	161	31,5	31,5
	3	97	19,0	19,0
	4	106	20,7	87,9
	5	42	8,2	96,1
	6	14	2,7	98,8
	7	4	,8	99,6
	8	1	,2	99,8
	10	1	,2	100,0
Total	511	100,0	100,0	

**W2\_W\_STRESS1 WERK/STUDIESTRESS: hoe vaak... - ...had je het gevoel dat je belangrijke zaken in je werk/studie niet onder controle had?**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 (Bijna) nooit	27	5,3	6,4
	2 Zelden	90	17,6	21,5
	3 Soms	143	28,0	34,1
	4 Vaak	113	22,1	27,0
	5 Heel vaak	46	9,0	11,0
	Total	419	82,0	100,0
Missing	System	92	18,0	
Total	511	100,0		

**W2\_W\_CONC3 WERK/STUDIECONCENTRATIE: in welke mate... - ...ben jij afgeleid terwijl je werkt?**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Helemaal niet	106	20,7	20,7
	2 Meerdere dagen, maar minder dan de helft	195	38,2	58,9
	3 Meer dan de helft van de dagen	101	19,8	78,7
	4 Bijna elke dag	109	21,3	100,0
Total	511	100,0	100,0	

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Veel minder dan gewoonlijk	48	9,4	11,5
	2 Minder dan gewoonlijk	168	32,9	40,1
	3 Net zoveel als gewoonlijk	172	33,7	41,1
	4 Meer dan gewoonlijk	28	5,5	6,7
	5 Veel meer dan gewoonlijk	3	,6	,7
	Total	419	82,0	100,0
Missing	System	92	18,0	
Total	511	100,0		

We tonen hier slechts een selectie van alle variabelen die gecontroleerd moeten worden op inconsistenties. Indien je ze allemaal opvraagt, zie je dat geen enkel item omgekeerd hoeft te worden voor we de metrische schaalvariabelen aanmaken, behalve het derde item van werk- of studieconcentratie: momenteel geldt dat hoe hoger de score op dat item is, hoe minder er sprake is van concentratie. Daarnaast valt op dat er telkens 92 ontbrekende waarden zijn bij de items rond werken of studeren: dat komt omdat 92 respondenten aangaven dat ze op pensioen zijn of inactief zijn (W2\_ACT) en deze vragen niet voor zich kregen. Die respondenten zullen door SPSS automatisch niet meegenomen worden in de regressieanalyse. Vooraf kun je het best nog nagaan of de schalen intern consistent zijn aan de hand van een Cronbach's alpha toets, maar dit valt buiten de scope van deze oefening.

COMPUTE W2\_W\_STRESS\_Gem=MEAN(W2\_W\_STRESS1,W2\_W\_STRESS2,W2\_W\_STRESS3).

VARIABLE LABELS W2\_W\_STRESS\_Gem 'Gemiddelde score op drie items over werk- of studiestress'.

EXECUTE.

COMPUTE W2\_Angst\_Gem=MEAN(W2\_Angst1,W2\_Angst2,W2\_Angst3,W2\_Angst4).

VARIABLE LABELS W2\_Angst\_Gem 'Gemiddelde score op vier items over angstgevoelens'.

EXECUTE.

RECODE W2\_W\_CONC3 (1=5) (2=4) (3=3) (4=2) (5=1) (**MISSING=SYSMIS**) INTO W2\_W\_CONC3\_Recode.

VARIABLE LABELS W2\_W\_CONC3\_Recode 'Omkering van W2\_W\_CONC3'.

EXECUTE.

COMPUTE W2\_W\_CONC\_Gem=MEAN(W2\_W\_CONC1,W2\_W\_CONC2,W2\_W\_CONC3\_Recode).

VARIABLE LABELS W2\_W\_CONC\_Gem 'Gemiddelde score op drie items over werk- of studieconcentratie'.

EXECUTE.

Assumptiecheck:

1. De afhankelijke variabele (of uitkomstvariabele) is van metrisch meetniveau. **IN ORDE**
2. Er zijn twee of meer onafhankelijke variabelen (of predictoren). **IN ORDE**

3. Onafhankelijkheid: de residuele termen zijn niet gecorreleerd met elkaar. **NAGAAN IN OUTPUT REGRESSIEANALYSE**
4. Lineariteit: de afhankelijke variabele heeft een collectief lineair verband met alle predictoren. **NAGAAN VIA RESIDUENANALYSE**
5. Normaliteit: de residuele termen zijn normaal verdeeld. **NAGAAN VIA RESIDUENANALYSE**
6. Homoscedasticiteit: de residuele termen hebben homogene (of anders gezegd: gelijke) varianties voor elke waarde van de afhankelijke variabele. **NAGAAN VIA RESIDUENANALYSE**
7. Er zijn geen uitschieters onder de residuele termen (de gestandaardiseerde residuele termen zijn niet groter dan drie standaardafwijkingen). **NAGAAN VIA RESIDUENANALYSE**
8. Geen multicollineariteit, waarbij twee of meer predictoren te sterk gecorreleerd zijn met elkaar. **NAGAAN IN OUTPUT REGRESSIEANALYSE**

**Stap 3.** Hypothesen formuleren. Dit moet hier voor (1) het volledige model met alle predictoren en (2) elke predictor van het model apart (maar hier doen we dit bij wijze van voorbeeld alleen voor aantal huisgenoten):

#### Verklaringskracht model

$H_0$ : Aantal huisgenoten, werk- of studiestress en angst hebben geen effect op werk- of studieconcentratie: de verklaringskracht van dit lineair model is gelijk aan nul.

$H_a$ : Aantal huisgenoten, werk- of studiestress en angst hebben wel een effect op werk- of studieconcentratie: de verklaringskracht van dit lineair model is groter of kleiner dan nul.

#### Verklaringskracht predictor aantal huisgenoten

$H_0$ : Aantal huisgenoten heeft geen effect op werk- of studieconcentratie: de verklaringskracht van relatietevredenheid is gelijk aan nul.

$H_a$ : Aantal huisgenoten heeft wel een effect op werk- of studieconcentratie: de verklaringskracht van relatietevredenheid is groter of kleiner dan nul.

**Stap 4.** Significantieniveau bepalen.

Het gaat om een steekproef met meer dan 400 onderzoekseenheden ( $n = 419$ ), dus hanteer  $\alpha = .01$  of 1 %.

**Stap 5.** Toetsingsgrootte en p-waarde berekenen.

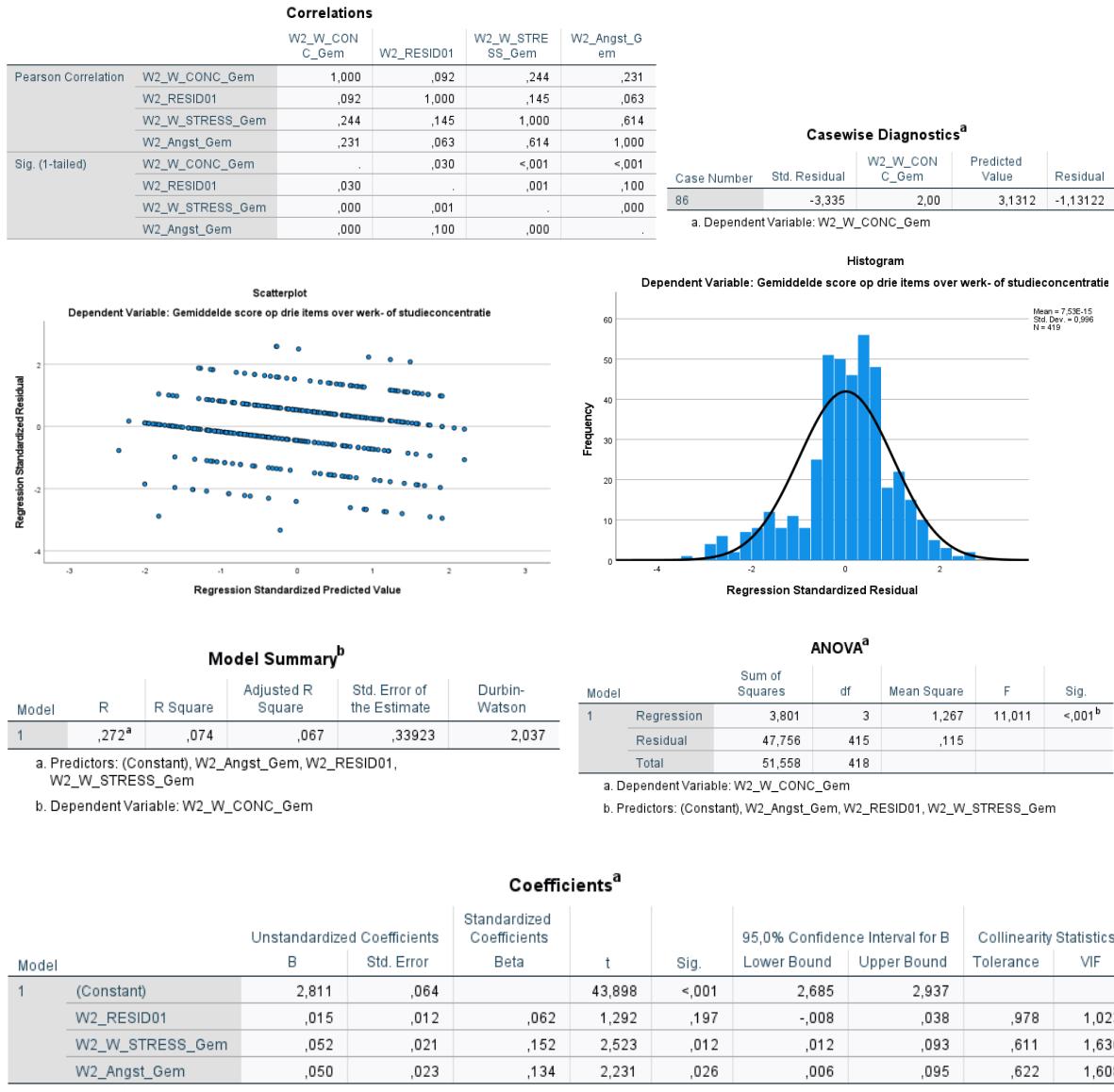
Doorloop de reeks SPSS-acties in het boek om het volgende commando te verkrijgen:

REGRESSION

```
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT W2_W_CONC_Gem
/METHOD=ENTER W2_RESID01 W2_W_STRESS_Gem W2_Angst_Gem
/SCATTERPLOT=(ZRESID, ZPRED)
/RESIDUALS DURBIN HISTOGRAM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

**Output.** Controleer eerst de overige zes assumpties aan de hand van de correlatietafel, VIF-waarden (assumptie 8) en de output voor residuenanalyse (assumptie 3-7). **Assumptie 8 is oké**, ook al is de correlatie tussen de predictoren werk- of studiestress en angst groter dan .60: geen enkele VIF-waarde is groter dan 5. Verder zijn  
© 2022, Toegepaste Statistiek – Wim Hardyns, Koen Ponnet en Pelckmans Uitgevers nv

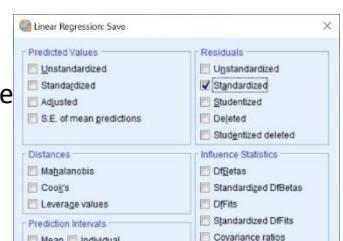
**assumptie 4 en 6 ook oké**, want de scatterplot vertoont geen non-lineair verband (lineariteit in orde) en de datapunten vormen een ovaal en geen kegel- of zandlopervorm (homoscedasticiteit in orde). **Over de normaliteit van residuen, assumptie 5, kun je als onderzoeker twijfelen:** het histogram van de residuen vertoont wel een soort normale verdelingsvorm, maar het histogram is erg gepiekt en vult links niet helemaal de normale curve die erop geprojecteerd werd. We kunnen als tweede stap kijken naar de z-scores van scheefheid en kurtosis van de residuen om een besluit te maken inzake normaliteit (zie hieronder). Bij *Casewise Diagnostics* zie je dat één uitschieter voor problemen kan zorgen, maar die is **lang niet extreem genoeg om assumptie 7 rond uitschieters niet als vervuld te beschouwen**. Tot slot ligt de toetsingsgrootte van de Durbin-Watson toets (zie *Model Summary* tabel) dicht bij 2, waardoor **assumptie 3 rond onafhankelijkheid van residuele termen ook in orde is**.



#### Assumptie 5. Normaliteit van residuen controleren a.d.h.v. z-scores scheefheid en kurtosis:

Actie 1. Vink tijdens het uitvoeren van het regressiecommando in de dialoogvensters onder de knop **Save** de optie **Standardized** aan in het kader **Residuals**. Je kunt ook eenvoudigweg in de **Syntax** het lijntje **/SAVE ZRESID** toevoegen bij het regressiecommando onderaan. Je vraagt hiermee aan SPSS om een nieuwe variabele aan te maken, waarin de gestandaardiseerde residuen van het regressiemodel opgeslagen worden van elke respondent in de dataset.

#### REGRESSION



```

/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT W2_W_CONC_Gem
/METHOD=ENTER W2_RESID01 W2_W_STRESS_Gem W2_Angst_Gem
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS DURBIN HISTOGRAM(ZRESID)
/CASEWISE PLOT(ZRESID) OUTLIERS(3)
/SAVE ZRESID.

```

Actie 2. Voer het bovenstaande regressiecommando uit. Een nieuwe variabele met de gestandaardiseerde residuen zal nu verschijnen in je databestand, genaamd 'ZRE\_1' (of een ander nummer dan 1, afhankelijk van of je al eerder andere residuen opgeslagen hebt).

Actie 3. Vraag via het *Frequencies*-commando scheefheid, kurtosis en hun standaardfouten op van de variabele 'ZRE\_1'.

Statistics		
ZRE_1		
	N	Valid
/FREQUENCIES VARIABLES=ZRE_1		419
/STATISTICS=SKEWNESS SESKEW KURTOSIS SEKURT		92
/ORDER=ANALYSIS.		
Skewness		-,610
Std. Error of Skewness		,119
Kurtosis		,869
Std. Error of Kurtosis		,238

Actie 4. Bereken de z-scores van scheefheid en kurtosis. Je ziet hier dat beide z-scores buiten de grenzen van +/- 1.96 vallen, waardoor we kunnen besluiten dat de residuen niet normaal verdeeld zijn.

$$z_{\text{scheefheid}} = -0.61 / 0.12 = -5.08 \quad z_{\text{kurtosis}} = 0.87 / 0.24 = 3.63$$

In principe kun je hier geen lineaire regressie toepassen, maar om didactische redenen zullen we in deze oefening de assumpties even als vervuld beschouwen en de resultaten interpreteren.

**Stap 6.** Conclusie en rapportering. De vrijheidsgraden van de t-toetsen van elke predictor vind je als volgt:  $df = n - k$ , met  $k$  = aantal variabelen in de regressieanalyse (hier 4). **LET OP, deze rapportering is bij wijze van voorbeeld. Aangezien de normaliteitsassumptie niet vervuld is van de lineaire regressieanalyse, is de analysetechniek niet helemaal betrouwbaar en interpreteer je de resultaten beter niet.**

"Het effect van het aantal huisgenoten, werk- of studiestress en angst als predictoren van werk- of studieconcentratie werden gezamenlijk getoetst aan de hand van een meervoudig lineair regressiemodel. Dit lineair model verklaart een significant deel van de variantie in werk- of studieconcentratie, met  $R^2 = .074$ ,  $R^2_{\text{Adjusted}} = .067$ ,  $F(3, 415) = 11.01$ ,  $p < .001$ . Hieruit blijkt dat alle predictoren samen een significant deel van de variantie in werk- of studieconcentratie verklaren, maar elk apart hebben ze niet voldoende verklaringskracht: zowel het effect van het aantal huisgenoten ( $\beta = -0.06$ ,  $t(415) = 1.29$ ,  $p = .197$ ) als het effect van werk- of studiestress ( $\beta = 0.15$ ,  $t(415) = 2.52$ ,  $p = .012$ ) en het effect van angst ( $\beta = 0.13$ ,  $t(415) = 2.23$ ,  $p = .026$ ) waren niet significant op het vooropgestelde significantieniveau ( $\alpha = .01$ ).