Methoden SPSS hoe en waarom

LES 1: Introductie SPSS

1. Basiskennis SPSS

Eerst 3 stappen uitvoeren in de databank:

1. Definitie van variabelen
   = variabelen invoegen en “values” definiëren
   = missing values definiëren
2. Data invoeren
   = invoeren vanuit bestand (SPSS-bestand of uit excel-bestand)
   = vanuit excel oppassen voor de definitie van variabelen
3. Analyse uitvoeren

Doe ook altijd aan data cleaning: ga na of er geen fouten gemaakt zijn.

1. Kijken naar de **frequentieverdeling** voor de verschillende variabelen.
2. Nagaan bij welke cases er juist fouten zijn gemaakt en aanpassen

LES 2: Data-cleaning en -transformatie

1. Frequentietabel

Analyse/Descriptive Statistics/Frequencies in SPSS, daarna gevraagde variabele ingeven

Stad

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Gent	699	37,3	37,3	37,3
	Brugge	494	26,4	26,4	63,7
	Antwerpen	679	36,3	36,3	99,9
	4	1	,1	,1	100,0
	Total	1873	100,0	100,0	

Deze waarde “4” is een onmogelijke waarde en moet gecontroleerd worden.

Als we kijken naar de beschikbare waarden in de frequentietabel, kunnen we nagaan of hier onmogelijke waarden bij zitten en wanneer nodig kunnen we dit aanpassen.

2. Waarden opzoeken

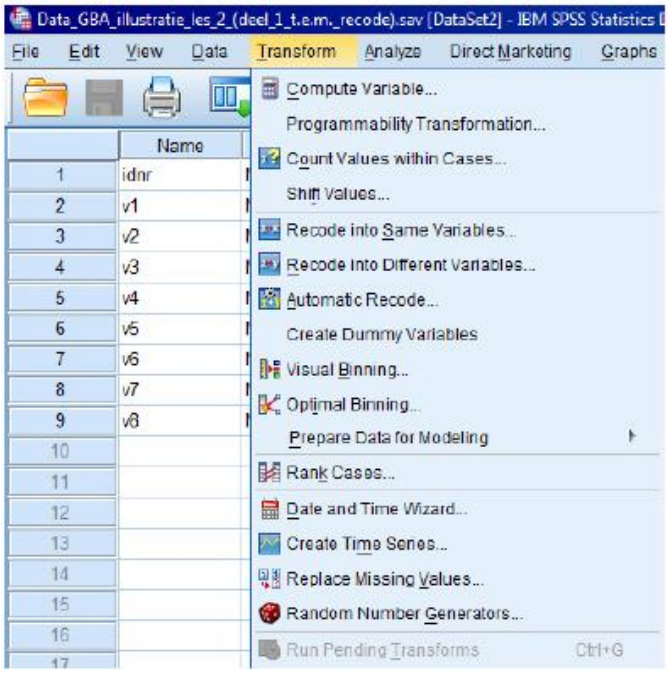
“Data View” in de “Data Editor”, “Find” in de juiste kolom
OF
“Edit/Find”

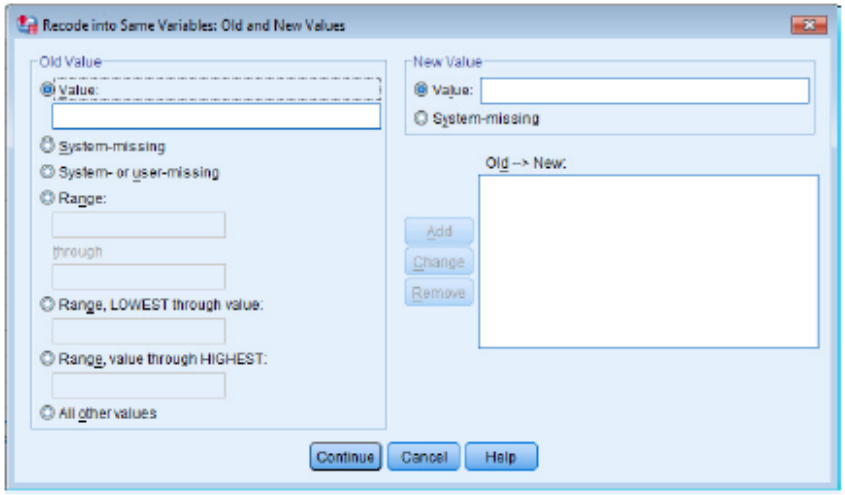
Zo kan de verkeerde waarde opgezocht worden en desnoods gedefinieerd worden als missing value. Lege cellen zoeken we door “*Data/Sort cases”.*

3. Data transformatie

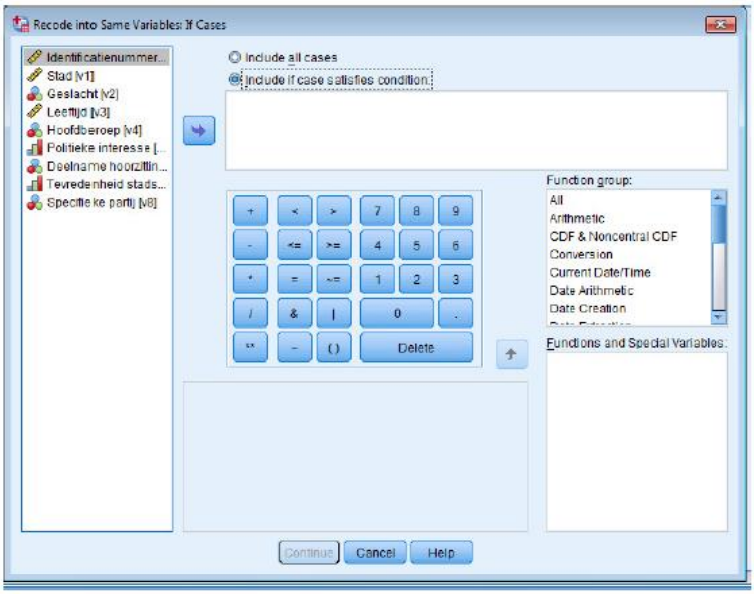
Via menu *"Transform"* → *"Recode" of "Compute"*

- **Recode**
  Bestaande variabele vervangen = *"Recode/into same variables"*

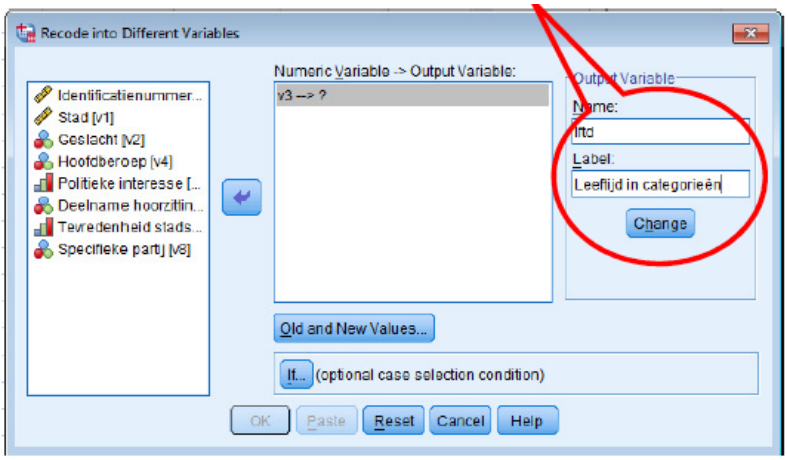




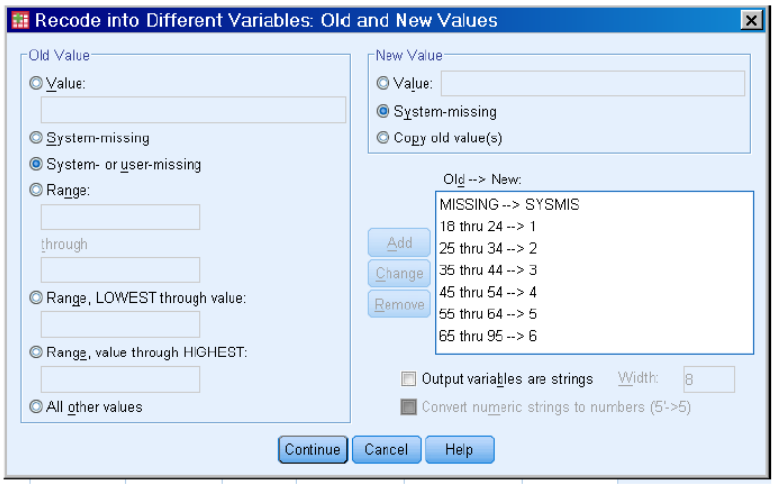
daarna via dialoogvenster *"Old and new values"* hercoderen
is ook mogelijk voor *"If cases"* (subgroep, alleen wanneer het aan een bepaald kenmerk voldoet)



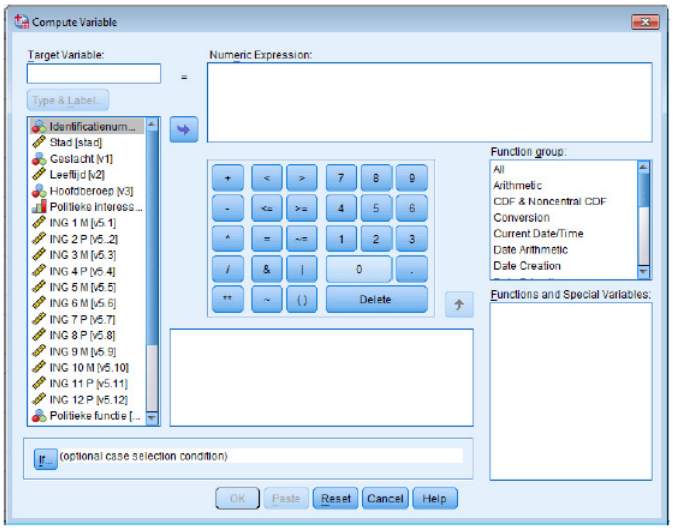
Op grond van een bestaande een andere, nieuwe variabele maken = *"Recode/Into Different Variables"*, Eerst de gewenste variabele selecteren en dan een nieuwe naam geven aan de nieuwe variabele



In het volgende dialoogvenster kan je dan aangeven hoe je de nieuwe variabele wil definiëren, ook mogelijk voor enkel subgroepen
Labels daarna nog aanpassen!
Goed om originele data te bewaren bij hercoderen



- **Compute**
  Wordt gebruikt bij schaalconstructies, omscoren en combinaties maken van variabelen
  "Transform/Compute Variable", daarna nog type en label ingeven



LES 3: Univariate statistiek

1. Inleiding univariate statistiek

Wat? → beschrijven van diverse types variabelen naargelang hun meetniveau

- Nominaal
- Ordinaal
- Interval/ratio

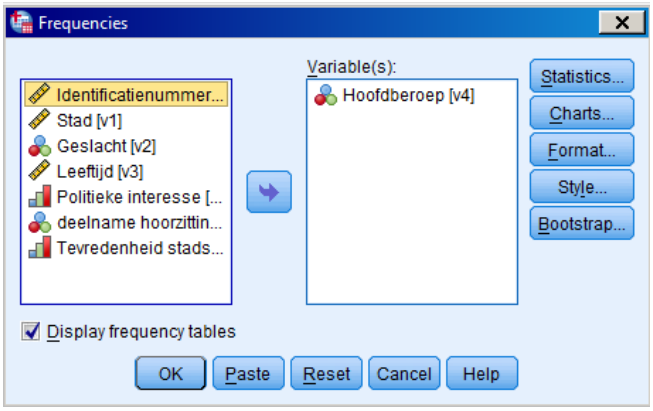
Onderscheid maken door 3 kenmerken

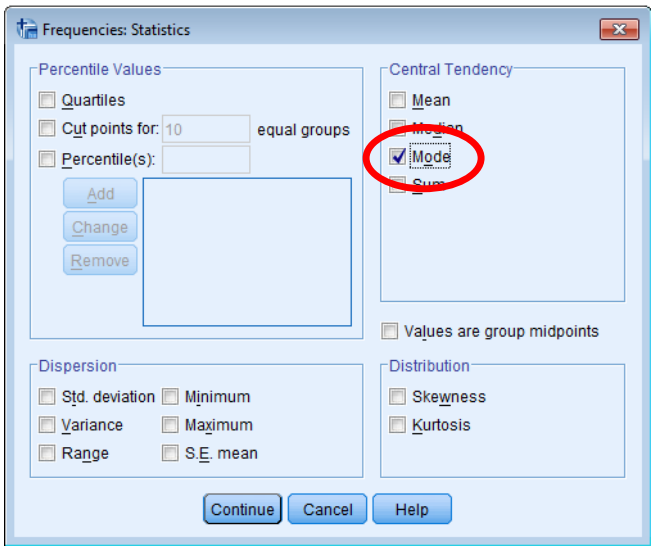
- Ordenbaarheid?
- Vaste meeteenheid?
- Absoluut nulpunt?

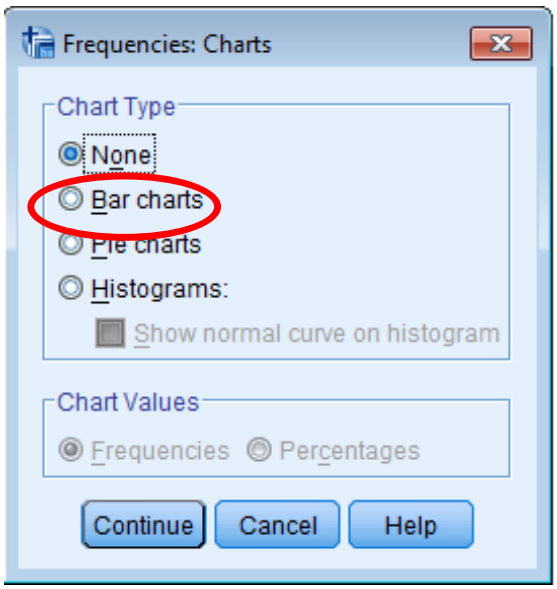
	Categorisch		Metrisch	
	NOMINAAL	ORDINAAL	INTERVAL	RATIO
Orde		x	x	x
Vaste meeteenheid			x	x
Absoluut nulpunt				x
Voorbeeld	Geslacht	Opleiding	Temp	Lengte
	M/V	L/M/H	°C	m

2. Beschrijven nominale variabele

"Analyse/Descriptive Statistics/Frequencies"

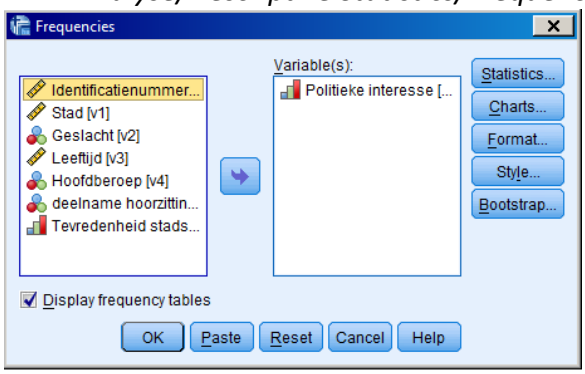


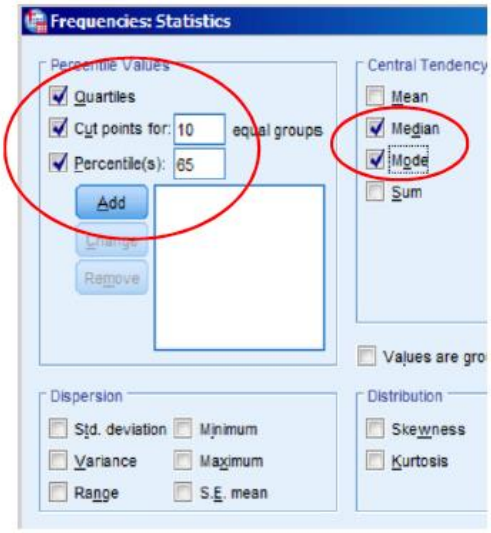


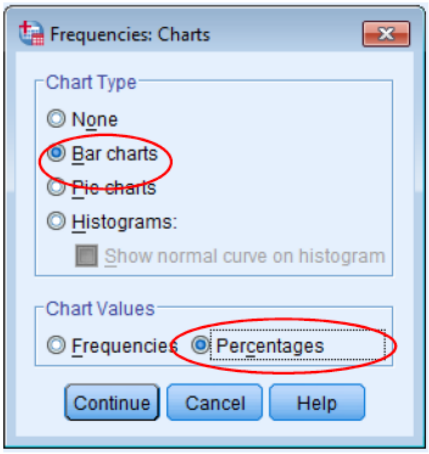


3. Beschrijven ordinale variabele

"Analyse/Descriptive Statistics/Frequencies"

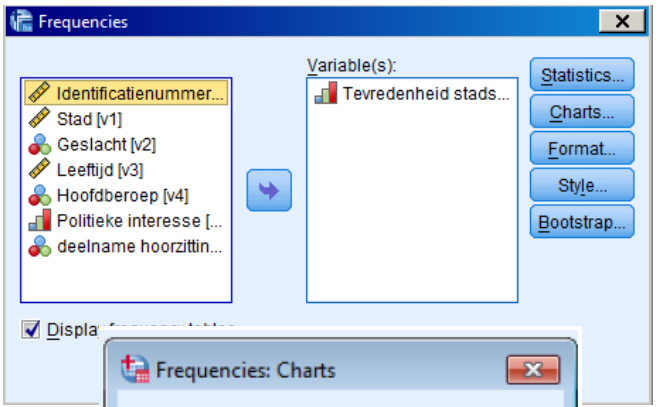


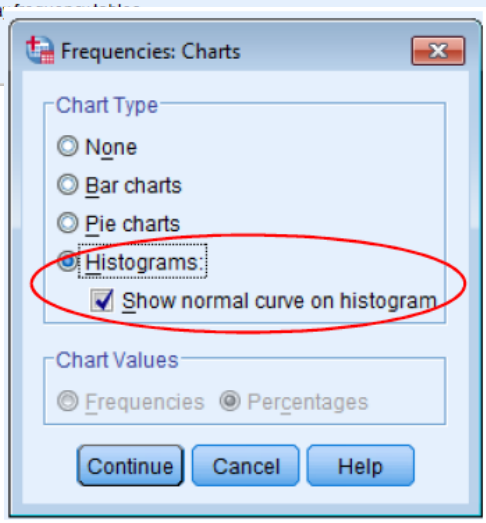


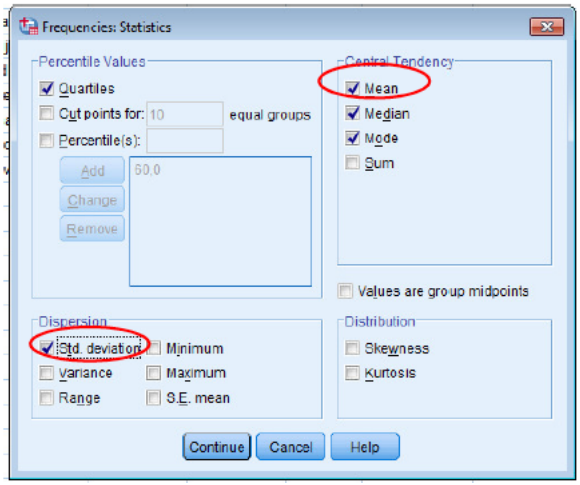


4. Beschrijven metrische variabele

"Analyse/Descriptive Statistics/Frequencies" Het histogram kan aangepast worden in de "*chart editor*"



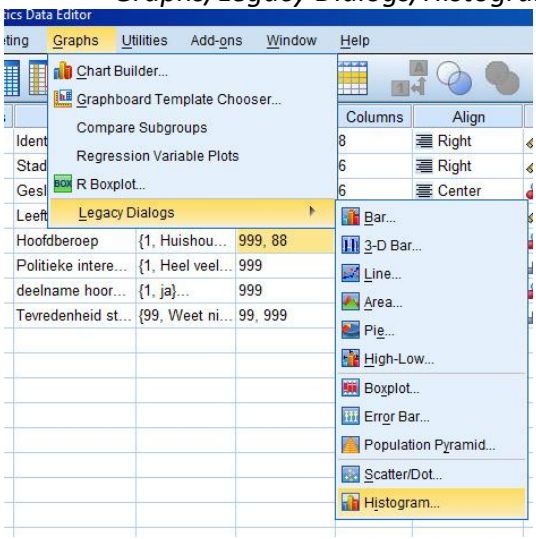


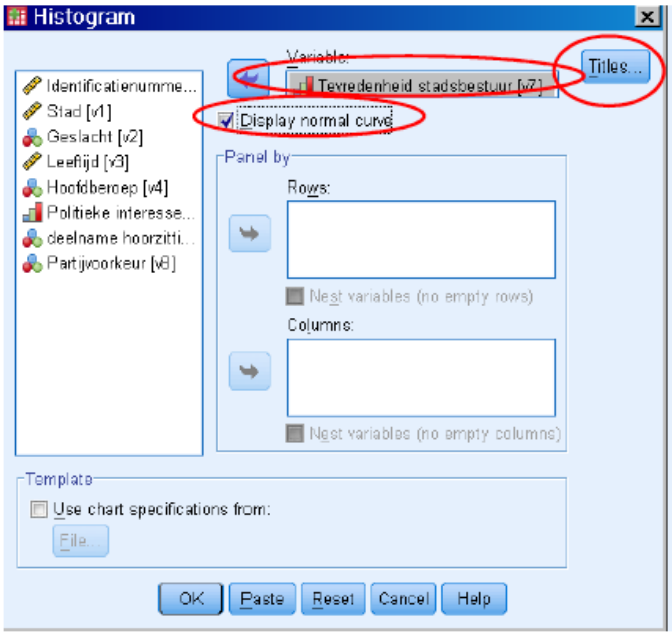


5. Figuren

1. Histogram

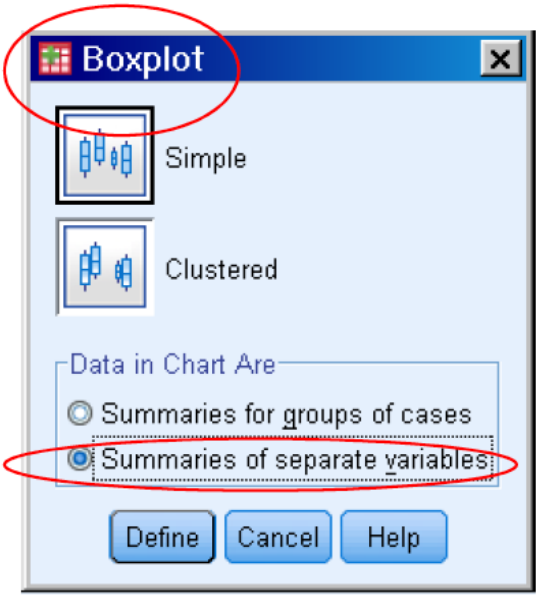
   –"Graphs/Legacy Dialogs/Histogram…"

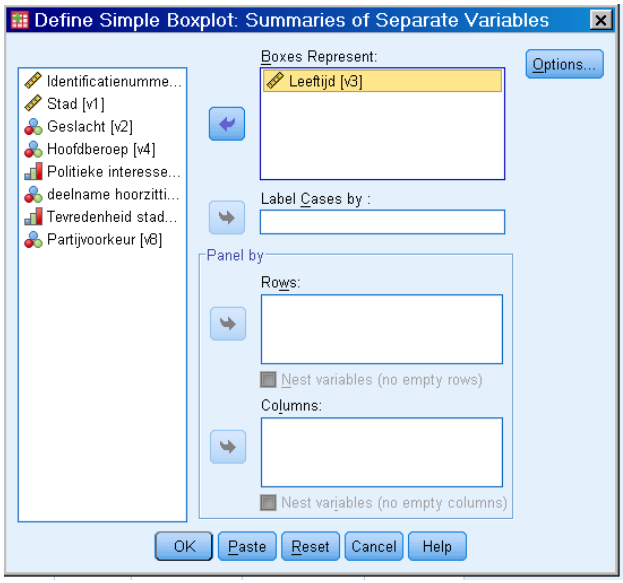




2. Boxplot

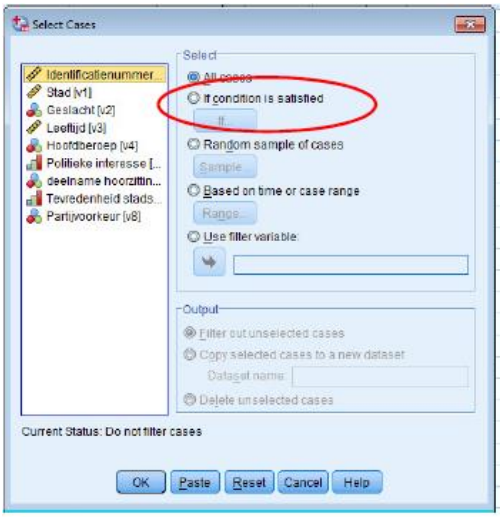
   "Graphs/Legacy Dialogs/Boxplot" (kan ook voor subgroepen, "groups of cases")

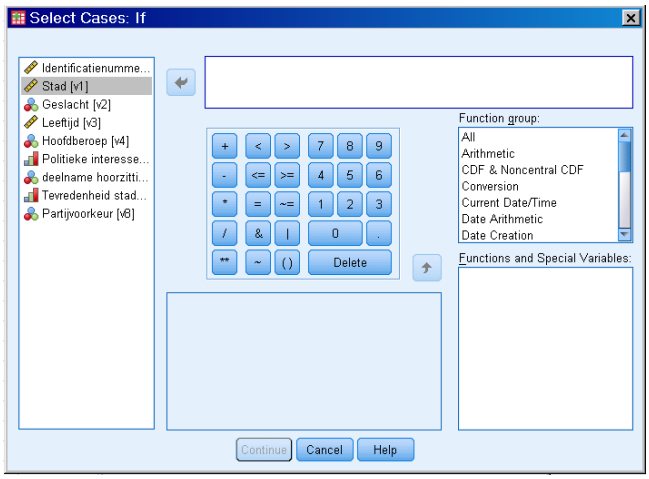


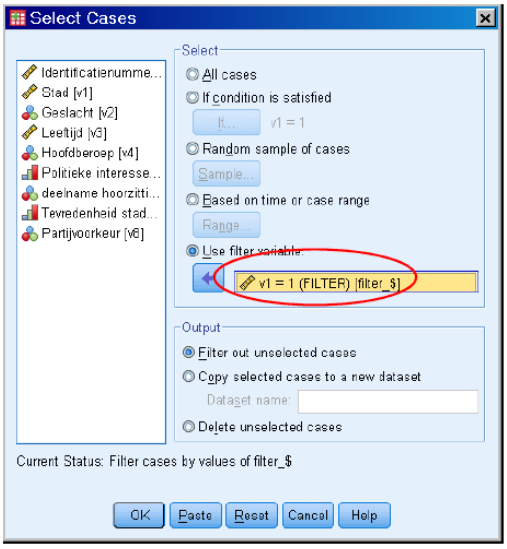


6. Mogelijkheden analyse subgroepen

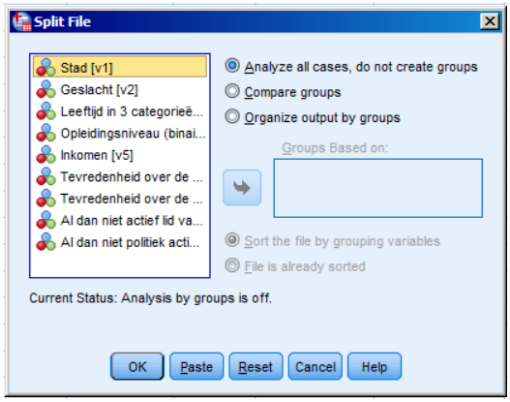
1. Select cases
   "Data/Select Cases"
   Na deze analyse is er een filter aanwezig op de data. Belangrijk om achteraf deze filter weer te verwijderen om verder te werken. Er verschijnt ook een filtervariabele die je achteraf kan gebruiken om het simpel weer te analyseren.

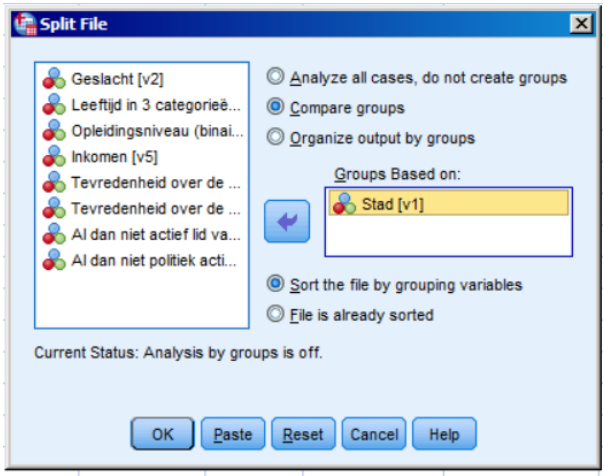






2. Split file
   "Data/Split file"
   Analyses van verschillende subgroepen vergelijken.





LES 4: Bivariate Statistiek

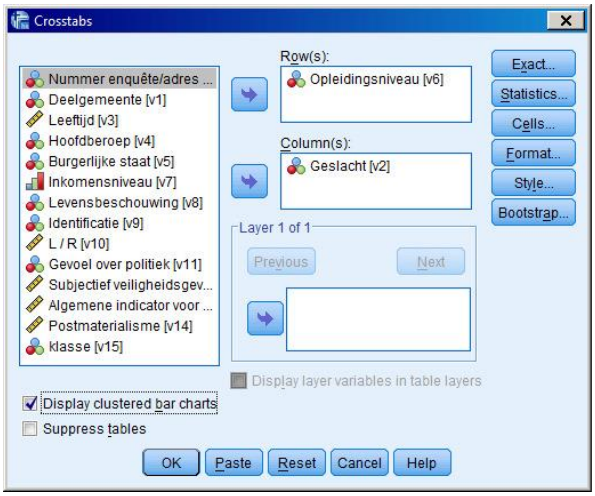
1. inleiding bivariate statistiek

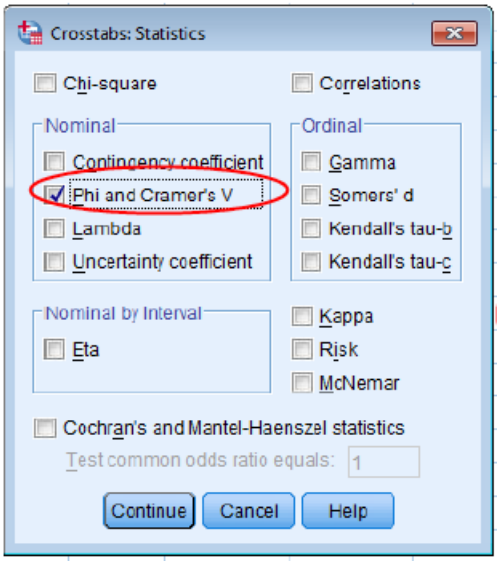
Analyseren van meerdere variabelen tegelijk
We moeten rekening houden met het meetniveau van de variabelen

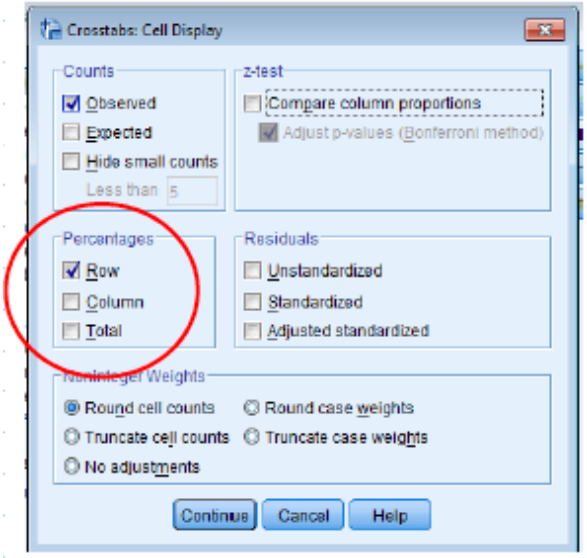
- 2 categorische variabelen: kruistabellen
- 2 metrische variabelen: correlatiecoëfficiënten
- 1 categorische en 1 metrische variabele: Compare means of variantieanalyse (ANOVA)

2. Verband twee categorische variabelen

"Analyse/Descriptive Statistics/Crosstabs"
In de kruistabellen kunnen we met rij- en kolompercentages werken, maar ook met totaalpercentages. Deze kunnen ook alle drie samen gecombineerd worden. *"Phi and Cramer's V"* is een maat om de samenhang tussen twee variabelen te constateren. De waarde ligt altijd tussen 0 en 1. Een waarde van 0 betekent dat er geen samenhang is. Een waarde van 1 betekent dat er een perfecte samenhang is. Belangrijk hierbij ook is de significantie. Om een significant verband te zijn, moet de waarde kleiner zijn dan 0,050.



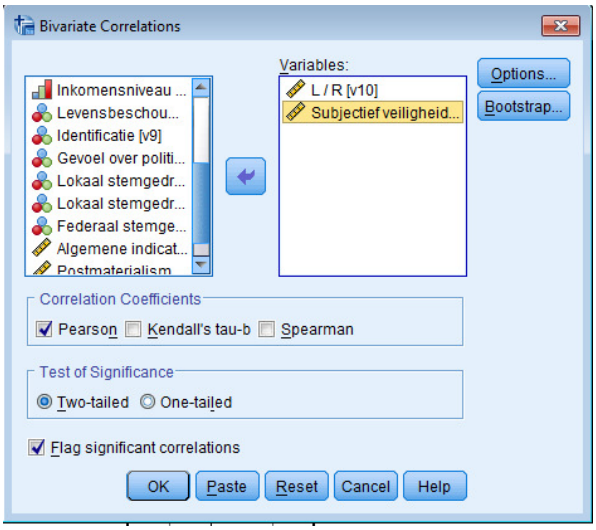




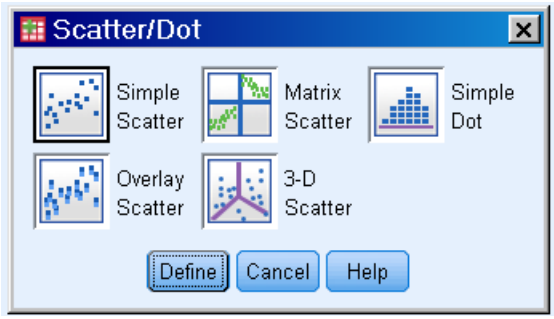
	Variabele X	
Variabele Y	X	niet X
Y	A	B
niet Y	C	D
Totaal	A+C	B+D

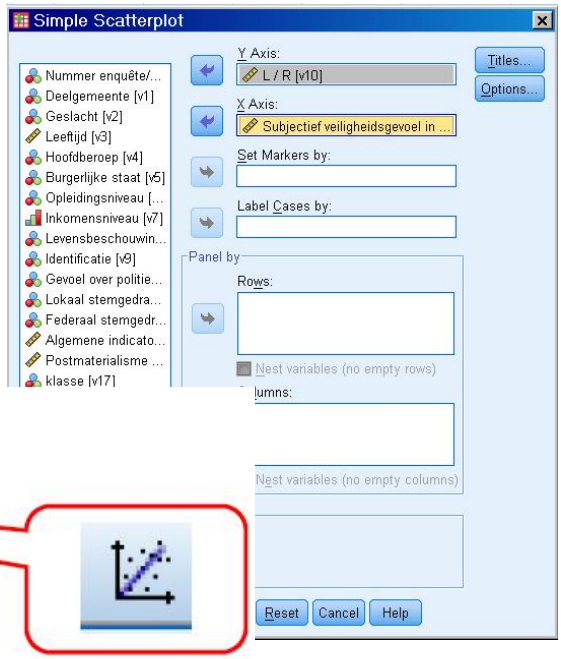
3. Verband twee metrische variabelen

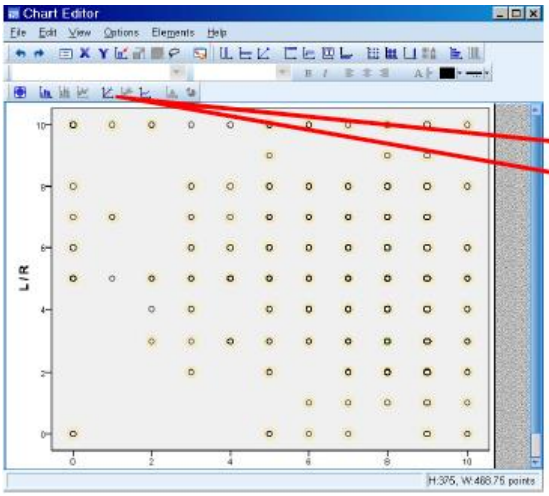
"Analyse/Correlate/Bivariate"
In deze analyse vinden we de *"Pearson Correlation".* Deze geeft een beeld van de samenhang tussen de variabelen. De waarden variëren tussen -1 en 1. Bij een waarde van 0 is er geen of een zeer zwak verband. Bij waarden van -1 of 1 is er een groot verband. Hierbij geldt dezelfde regel van significantie als bij de kruistabellen.

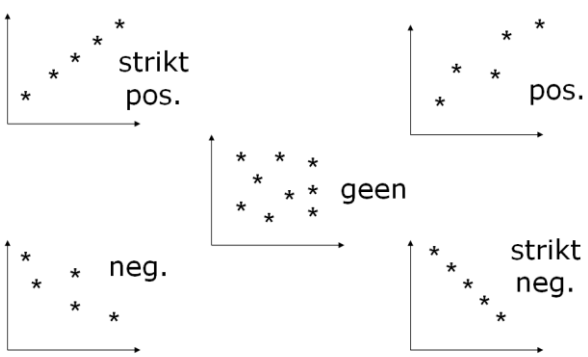


Grafische weergave: scatterplot
"Graphs/Legacy Dialogs/Scatter/Dot…"
In de grafiek best ook altijd een passende rechte (linear) toevoegen zodat men een algemene verdeling kan bekijken









LES 5: Bivariate statistiek en Variantieanalyse

1. Verband categorische en metrische variabele

- Vergelijken van gemiddelden per categorie
- Nagaan statistische verschillen door variantieanalyse (ANOVA)
- Totale variantie verklaarbaar door onderscheid **tussen** groepen of onderscheid **binnen** groepen?

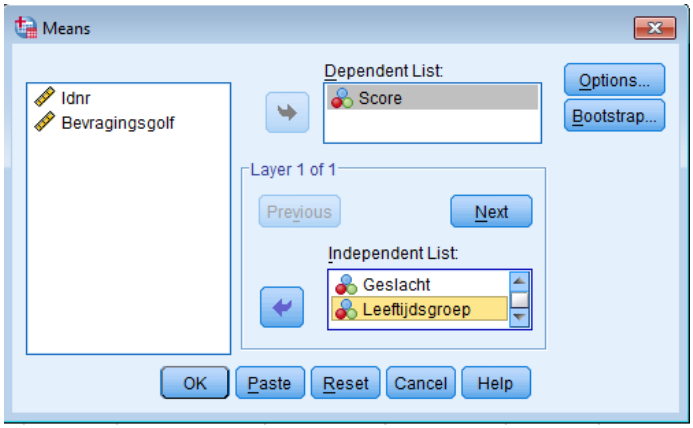
Veronderstellingen:

- Frequenties binnen groepen normaal verdeeld
- Varianties binnen groepen ongeveer gelijk: vergelijken standaardafwijkingen en testen door *"Levene's test"*

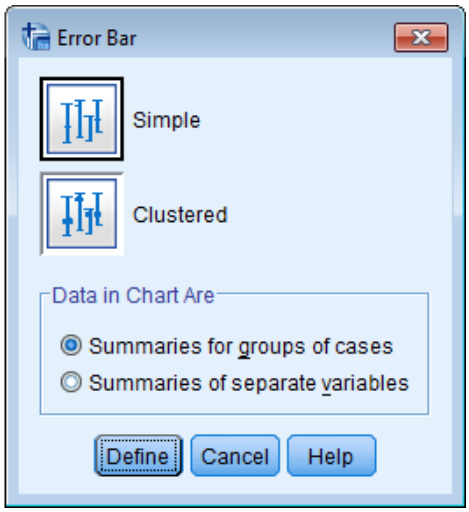
Rekenkundige uitleg op dia 9-10-11

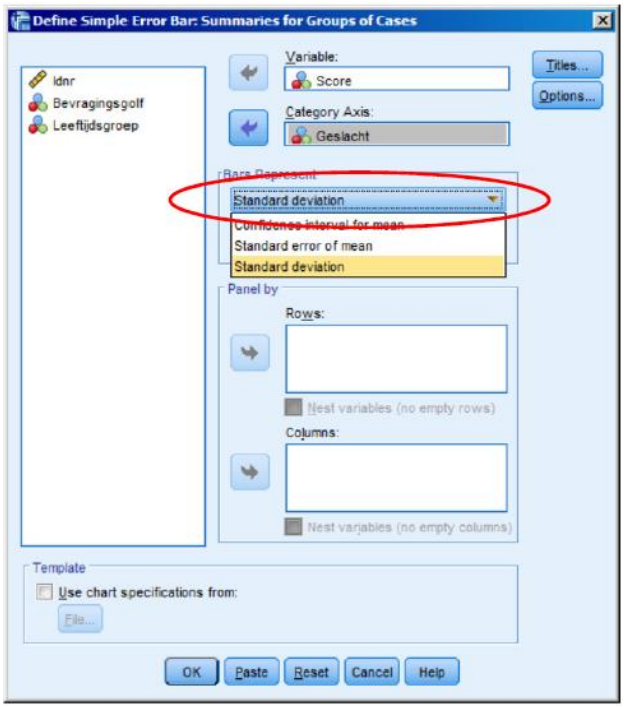
2. Means-procedure

Manier binnen SPSS om categorische en metrische variabelen te analyseren
"Analyse/Compare Means/Means"



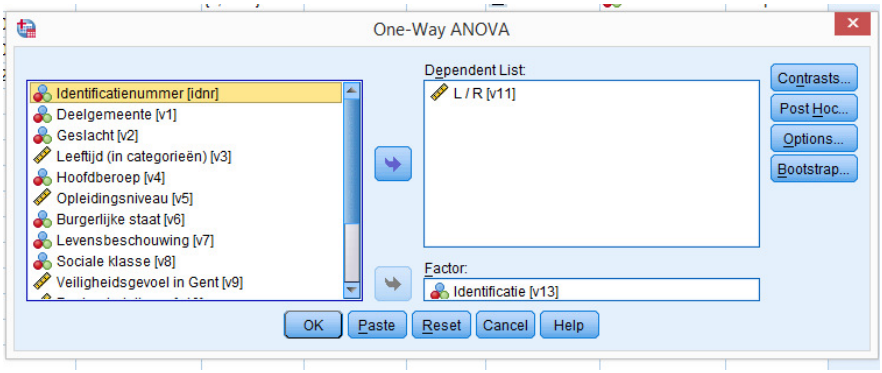
Grafische weergave door *"Errorbars"*
"Graphs/Legacy Dialogs/Error Bar…"

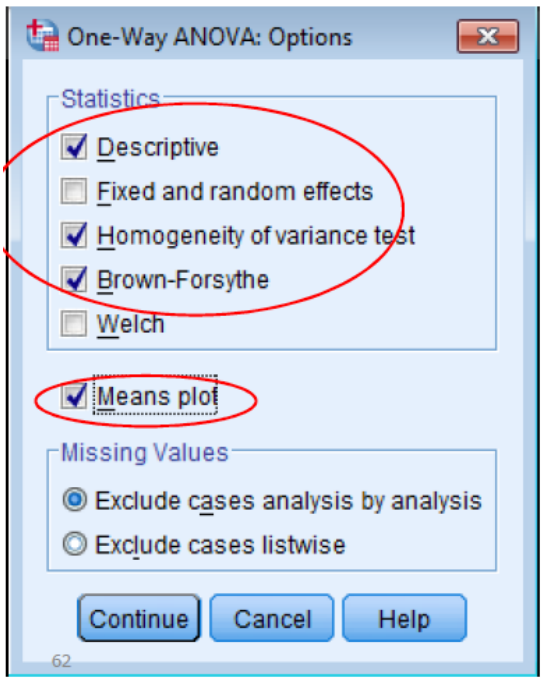




3. Variantieanalyse of ANOVA

Zijn deze verschillen nu statistisch significant? → ANOVA
"Analyse/Compare M eans/One Way ANOVA"





4. Interpretatie van ANOVA

Hoe oordelen over significantie?

- Kijken naar homogeniteit binnen groepen, zijn de standaardafwijkingen ongeveer gelijk?
- Testen door *"Levene's test"*
- Bepalend voor keuze tussen ANOVA en alternatief (Brown-Forsythe)

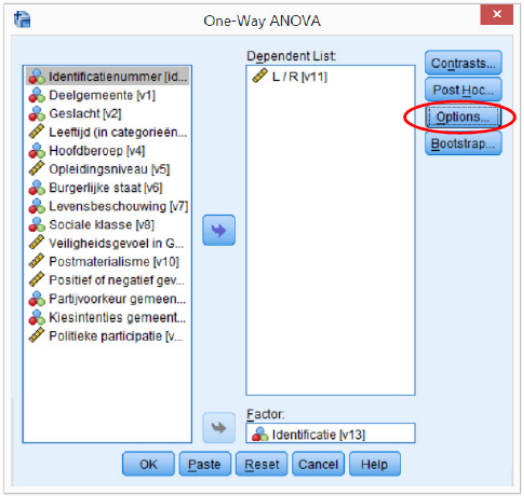
Levene's test met p-waarde kleiner dan 0,050

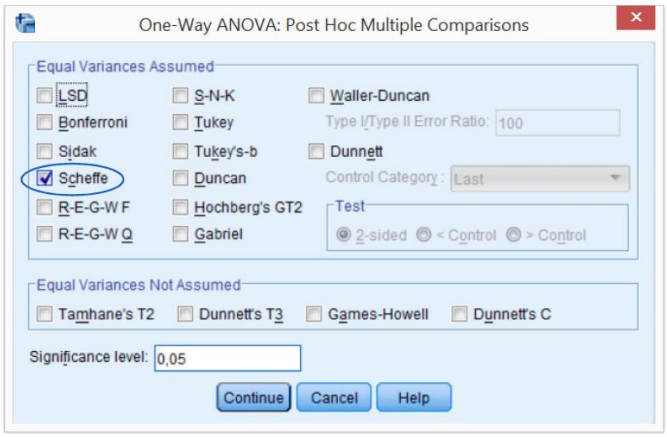
- Varianties **NIET** homogeen **binnen** groepen
- ANOVA op zich niet gebruiken
- Alternatief: Brown-Forsythe test
  → p-waarde kleiner dan 0,050: verschillen tussen groepen zijn significant
  → p-waarde groter dan 0,050: verschillen tussen groepen zijn **NIET** significant

Levene's test met p-waarde groter dan 0,050

- Varianties zijn homogeen **binnen** groepen
- ANOVA gebruiken
  → p-waarde kleiner dan 0,050: verschillen tussen groepen zijn significant
  → p-waarde groter dan 0,050: verschillen tussen groepen zijn **NIET** significant

Op basis van deze testen beoordelen we enkel het significant zijn van de verschillen over het algemeen. Om een meer specifieke analyse te maken moeten we gebruik maken van “post-hoc” testen. Daar kunnen we individueel de verschilen zien tussen elk van de variabelen.





5. Van twee variabelen één maken

Reeds gezien in les 2, voor extra voorbeeld zie dia 28 t/m 31

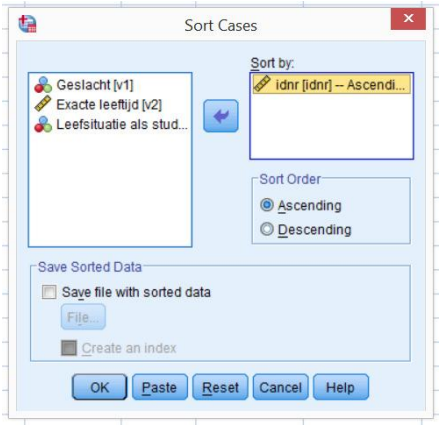
6. Databanken samenvoegen

Twee gevallen

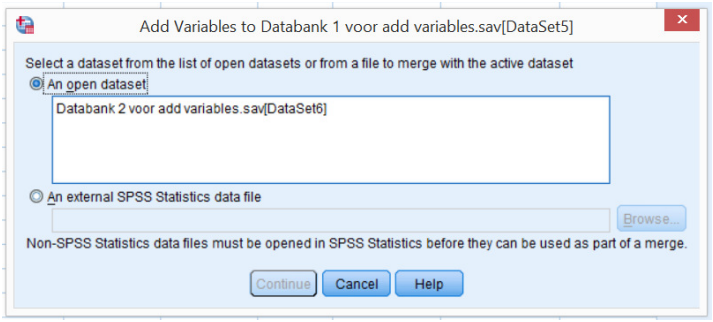
1. Additionele variabelen voor eenzelfde groep cases = *“add variables”*
   bv: we hebben al een aantal kenmerken van de studenten en willen er nog een aantal toevoegen
2. Gelijkaardige variabelen voor verschillende groepen = *“add cases”*
   bv: data over studenten en gelijkaardige data over een algemene populatie en willen die samenvoegen

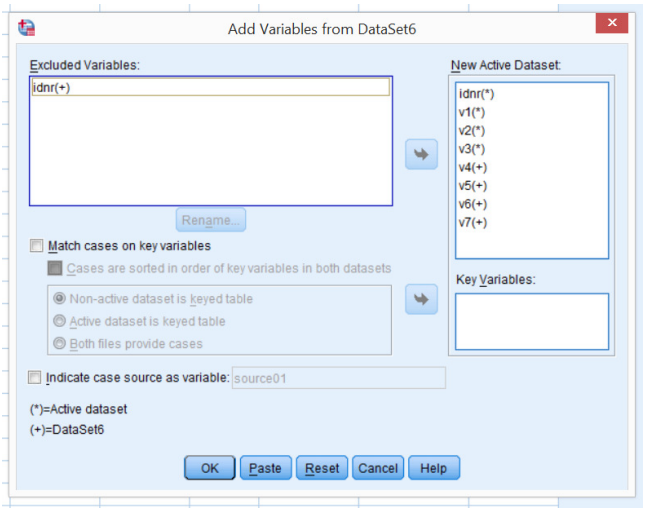
1. Add variables

STAP 1: **Beide** databanken sorteren op basis van de identificatie variabele
“Data/sort cases”



STAP 2: samenvoegen
"Data/Merge Files/Add Variables …"





- Overzicht van de variabelen uit de actieve databank (*) en de tweede databank (+)
- Best "matchen" op een identificatievariabele, hier de "key variable" genoemd, i.c. "idnr" (identificatienummer)
- Deze "key variable" mag enkel unieke waarden hebben
- In dialoogvenster aanduiden welke deze identificatievariabele is
- Default is "non-active dataset is keyed table": hiermee geef je aan dat enkel voor de cases in de huidige ("working") dataset er variabelen worden toegevoegd (in ons geval nu minder relevant, gezien in beide databanken dezelfde cases zijn opgenomen)
- Vervolgens ook aanduiden dat in beide databanken de cases gesorteerd zijn o.g.v. die sleutelvariabele
- De default is nu "both cases provide cases".
- In ons geval nu niet relevant, gezien in beide databanken dezelfde cases zijn opgenomen
- Op "OK" klikken
- Je krijgt nog een melding dat uw cases gesorteerd moeten zijn (maar dat deed je al)
- In de "working file" heb je nu de extra variabelen toegevoegd!

LES 6 & 7: Schaalconstructies

1. Factoranalyse

WAT?

- Patronen in de correlaties worden geanalyseerd en geïdentificeerd om een indicatie te vormen van een onderliggende theorie (cf. factoranalyse) of om gewoon de data te beschrijven (principale componentenanalyse)
- Veel variabelen bestuderen en nagaan of we deze kunnen samenvatten door een beperkt aantal factoren of componenten
- Sterk gecorreleerde variabelen worden gegroepeerd en afgezonderd van andere variabelen waarmee ze minder of niet correleren
- Assumptie = gebruik maken van achterliggende dimensies om complexe fenomenen te meten
- Enkel bij metrische variabelen

WANNEER?

- Ontwikkeling van schalen
- Datareductie
- Evaluatie van psychometrische kwaliteiten van een meetinstrument
- Dimensionaliteit in een set variabelen te analyseren

WELKE VRAGEN?

- Hoeveel achterliggende dimensies zijn er in onze data?
- Hoeveel dimensies hebben we nodig om de patronen in de samenhang/correlatie tussen variabelen weer te geven?
- Wat betekent elke dimensie/factor? Hoe moeten we die interpreteren?
- Welk aandeel van de variantie/informatie kunnen we verklaren/toeschrijven aan de factoren/componenten?
- Welke factoren verklaren de hoogste mate van variantie?
- In welke mate komt de factorstructuur overeen met de vooropgestelde theorie?
- Wat zou de score zijn van elke respondent indien we de concepten zouden meten door middel van de factoren/componenten?

DOEL?

- Weergeven relaties tussen variabelen die betekenisvol zijn
- Zowel statistische als inhoudelijke is van belang: gaat over de samenhang
- Geobserveerde relaties komen door hun gedeelde achterliggende dimensies

TOEPASSINGEN?

- Selectie van items/schalen die we willen opnemen in een meetinstrument → helpt kiezen wat we betrekken en weglaten om iets te meten
- Combinatie van statistische analyse met inzichten uit de theorievorming
- Validiteit, betrouwbaarheid, missing values!

Proces in vier stappen

1. Correlatiematrix
   - Matrix met alle variabelen
   - Variabelen identificeren die niet samenhangen met andere variabelen
   - Zwakke samenhang → weinig waarschijnlijk dat zij een gemeenschappelijke factor meten
   - Kijken naar relatieve sterkte van de correlaties
   - Goed om zicht te krijgen op mate waarin de uiteindelijke factoroplossing geschikt is

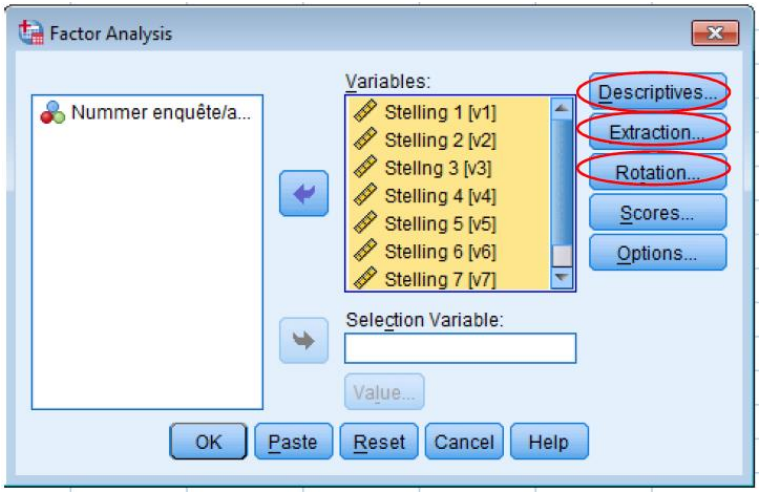
2. Factorextractie
   - Achterliggende dimensies? Hoeveel?
   - Werking principale componentenanalyse: 1e en 2e component verklaren nog de variantie maar de volgende steeds progressief minder

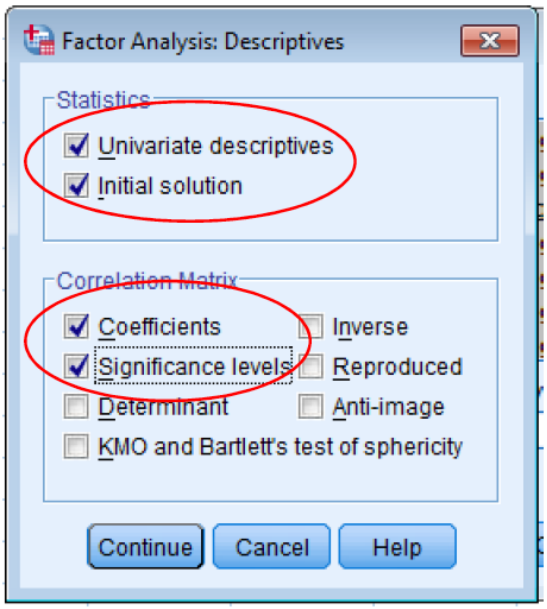
3. Factorrotatie
   - Kwestie van interpretatie van dimensies
   - Niet geroteerd → vaak moeilijk te interpreteren
   - Verschillende methoden om te roteren → vaak varimax-rotatie, probeert aantal variabelen met een hoge lading op een dimensie te minimaliseren, zodat de interpretatie van de dimensies ten goede komt

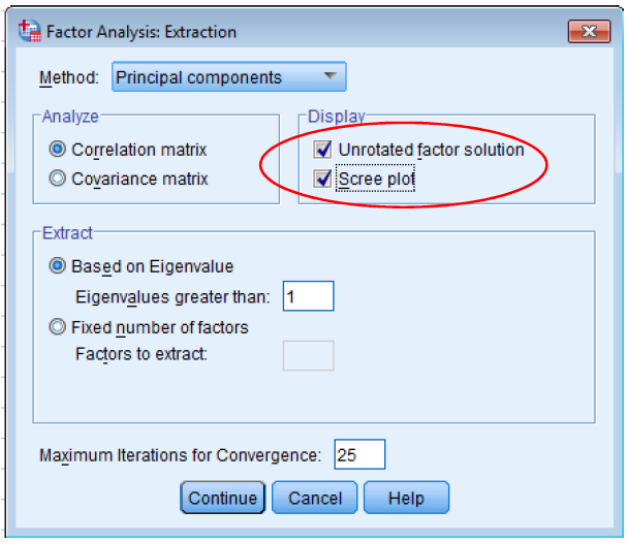
4. Beslissing
   - Aantal achterliggende dimensies?
   - Inhoudelijke verantwoording

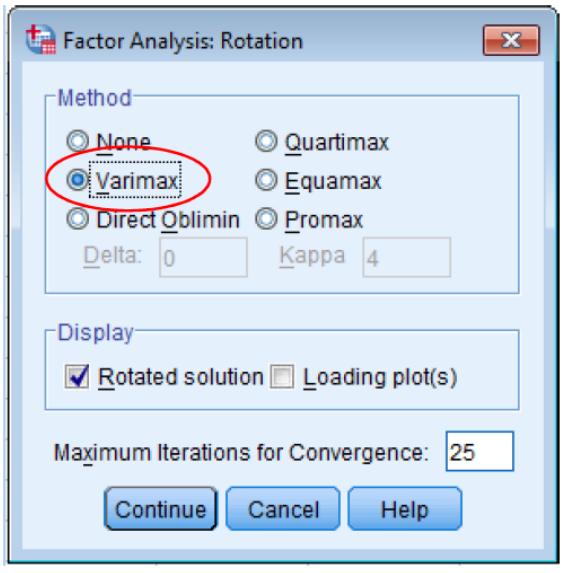
2. Factoranalyse met SPSS

"Analyse/Dimension Reduction/Factor…"









Output

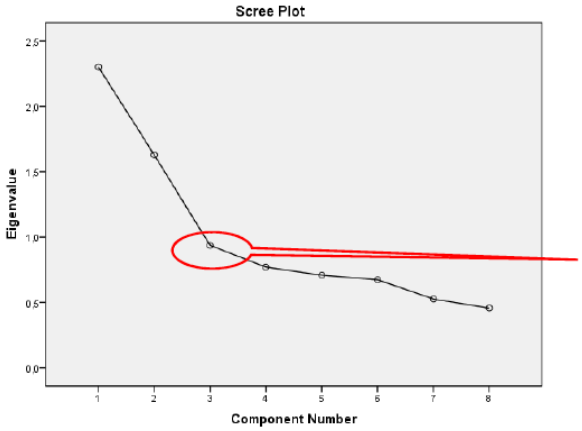
- Communalitetiten
  Deze tabel geeft ons informatie over de mate waarin de factoren de variantie binnen de verschillende variabelen kunnen verklaren. Bv. 61,0% van de variantie van stelling 1 wordt verklaard door alle factoren uit de analyse Voor stelling 5 is dit slechts 27,7%. Dus voor deze 5e stelling op het eerste zicht geen goede factoroplossing

Communalities

	Initial	Extraction
Stelling 1	1,000	,610
Stelling 2	1,000	,622
Stellng 3	1,000	,529
Stelling 4	1,000	,605
Stelling 5	1,000	,277
Stelling 6	1,000	,432
Stelling 7	1,000	,455
Stelling 8	1,000	,399

Extraction Method: Principal Component Analysis.

- Scree plot
  Dit is een grafische weergave van de eigenwaarde van elke component. Het is een alternatieve methode om te kijken hoeveel factoren men moet weerhouden. Hier moeten we kijken wanneer de lijn begint af te vlakken.



- Component matrix
  Deze tabel geeft de factorladingen weer vóór de factor rotatie

Component Matrix[a]

	Component 1	Component 2
Stelling 1	,770	,128
Stelling 2	,756	,225
Stellng 3	,630	,364
Stelling 4	-,772	,092
Stelling 5	-,169	,499
Stelling 6	-,141	,642
Stelling 7	-,279	,614
Stelling 8	-,130	,618

Extraction Method: Principal Component Analysis.

- Geroteerde component matrix
- Deze tabel geeft de factorladingen weer na de factor rotatie. Deze tabel helpt ons bij het “benoemen” van de factoren. Bv. de eerste vier stellingen leunen het dichts aan bij de eerste component, de laatste vier bij de tweede component. Ook uit deze analyse blijkt de vijfde stelling “zwakker” te zijn
  → aanduiding om deze stelling te schrappen

Rotated Component Matrix[a]

	Component 1	Component 2
Stelling 1	,777	-,076
Stelling 2	,789	,021
Stellng 3	,703	,188
Stelling 4	-,722	,290
Stelling 5	-,034	,526
Stelling 6	,030	,657
Stelling 7	-,110	,666
Stelling 8	,035	,631

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Eerder economische items	1. De klassenverschillen zouden kleiner moeten zijn dan nu het geval is.
	2. De overheid moet tussenkomen om de verschillen tussen de inkomens te verkleinen.
	3. Arbeiders moeten nog steeds strijden voor een gelijkwaardige positie in de maatschappij.
	4. De verschillen tussen hoge en lage inkomens moeten blijven zoals ze zijn.
Eerder sociaal-culturele items	5. Van hard werken word je een beter mens.
	6. Abortus moet in alle omstandigheden verboden blijven.
	7. Een vrouw is geschikter om kleine kinderen op te voeden dan een man.
	8. Een goed ouder zorgt ervoor dat zijn kinderen de gewoonte krijgen om te gehoorzamen.

Types factoranalyse

1. Exploratorieve factoranalyse
   - verkennend
   - data laten spreken en zien welke dimensies naar boven komen
2. Confirmatorische factoranalyse
   - meer geavanceerd
   - factorstructuur al gekend of afleidbaar uit theorie
   - gebruikt om na te gaan of gekende factorstructuren ook van toepassing zijn op nieuwe data

Wat maakt een goede factoroplossing?

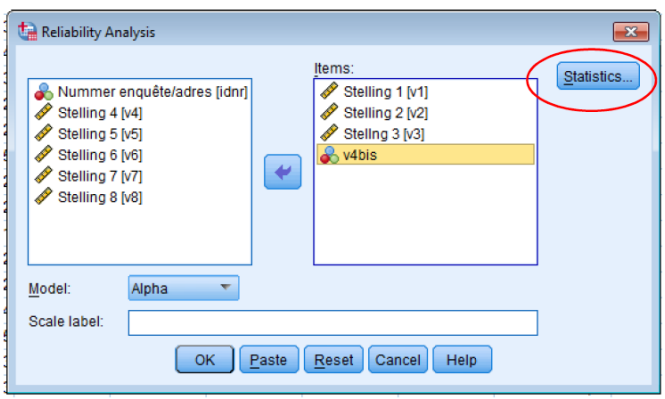
- Zinvol
- Makkelijk te interpreteren
- Simpele structuur
- Niet te complex qua ladingen

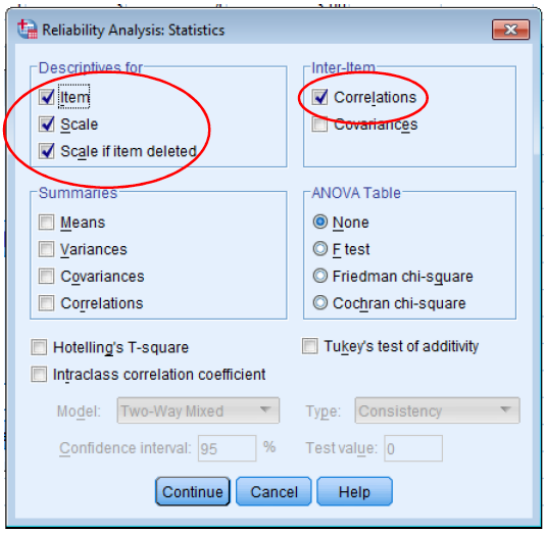
Hoe beoordelingen in dezelfde richting krijgen?

1. Kijken naar inhoud variabelen
2. Hoge/lage scores → wijzen zij naar eenzelfde richting?
3. Indien NIET → hercoderen met *"recode into different variables"*
4. Alternatief: compute

Betrouwbaarheid

- Omschrijving
- Analyse voor verschillende factoren afzonderlijk → voor de analyse moeten de hoge en lage scores wel dezelfde richting uit wijzen (zie vb dia 76)
- Cronbach's alfa (minimum 0,700 om betrouwbaar te zijn)
  "Analyse/Scale/Reliability Analysis"





Eigenlijke waarden voor Cronbach's alfa

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,749	,751	4

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Stelling 1	10,5361	4,995	,581	,341	,672
Stelling 2	10,7432	4,506	,588	,353	,665
Stelling 3	10,7689	4,956	,483	,250	,727
v4bis	10,4526	5,200	,534	,318	,697

→ In welke mate zou het weglaten van een variabele de waarde beïnvloeden en dus ook de betrouwbaarheid

LES 8: Regressie

1. Regressieanalyse

WAT?

- Verband tussen variabelen gebruiken om de waarde van één van de variabelen te voorspellen op grond van de kennis van de andere variabele
- Enkelvoudige vs. Meervoudige lineaire regressie

WAAROM?

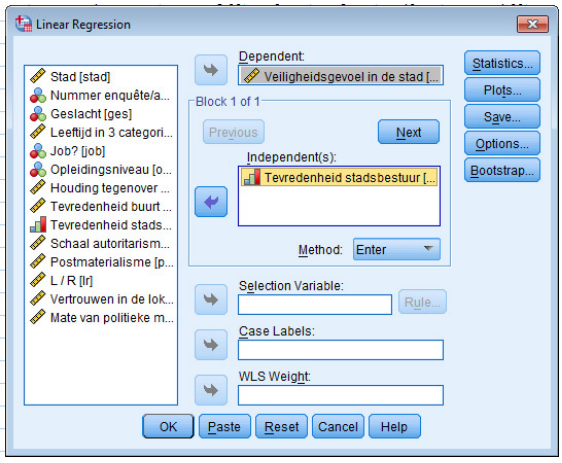
- Geeft beeld over sterkte van verbanden
- Beeld van de verklarende waarde van variabelen
- Meervoudige regressie: houdt met meerdere variabelen rekening en laat ons toe effecten te isoleren

VERONDERSTELLINGEN

- Meetniveau van de variabelen
  afhankelijk: metrisch
  onafhankelijk: metrisch of dummy
- Verbanden tussen afhankelijk en onafhankelijk moeten lineair zijn
- Meervoudig: onafhankelijken mogen niet sterk correleren

2. Enkelvoudige regressieanalyse met SPSS

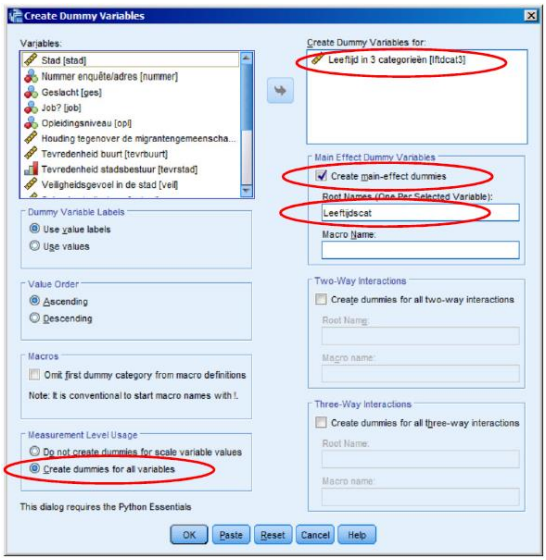
"Analyze/Regression/Linear"



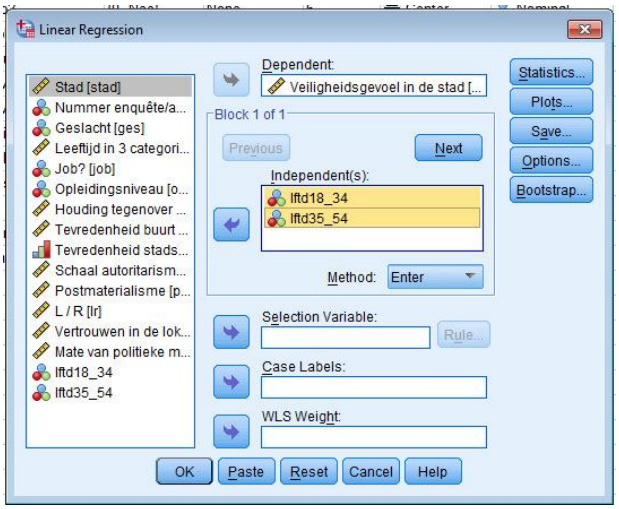
Altijd nagaan in ANOVA test of de significantie onder 0,050 ligt, enkel dan is het geen toevallig verband.

Categorische variabelen als onafhankelijke

- Kan geen onafhankelijke zijn → dummy (dichotome variabele, ja of nee)
- Maakt niet uit welke voor de verklarende waarde
- Moeten allemaal opgenomen worden
- Interpretatie van de rico zal enkel veranderen omdat de referentiecategorie anders is
- Dummy in SPSS
  "Transform/Create Dummy Variables"



- Daarna kan je de dummy opnemen als onafhankelijke



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,174[a]	,030	,029	2,236

a. Predictors: (Constant), lftdcat3=55+, lftdcat3=18-34

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	288,337	2	144,169	28,828	,000[b]
	Residual	9281,739	1856	5,001		
	Total	9570,076	1858			

a. Dependent Variable: Veiligheidsgevoel in de stad

b. Predictors: (Constant), lftdcat3=55+, lftdcat3=18-34

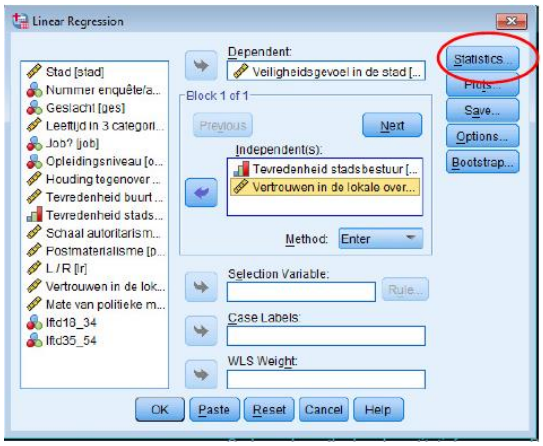
R-Square: 3% van de variantie in het veiligheidsgevoel kunnen we verklaren door de leeftijd

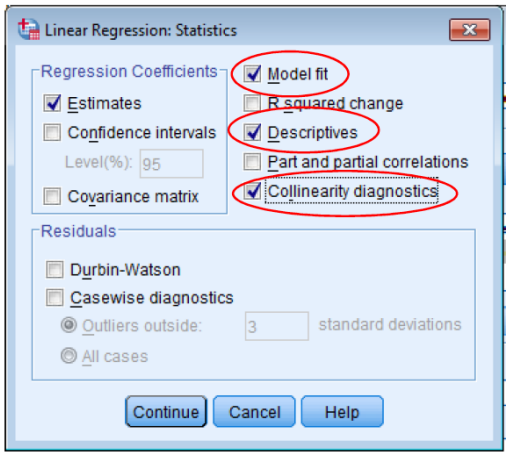
Adjusted R-square: houdt rekening met het aantal onafhankelijke variabelen, kleiner dan R-square, beter geschikt om te vergelijken

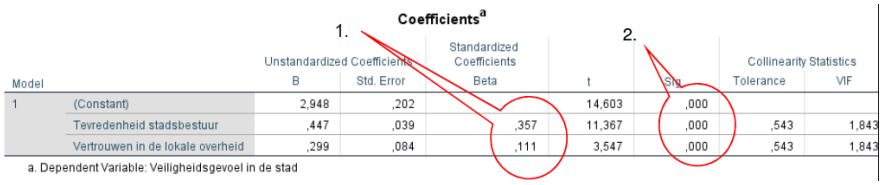
Ook kijken naar significantie!

3. Meervoudige regressieanalyse met SPSS

"Analyze/Regression/Linear"







Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	2,948	,202		14,603	,000		
	Tevredenheid stadsbestuur	,447	,039	,357	11,367	,000	,543	1,843
	Vertrouwen in de lokale overheid	,299	,084	,111	3,547	,000	,543	1,843

a. Dependent Variable: Veiligheidsgevoel in de stad

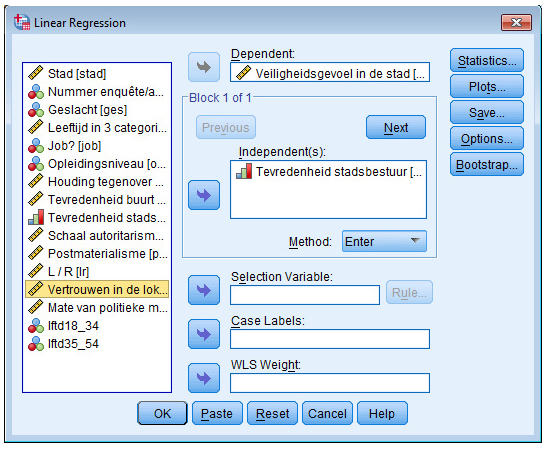
1. BETA-waarden: hoe groter (positief of negatief), hoe belangrijker een waarde is in het verklaren van de unieke variantie in de afhankelijke variabele
2. Significantie

Tolerance & VIF: indicatoren voor multicollineariteit, wij gebruiken enkel VIF. Indien lager dan 10 → geen probleem

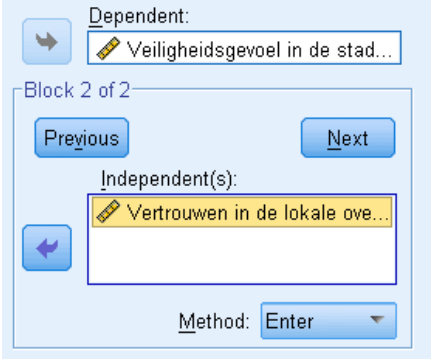
Multicollineariteit
= hoge samenhang tussen twee of meer onafhankelijke variabelen = geen correcte info in regressiemodel
→ Kunnen we zien door grote veranderingen in waarden van coëfficiënten of significantiewaarden bij toevoeging variabele
→ Tegengaan door elimineren van een variabele of samenvoegen verschillende variabelen

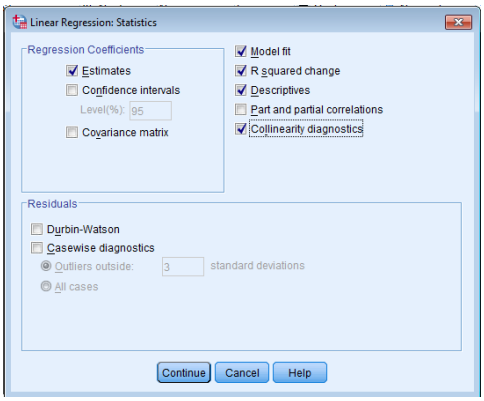
3. Hiërarchische regressieanalyse met SPSS

type meervoudige regressie: volgorde bepalen we nu zelf → kijken of variabelen additioneel kunnen bijdragen tot verklaring afhankelijke variabele
"Analyze/Regression/linear"



→ *NEXT* →





Output is helemaal hetzelfde als de meervoudige analyse, enkel voor 2 gevallen zodat we kunnen vergelijken.