

SPSS

- hoofdstuk 1 : [1.4. fase 4 : verrichten van metingen en / of verzamelen van gegevens](#)

Gegevens gevonden bij een onderzoek worden systematisch weergegeven in een datamatrix bij SPSS

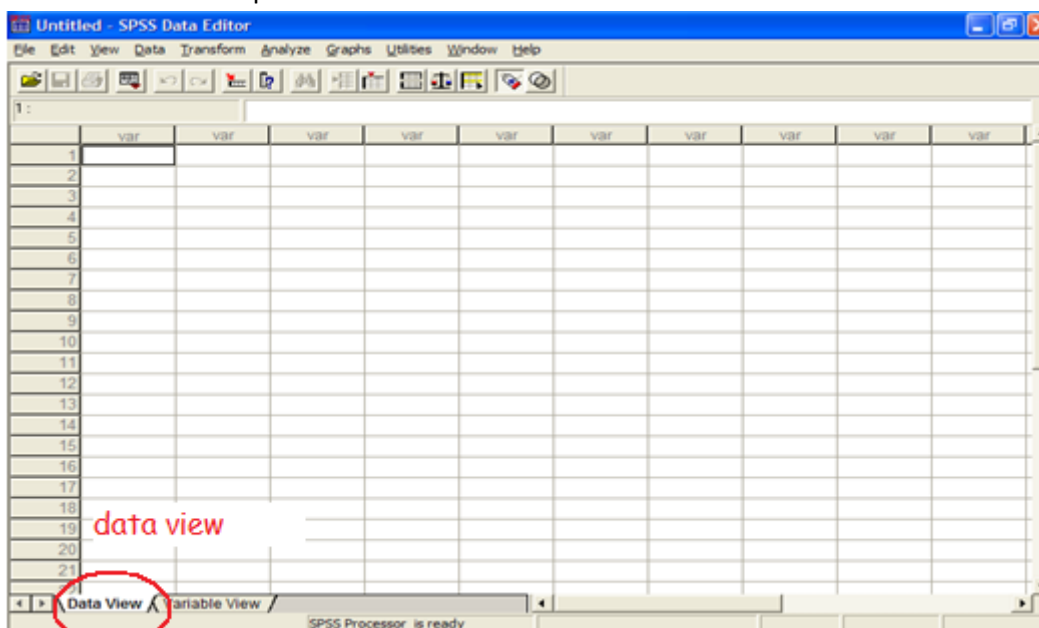
De datamatrix

- Gebruik van Excel: zeer beperkte mogelijkheden voor inductieve statistiek
- Gebruik van SPSS: zeer handig.

Meest gebruikte software voor statistische analyse.

SPSS: Statistical Package for the Social Science (SPSS)

- Twee tabs:
 - variable view
 - data view
- Twee windows:
 - input
 - output



SPSS Variable view

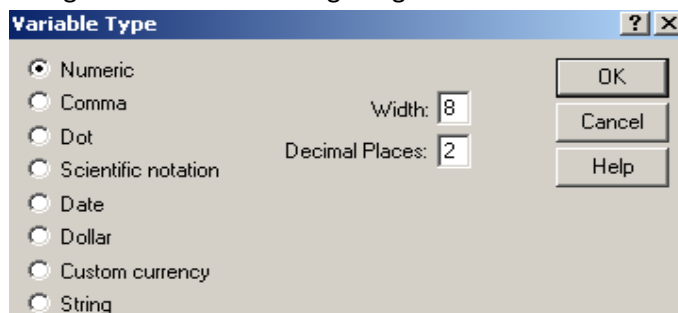
- Name: naam van de variabele; dit gegeven komt in de data file boven de variabele staan, beperkt aantal karakters, geen spaties,...
- Type: kies standaard voor 'numeriek' en u kunt voor deze variabele getallen invoeren. U kunt voor 'string' kiezen, om bv. de namen van de proefpersonen in te geven, of om bemerkingen te noteren, bv. persoon geeuwde vaak, vermoeidheid?
- Width: hoeveel posities heeft u nodig voor deze variabele? Standaard is 8.
- Decimals: indien u geen gebruik maakt van decimale getallen kiest u voor nul
- Label: de naam van de variabele die u nu ingeeft zal gebruikt worden als titel boven een tabel, grafiek...; minder beperkingen dan bij name
- Values: hier geeft u aan welke getallen voor welke groepen staan. (niet te gebruiken bij scale metingen), bv. 1 is man en 2 staat voor vrouw
- Missing value: op welke wijze geeft u aan dat een persoon dit niet ingevuld heeft? U gebruikt hiervoor een getal dat onmogelijk als waarde gescoord kan worden.
- Measure: welk is het niveau van meting: nominal, ordinal, scale.

Begin ALTIJD met het aanmaken van de variable view

Type :

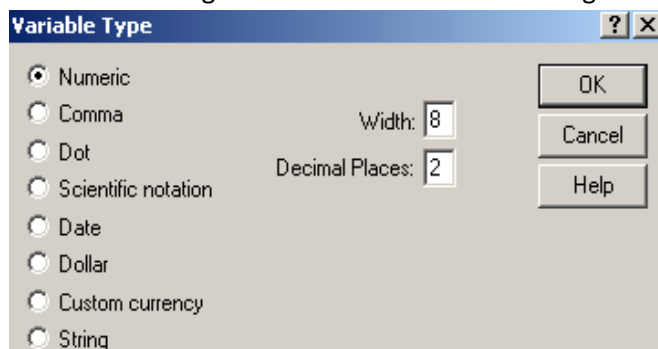
numeric = getallen --> kan dus een gemiddelde worden berekent

string = letters --> kan dus geen gemiddelde worden berekent



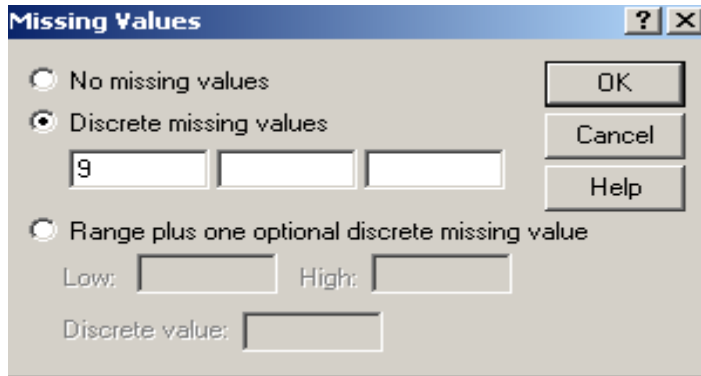
Values :

moet worden ingevuld als de variabele een categorie is



Missing:

wat niet in gevuld is uit de enquête , daar wordt een getal voor geplaatst dat geen waarde kan zijn

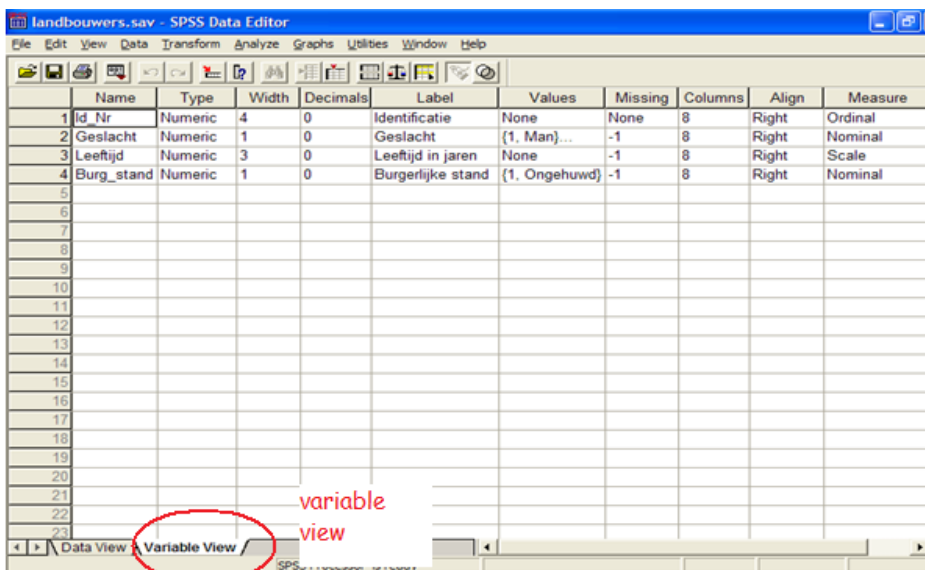


Scale : meetbaar, metingen

ordinaal : volgorde vb.: niet akkoord, beetje akkoord, grotendeels akkoord, helemaal akkoord

nominaal : groepen, categorieën : studenten psychologie, studenten logopedie

Columns	Align	Measure
8	Right	Scale
		Scale
		Ordinal
		Nominal



landbouwers.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1: Id_Nr 1

	Id_Nr	Geslacht	Leeftijd	Burg_stand	var	var	var	var	var	var
1	1	1	36	1						
2	2	2	64	2						
3	3	2	71	4						
4	4	2	48	2						
5	5	2	39	2						
6	6	1	44	2						
7	7	1	19	-1						
8	8	1	29	3						
9	9	1	53	2						
10	10	2	44	2						
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

Gegevens van persoon nr. 3

Gegevens over leeftijd

Data View Variable View

SPSS Processor is ready

landbouwers.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1: Id_Nr 1

	Id_Nr	Geslacht	Leeftijd	Burg_stand	var	var	var	var	var	var
1	1	Man	36	Ongehuwd						
2	2	Vrouw	64	Gehuwd						
3	3	Vrouw	71	Weduwe						
4	4	Vrouw	48	Gehuwd						
5	5	Vrouw	39	Gehuwd						
6	6	Man	44	Gehuwd						
7	7	Man	19	-1						
8	8	Man	29	Gescheide						
9	9	Man	53	Gehuwd						
10	10	Vrouw	44	Gehuwd						
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

Data View Variable View

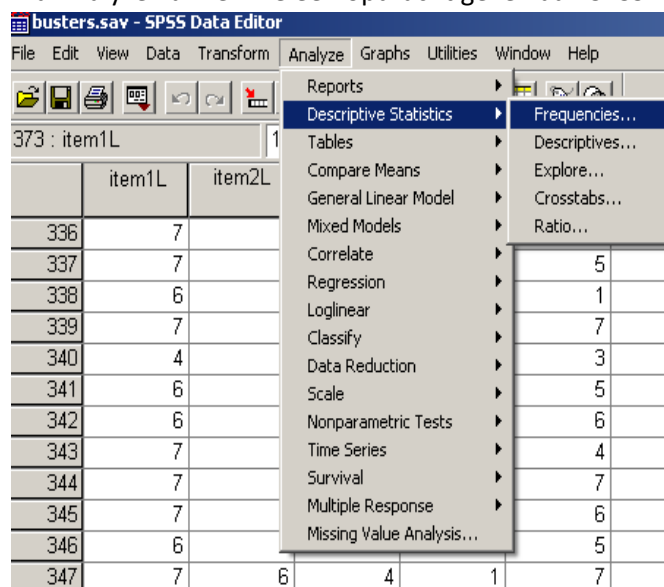
SPSS Processor is ready

SPSS. Data View **Zeer belangrijk**

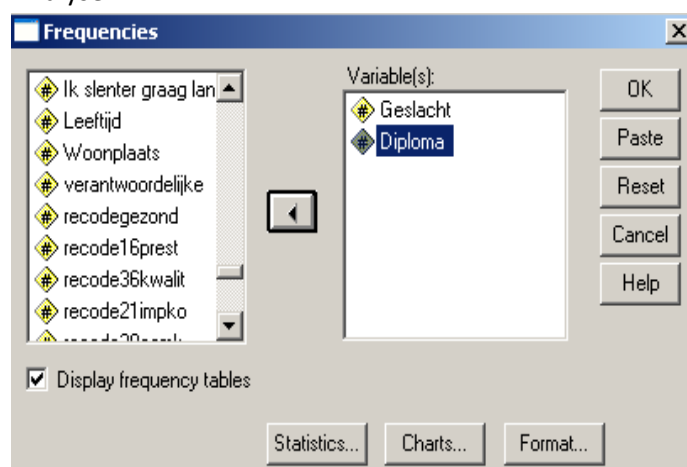
- Resultaten van één proefpersoon vinden we in één en slechts één rij. Alle gegevens van één persoon staan in één rij. In één rij vinden we de gegevens van slechts één proefpersoon.
- Resultaten van één variabele vinden we in één en slechts één kolom. Alle gegevens van één variabele vinden we in één kolom en de gegevens in één kolom hebben betrekking op slechts één variabele.

SPSS : output viewer

Via Analyze kunnen we een opdracht geven aan SPSS



Analyse



Output viewer

The screenshot shows the SPSS Output Viewer window. The left pane displays a tree view of the output, with 'Frequencies' expanded to show 'Statistics', 'Frequency Table', 'Geslacht', and 'Diploma'. The main area displays the following tables:

Statistics

	Geslacht	Diploma
N	Valid 402	402
	Missing 0	0

Frequency Table

Geslacht

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid man	142	35,3	35,3	35,3
vrouw	260	64,7	64,7	100,0
Total	402	100,0	100,0	

Diploma

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid lager onderwijs	15	3,7	3,7	3,7
lager middelbaar	67	16,7	16,7	20,4
hoger middelbaar	150	37,3	37,3	57,7
hoger niet-universitair onderwijs	127	31,6	31,6	89,3
universitair onderwijs	43	10,7	10,7	100,0
Total	402	100,0	100,0	


Opslaan van gegevens in SPSS


- Gegevens kunnen bewaard worden via File, Save as...
- Datafile krijgt automatisch extensie .sav
Bv. busters.sav
- Output gegevens krijgen een extensie .spo
Bv. busters.spo
- Gegevens kunnen opgevraagd worden via File, Open, Data...


[HOOFDSTUK 2 : 2.3. betekenis van dit meetniveau](#)

THEORIE : ZIE SAMENVATTING

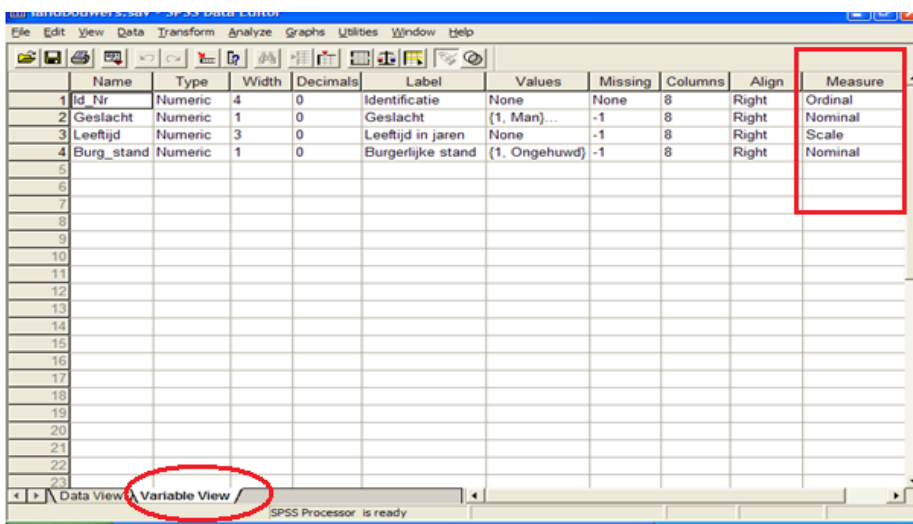
- In de psychologie vooral nominale, ordinale en interval variabelen, weinig ratio schalen. Onderscheid tussen interval en niet interval is het belangrijkste onderscheid, met oog op de analyse van de gegevens. SPSS maakt geen onderscheid tussen interval en ratio: nominal, ordinal en scale

 SCALE (= interval en ratio) (vb: leeftijd)

 NOMINAAL (vb: geslacht)

 ORDINAAL (vb: diploma)

-> dit is in te vullen bij variable view -> measure = niveau van variabelen



Soms berekent spss zinloze waarden :

Descriptive Statistics

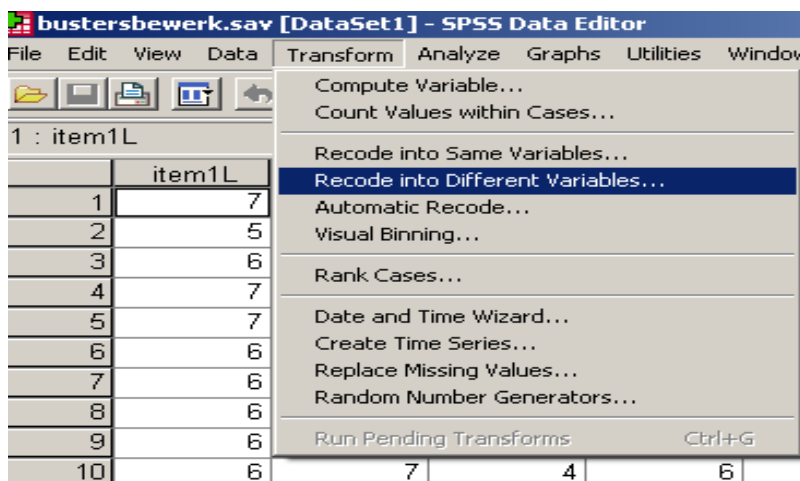
	N	Minimum	Maximum	Mean	Std. Deviation
Identificatie	10	1	10	5,50	3,028
Geslacht	10	1	2	1,50	,527
Leeftijd in jaren	10	19	71	44,70	15,492
Burgerlijke stand	9	1	4	2,22	,833
Valid N (listwise)	9				

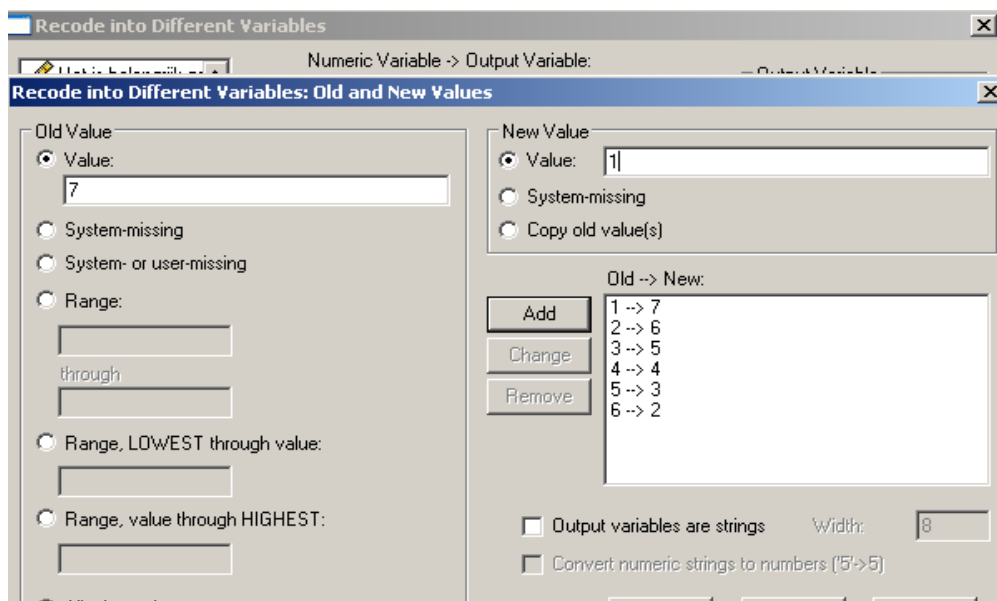
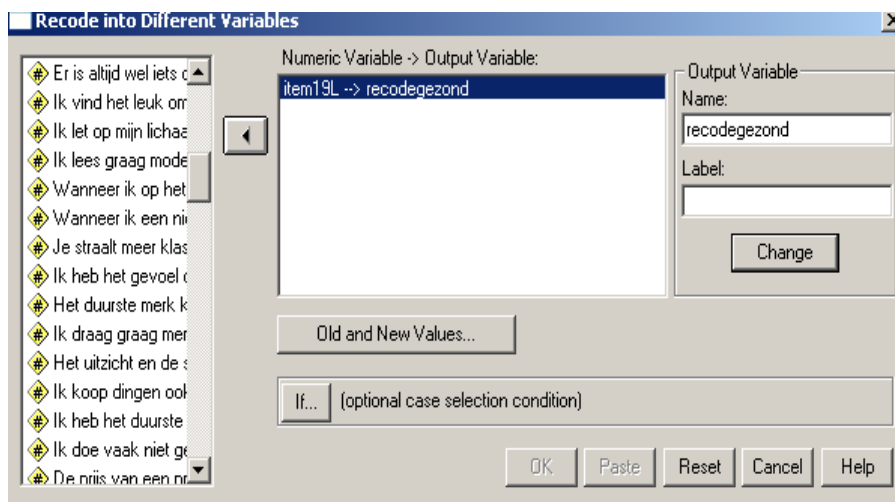
--> zie theorie nu van 2.3

Hoort bij hoofdstuk 2 : SPSS TRANSFORM

- U kunt variabelen bewerken via het menu transform
 - compute maakt berekeningen op variabelen, bv. optellen, keer, etc...
 - via recode kunt u variabelen hercoderen
 - via count u bepaalde scores tellen per persoon.

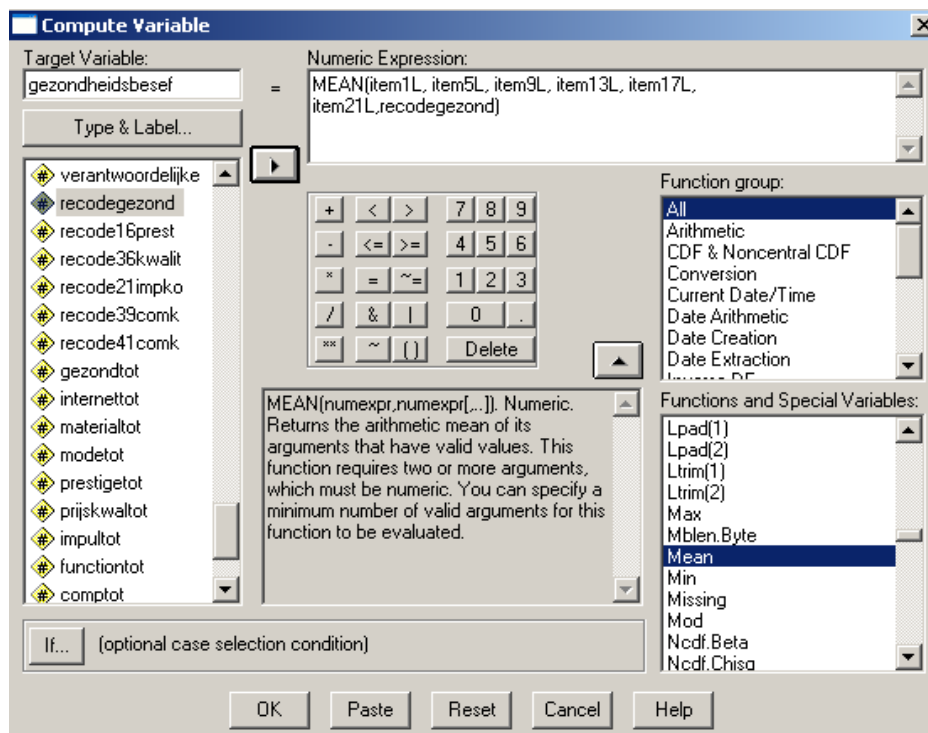
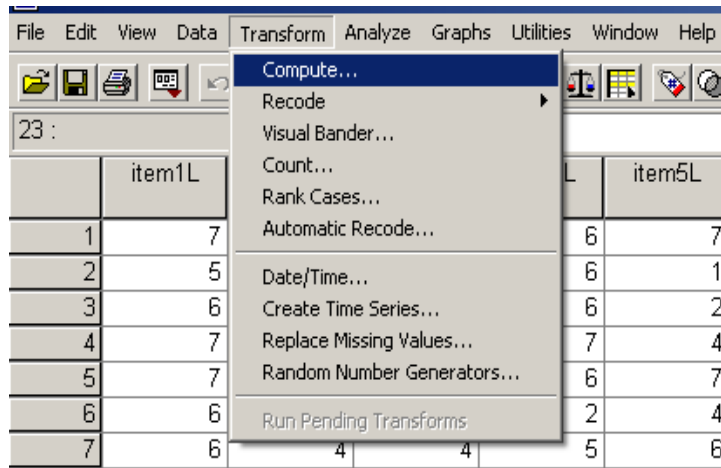
SPSS transform – RECODE





- Op deze wijze ontstaat een nieuwe variabele met als titel 'recodegezond'. Deze variabele is precies het omgekeerde van item 19L.
Tip: maak gebruik van recode 'into a different var';
- Voor de totaalscore van gezondheidsbesef moeten we gebruik maken van 'recodegezond' en niet van het item 19L
- Recode IF kan ook gebruikt worden. U definieert een conditie waaraan voldaan moet worden om de recode uit te voeren.

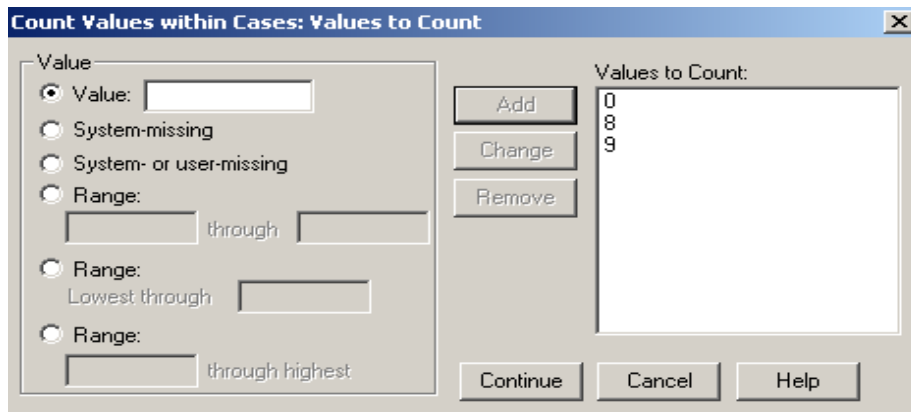
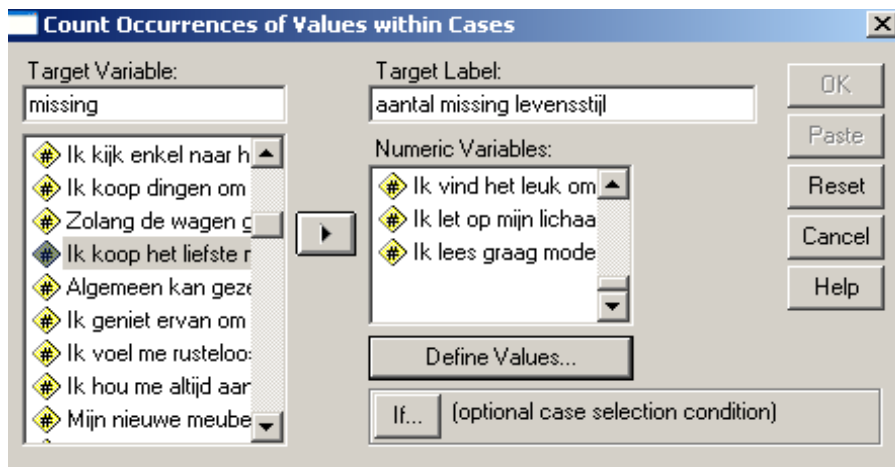
SPSS transform – COMPUTE



- We zouden ook gebruik kunnen maken van de som van de items. ???
- Op deze wijze ontstaat een nieuwe variabele 'gezondheidsbesef', waarmee we zullen verder werken, en niet meer met de afzonderlijke items.
- U kunt ook gebruik maken van de mogelijkheid om een conditie aan te geven (compute IF) om deze compute uit te voeren

SPSS transform -> COUNT

tel het aantal missing values per persoon



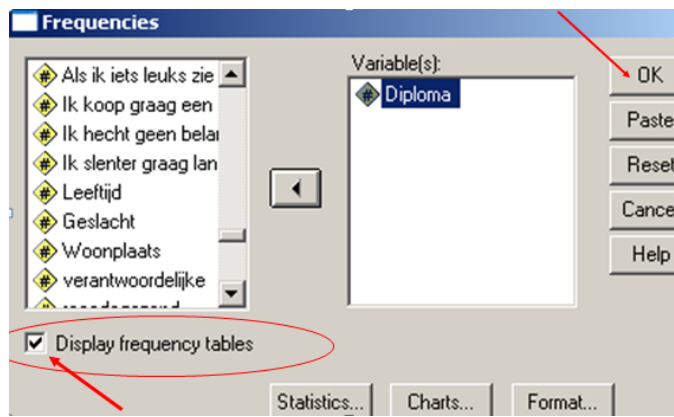
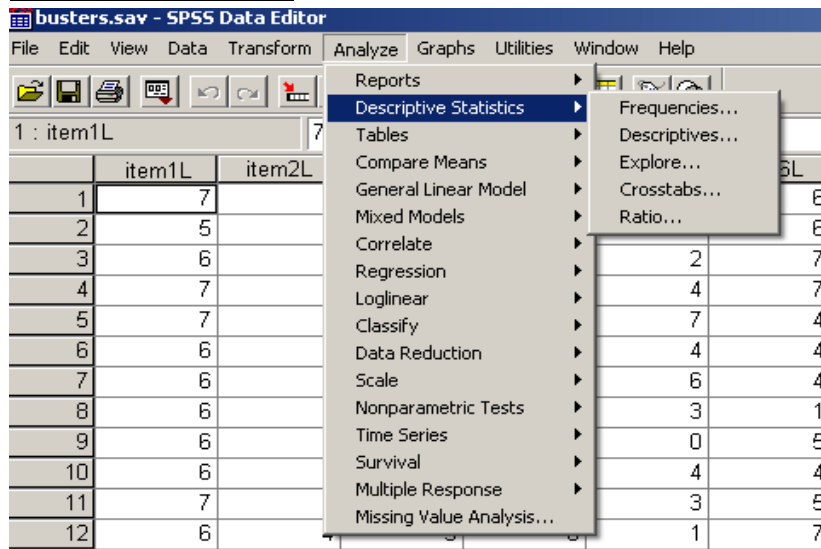
- Op deze wijze ontstaat een nieuwe variabele waarin het aantal missing values per persoon samengeteld wordt.
- U kunt hierbij ook gebruik maken van de optie Count IF.

HOOFDSTUK 3 : [Frequentietabel en histogram met SPSS !!](#)

Frequentietabel en histogram met SPSS

- SPSS zal voor een frequentietabel niet automatisch een gegroepede tabel maken. Dit is wel mogelijk via recode van de gegevens. Eventueel een stem and leaf plot.
- Als histogram zal SPSS bij een grote diversiteit van waarden automatisch de waarden in klassen indelen

AANMAKEN VAN TABEL



Spss output van een tabel

Diploma

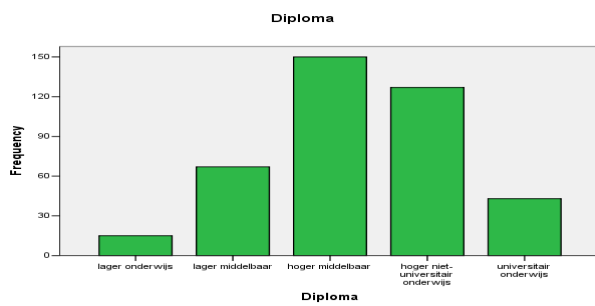
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid lager onderwijs	15	3,7	3,7	3,7
lager middelbaar	67	16,7	16,7	20,4
hoger middelbaar	150	37,3	37,3	57,7
hoger niet-universitair onderwijs	127	31,6	31,6	89,3
universitair onderwijs	43	10,7	10,7	100,0
Total	402	100,0	100,0	

Spss aanmaken van een staafdiagram

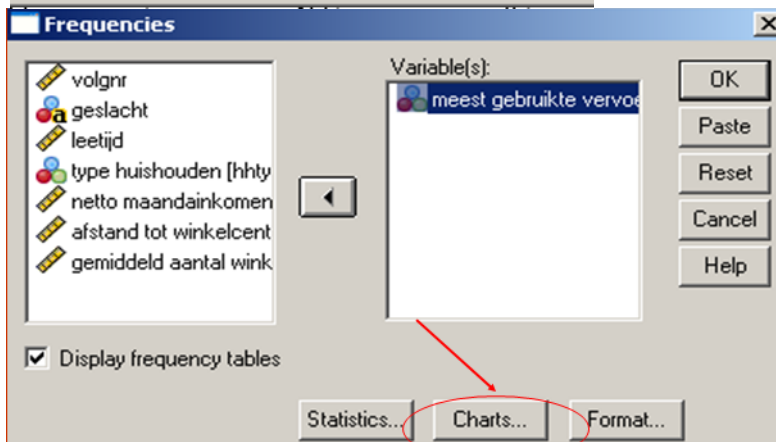
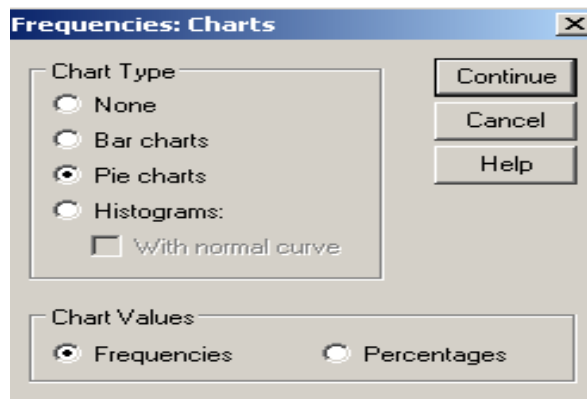
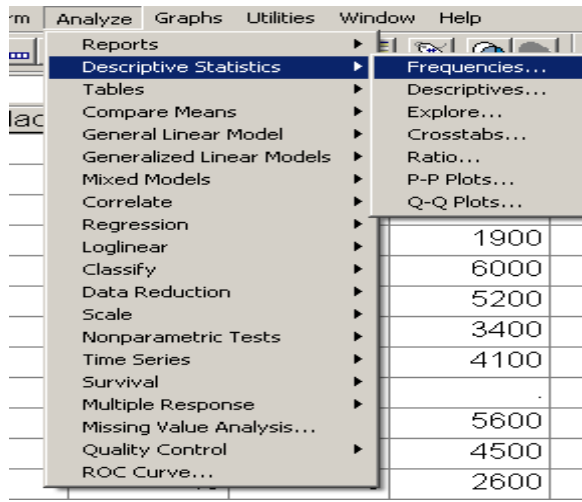
The screenshot shows the SPSS interface with the 'Frequencies' dialog box open. The 'Variable(s):' list contains 'Diploma'. The 'Charts' sub-dialog box is also open, showing 'Bar charts' selected under 'Chart Type' and 'Frequencies' selected under 'Chart Values'. The background shows a data table with columns 'item4L' through 'item10L' and rows of data.

item4L	item5L	item6L	item7L	item8L	item9L	item10L	it
6							
6							
7							
3							

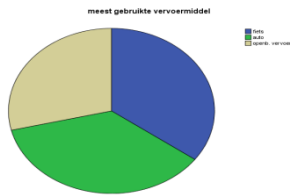
Spss output van een staafdiagram



SPSS en een taartdiagram



grafiek van nominale waarden :



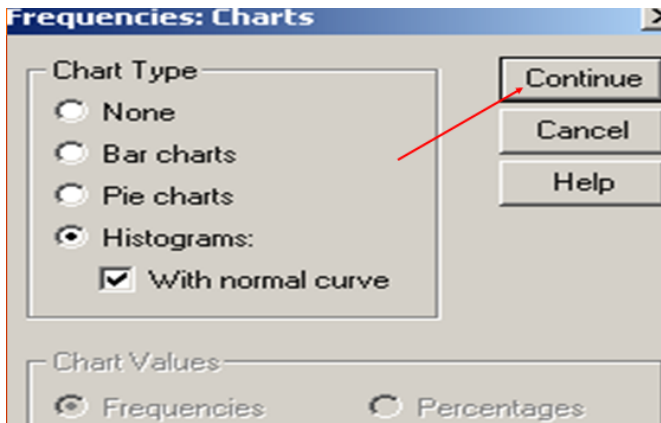
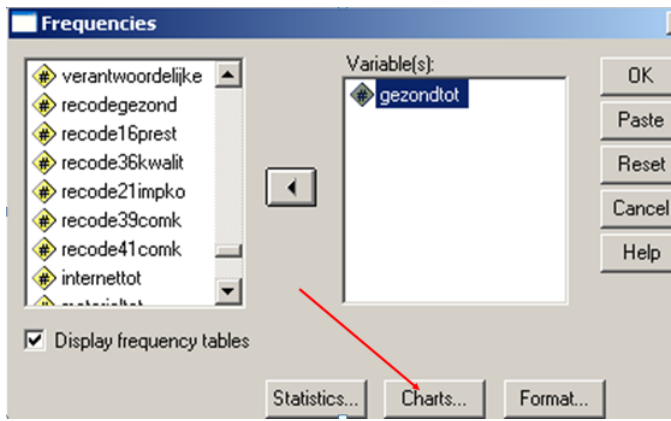
SPSS tabel en grafiek maken van interval waarden

	item1L	item2L
1	7	
2	5	
3	6	
4	7	
5	7	
6	6	
7	6	
8	6	
9	6	
10	6	
11	7	
12	6	

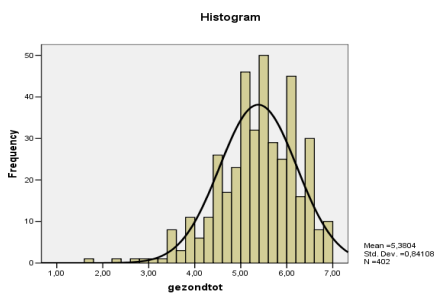
SPSS maakt niet automatisch een gegroepede frequentietabel, dus een gegroepede frequentietabel maken

	impultot	fur
53	2.75	
54	2.00	
55	3.50	

SPSS aanmaken van een histogram



SPSS output van een histogram



SPSS : boxplot maken

.sav [DataSet2] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Chart Builder...

Interactive
Legacy Dialogs
Map

Bar...
Dot...
Line...
Ribbon...
Drop-Line...
Area...
Pie
Boxplot...
Error Bar...
Histogram...
Scatterplot...

gezondtot	internetto	pt
4,86	7	
5,43	6,67	
4,86	5,00	
5,71	7,00	
6,71	4,00	
5,29	5,00	
6,00	4,00	
4,71	1,00	
4,50	5,00	
5,29	3,33	3,75
5,71	5,33	4,25
4,43	6,33	4,50
5,00	7,00	4,00
6,14	3,67	5,50
5,57	4,67	3,50
6,14	4,00	4,00
7,00	5,00	3,00
6,57	7,00	4,50
5,71	6,00	5,25
5,43	1,00	4,00

Create Boxplot

Assign Variables Boxes Titles Options

Algemeen kan c
Als ik iets leuk
Als ik iets nieu
Anderen vragen
[comptot]
De aankoop vai
De kwaliteit is ni
De prijs van eer
[Diploma]
[diplomahogerr
[diplomahogesc
[diplomalager]
[diplomalagermi
Er is altijd wel iel
[functiontot]
Geld maakt gelu
Het dragen van
Het duurste mer
Het is belangrijk
Het is belangrijk

[gezondtot]

[Geslacht]

Legend Variables

Color:

Style:

Panel Variables

Label Cases By:

OK Paste Reset Cancel Help

OF !!!! (hoofdstuk 4 : na de mediaan)

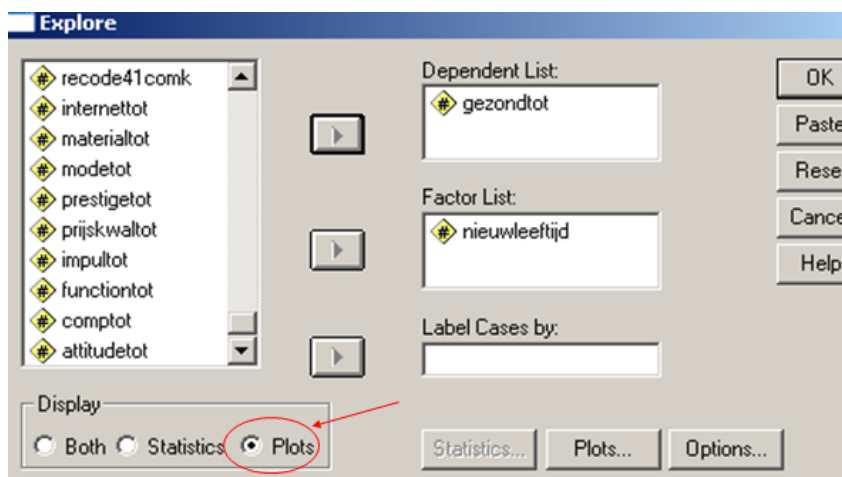
busters.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

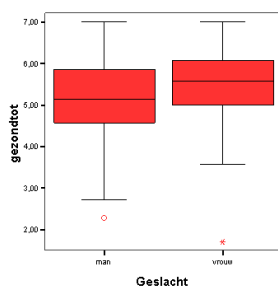
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...

Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...

373 : item1L	item1L	item2L
336	7	
337	7	
338	6	
339	7	
340	4	
341	6	
342	6	
343	7	
344	7	
345	7	
346	6	
347	7	6



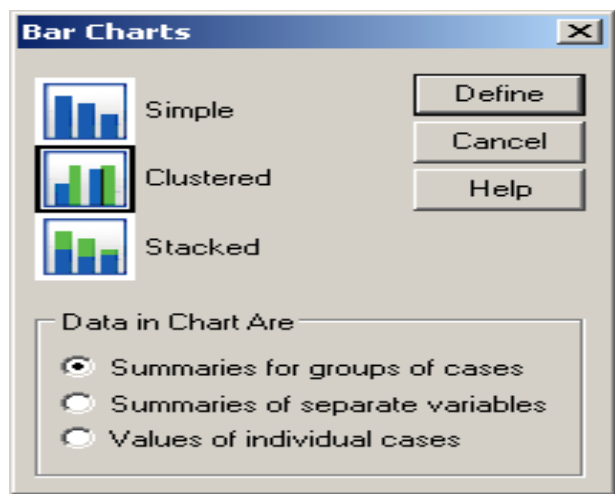
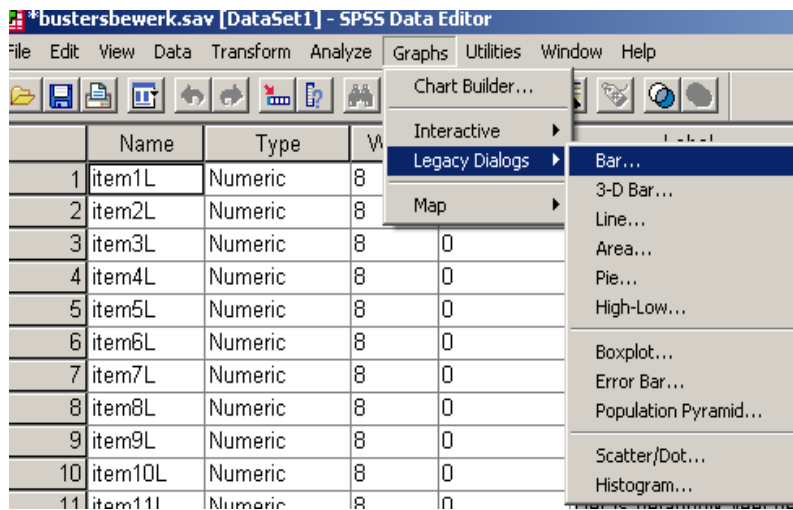
BOXPLOT :

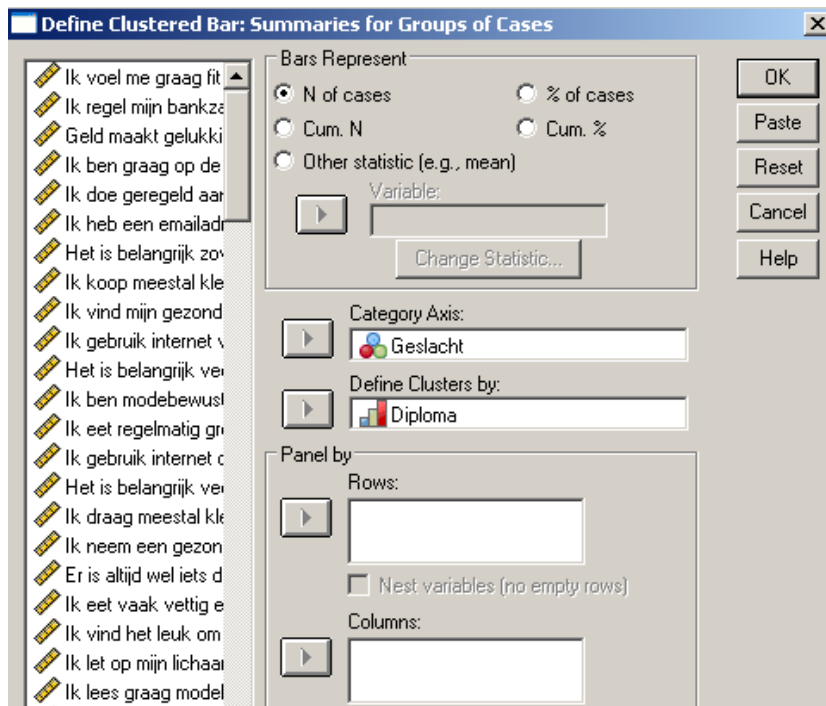


SPSS werken interactieve grafiek

Hoe zou de verdeling zijn van de diploma's bij dames en heren in deze steekproef?

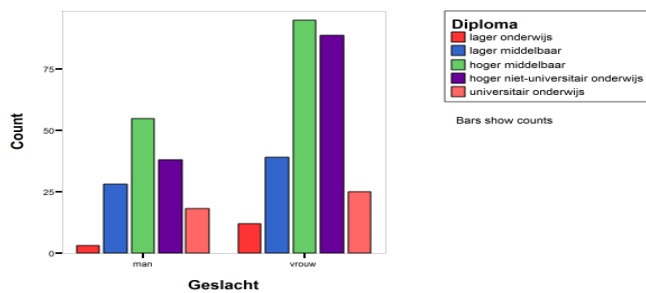
GECLUSTERD STAAFDIAGRAM





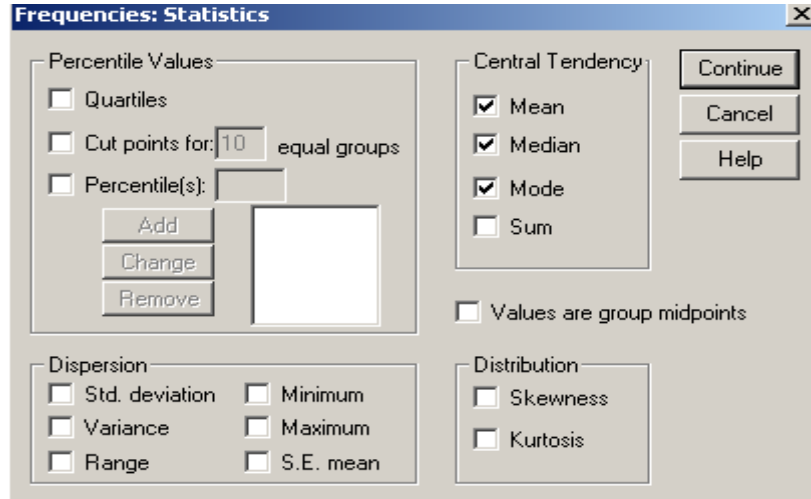
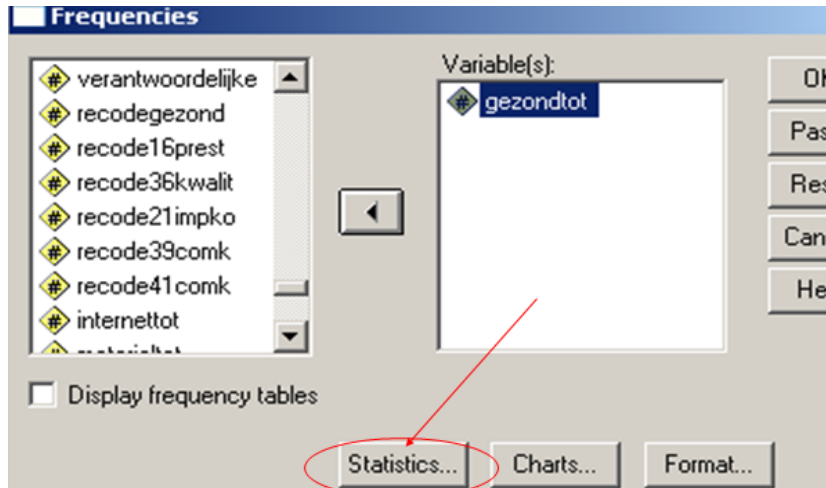
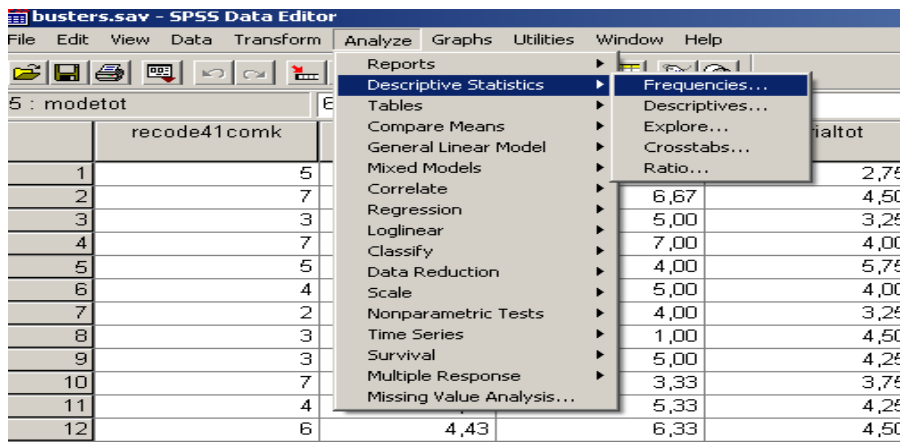
Diploma

- lager onderwijs
- lager middelbaar
- hoger middelbaar
- hoger niet-universitair onderwijs
- universitair onderwijs



[HOOFDSTUK 4 : centrummaten](#)

[SPSS en centrummaten](#)



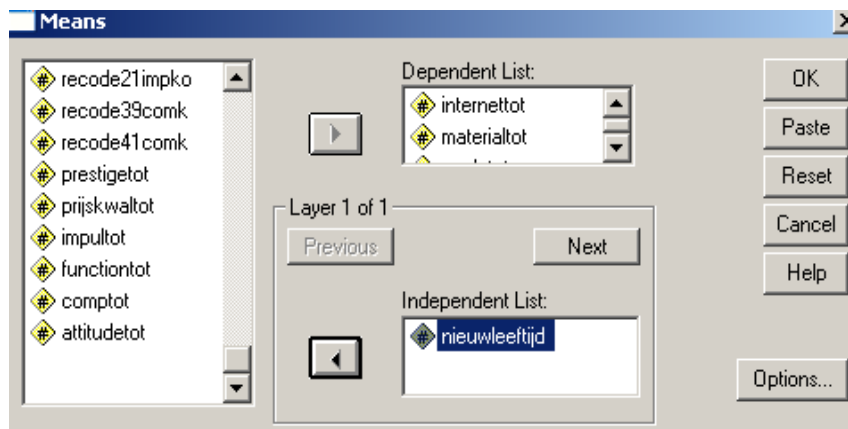
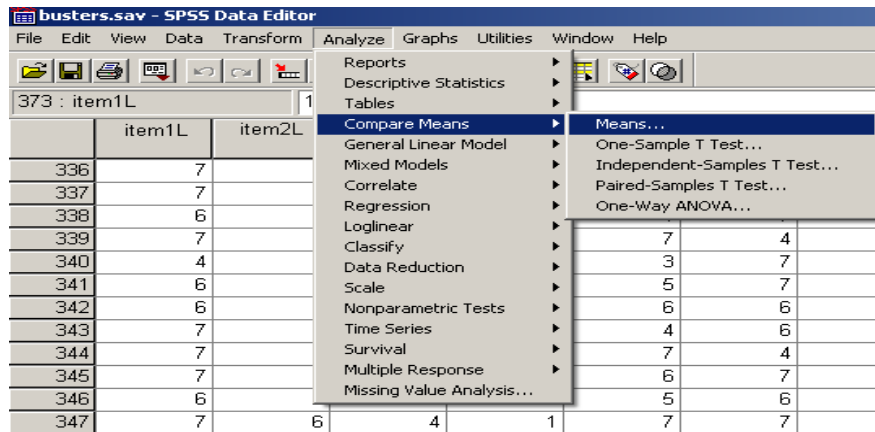
Spss output

Statistics		
gezondtot		
N	Valid	402
	Missing	0
Mean		5,3804
Median		5,4286
Mode		5,57

spss en het rekenkundig gemiddelde

- Om subgroepen te vergelijken maken we vaak gebruik van het rekenkundig gemiddelde.
- Maak uitgaande van het bestand busters.sav een vergelijking tussen de beide leeftijdsgroepen voor wat betreft de levensstijl variabelen (op grond van de gemiddelden)

Spss maak een vergelijking tussen de subgroepen



spss output

Report

nieuwleeftijd		gezondtot	internettot	materialtot	modetot
buster	Mean	5,0967	5,2650	4,2222	4,5385
	N	117	117	117	117
	Std. Deviation	,88092	1,41593	1,07016	1,24979
boomer	Mean	5,5261	4,6196	3,9212	4,1547
	N	276	276	276	276
	Std. Deviation	,75998	1,77200	1,08493	1,19285
Total	Mean	5,3982	4,8117	4,0108	4,2690
	N	393	393	393	393
	Std. Deviation	,82065	1,69805	1,08795	1,22118

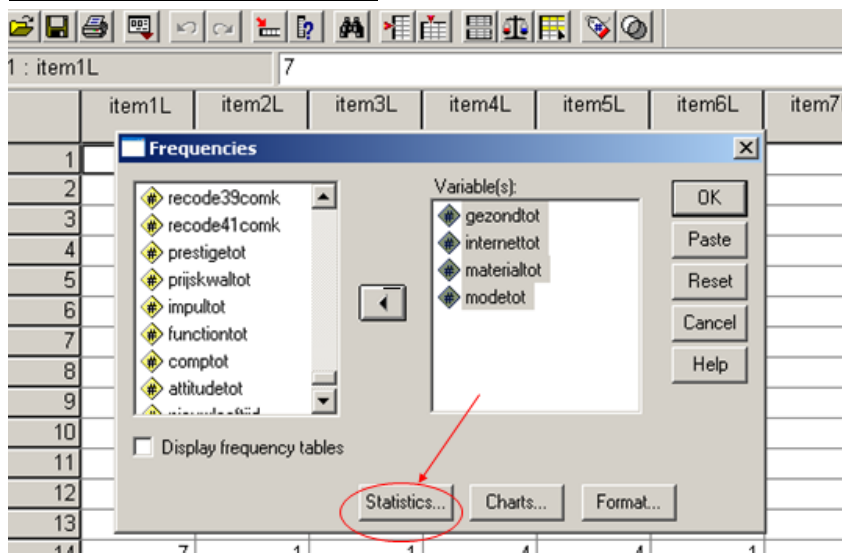
BESLUIT centrummaten :

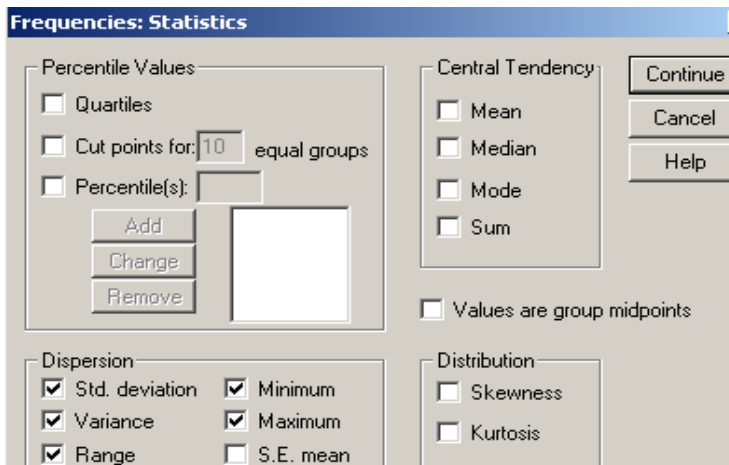
Bij centrummaten gaat het om het centrum van de geobserveerde scores. We onderscheiden modus, mediaan en gemiddelde. De mediaan is niet gevoelig voor extreme waarden, het gemiddelde wel. Het gemiddelde varieert minder (in vergelijking met de mediaan) wanneer men uit een populatie meerdere steekproeven trekt.

[HOOFDSTUK 5 : spreidingsmaten](#) [standaardafwijking ...](#)

Analoog voor de variantie; SPSS maakt altijd gebruik van N-1 (schatting van de sigma)

SPSS EN STANDAARD DEVIATIE





Output1 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Output

- Frequencies
 - Title
 - Notes
 - Statistics

Frequencies

Statistics

		gezondtot	internettot	materialtot	modetot
N	Valid	402	402	402	402
	Missing	0	0	0	0
Std. Deviation		,84108	1,71104	1,10038	1,25498
Variance		,707	2,928	1,211	1,575
Range		5,29	6,00	6,00	6,00
Minimum		1,71	1,00	1,00	1,00
Maximum		7,00	7,00	7,00	7,00

SPSS en standaardscores

busters.sav - SPSS Data Editor

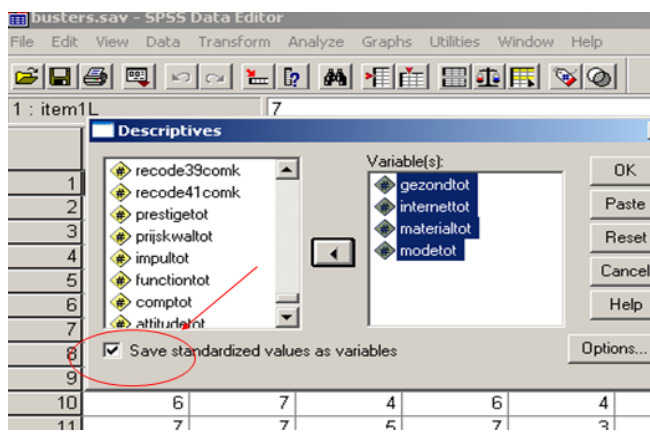
File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : item1L

	item1L	item2L			
1	7				
2	5				
3	6				
4	7				
5	7				
6	6				
7	6				
8	6				
9	6				
10	6				
11	7				
12	6	4	5	6	1

Analyze

- Reports
- Descriptive Statistics
 - Frequencies...
 - Descriptives...
 - Explore...
 - Crosstabs...
 - Ratio...
- Tables
- Compare Means
- General Linear Model
- Mixed Models
- Correlate
- Regression
- Loglinear
- Classify
- Data Reduction
- Scale
- Nonparametric Tests
- Time Series
- Survival
- Multiple Response
- Missing Value Analysis...



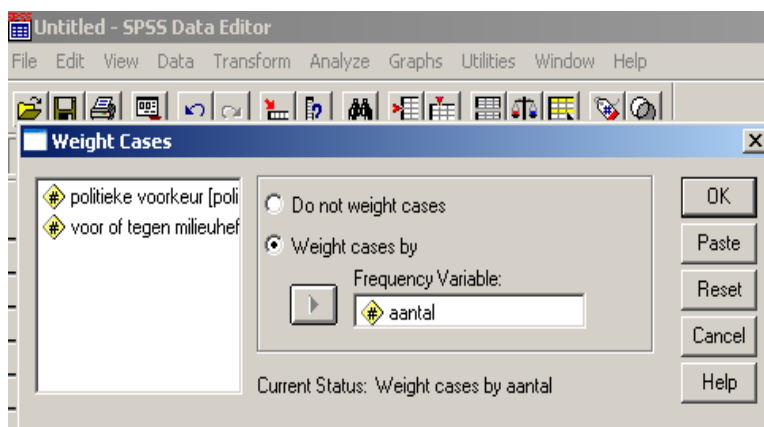
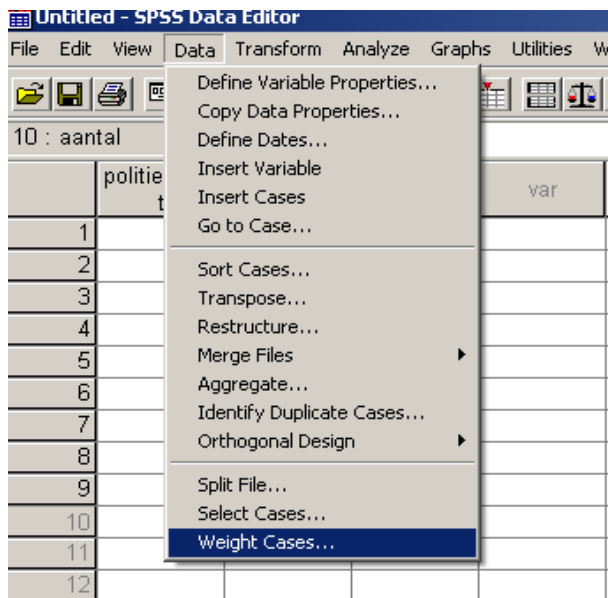
[Hoofdstuk 8 : kruistabellen en spss crosstabs](#)
snelle invoer van een kruistabel

Untitled - SPSS Data Editor

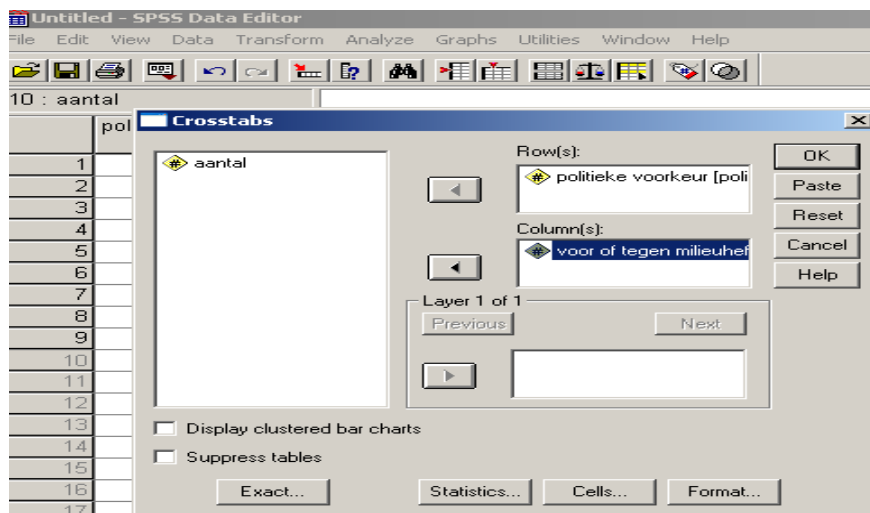
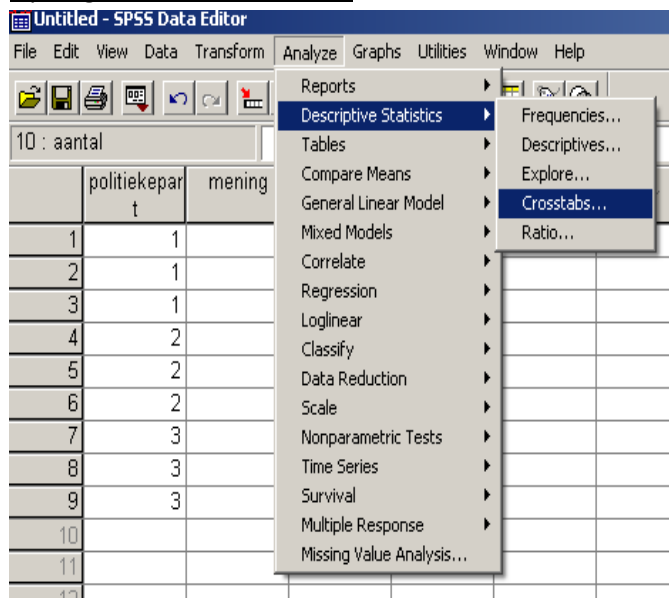
File Edit View Data Transform Analyze Graphs Utilities W

10 : aantal

	politiekepar t	mening	aantal	var
1	1	1	12	
2	1	2	5	
3	1	3	8	
4	2	1	5	
5	2	2	12	
6	2	3	23	
7	3	1	12	
8	3	2	7	
9	3	3	1	
10				
11				



Opvragen van een kruistabel



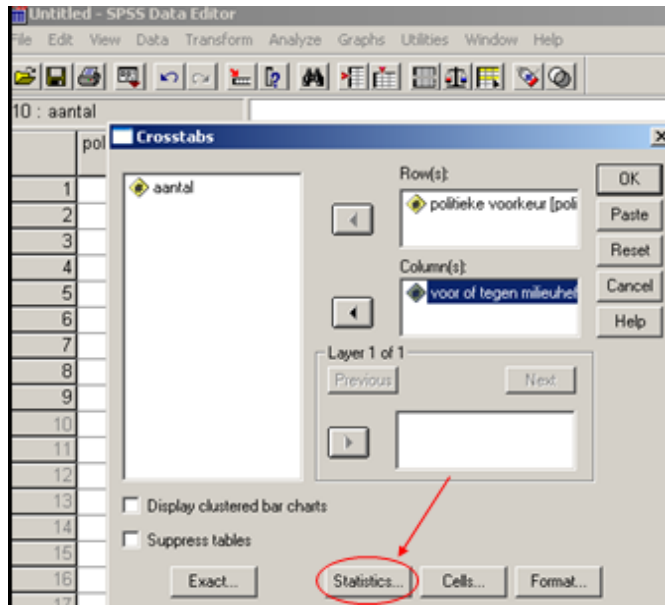
output van de kruistabel

politieke voorkeur * voor of tegen milieueffing Crosstabulation

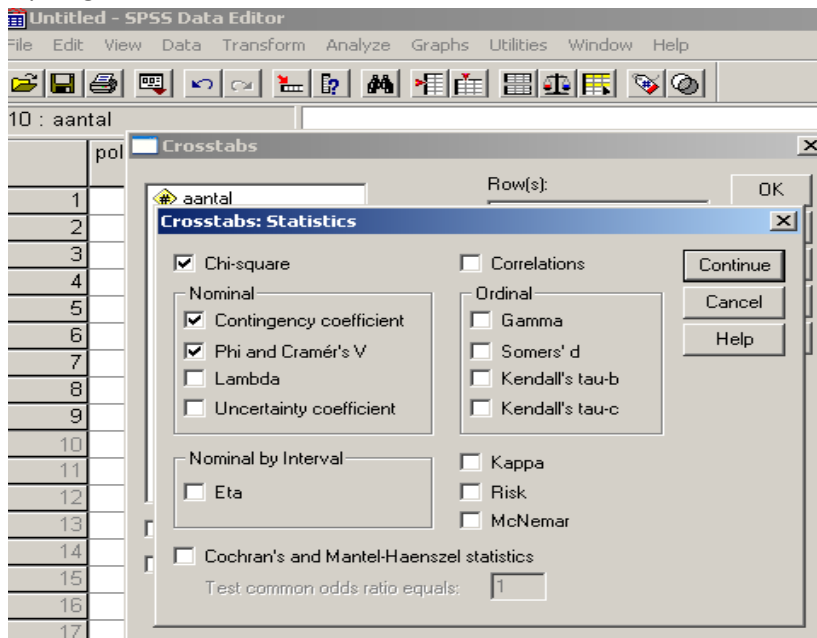
Count

		voor of tegen milieueffing			Total
		voor	neutraal	tegen	
politieke voorkeur	Nieuw groen	12	5	8	25
	Notax	5	12	23	40
	Nova	12	7	1	20
Total		29	24	32	85

Opvragen van de Chi-kwadraat



Opvragen van de Chi-kwadraat en associatiematen



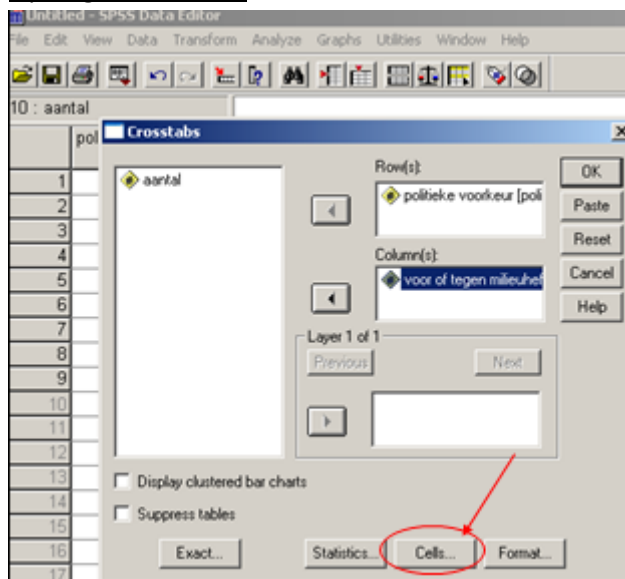
output van de Chi-kwadraat via spss

Chi-Square Tests

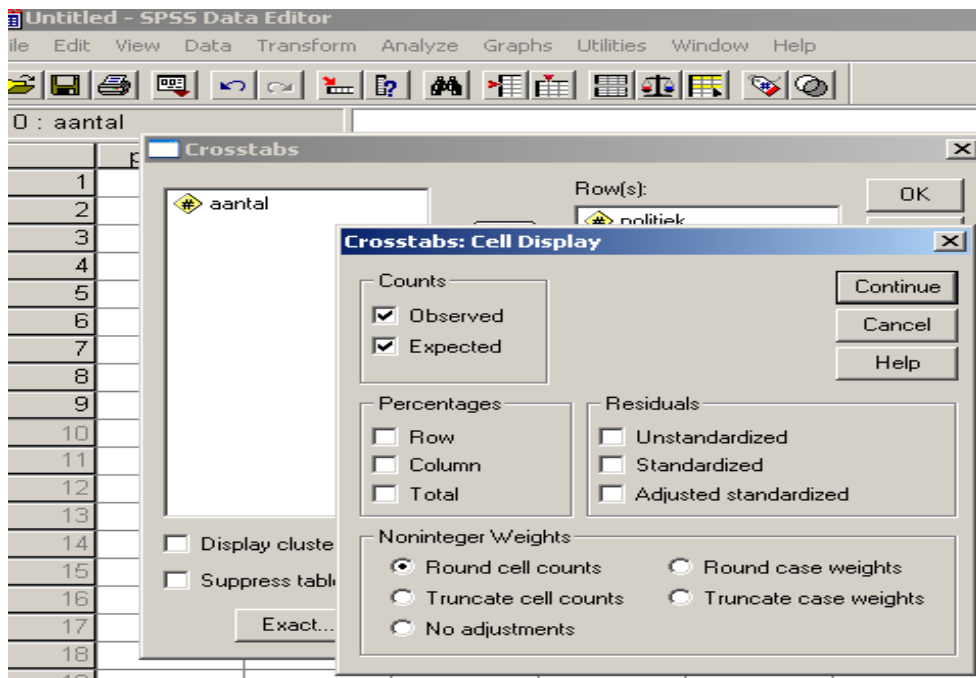
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21,849 ^a	4	,000
Likelihood Ratio	25,559	4	,000
Linear-by-Linear Association	1,437	1	,231
N of Valid Cases	85		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5,65.

Opvragen van de Fe



Opvragen van de geobserveerde en verwachte aantallen



output van een kruistabel via SPSS

partij * attitude Crosstabulation

			attitude			Total
			voor	weet niet	tegen	
partij	Nieuw Groen	Count	12	5	8	25
		Expected Count	8,5	7,1	9,4	25,0
	NOTAX	Count	5	12	23	40
		Expected Count	13,6	11,3	15,1	40,0
	NOVA	Count	12	7	1	20
		Expected Count	6,8	5,6	7,5	20,0
Total		Count	29	24	32	85
		Expected Count	29,0	24,0	32,0	85,0

Relatie tussen roken en hart- en vaatziekten

Chi-Square Tests

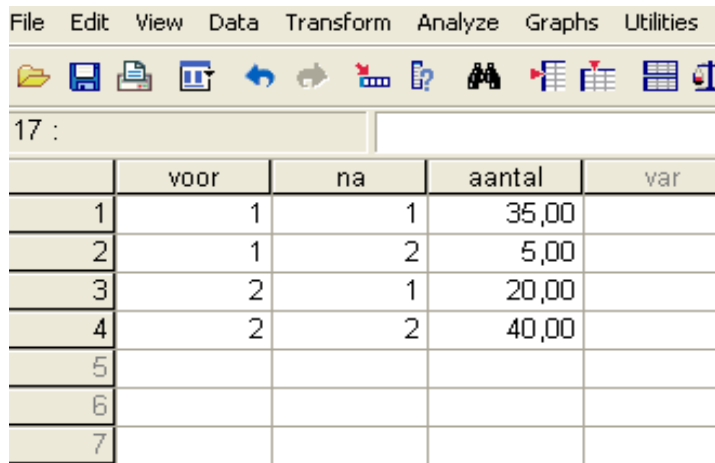
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	20,211 ^b	1	,000		
Continuity Correction ^a	15,474	1	,000		
Likelihood Ratio	13,950	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	20,042	1	,000		
N of Valid Cases	120				

a. Computed only for a 2x2 table

b. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 1,00.

HOOFDSTUK 8 : (2^{de} keer : Mc Nemar toets)

Input data via weight cases



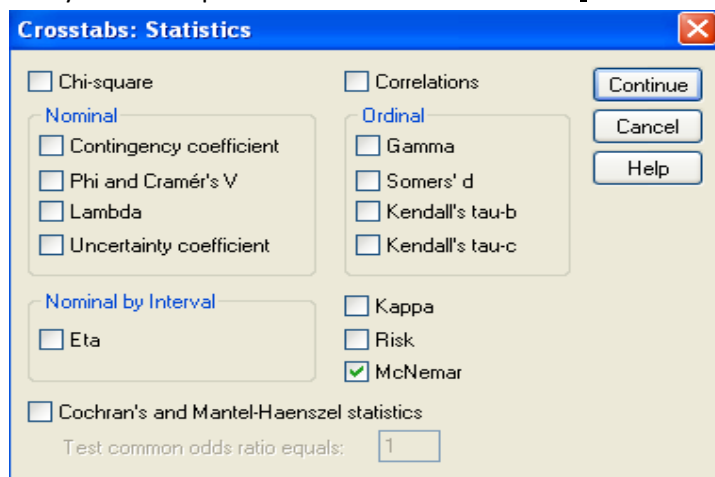
File Edit View Data Transform Analyze Graphs Utilities

17 :

	voor	na	aantal	var
1	1	1	35,00	
2	1	2	5,00	
3	2	1	20,00	
4	2	2	40,00	
5				
6				
7				

Mc Nemar toets via SPSS

Analyze – descriptive stat – crosstabs - statistics_



output Mc Nemar

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		,004 ^a
N of Valid Cases	100	

a. Binomial distribution used.

(terug naar samenvatting- voor goodness of fit)

[hoofdstuk 8 deel 3 : goodness of fit](#)

SPSS Chi-kwadraat goodness-of-fit

Invoeren van de data, via data weight cases_

Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graph

	politiek	aantal	var
1	1	60	
2	2	5	
3	3	25	
4			
5			
6			
7			

SPSS output data

politiek

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	60	66,7	66,7	66,7
2	5	5,6	5,6	72,2
3	25	27,8	27,8	100,0
Total	90	100,0	100,0	

SPSS Chi-kwadraat goodness-of-fit

Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	politiek	aantal	var	var
1	1	6		
2	2			
3	3	2		
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

Analyze menu options:

- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Mixed Models
- Correlate
- Regression
- Loglinear
- Classify
- Data Reduction
- Scale
- Nonparametric Tests**
 - Chi-Square...
 - Binomial...
 - Runs...
 - 1-Sample K-S...
 - 2 Independent Samples...
 - K Independent Samples...
 - 2 Related Samples...
 - K Related Samples...
- Time Series
- Survival
- Multiple Response
- Missing Value Analysis...

SPSS output Chi-kwadraat goodness-of-fit

politiek

	Observed N	Expected N	Residual
1	60	30,0	30,0
2	5	30,0	-25,0
3	25	30,0	-5,0
Total	90		

Test Statistics

	politiek
Chi-Square ^a	51,667
df	2
Asymp. Sig.	,000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 30,0.

(terug naar samenvatting)

[na uitleg van cramer's V, C, en phi-coëfficiënt staat :](#)

Output van associatiematen in SPSS politieke voorkeur t.o.v. attitude milieutaks

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,507	,000
	Cramer's V	,358	,000
	Contingency Coefficient	,452	,000
N of Valid Cases		85	

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.

(terug naar samenvatting)

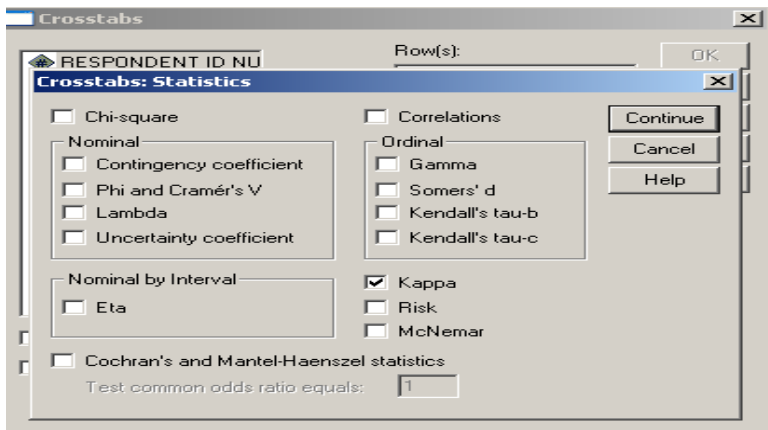
[hoofdstuk 8 : deel 4](#)

Input in SPSS via weight cases

	eerste	tweede	aantal	vs
1	1	1	14	
2	1	2	1	
3	1	3	0	
4	1	4	0	
5	2	1	3	
6	2	2	10	
7	2	3	0	
8	2	4	2	
9	3	1	0	
10	3	2	0	
11	3	3	10	
12	3	4	0	
13	4	1	3	
14	4	2	4	
15	4	3	0	
16	4	4	3	
17				
18				

spss en Kappa

Analyze – descriptive stat – crosstabs – statistics -



spss en Kappa output

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement Kappa	,644	,083	7,771	,000
N of Valid Cases	50			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Toepassingen :

toepassing 1

- Bestaat er een verband tussen geslacht en dyslexie?
Bereken de samenhang handmatig en controleer het resultaat aan de hand van SPSS.
- **Verband tussen geslacht en dyslexie?**

	dyslexie	niet dyslexie
Jongens	10	90
Meisjes	2	98

- Geslacht en dyslexie: twee vragen
- Bestaat er een significant verband?
- Hoe sterk is dit verband?

De kruistabel:

geslacht * dyslexie Crosstabulation

Count		dyslexie		Total
		1	2	
geslacht	1	10	90	100
	2	2	98	100
Total		12	188	200

SPSS output: de Chi-kwadraat

SPSS output: de associatiematen

Toepassing 2 : zie samenvatting

[Hoofdstuk 9 : correlatie en regressie na **Correlatie en causaliteit.**](#)

Verband tussen IQ en schooluitslag

● IQ	Schooluitslag
● 100	70
● 120	80
● 130	85
● 140	85
● 112	82
● 90	60
● 97	65
● 111	70

spss input :

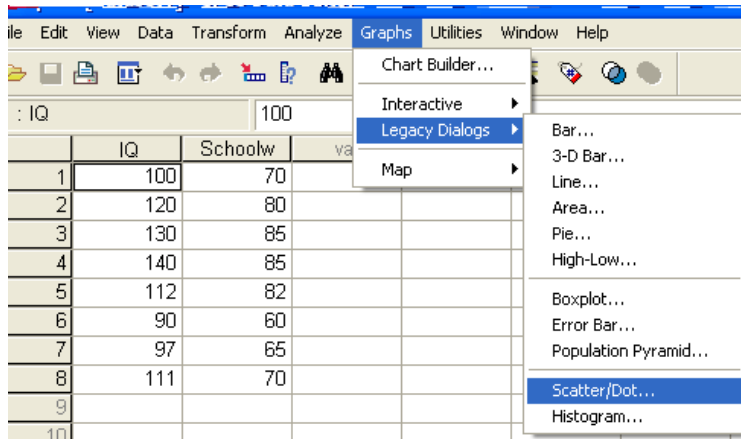
Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

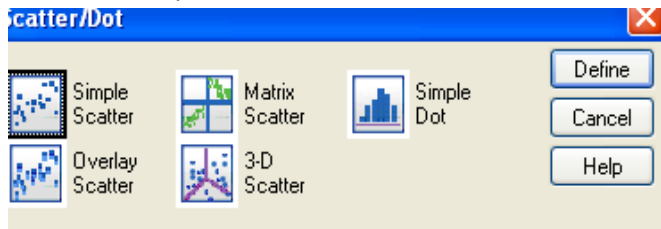
10 :

	IQ	schooluitslag	var	var	var
1	100	70			
2	120	80			
3	130	85			
4	140	85			
5	112	82			
6	90	60			
7	97	65			
8	111	70			
9					
10					
11					

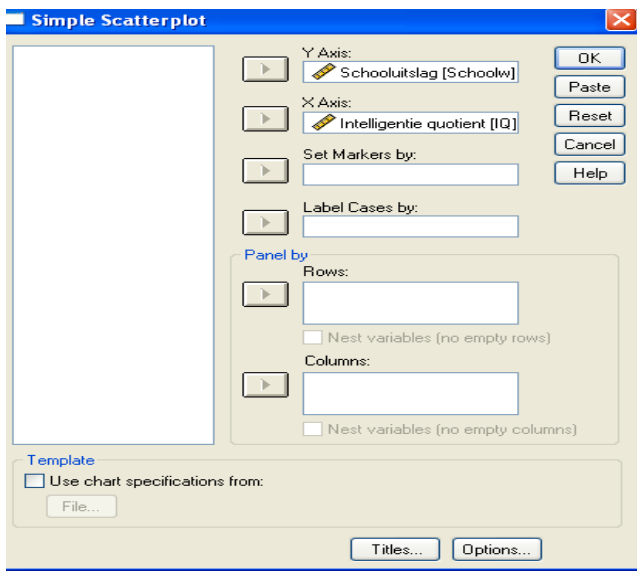
SPSS scatterplot



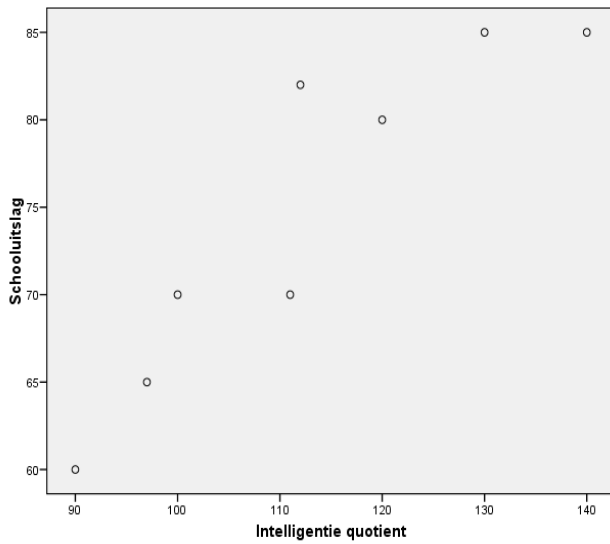
Kies voor 'simple scatter'



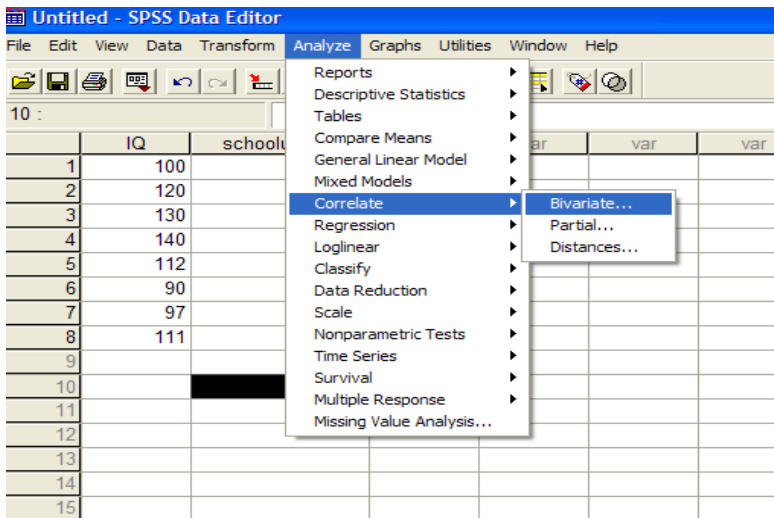
Afhankelijke variabele op de Y-as
 Onafhankelijke variabele op de X-as

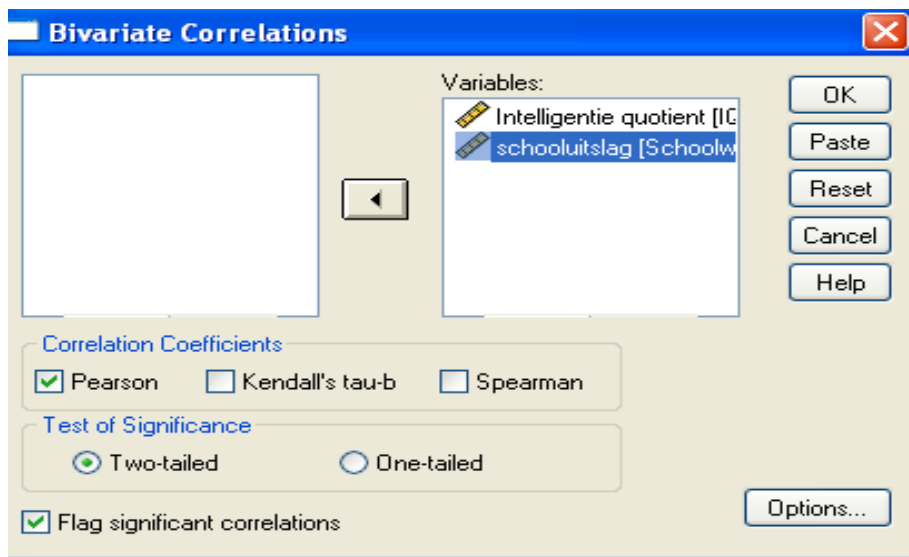


het spreidingsdiagram



spss analyse





spss output :

Correlations

		Intelligentie quotient	schooluitslag
Intelligentie quotient	Pearson Correlation	1	,914**
	Sig. (2-tailed)		,001
	N	8	8
schooluitslag	Pearson Correlation	,914**	1
	Sig. (2-tailed)	,001	
	N	8	8

** . Correlation is significant at the 0.01 level (2-tailed).

(,914 = ,91 / 0,91)

de 0-hypothese zegt : geen samenhang (= een veronderstelling)

de alternatieve hypothese zegt : wel samenhang (= feitelijk)

dus -> de 0-hypothese = FOUT

sig.(2-tailed) :

- kleiner dan 0,05 : significantie – de 0-hypothese wordt verworpen

- groter dan 0.05 : geen significantie – de 0-hypothese blijft

De 0-hypothese veronderstelt altijd : geen samenhang.

(terug naar samenvatting : lineaire regressie)

[Hoofdstuk 9 : lineaire regressie](#)

[De regressielijn een eenvoudig voorbeeld: het salaris](#)

De regressielijn een eenvoudig voorbeeld: het salaris

Maandelijks salaris	Jaarlijks inkomen
4.000	48.000
4.500	54.000
5.000	60.000
5.750	69.000
6.000	72.000
6.250	75.000
6.500	78.000
6.700	80.400

spss input :

sal.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities W

1 : maandsal 4000

	maandsal	jaarsal	var	var
1	4000	48000		
2	4500	54000		
3	5000	60000		
4	5750	69000		
5	6000	72000		
6	6250	75000		
7	6500	78000		
8	6700	80400		
9				

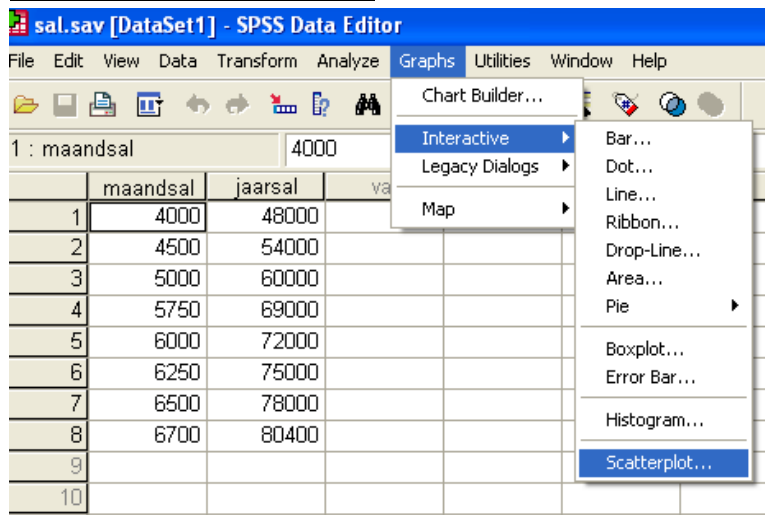
spss output de correlatie :

Correlations

		maandsal	jaarsalaris
maandsal	Pearson Correlation	1	1,000**
	Sig. (2-tailed)		,000
	N	8	8
jaarsalaris	Pearson Correlation	1,000**	1
	Sig. (2-tailed)	,000	
	N	8	8

** - Correlation is significant at the 0.01 level (2-tailed).

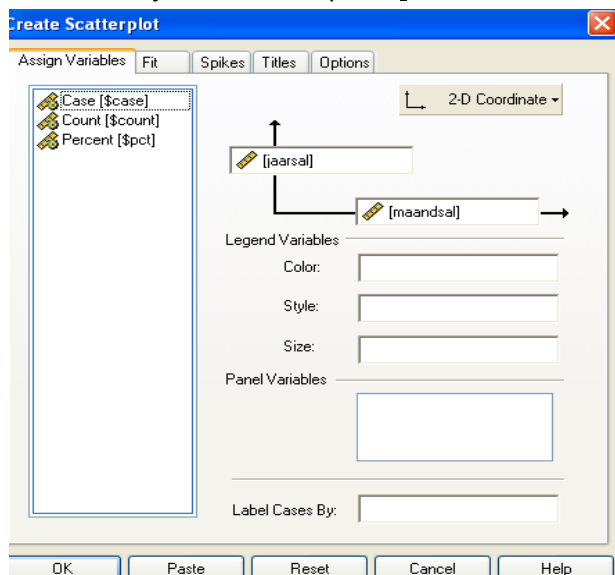
Teken een scatterplot via SPSS



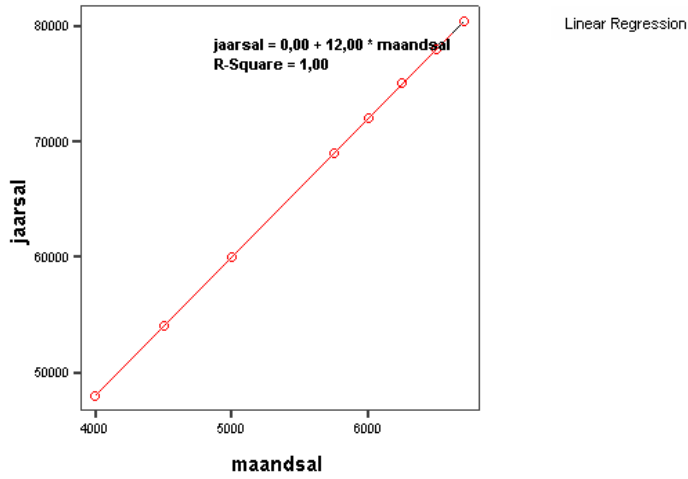
SPSS aanmaken van een scatterplot

Afhankelijke variabele op Y-as

Onafhankelijke variabele op X-as



SPSS output: de scatterplot



Maandelijks salaris	Jaarlijks inkomen
4.000	48.000
4.500	54.000
5.000	60.000
5.750	69.000
6.000	72.000
6.250	75.000
6.500	78.000
6.700	80.400

De regressielijn: $Y = a + bX$

$$Y = 12 * X$$

Fouten bij de voorspelling?

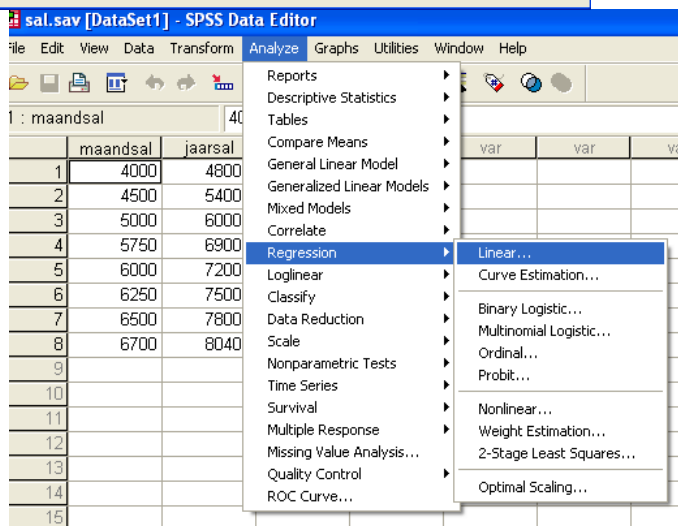
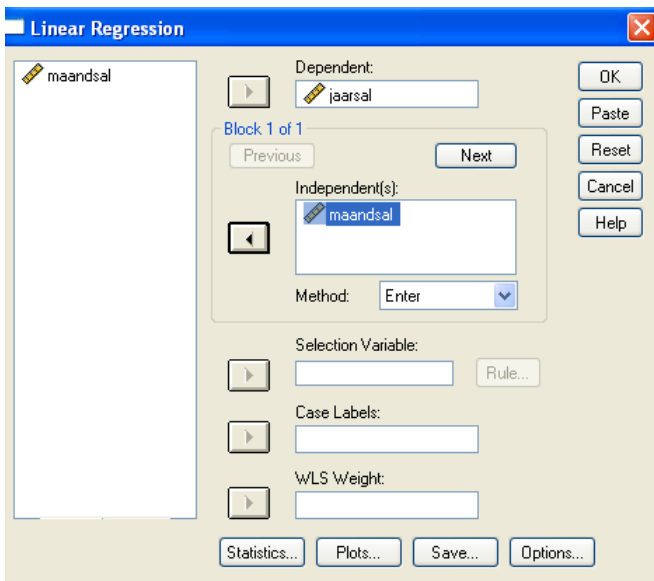
B = +/- (+12,00)

A = 0

R – square = determinatie coëfficiënt = de maat hoe goed uw regressie lijn is : hier perfect = 1

Fouten bij de voorspelling?

nee – determinatie coëfficiënt = 100%



SPSS regressie analyse

Dependent =afhankelijke variabelen

Independent =onafhankelijke variabelen

-

De regressielijn: een voorbeeld, nu met eindejaarspremie € 1000

Maandelijks salaris	Jaarlijks inkomen
4.000	49.000
4.500	55.000
5.000	61.000
5.750	70.000
6.000	73.000
6.250	76.000
6.500	79.000
6.700	82.400

Welk is de r? en de scatterplot?
En de regressielijn?
Fouten bij de voorspelling?

$y = 1000 + 12,00 X$

$a = 1000$

$b = 12,00$

X = maandloon

2 vragen :

- Hoe vinden we de parameters van de regressielijn ($Y = a + bX$)?
- Hoe goed kunnen we de Y waarden voorspellen op basis van dit model?

hoe dichter de punten bij regressie lijn : hoe meer juist het zal zijn.

Hoe vinden we deze a en b coëfficiënten?

b: dit is de richtingscoëfficiënt; deze is functie van de r en de verhouding tussen de twee SD.

$b = r_{YX} \cdot SD_Y / SD_X$ (slope)

a: dit is de uitslag van Y indien X nul bedraagt;

$a = Y - bX$ (intercept)

b = de verandering in y wanneer x met 1 eenheid toeneemt

Op grond van deze vergelijking kunnen we voor elke score van X een verwachting voor Y formuleren.

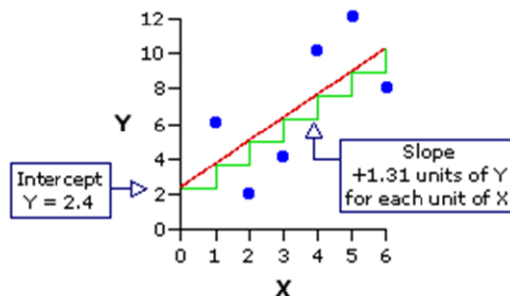
$Y = a + bX$

$Y = 2,4 + 1,31X$

bv. $X = 3, Y = ?$

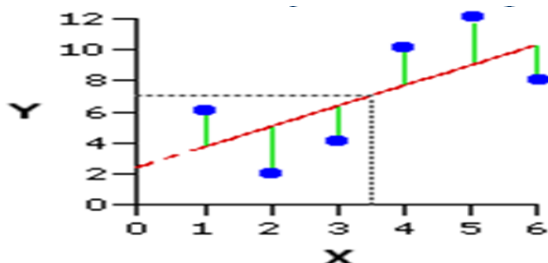
$Y = 2,4 + 1,31 \cdot 3$

$Y = 6,31$



De regressielijn voldoet aan het criterium van het kleinste kwadraat. D.w.z. dat de gekwadrateerde afwijking van de verwachte uitslag t.o.v. de

feitelijke uitslag minimaal is.



- Algemene werkformule van de regressielijn:

$$Y = \bar{Y} + r_{XY} (X_i - \bar{X}) * SD_Y / SD_X$$
- Vergelijking van de **best** passende lijn, waarbij de Y waarden zo goed mogelijk geschat kunnen worden op grond van de X waarden. De regressielijn is **niet** symmetrisch.

Hoe goed voldoet dit model om de werkelijkheid te voorspellen?

Waarde van de regressielijn I

Het verschil tussen de verwachte en de feitelijke score van Y is de schattingsfout.

De SD van deze fout is de standaardschattingsfout. Dit komt overeen met de SD van de verschillen tussen de verwachte en feitelijke uitslag.

- $SD_{Y,X} = SD_Y * \sqrt{1-r^2}$;

Standaardschattingsfout

- In SPSS wordt deze standaardfout aangeduid middels 'std. error of the estimate'.
- Deze schattingsfout geeft een indicatie van de (on)nauwkeurigheid van de voorspelling.
- (omgekeerde) relatie met r_{XY}

de standaardschattingsfout : standaard fout of estimate geeft aan hoe onnauwkeurig uw voorspelling van y is op grond van de regressielijn.

(berekening : standaarddeviatie (SD) $SD_{y \cdot x} = SD_y \cdot \sqrt{1-r^2}$)

= de standaard deviatie van uw verdeling van fouten (wat u verwacht <-> feiten)

in 2/3^{de} van de gevallen zal de fout kleiner zijn dan standaard schattingsfout.

Waarde van de regressielijn II

- Hoe goed verklaart het model de werkelijke gegevens?
- Proportie niet-verklaarde variantie: $(1- r^2_{XY})$
- Proportie verklaarde variantie: r^2_{XY}

Deze determinatiecoëfficiënt geeft de gemeenschappelijk variantie weer.

Kan berekend worden via het kwadraat van de r. Voorstelling: R^2

Mijn correlatie coëfficiënt zou al .70 moeten zijn, dan kan ik 50% verklaren

want : $.70^2 = .50 = 50$ procent

bv.: correlatiecoëfficiënt = .40 ($.40^2 = 0,16$) dit betekent dat ik 16% zou kunnen verklaren, en dus 84 % niet.

je zou al de helft, of meer van de verschillen moeten kunnen verklaren.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,914 ^a	,835	,808	4,219

a. Predictors: (Constant), Intelligentie quotient

R Square : determinantiecoëfficiënt = $0,914^2 = 0,835$

Adjusted R Square : niet belangrijk bij 1 regressielijn, wel bij meerdere!

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16,457	10,654		1,545	,173
	Intelligentie quotient	,517	,094	,914	5,514	,001

a. Dependent Variable: Schooluitslag

16,46 = constante

0,52 = B₁

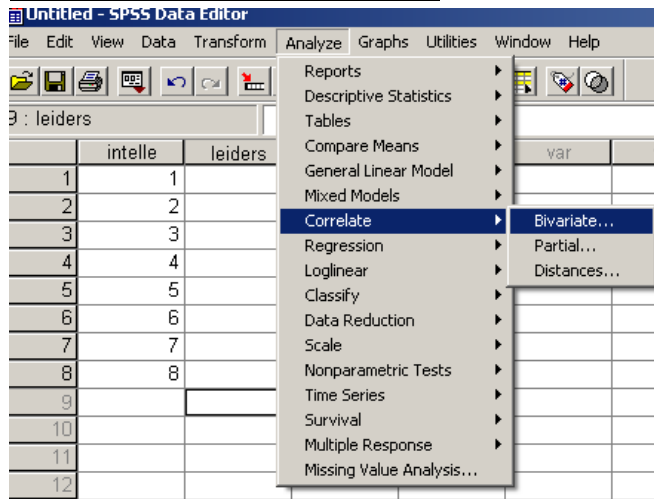
--> zie uitgaven restaurant en inkomen de samenvatting!

Hoofdstuk 10 : De rangcorrelatiecoëfficiënt

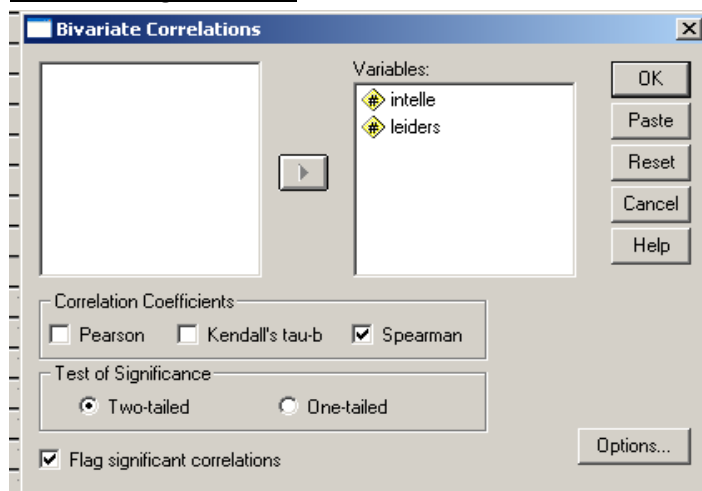
spss input :

	intelle	leiders	var
1	1	8	
2	2	5	
3	3	7	
4	4	6	
5	5	4	
6	6	2	
7	7	3	
8	8	1	
9			
10			

SPSS analyse – correlatie – bivariate



SPSS. De rangcorrelatie:



output :

Correlations

		intelle	leiders
Spearman's rho	intelle	Correlation Coefficient	1,000
		Sig. (2-tailed)	,002
		N	8
leiders		Correlation Coefficient	-,905**
		Sig. (2-tailed)	,002
		N	8

** . Correlation is significant at the 0.01 level (2-tailed).

9,05 = significant

--> terug naar samenvatting : probleem :knopen

Hoofdstuk 10 : De Puntbiseriële coëfficiënt

hoe doet ge itemanalyse:

analyse – scale – betrouwbaarheid – chromer alfa

en dan bekomt ge dit:

Cronbach's Alpha	Nof Items
,704	18

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Vraag1_nummer	8,37	9,375	,107	,707
Vraag2_nummer	8,03	8,820	,227	,699
Vraag3_nummer	7,98	8,538	,329	,687
Vraag4_nummer	7,62	8,640	,580	,671
Vraag5_nummer	8,27	9,181	,141	,706
Vraag6_nummer	8,05	8,594	,307	,689
Vraag7_nummer	8,43	9,655	,004	,712
Vraag8_nummer	7,74	8,527	,437	,676
Vraag9_nummer	8,22	9,268	,094	,712
Vraag10_nummer	7,72	8,500	,469	,674
Vraag11_nummer	7,64	8,660	,515	,674
Vraag12_nummer	7,95	8,429	,373	,681
Vraag13_nummer	8,17	8,940	,201	,701
Vraag14_nummer	8,32	9,470	,044	,714
Vraag15_nummer	7,72	8,467	,490	,672
Vraag16_nummer	8,07	8,647	,290	,691
Vraag17bis2	8,29	8,701	,345	,685
Vraag18bis2	8,32	8,781	,340	,686

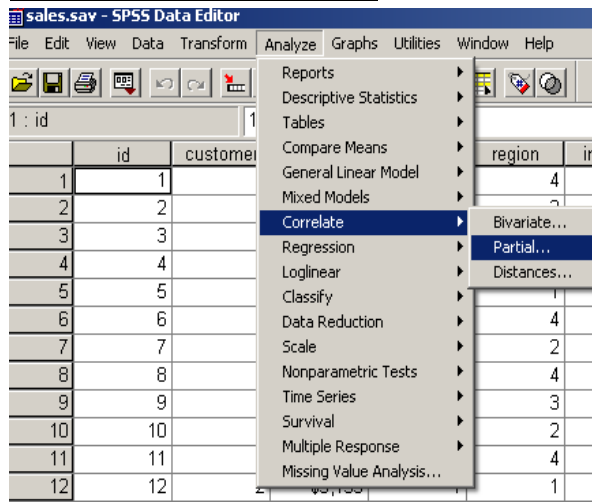
Cronbach's Alpha	Nof Items
,746	13

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Vraag2_nummer	7,02	7,073	,246	,746
Vraag3_nummer	6,97	6,764	,372	,730
Vraag4_nummer	6,61	6,987	,573	,716
Vraag6_nummer	7,03	6,888	,320	,737
Vraag8_nummer	6,72	6,893	,425	,724
Vraag10_nummer	6,71	6,850	,466	,720
Vraag11_nummer	6,63	7,012	,503	,720
Vraag12_nummer	6,93	6,771	,374	,730
Vraag13_nummer	7,16	7,213	,209	,749
Vraag15_nummer	6,70	6,877	,458	,721
Vraag16_nummer	7,06	6,916	,310	,738
Vraag17bis2	7,27	6,942	,383	,728
Vraag18bis2	7,31	7,058	,358	,731

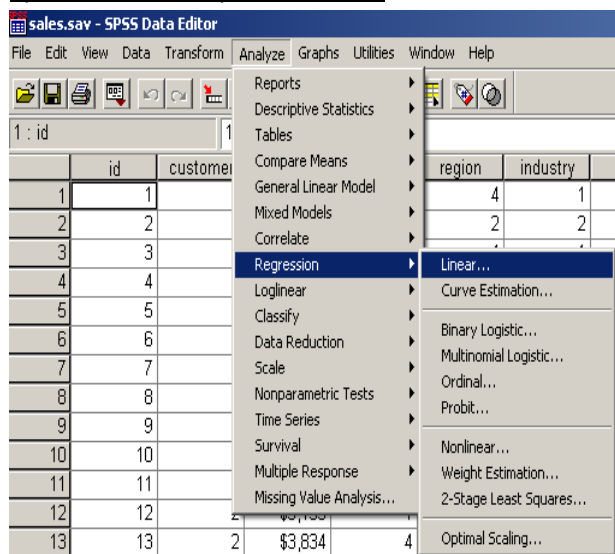
-> terug naar samenvatting : 10.3

10.3 verband tussen meer dan 2 (interval) variabelen.

spss : en de partiële correlatie:



Spss : en de multiple correlatie



een voorbeeld :

- Welk is het verband tussen een leesvoorwaardentest en leesniveautest, met constant houding van het IQ?

Het betreft gegevens van 6 proefpersonen.

Overzicht van de correlaties:

Correlations

		leesvoor waarden	leesnivea	intelligente
leesvoorwaarden	Pearson Correlation	1	,489	,044
	Sig. (2-tailed)	.	,325	,935
	N	6	6	6
leesnivea	Pearson Correlation	,489	1	,714
	Sig. (2-tailed)	,325	.	,111
	N	6	6	6
intelligente	Pearson Correlation	,044	,714	1
	Sig. (2-tailed)	,935	,111	.
	N	6	6	6

Samenhang tussen leesvoorwaarden en leesniveau met constant houding van IQ

Correlations

Control Variables			leesvoor waarden	leesnivea
intelligente	leesvoorwaarden	Correlation	1,000	,655
		Significance (2-tailed)	.	,231
		df	0	3
	leesnivea	Correlation	,655	1,000
		Significance (2-tailed)	,231	.
		df	3	0

--> terug naar samenvatting (samenhang tussen meerdere interval variabelen)