Methodologieën voor Politiek-wetenschappelijk onderzoek

Module 2: kwantitatieve databeheers- en analyse-technieken

Integrale presentatie

Lesgever: Tom Schamp

Module 2:
Databeheer en –analyse

Interactieve inleiding tot SPSS Statistical Software

E-bronnen:
Elizabeth Bigham, Ph.D.
California State University San Marcos May 2007 - Funded by SSRIC
HOGent: http://habe.hogent.be/stat/statistiek/spss/spss.html

Deel 2

SPSS opstarten

SPSS – wat ?

- SPSS is **een statistisch programma** (voorheen tevens de naam van het bedrijf dat dit op de markt bracht). Sinds medio 2009 is de onderneming eigendom van IBM. (verkoopprijs: 840 miljoen euro)
- SPSS : Statistical Package for the Social Sciences
- De eerste versie van het programma SPSS stamt **uit 1968**.
- Met SPSS kunnen **data** worden:
  - gemaakt - gegevensinvoer,
  - bewerkt - bijvoorbeeld het samenvoegen van mensen in een aantal leeftijdscategorieën,
  - geanalyseerd - bijvoorbeeld het berekenen van de samenhang tussen twee variabelen of het bepalen van de statistische significantie van een verschil tussen twee groepen.

- Het programma is opgebouwd uit **verschillende modules**:
  - De 'Base Module' is onmisbaar om het programma te laten draaien.
  - De basis kan worden uitgebreid met onderdelen die meer geavanceerde analyses mogelijk maken, zoals speciale vormen van regressie-analyse (SPSS Regression Models), tabellen (SPSS Tables) en SPSS Trends™ voor trendanalyses.

Leerdoelstellingen

Studenten leren:

- Een **data-entry pagina** op te zetten
- **Werken met variabelen** (toevoegen, verplaatsen, hercoderen)
- **Beschrijvende statististische analyses** uitvoeren (descriptieve analyse)
- **Eenvoudige correlaties uitvoeren** (om samenhang tussen twee variabelen te analyseren)
- Resultaten voor **twee groepen met elkaar vergelijken**

Opzet en reikwijdte

- In de module 1 hebben we gezien dat dataverzamelingstechnieken veel van elkaar kunnen verschillen.
- Bovendien zijn de mogelijkheden van SPSS om data te analyseren heel uitgebreid.
- In deze 'inleidende' module hanteren we **eenvoudige voorbeelden**, een eenvoudige vragenlijst als onderdeel van een louter hypothetische studie naar politieke en maatschappelijke attitudes.

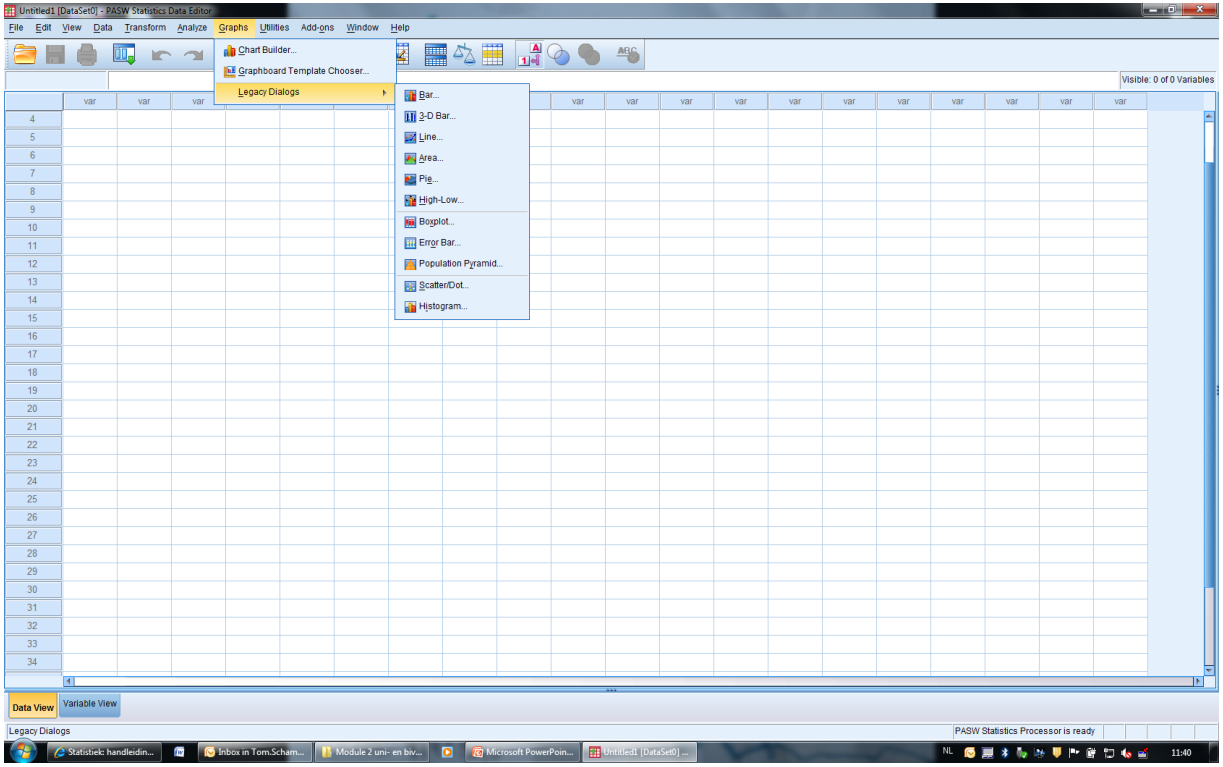
Welke vensters in SPSS ?

- SPSS is net zoals andere Windows-toepassingen (vb. Word) gestructureerd aan de hand van vensters en menu's.

- Drie types van vensters onderscheiden met elk hun eigen functie:
  - **Applicatievenster**
  - **Data editor venster**
  - **Output-venster**

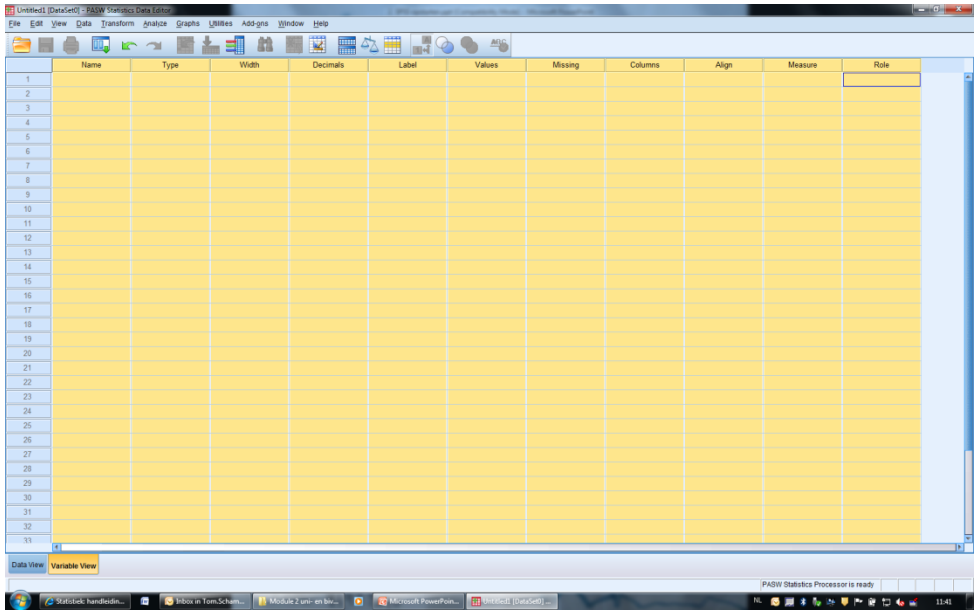
Het applicatievenster

- Dit is het basisvenster van SPSS waarin alle andere vensters verschijnen :



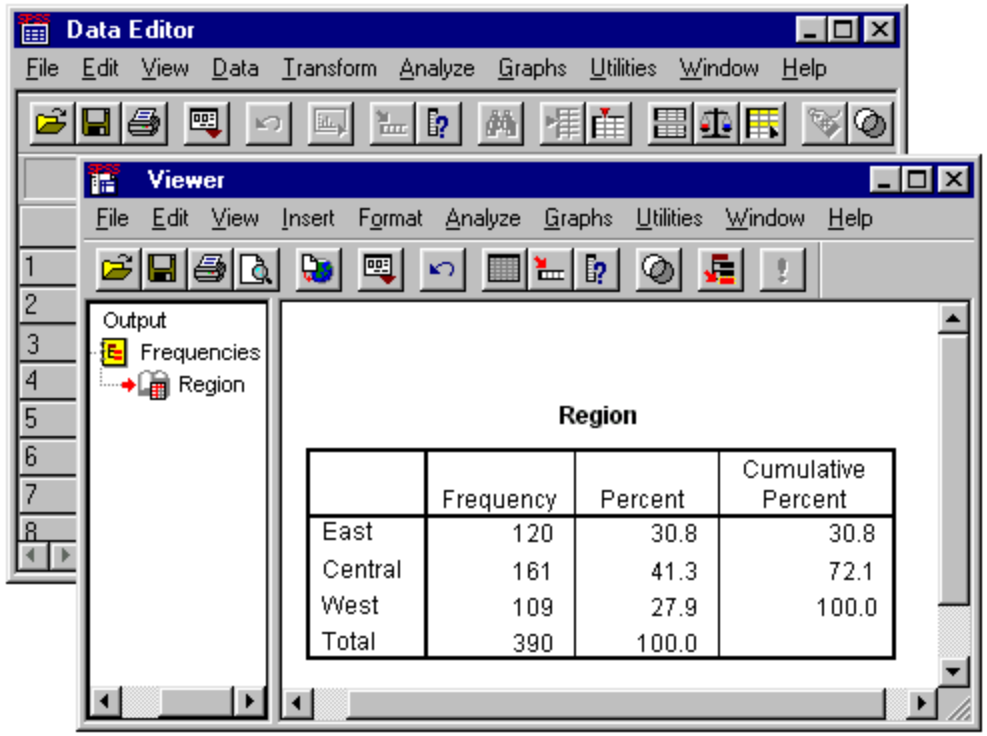
Het data editor venster

- In deze spreadsheetachtige datamatrix wordt de inhoud van je datafile getoond, je kan dit venster opsplitsen in twee mappen. In de map **Data View** kan je de data zien van het actieve bestand. In de map **Variabele View** zie je de definiëring van de verschillende variabelen uit je data-bestand:



Het output venster

- In dit venster verschijnen alle outputs van je analyses met name tabellen, statistieken en grafieken.



Menustructuur

Het menusysteem in het applicatievenster (zowel de data-editor als de output applicatie) ziet er als volgt uit:

- **File**: Het file-menu bevat een aantal opdrachten om bestanden te bewaren en te activeren. Je kan er ook SPSS mee afsluiten.
- **Edit**: Hiermee kunnen teksten en afbeeldingen worden gekopieerd, gewijzigd of verwijderd.
- **View**: In dit menu kan je bepaalde toolbars selecteren zodat deze wel of niet in je applicatievenster verschijnen.
- **Data**: Onder dit menu zitten opdrachten waarmee de globale actieve dataset kan worden gewijzigd (vb. *het trekken van een steekproef alle mannen uit de dataset*). De aangebrachte wijzigingen zijn slechts tijdelijk.

Menustructuur - 2

- **Transform**: Met de opties in dit menu kunnen wijzigingen worden aangebracht in de individuele variabelen (vb. hercoderen van variabelen of nieuwe variabelen creëren op basis van bestaande). Ook deze wijzigingen zijn slechts tijdelijk en moeten worden bewaard als je ze definitief wil vastleggen op schijf.
- **Analyze**: Via dit menu kunnen de verschillende statistische analyses worden uitgevoerd.
- **Graph**s: Dit biedt een overzicht van het groot aantal grafieken dat je kan maken met SPSS.
- **Utilities**: Via dit menu kan je een overzicht verkrijgen van al je variabelen in je data-file.
- **Windows**: Hiermee kan van het ene venster naar het andere worden overgegaan en kunnen de geopende vensters in een bepaalde volgorde worden geplaatst

Menustructuur - 3

- **Help**: Via dit menu kan hulp worden gevraagd over de algemene werkwijze van SPSS en over de verschillende statistische technieken.

De statusbalk in het applicatievenster

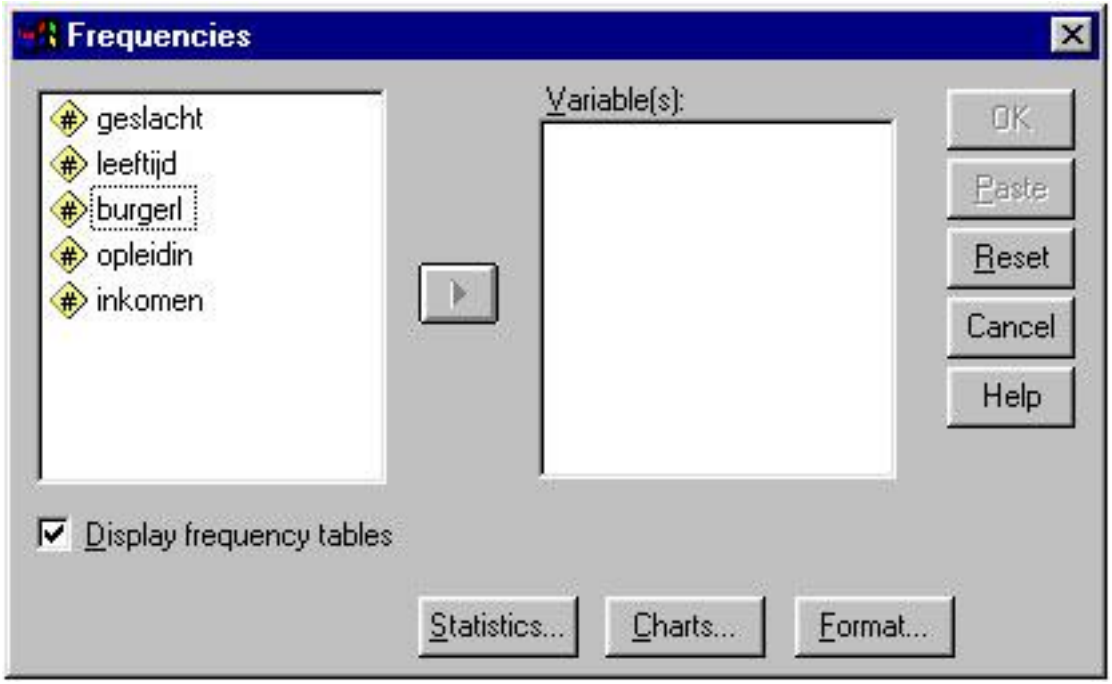
Onderaan in het applicatievenster is de statusbalk zichtbaar.

Wanneer de processor bezig is met het uitvoeren van een commando verschijnt hierop de naam van dit commando. Wanneer je slechts met een beperkt aantal cases werkt doordat je bvb. een steekproef hebt getrokken dan verschijnt op de statusbalk 'filter on'.

Dialoogvensters

- Een dialoogvenster is een venster dat verschijnt wanneer **een bepaalde optie uit het menu** is geselecteerd. Het dialoogkader wordt gebruikt om het gevraagde commando te specifiëren.
  - Wanneer er bijvoorbeeld wordt gevraagd om een frequentietabel te berekenen van een variabele uit de datafile dan verschijnt een volgend dialoogvenster.

Dialoogvenster FREQ



Dialoogvenster FREQ

- In het bovenstaande dialoogkader ‘Frequencies’ zijn de volgende elementen aanwezig:
  - **Titelbalk**: de naam van het dialoogvenster
  - **Commandoknop**: Dit zijn de grote rechthoekige knoppen rechts in het dialoogkader. Er zijn verschillende commando's:
  - **OK**: Deze opdracht voert het gevraagde commando uit met alle opgegeven specificaties. (Opmerking: in een subdialoogvenster staat in plaats van OK, Continue vermeld.)
  - **Paste**: Dit commando kopieert het geselecteerde commando in tekstvorm met alle opties naar het syntax-venster.

- **Reset**: Door deze knop aan te klikken worden alle reeds gemaakte selecties in het dialoogkader gewist.
- **Cancel**: Dit commando sluit het dialoogvenster zonder de functie uit te voeren.
- **Help**: Hiermee kan je de hulp inroepen van SPSS: je krijgt dan informatie over wat de verschillende mogelijkheden betekenen (vb. in dit dialoogkader krijg je informatie over wat je kan berekenen met statistics, charts en format).
- **Bronlijst**: in deze lijst vind je een overzicht van alle variabelen, wanneer je een analyse wil uitvoeren selecteer je een variabele uit deze lijst en breng je ze met behulp van de pijl over naar de 'doel'-kolom.
- **Doellijst**: in deze lijst staan alle variabelen waarvoor je de geselecteerde analyse zal uitvoeren

Deel 3

Een data-invoer pagina aanmaken

Twee opties

- Maak een dataset aan
- Laadt een bestaande dataset op

Enkele weetjes

- Woorden in *italics* verwijzen naar een locatie.
- Woorden in “quotes” zijn woorden die je zo moet typen of ingeven.
- **Vet** betekent een selectie (om op te klikken).
- Het symbool > (of “>>”) betekent dat je de volgende stap moet nemen.

Voorbeeldvragenlijst

Deelnemer #: __

Geslacht: M F

Leeftijd: ___

Huidig studiejaar: ___

Onderwijsinstelling: ___

Op een schaal van 1 tot 10 hoe zeker ben je dat je na deze module zal slagen voor de test?

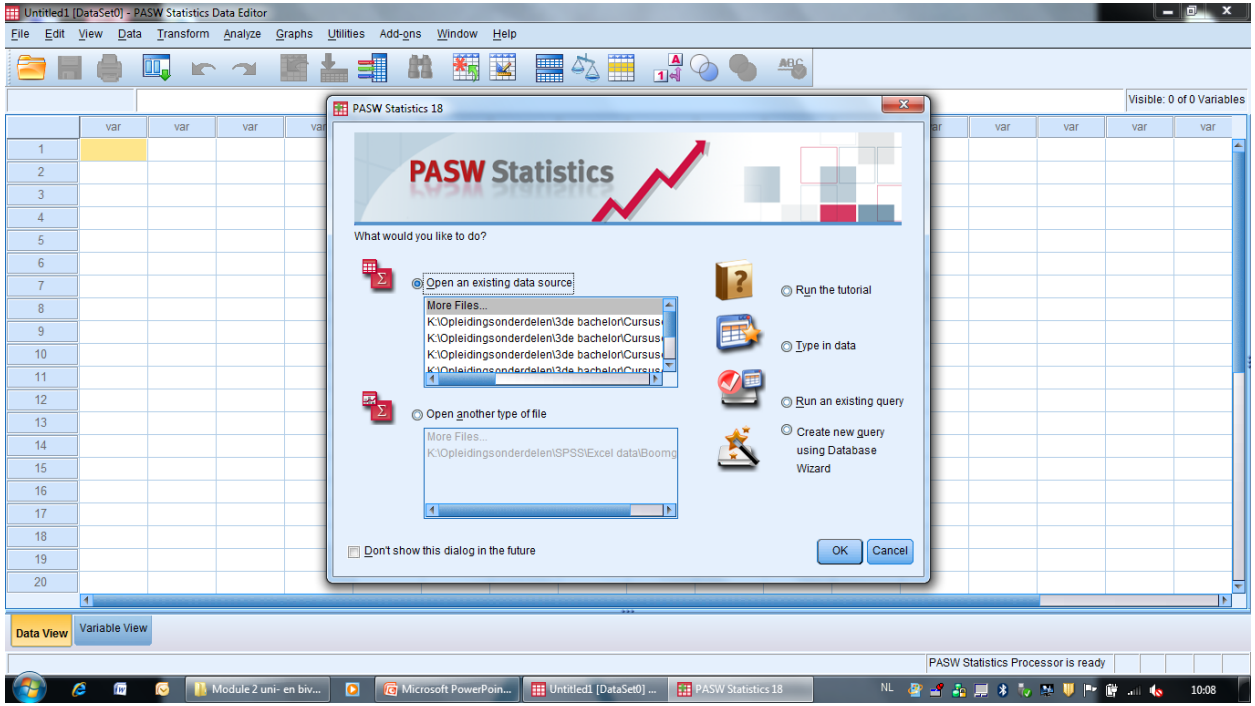
1 2 3 4 5 6 7 8 9 10

Helemaal niet Heel zeker

Jaren PC-ervaring? ___

Open SPSS

- Open SPSS: **Start > All Programs > PASW Statistics 18.0 for Windows**
- Dit venster opent zich

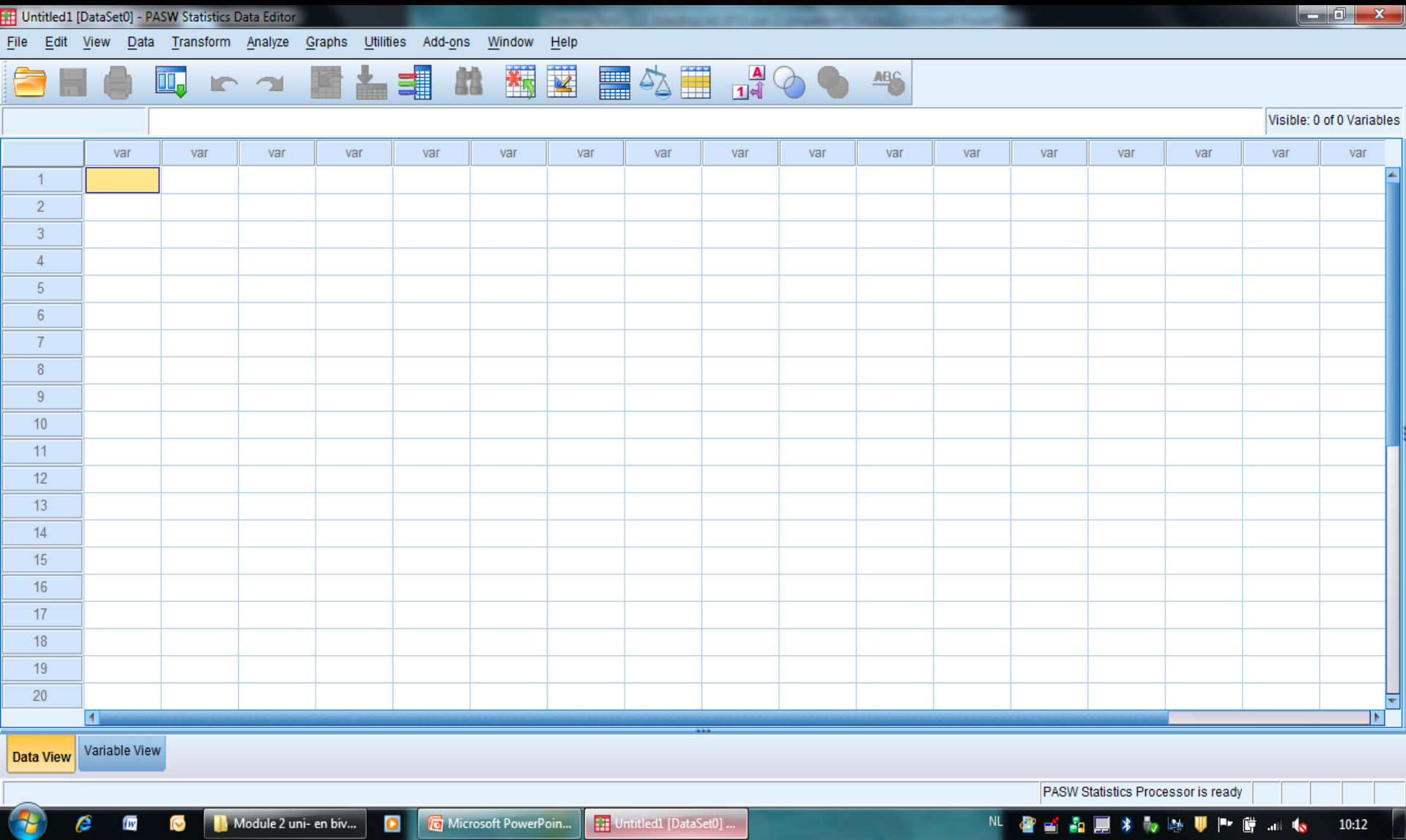


- We hebben **5 keuzemogelijkheden**:
  - doorloop de handleiding
  - **typ data in**
  - laat **een bestaande databank** lopen
  - creëer een nieuw databank aan de hand van de Database Wizard
  - open een bestaande file (bvb. Excel-file)

Type in data

- Het eerste scherm (*Data View*) is de plaats waar je de data ingeeft
  - **een deelnemer per rij**
  - **een variabele per kolom**.

- Tip: als het venster niet helemaal opent zoals op de volgende slide wordt afgebeeld: tik dan op de full screen button in de hoek rechtsbovenaan.

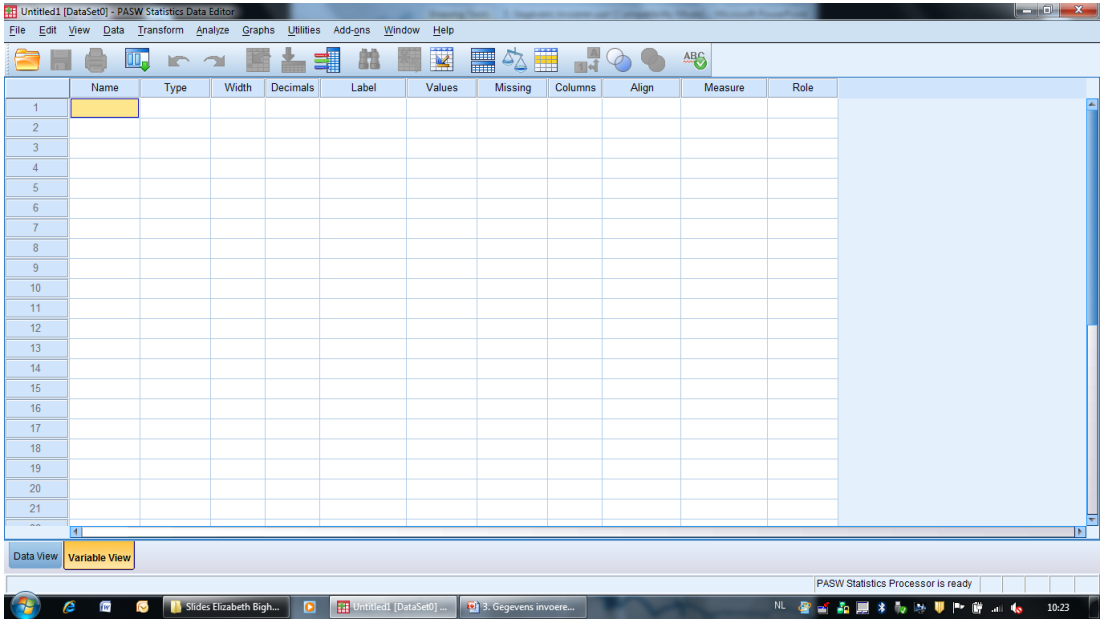

Untitled1 [DataSet0] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 0 of 0 Variables
var
Data View
Variable View
PASW Statistics Processor is ready
Module 2 uni- en biv...
Microsoft PowerPoin...
Untitled1 [DataSet0] ...
NL
10:12

De data editor

- De gegevens worden in SPSS ingebracht via de Data Editor. Dit venster is een spreadsheet-achtige matrix om data in te tikken, te bekijken en te bewerken. Je komt op dit venster door bij het opstarten van SPSS te kiezen voor de optie *'type in new data*'.
- Is SPSS reeds opgestart dan kan je dit data-editor venster ook bekomen door de volgende selecties te maken: **File, New, Data**.

Set Up kolommen

- Switch naar *Variable View.*
- Select: **Variable View**.



(Dit is het scherm waar je informatie over de gebruikte variabelen ingeeft; **een variabele per rij**)

Hoe data invoeren ?

- Cases worden weergegeven in de rijen: op iedere rij staan de antwoorden van één respondent
- Variabelen worden weergegeven in de kolommen: elke variabele staat in een aparte kolom
- Cellen bevatten waarden: elke cel is de kruising tussen de cases en de variabelen. M.a.w. elke cel bevat de waarde van één bepaald element uit de populatie op één bepaalde variabele.
- De waarden van de cellen zijn **bij voorkeur numeriek**! Veel statistische toetsen werken in SPSS niet als er geen gebruik werd gemaakt van de numerieke codes. Bijvoorbeeld: voor de variabele geslacht tik je best de codes 0, 1 of 9 in in de datamatrix.

Waarden geven in cellen

- Klik in de cel waar je waarden wil intikken: zo wordt de gewenste cel actief (er komt een dikker rand rond de cel die actief is), enkel in deze actieve cel kan je gegevens intikken (de variabele naam en het casenummer wordt in de actieve cel steeds vermeld in de linker-bovenhoek van het data-editor venster.
- Tik de waarde of numerieke code in

- Druk op:
  - **Enter**: om naar de cel onder aan de actieve cel te gaan
  - **Tab**: om naar de cel rechts van de actieve cel te gaan (is het meest gebruikt omdat het eenvoudigst is om case per case in te vullen)
  - Gebruik **de cursor** om de respectievelijke richting uit te gaan

Variabelen definiëren

- Onderaan het venster zie je dat de Data Editor onderverdeeld is in twee mappen: '**Data View'** en Variabele View.
- Om data in te voeren moet je eerst je variabele definiëren. Dit kan je doen door te klikken op de map **'Variabele View'** of je dubbelklikt op de variabele naam '**var**' aan het hoofd van elke kolom en de 'Variabele View' map wordt zichtbaar. In de kolom variabele namen klik je in de gewenste rij van de variabele en geef je de nieuwe variabele naam in.

Variabelen definiëren - 2

- Een *variabele naam* moet aan een aantal voorwaarden voldoen:
  - de variabele naam is maximum 8 karakters
  - de naam moet beginnen met een letter (dus niet met een cijfer!)
  - speciale tekens zoals !,? en * zijn niet wenselijk
  - geen spaties !!!
  - de naam mag niet één van de sleutelwoorden zijn waarmee SPSS zijn statistische berekeningen uitvoert (zoals AND, NOT, EQ, BY en ALL)

Stap 1 en 2

- Omdat we getallen of woorden wensen in te voeren in SPSS moeten we ook aangeven wat voor type van gegevens de variabele is. Dit doe je door in de tweede kolom '**Type**' de cel te selecteren. Er verschijnt dan een grijs blokje. Als je op dit blokje klikt dan komt er een dialoogkader te voorschijnt voor het ingeven van het type van variabele. Je hebt een keuzemogelijkheid uit diverse types van data zoals *numeriek, munteenheid, comma, datum en tekstwoord (string)*.

Stap 3, 4 en 5

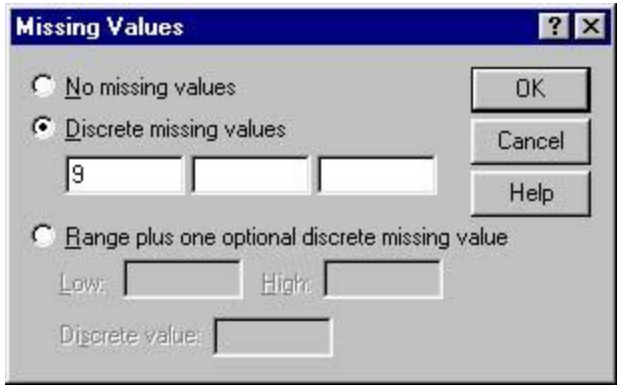
- In de derde kolom ('**width**') duid je de grootte van je variabele naam aan.
- Voor je vierde stap bepaal als je variabele een decimaal getal is, hoeveel decimalen er moeten zijn. ('**decimals**')
- In de vijfde kolom '**label**' kan je een omschrijving geven van je variabelenaam. Dit is vooral belangrijk bij de interpretatie van je statistische outprint. In de tabellen van je output vind je dan niet de variabele naam terug (bvb. opleiding) maar wel het volledige label (bvb. hoogst behaalde diploma).

Stap 6 en 7

- In de zesde kolom ‘values’ moet je ingeven wat de waarden zijn **van je labels** (bvb. variabele geslacht: code 0 voor man en code 1 voor vrouw). Ook dit is zeer belangrijk om je output interpreteerbaar te maken. Anders krijg je tabellen met cijfers waarvan je niet meer weet waar elk cijfer voor staat. Hier wordt nog op ingegaan verder in deze handleiding.
- Het invoeren van **missing values** doe je in kolom zeven. Meestal wordt er gebruik gemaakt van 9, 99, 999 of 9999 in de optie 'discrete missing values'. Dit wordt in de volgende alinea uitgelegd.

Missing values

- In SPSS roep je het dialoogvenster op door in de map 'Variabele View' de cel bij een bepaalde variabele aan te klikken in de kolom 'missing'. Dan verschijnt er in deze cel een rechthoekje waarop je nog eens klikt. Je hebt drie mogelijke opties voor het definiëren van de missing value:



Missing values - 2

- **no missing values**: op de variabele zijn van alle respondenten een valabel antwoord bekend.
- **discrete missing values**: de ontbrekende waarde wordt weergegeven door middel van een discreet getal. In SPSS kiest men meestal voor 9, 99, 999 of 9999.
- **range plus one optional discrete missing value**: bij deze optie kan men een volledig interval opgegeven als missing value en optioneel een discrete waarde buiten dit interval (bvb. voor onmogelijke waarden opgegeven voor de variabele: aantal uren dat je gemiddeld per week werkt)

- Meestal wordt de optie 'discrete missing value' gebruikt.

Stap 8 en 9

- Je kan in de achtste kolom ook nog de kolomwijdte opgeven.
- Kolom negen (**aligneren**): geen actie
- De allerlaatste stap (kolom tien) voor het definiëren van je variabele is het opgeven van **het meetniveau**. Omdat hier regelmatig fouten tegen worden gemaakt, is dit een cruciale stap ! Zonder kan je de data niet correct analyseren!!
- Laatste kolom (**role input**): geen actie

- Zo, klaar voor een oefening

Voorbeeldvragenlijst

Deelnemer #: __

Geslacht: M F

Leeftijd: ___

Huidig studiejaar: ___

Onderwijsinstelling: ___

Op een schaal van 1 tot 10 hoe zeker ben je dat je na deze module zal slagen voor de test?

1 2 3 4 5 6 7 8 9 10

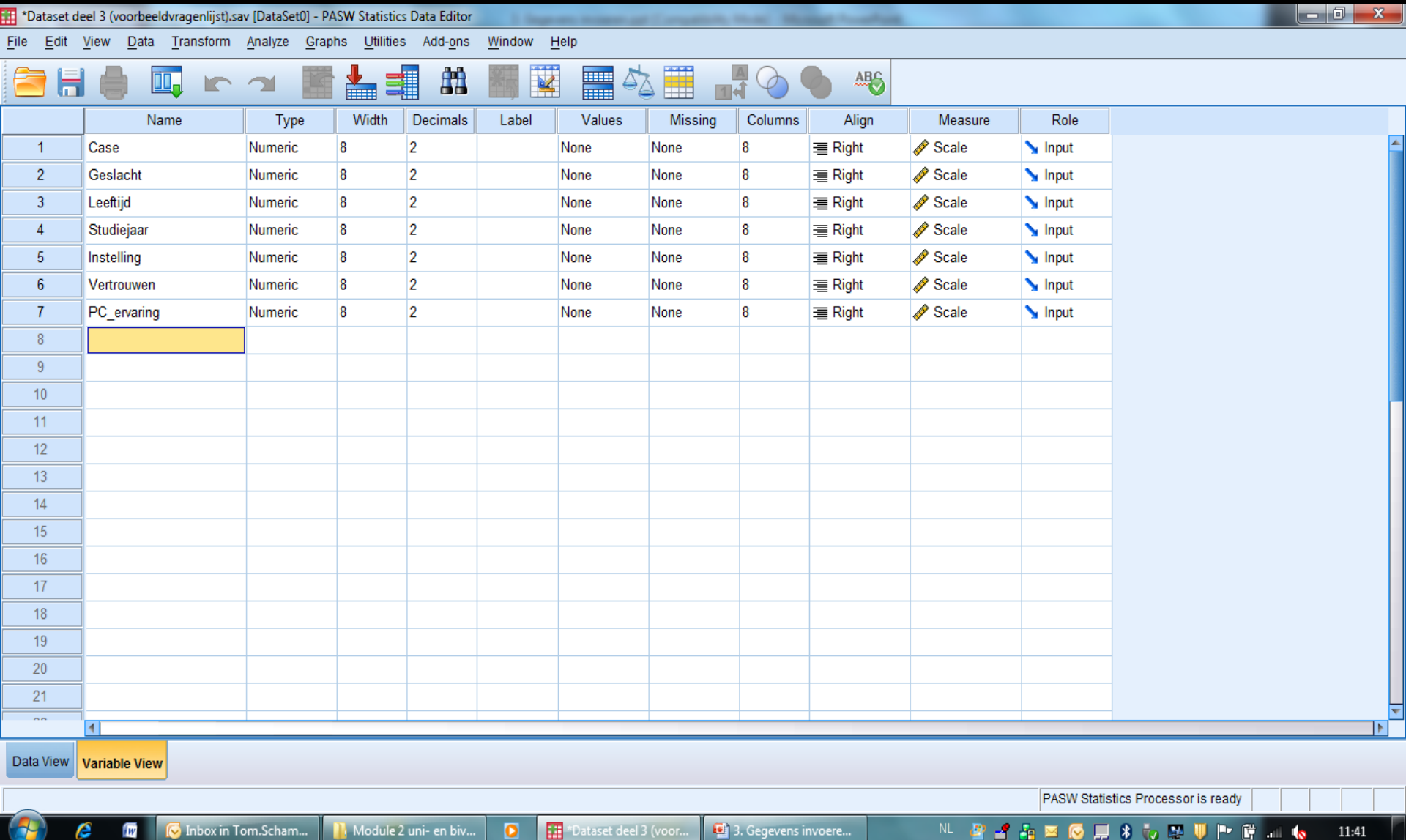
Helemaal niet Heel zeker

Jaren PC-ervaring? ___

Een oefening

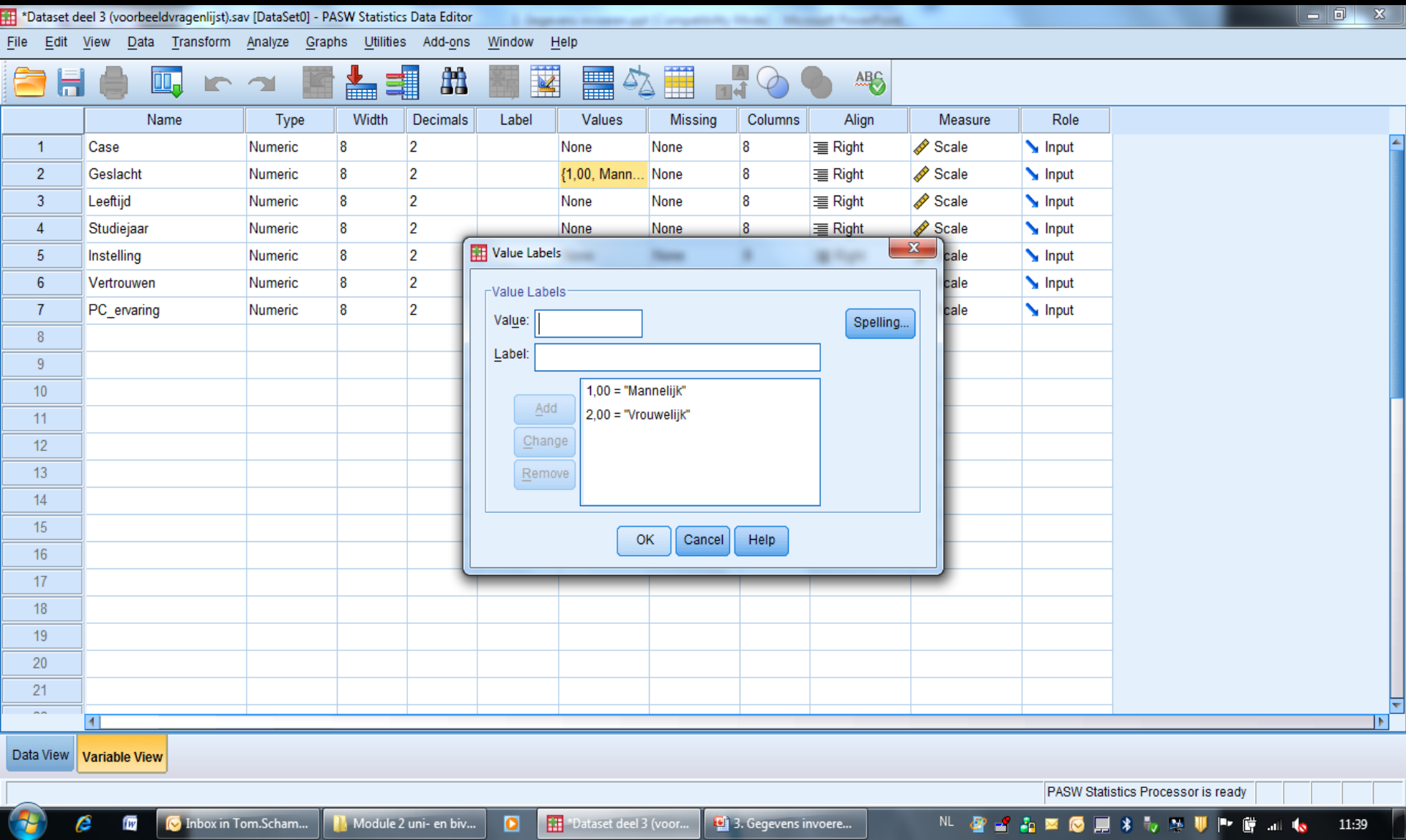
- De eerste kolom is *Name*. Geef hier de naam in van je eerste variabele (gebruik geen spaties!). Typ "Case" en ga naar de volgende lijn.
  - Bij de overgang van de ene naar de andere variabele vult de hele lijn na de aangemaakte variabelenaam zich met default settings. Deze worden later aangepast.
- Volgende variabele. Typ: "Geslacht"
- Volgende, type: "Leeftijd" etc.

- Volgende, typ: “Studiejaar”
- Volgende en typ: “Instelling”
- Volgende en typ: “Vertrouwen” (lees: mate van vetrouwen dat die al dan niet bereikt wordt)
- Volgende en typ: “PC_ervaring” (opgelet: niet alle niet-alpha of niet-numerieke tekens worden aanvaard; ook geen spaties vandaar *underscore*)

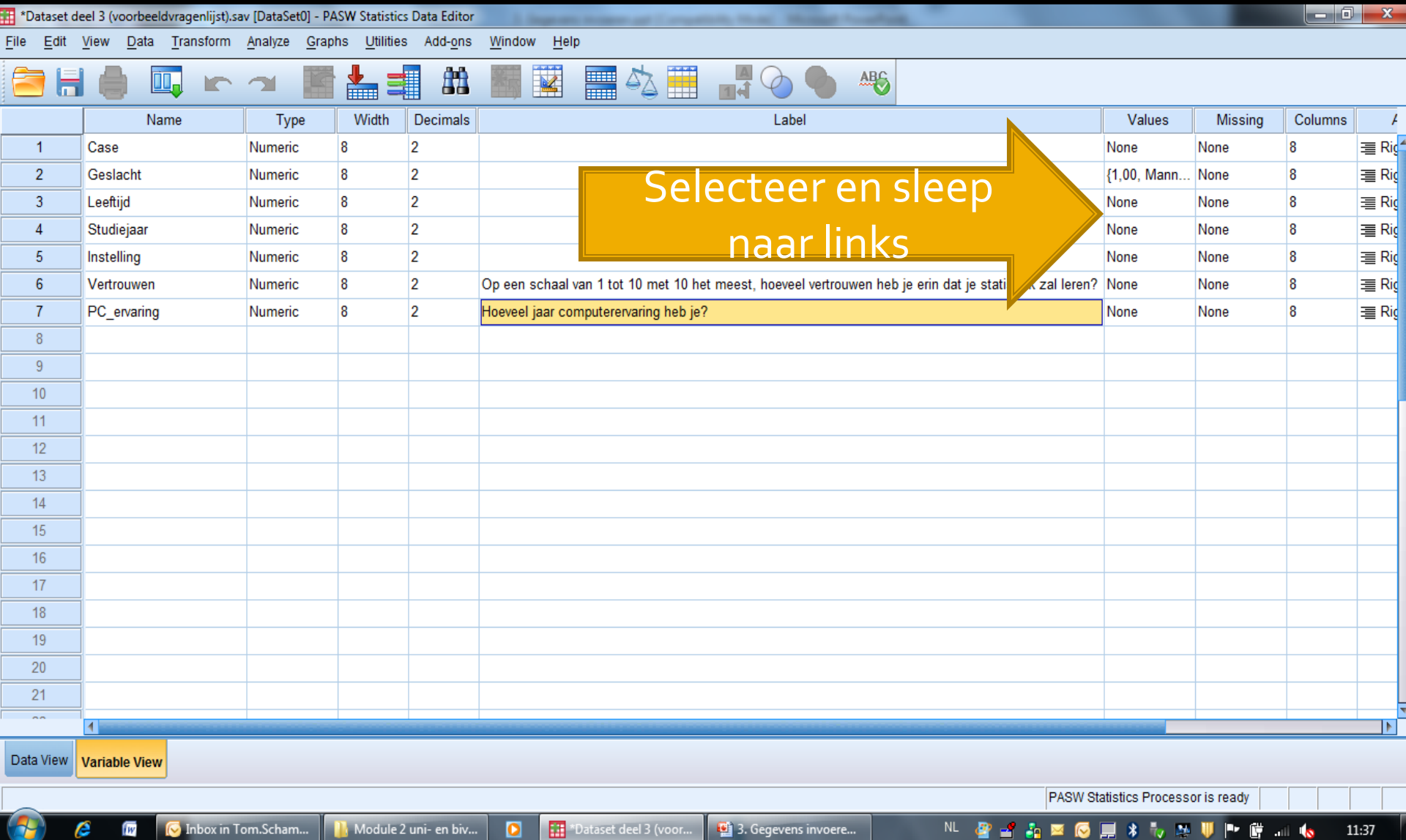

*Dataset deel 3 (voorbeeldvragenlijst).sav [DataSet0] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 Case Numeric 8 2 None None 8 Right Scale Input
2 Geslacht Numeric 8 2 None None 8 Right Scale Input
3 Leeftijd Numeric 8 2 None None 8 Right Scale Input
4 Studiejaar Numeric 8 2 None None 8 Right Scale Input
5 Instelling Numeric 8 2 None None 8 Right Scale Input
6 Vertrouwen Numeric 8 2 None None 8 Right Scale Input
7 PC_ervaring Numeric 8 2 None None 8 Right Scale Input
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Data View
Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham...
Module 2 uni- en biv...
*Dataset deel 3 (voor...
3. Gegevens invoere...
NL
11:41

Waarden definiëren

- Het antwoord op de eerste vraag is of "M" of "V". SPSS heeft een nummer nodig dat M aanduidt en één dat V aanduidt.
- Ga naar lijn 2 (Geslacht) en schuif naar de *Values* kolom, klik op de cel vervolgens op de 3 dots in grijs. Een *Value Labels* window komt te voorschijn. Geel "1" in de *Value* box en "Mannelijk" in de *Label* box. Selecteer: **Add.** Vervolgens typ "2" en "Vrouwelijk." Selecteer: **Add > OK.**

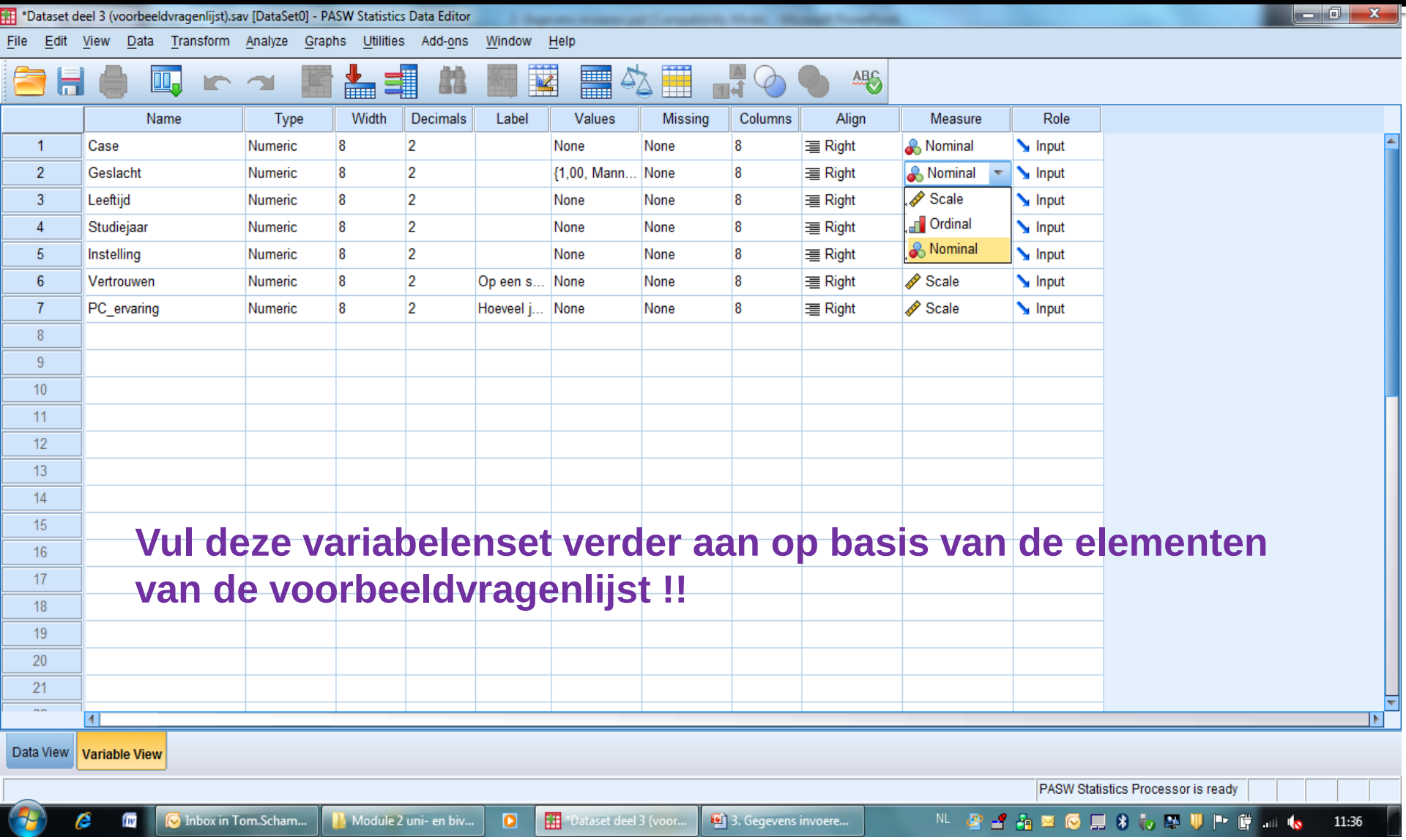

*Dataset deel 3 (voorbeeldvragenlijst).sav [DataSet0] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 Case Numeric 8 2 None None 8 Right Scale Input
2 Geslacht Numeric 8 2 {1,00, Mann... None 8 Right Scale Input
3 Leeftijd Numeric 8 2 None None 8 Right Scale Input
4 Studiejaar Numeric 8 2 None None 8 Right Scale Input
5 Instelling Numeric 8 2 Input
6 Vertrouwen Numeric 8 2 Input
7 PC_ervaring Numeric 8 2 Input
Value Labels
Value Labels
Value:
Label:
Spelling...
Add
Change
Remove
1,00 = "Mannelijk"
2,00 = "Vrouwelijk"
OK Cancel Help
Data View Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham... Module 2 uni- en biv... *Dataset deel 3 (voor... 3. Gegevens invoere...
NL 11:39

- Informatie die uitleg geeft bij de variabele schrijf je in de *Label* column.
- Ga naar lijn 6 (Vertrouwen) en schuif naar de *Label* kolom en typ: "Op een schaal van 1 tot 10 met 1 als het minst en 10 het meest, hoeveel vertrouwen heb je erin dat je zal slagen voor de test?"
- Ga naar de *Label* kolom van lijn 7 (PC_ervaring) en typ: "Hoeveel jaar computerervaring heb je?"


*Dataset deel 3 (voorbeeldvragenlijst).sav [DataSet0] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns
1 Case Numeric 8 2 None None 8
2 Geslacht Numeric 8 2 {1,00, Mann... None 8
3 Leeftijd Numeric 8 2 None None 8
4 Studiejaar Numeric 8 2 None None 8
5 Instelling Numeric 8 2 None None 8
6 Vertrouwen Numeric 8 2 Op een schaal van 1 tot 10 met 10 het meest, hoeveel vertrouwen heb je erin dat je stati zal leren? None None 8
7 PC_ervaring Numeric 8 2 Hoeveel jaar computerervaring heb je? None None 8
Selecteer en sleep naar links
Data View
Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham...
Module 2 uni- en biv...
*Dataset deel 3 (voor...
3. Gegevens invoere...
NL
11:37

- In de *Measure* kolom geef je het meetniveau van de variabele aan. Mogelijkheden zijn: **Nominal, Ordinal, or Scale** (Interval or Ratio).
  - Ga naar lijn Case en klik op de *Measure* kolom. Klik op de neerwaartse pijl en selecteer: **Nominal**.
  - Ga naar lijn Geslacht , selecteer *Measure* : **Nominal**.
  - Laat alle andere lijnen voorlopig bij de default value (Scale).

Opdracht


*Dataset deel 3 (voorbeeldvragenlijst).sav [DataSet0] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 Case Numeric 8 2 None None 8 Right Nominal Input
2 Geslacht Numeric 8 2 {1,00, Mann... None 8 Right Nominal Input
3 Leeftijd Numeric 8 2 None None 8 Right Scale Input
4 Studiejaar Numeric 8 2 None None 8 Right Ordinal Input
5 Instelling Numeric 8 2 None None 8 Right Nominal Input
6 Vertrouwen Numeric 8 2 Op een s... None None 8 Right Scale Input
7 PC_ervaring Numeric 8 2 Hoeveel j... None None 8 Right Scale Input
Vul deze variabelenset verder aan op basis van de elementen van de voorbeeldvragenlijst !!
Data View
Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham...
Module 2 uni- en biv...
*Dataset deel 3 (voor...
3. Gegevens invoere...
NL
11:36

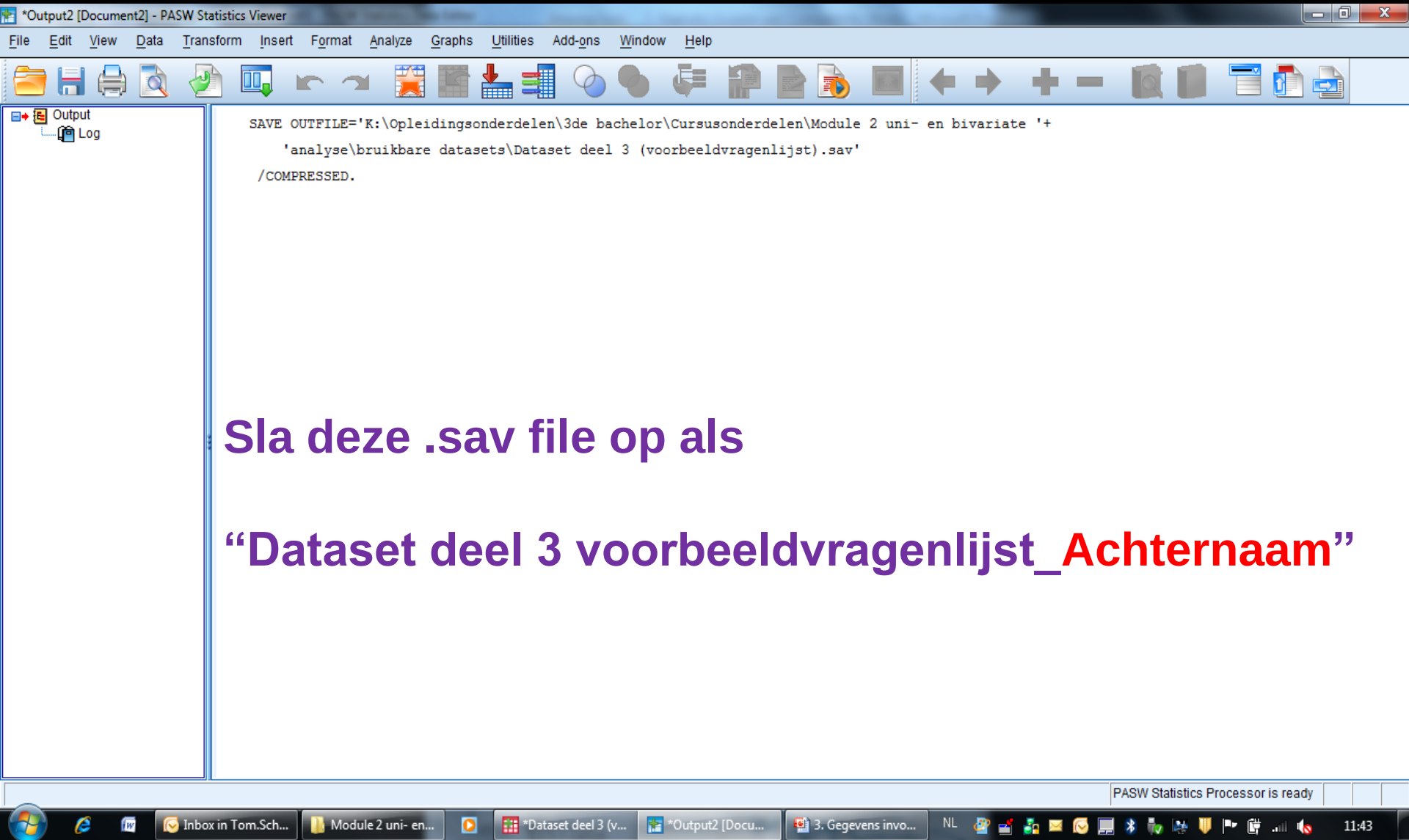
Variabelenlijst opslaan

- Om je data-file te saven ga je te werk net zoals bij alle andere Windows toepassingen. Kies achtereenvolgens in de menustructuur: **File, Save as**. De 'Save as' dialoogkader wordt geopend. In dit venster *geef je het bestand een naam* en *definieer op welke locatie je de file wil opsl*aan, je duidt het type van je file aan (in spss is dit .sav) en klikt op **OK**.

- Hou er wel rekening mee dat je data-editor venster actief moet zijn.

- Indien je reeds statistische analyses hebt gemaakt en je output-venster is actief dan zal je je output file opslaan en niet je data-file.

Sla deze aanpassingen op



Twee opties

- Maak een dataset aan
- Laadt een bestaande dataset op

Bestaande dataset selecteren

- In ons geval: data van 'Politieke besluitvorming' (survey) in de vorm van een excel-bestand geëxporteed via *Surveymonkey*
- Demonstratie
- Vragenlijst via Minerva beschikbaar gesteld (+ codeboek voor de selectie van 20 vragen)

Opdracht 1

- Definieer de volgende 20 variabelen van de dataset 'Politieke besluitvorming':

Naam variabele	Antwoordcategorieën	Waar in survey
VAR1 geslacht (m/v),	M(1)/V(2)	D2, VR1
VAR2 postcode domicilie	(getal)	D2, VR2
VAR3 postcode verblijfplaats	(getal)	D2, VR2
VAR4 geboortejaar	(idem) Opgelet: keuze 28 = 1987, 29 = 1988 enz.	D2, VR3
VAR5 partijkeuze (KvV 2010)	partijen oplopend van 1-11 (overige)	D3, VR4
VAR 6 uitgenodigd volksraadpleging (vrp)	J(1)/N(2)/WN(3)	D4, VR1
VAR7 deelname vrp overwegen	J(1)/N(2)/GM(3)	D5, VR2
VAR8 belangrijkheid vrp	eb(1)/b(2)/mb(3)/hnb(4)/WN(5)	D5, VR5
VAR9 welke machtigst: kabinetten of administratie	vk(1)/ek(2)/em(3)/ea(4)/va(5)/WN(6)/nvt(7)	D7, VR1

Naam variabele	Antwoordcategorieën	Waar in codeboek
VAR10 houding tav. beheersing v. cumul	a(1)/b(2)/c(3)/d(4)/GM(5)	D8, VR1
VAR11 houding tav. beheersing v. schijnkadidaturen	a(1)/b(2)/c(3)/d(4)/GM(5)	D8, VR1
VAR12 wetgevend initiatief mbt. verhindering?	js(1)/jc(2)/jb(3)/n(4)/GM(5)	D8, VR2
VAR13 welk initiatief?	a(1)/b(2)/c(3)/d(4)/e(5)GM(6)	D9, VR1
VAR14 sanctioneren?	J(1)/N(2)/GM(3)	D9, VR2
VAR15 welke sanctie	a(1)/b(2)/c(3)/d(4)/e(5)	D10, VR1
VAR16 doeltreffendheid instrumenten beleidsbeïnvloeding (enkel lobbyen bij kabinetten)	hd(1)/d(2)/md(3)/wd(4)/hwd(5)/WN(6)	D11, VR1
VAR17 welke organisatie heeft meest invloed	organisaties oplopend van 1-12 (andere)	D11, VR2
VAR18 politici en gebruik van de media	tv(1)/g(2)/tw(3)/GM(4)	D13, VR1
VAR19 media en invloed op de politieke besluitvorming_klassieke	zni(1)/eni(2)/gn(3)/epi(4)/zpi(5)/WN(6)	D13, VR2
VAR20 media en invloed op de politieke besluitvorming_nieuwe	zni(1)/eni(2)/gn(3)/epi(4)/zpi(5)/WN(6)	D13, VR2

Opgelet voor foute definiëring van variabelen

- Bijvoorbeeld vastgesteld via FREQUENCIES

Bent u man of vrouw?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	VROUW	31	50,8	50,8	50,8
	2	30	49,2	49,2	100,0
	Total	61	100,0	100,0	

- Via het Codebook

Bentumanofvrouw

		Value	Count	Percent
Standard Attributes	Position	2		
	Label	Bent u man of vrouw?		
	Type	Numeric		
	Format	F11		
	Measurement	Nominal		
	Role	Input		
Valid Values	0	MAN	0	,0%
	1	VROUW	31	50,8%
	2		30	49,2%

Indiening opdracht 1

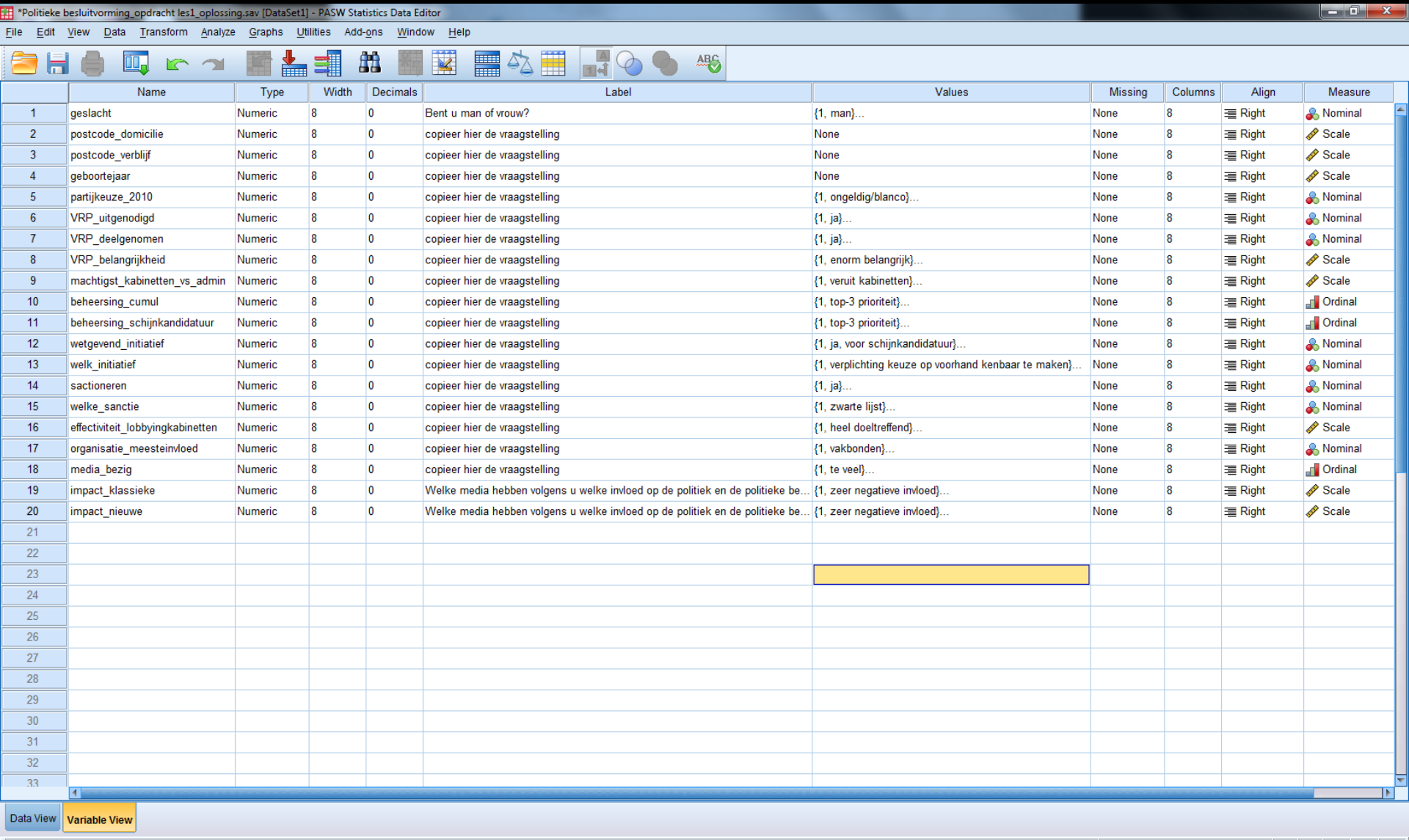
Als de volledige definiëring van alle VAR klaar is, sla deze file op als :

"Dataset politieke besluitvorming_Achternaam"

Dien het Codeboek in tegen ten laatste wo 6 april, 12:00pm via mail tom.schamp@ugent.be

Via: **Analyse** > **Reports** > **Codebook**

Oplossing



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	geslacht	Numeric	8	0	Bent u man of vrouw?	{1, man}...	None	8	Right	Nominal
2	postcode_domicilie	Numeric	8	0	copieer hier de vraagstelling	None	None	8	Right	Scale
3	postcode_verblijf	Numeric	8	0	copieer hier de vraagstelling	None	None	8	Right	Scale
4	geboortejaar	Numeric	8	0	copieer hier de vraagstelling	None	None	8	Right	Scale
5	partijkeuze_2010	Numeric	8	0	copieer hier de vraagstelling	{1, ongeldig/blanco}...	None	8	Right	Nominal
6	VRP_uitgenodigd	Numeric	8	0	copieer hier de vraagstelling	{1, ja}...	None	8	Right	Nominal
7	VRP_deelgenomen	Numeric	8	0	copieer hier de vraagstelling	{1, ja}...	None	8	Right	Nominal
8	VRP_belangrijkheid	Numeric	8	0	copieer hier de vraagstelling	{1, enorm belangrijk}...	None	8	Right	Scale
9	machtigst_kabinetten_vs_admin	Numeric	8	0	copieer hier de vraagstelling	{1, veruit kabinetten}...	None	8	Right	Scale
10	beheersing_cumul	Numeric	8	0	copieer hier de vraagstelling	{1, top-3 prioriteit}...	None	8	Right	Ordinal
11	beheersing_schijnkandidatuur	Numeric	8	0	copieer hier de vraagstelling	{1, top-3 prioriteit}...	None	8	Right	Ordinal
12	wetgevend_initiatief	Numeric	8	0	copieer hier de vraagstelling	{1, ja, voor schijnkandidatuur}...	None	8	Right	Nominal
13	welk_initiatief	Numeric	8	0	copieer hier de vraagstelling	{1, verplichting keuze op voorhand kenbaar te maken}...	None	8	Right	Nominal
14	sactioneren	Numeric	8	0	copieer hier de vraagstelling	{1, ja}...	None	8	Right	Nominal
15	welke_sanctie	Numeric	8	0	copieer hier de vraagstelling	{1, zwarte lijst}...	None	8	Right	Nominal
16	effectiviteit_lobbyingkabinetten	Numeric	8	0	copieer hier de vraagstelling	{1, heel doeltreffend}...	None	8	Right	Scale
17	organisatie_meesteinvloed	Numeric	8	0	copieer hier de vraagstelling	{1, vakbonden}...	None	8	Right	Nominal
18	media_bezig	Numeric	8	0	copieer hier de vraagstelling	{1, te veel}...	None	8	Right	Ordinal
19	impact_klassieke	Numeric	8	0	Welke media hebben volgens u welke invloed op de politiek en de politieke be...	{1, zeer negatieve invloed}...	None	8	Right	Scale
20	impact_nieuwe	Numeric	8	0	Welke media hebben volgens u welke invloed op de politiek en de politieke be...	{1, zeer negatieve invloed}...	None	8	Right	Scale
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										

Deel 4

Invoeren van data

Data coderen en het codeboek

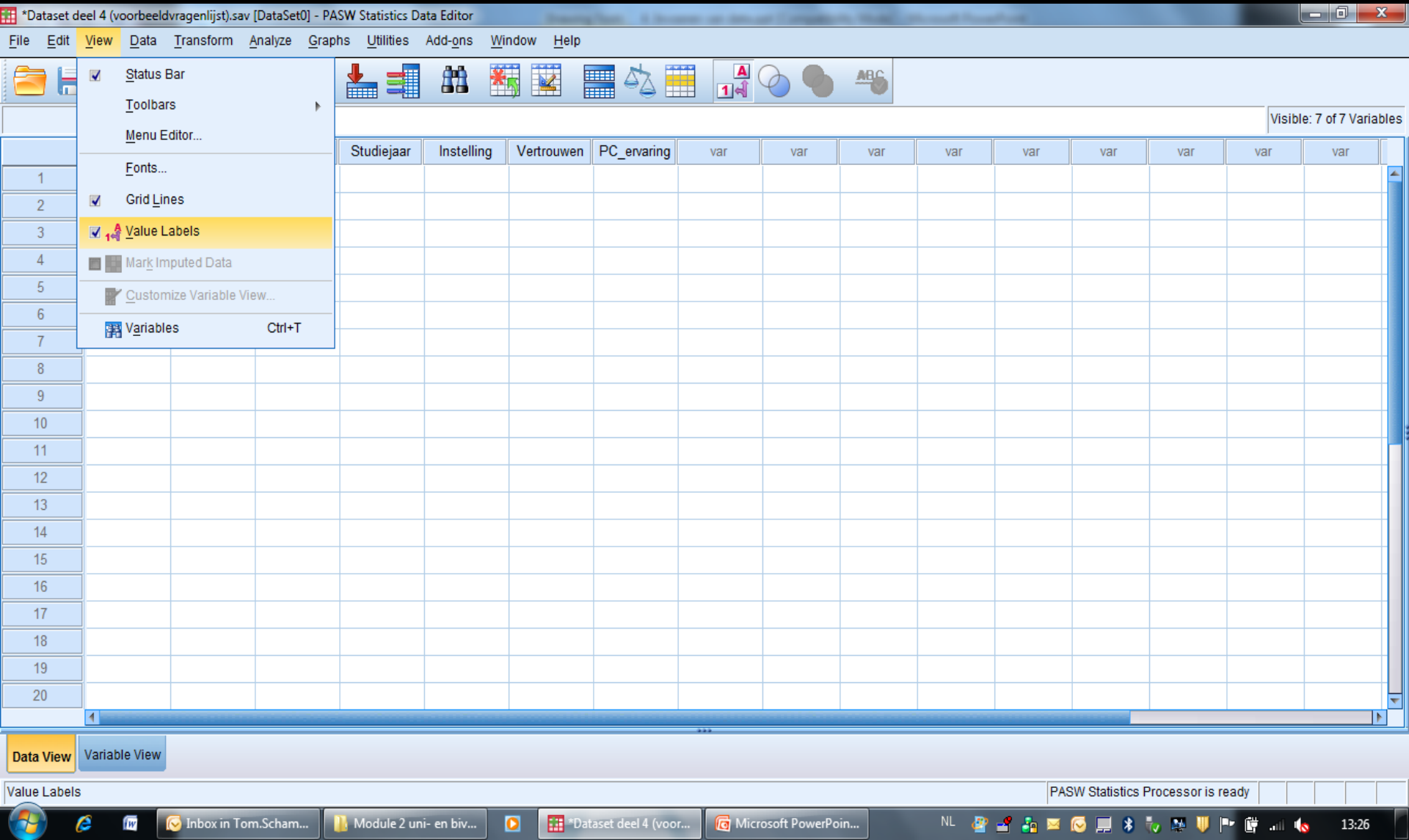
- **Data-invoer**: verwijst naar het maken van een computerbestand met de ruwe data van alle vragenlijsten die geschikt worden geacht voor analyse.

- **Data coderen**: codes toekennen aan de mogelijke antwoorden op elke vraag in de vragenlijst.

- **Codeboek**: boek waarin alle namen van variabelen zijn opgenomen en de codes die aan elk mogelijk antwoord op elke vraag zijn toegekend.

Data invoeren

- Switch van Variable view naar *Data View*.
- Selecteer in de hoek linksboven het scherm **View** en, als *Value Labels* niet aangevinkt is, selecteer **Value Labels.**


*Dataset deel 4 (voorbeeldvragenlijst).sav [DataSet0] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Status Bar
Toolbars
Menu Editor...
Fonts...
Grid Lines
Value Labels
Mark Imputed Data
Customize Variable View...
Variables Ctrl+T
Visible: 7 of 7 Variables
Studiejaar
Instelling
Vertrouwen
PC_ervaring
var
Data View
Variable View
Value Labels
PASW Statistics Processor is ready
Inbox in Tom.Scham...
Module 2 uni- en biv...
*Dataset deel 4 (voor...
Microsoft PowerPoin...
NL
13:26

Antwoorden (Participant #1)

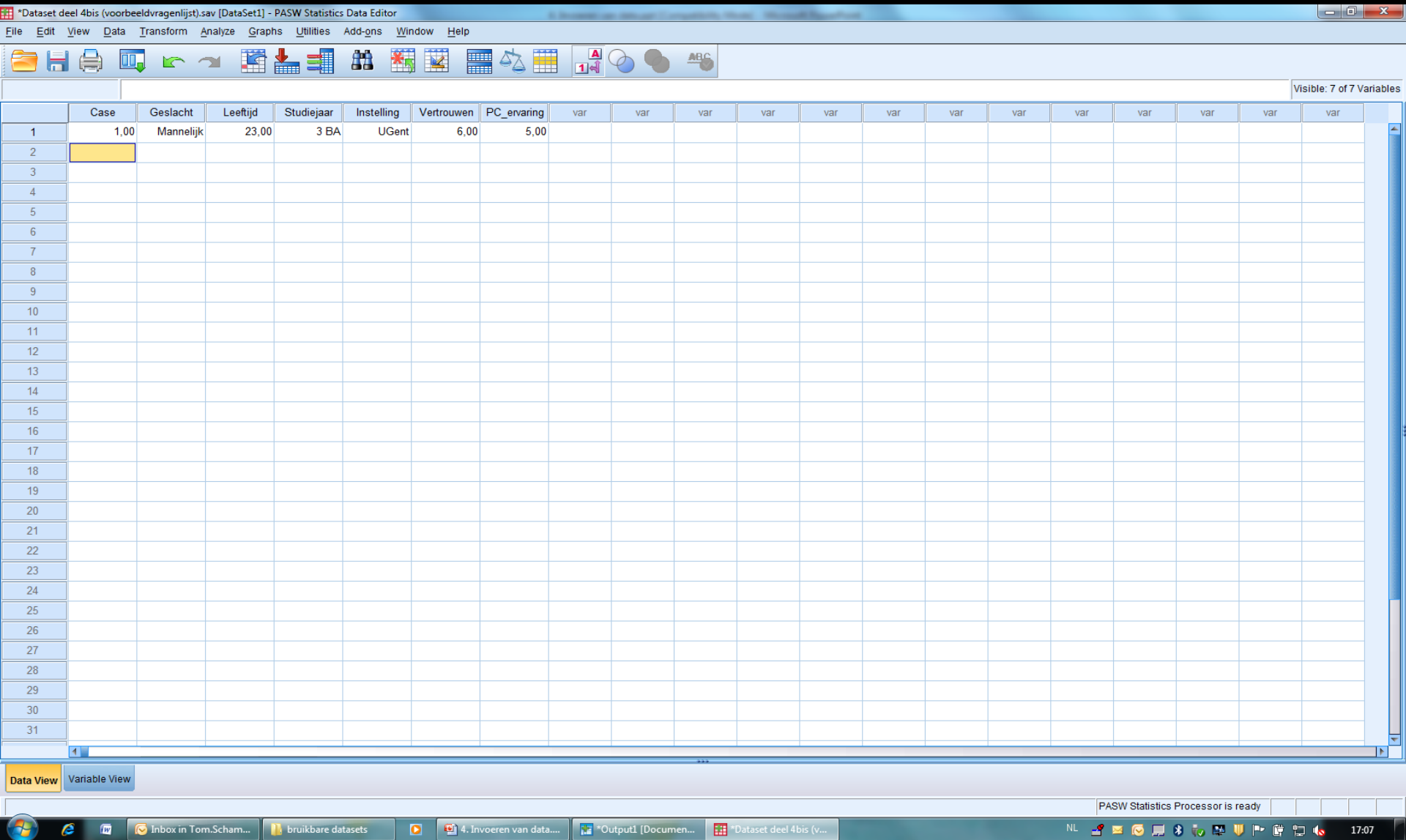
Deelnemer #: _1_
Geslacht: M F
Leeftijd: _23__
Huidig studiejaar: _3 BA__
Onderwijsinstelling: _UGent__
Op een schaal van 1 tot 10 hoe zeker ben je dat je na deze module zal slagen voor de test?

1 2 3 4 5 6 7 8 9 10

Helemaal niet Heel zeker

Jaren PC-ervaring? _5__

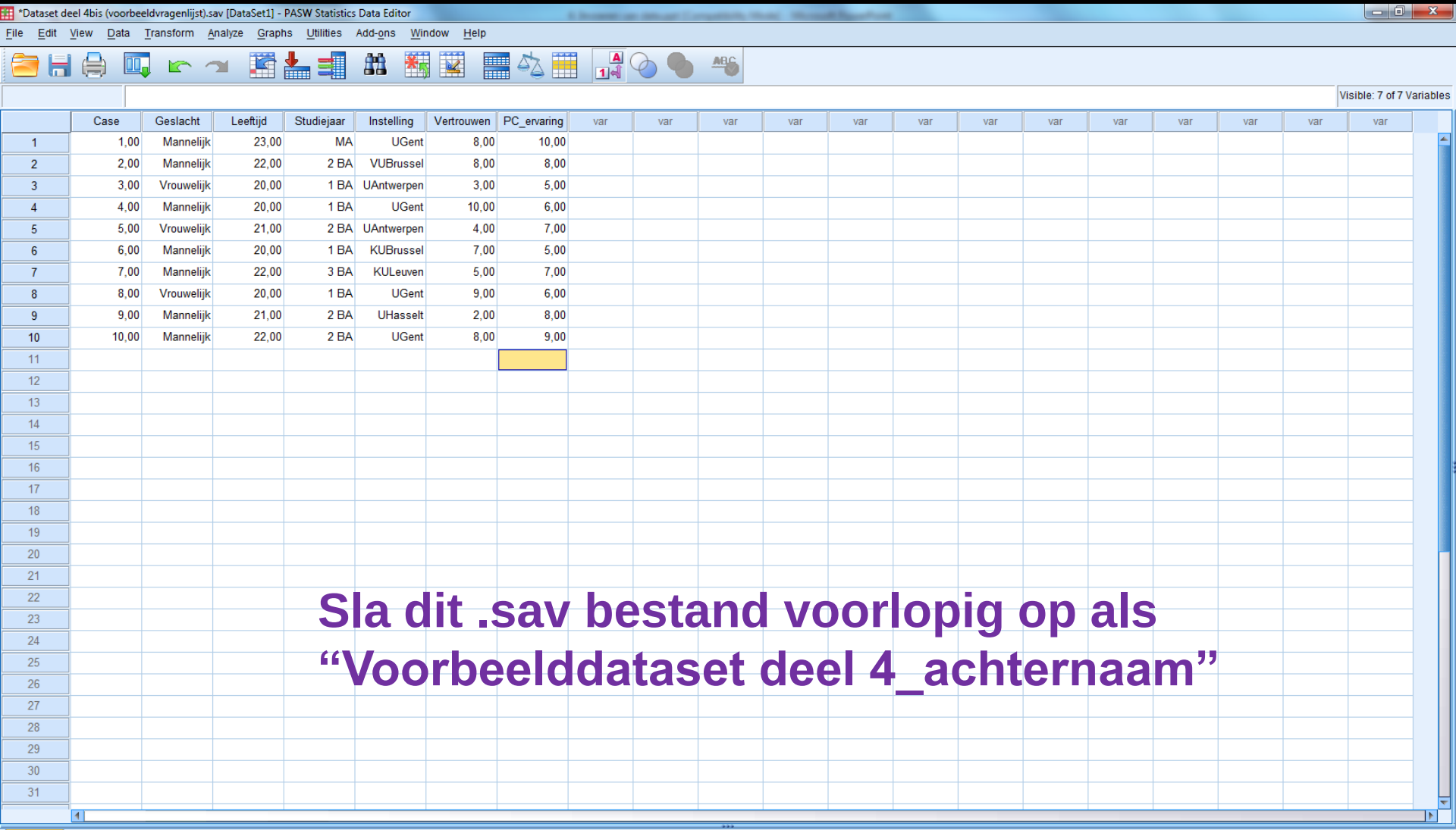
- Typ “1” in de eerste lijn van de *Case* kolom.
- Op naar de *Geslacht* kolom en klik op de neerwaartse cursor. Selecteer **Mannelijk**
- Ga naar *Leeftijd* en typ 23
- Ga naar *Studiejaar* en selecteer keuzemogelijkheid 3
- Ga naar *Instelling* en selecteer keuze 1
- Ga naar *Vertrouwen* en typ 6
- Ga naar *PC_ervaring* en typ 5


*Dataset deel 4bis (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 7 of 7 Variables
Case Geslacht Leeftijd Studiejaar Instelling Vertrouwen PC_ervaring var
1 1,00 Mannelijk 23,00 3 BA UGent 6,00 5,00
Data View
Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham...
bruikbare datasets
4. Invoeren van data....
*Output1 [Documen...
*Dataset deel 4bis (v...
NL
17:07

Sla deze file op

- Wanneer je met datafiles werkt, sla je best vaak op.
- Ga naar **File** > **Save As** > (kies een locatie)
- Typ een *File Name*.
- Selecteer: **Save**

Dataset deel 4



	Case	Geslacht	Leeftijd	Studiejaar	Instelling	Vertrouwen	PC_ervaring	var	var	var	var	var	var	var	var	var	var	var	var
1	1,00	Mannelijk	23,00	MA	UGent	8,00	10,00												
2	2,00	Mannelijk	22,00	2 BA	VUBrussel	8,00	8,00												
3	3,00	Vrouwelijk	20,00	1 BA	UAntwerpen	3,00	5,00												
4	4,00	Mannelijk	20,00	1 BA	UGent	10,00	6,00												
5	5,00	Vrouwelijk	21,00	2 BA	UAntwerpen	4,00	7,00												
6	6,00	Mannelijk	20,00	1 BA	KUBrussel	7,00	5,00												
7	7,00	Mannelijk	22,00	3 BA	KULeuven	5,00	7,00												
8	8,00	Vrouwelijk	20,00	1 BA	UGent	9,00	6,00												
9	9,00	Mannelijk	21,00	2 BA	UHasselt	2,00	8,00												
10	10,00	Mannelijk	22,00	2 BA	UGent	8,00	9,00												
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			
23																			
24																			
25																			
26																			
27																			
28																			
29																			
30																			
31																			

Sla dit .sav bestand voorlopig op als “Voorbeelddataset deel 4_achternaam”

Codeboek aanmaken

- Ga naar **Analyse** > **Reports** > **Codebook**

Sla dit codeboek op als
“Codeboek voorbeelddataset deel 4_achternaam”

Het codeboek

*Output1 [Document1] - PASW Statistics Viewer

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

Output
- Codebook
  - Title
  - Notes
  - Active Dataset
  - Case
  - Geslacht
  - Leeftijd
  - Studiejaar
  - Instelling
  - Vertrouwen
  - PC_ervaring

Codebook

[DataSet1] K:\Opleidingsonderdelen\3de bachelor\PWM\Module 2 uni- en bivariate analyse\bruikbare datasets\Bigham (stap voor stap)\Dataset deel 4bis (voorbeeldvragenlijst).sav

Case

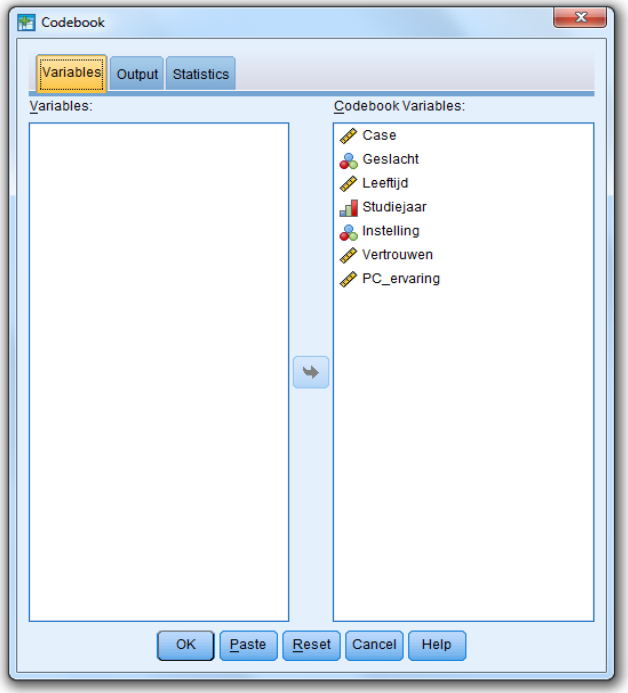
		Value
Standard Attributes	Position	1
	Label	<none>
	Type	Numeric
	Format	F8.2
	Measurement	Scale
	Role	Input
N	Valid	10
	Missing	0
Central Tendency and Dispersion	Mean	5,5000
	Standard Deviation	3,02765
	Percentile 25	3,0000
	Percentile 50	5,5000
	Percentile 75	8,0000

Geslacht

		Value	Count	Percent
Standard Attributes	Position	2		
	Label	<none>		
	Type	Numeric		
	Format	F8.2		
	Measurement	Nominal		
	Role	Input		
Valid Values	1,00	Mannelijk	6	60,0%
	2,00	Vrouwelijk	3	30,0%
	7,00		1	10,0%

Leeftijd

		Value
Standard Attributes	Position	3
	Label	<none>



PASW Statistics Processor is ready

Inbox in Tom.Scham... | Bigham (stap voor st... | Microsoft PowerPoin... | *Output1 [Documen... | *Dataset deel 4bis (v... | NL | 14:10

Data controleren

- Eens je codeboek is opgesteld, kan je beginnen met **het coderen van alle ingevulde vragenlijsten én nadien het controleren van de dataset**.

- **Tijdens deze fase wordt de datamatrix gecontroleerd op mogelijke typ- en/of codeerfouten.**

Data controleren

- Om tikfouten op het spoor te komen zijn er twee mogelijke strategieën.
  - Deze abnormaliteiten kunnen twee zaken zijn:
    - onmogelijke waarden
    - onmogelijke combinaties van waarden (bv. doorverwijsvragen)
  - Bij de eerste mogelijkheid neemt men weer alle ingevulde vragenlijsten in de hand en controleert men de volledig ingetikte datamatrix met de oorspronkelijke gegevens.
    - Het nadeel van deze werkwijze is dat ze zeer lang duurt
    - Het is maar de vraag of het zinvol is om zoveel energie te steken in deze methode
    - Opnieuw mogelijkheden om fouten te maken

- Voorbeeld: Stel je typt de gegevens van 300 enquêtes in. Bij het opvragen van je frequentietabel krijg je voor **geslacht** een onmogelijke waarde (bvb. "**7**"). Je hebt dus duidelijk een typfout gemaakt. Wat nu? Indien je een identificatienummer hebt aangebracht kan je deze vragenlijst/case opnieuw opzoeken en je fout verbeteren. Zonder identificatienummer is deze case niet meer correct. Je kan nu wel willekeurig kiezen of je "7" een 0, 1 of 9 wordt maar dit is uiteraard een te vermijden praktijk.

Data controleren

- Daarom wordt in de praktijk meestal de andere methode gehanteerd waarbij men probeert de fouten te vinden door abnormaliteiten in de datamatrix op te sporen.

DEZE CHECK IS EEN ABSOLUTE MUST !!

Data controleren

- De eenvoudigste manier om deze onmogelijke waarden op te sporen is door **een frequentietabel** op te vragen.
  - **Analyze** > **Descriptive Statistics** > **Frequencies**.
  - Je krijgt dan het dialoogkader voor het opgeven van de frequentietabel. In het bovenstaande voorbeeld van onze korte vragenlijst krijg je het volgende dialoogvenster.

*Dataset deel 4bis (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

2 : Geslacht 7,00 Visible: 7 of 7 Variables

	Case	Geslacht	Leeftijd	Studiejaar	Instelling	Vertrouwen	PC_ervaring	var	var	var	var	var	var	var	var	var	var	var	var
1	1,00	Mannelijk	23,00	MA	UGent	8,00	10,00												
2	2,00	7,00	22,00	2 BA	VUBrussel	8,00	8,00												
3	3,00	Vrouwelijk	20,00	1 BA	UAntwerpen	3,00	5,00												
4	4,00	Mannelijk	20,00	1 BA	UGent	10,00	6,00												
5	5,00	Vrouwelijk	21,00	2 BA	UAntwerpen	4,00	7,00												
6	6,00	Mannelijk	20,00	1 BA	KUBrussel	7,00	5,00												
7	7,00	Mannelijk	22,00	3 BA	KULeuven	5,00	7,00												
8	8,00	Vrouwelijk	20,00	1 BA	UGent	9,00	6,00												
9	9,00	Mannelijk	21,00	2 BA	UHasselt	2,00													
10	10,00	Mannelijk	22,00	2 BA	UGent	8,00													
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			
23																			
24																			
25																			
26																			
27																			
28																			
29																			
30																			
31																			

Frequencies

Case
Leeftijd
Studiejaar
Instelling
Vertrouwen
PC_ervaring

Variable(s):
Geslacht

Statistics...
Charts...
Format...

Display frequency tables

OK Paste Reset Cancel Help

Data View Variable View

PASW Statistics Processor is ready

Inbox in Tom.Scham... Bigham (stap voor st... Microsoft PowerPoin... *Output1 [Documen... *Dataset deel 4bis (v...

NL 14:07

*Output1 [Document1] - PASW Statistics Viewer

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

- Output
  - Log
  - Frequencies
    - Title
    - Notes
    - Active Dataset
    - Statistics
    - Geslacht
  - Log

```
GET
  FILE='K:\Opleidingsonderdelen\3de bachelor\PWM\Module 2 uni- en bivariate analyse\bruikbare datasets\Bigham (stap voor stap)\Dataset deel 4bis (voorbeeldvragenlijst).sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
FREQUENCIES VARIABLES=Geslacht
  /ORDER=ANALYSIS.
```

Frequencies

[DataSet1] K:\Opleidingsonderdelen\3de bachelor\PWM\Module 2 uni- en bivariate analyse\bruikbare datasets\Bigham (stap voor stap)\Dataset deel 4bis (voorbeeldvragenlijst).sav

Statistics

Geslacht

N	Valid	10
	Missing	0

Geslacht

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mannelijk	6	60,0	60,0	60,0
	Vrouwelijk	3	30,0	30,0	90,0
	7,00	1	10,0	10,0	100,0
	Total	10	100,0	100,0	

```
FREQUENCIES VARIABLES=Geslacht
  /ORDER=ANALYSIS.
```

PASW Statistics Processor is ready

Inbox in Tom.Scham... | Bigham (stap voor st... | Microsoft PowerPoin... | *Output1 [Documen... | *Dataset deel 4bis (v...

NL 14:06

Deel 4 bis

Aanpassen van data

Data veranderen

- Het wijzigen van data in SPSS is afhankelijk van waarom je wijzigingen wil aanbrengen.
  - Wil je de waarde van 1 cel wijzigen? of
  - Moeten er een hercodering van variabelen gebeuren?

(1) Het veranderen van individuele waarden

Het wijzigen van een waarde uit één enkele cel is zeer eenvoudig. Je klikt op de cel die je wil wijzigen, typ de nieuwe waarde en druk op enter.

Data veranderen - 2

(2) Kopiëren en plakken

Het is mogelijk om een blok cellen te selecteren en te kopiëren. Net zoals bij Word of Excel maak je dan gebruik van de methode 'Copy' en 'Paste'.

- Seleceer het gebied dat je wil kopiëren, Edit, Copy, nadien selecteer het gebied waar je de nieuwe waarden wil plaatsen, Edit, Paste
- Seleceer het gebied dat je wil kopiëren, Ctrl + C, nadien selecteer het gebied waar je de nieuwe waarden wil plaatsen, Ctrl + V

Data veranderen - 3

(3) Het verwijderen van waarden

Om een waarden in een cel of in een blok van cellen te verwijderen:

- Om de waarde van één cel te verwijderen: selecteer de cel en druk op de knop Delete
- Voor het verwijderen van een volledige rij van waarden: klik op het grijze hokje dat het rij nummer bevat (links van de data editor window), hierdoor worden alle cellen op deze rij opgelicht, klik op delete.
- Voor het verwijderen van een volledige kolom: idem maar voor een geselecteerde kolom.

Data veranderen - 4

(4) Het hercoderen van variabelen

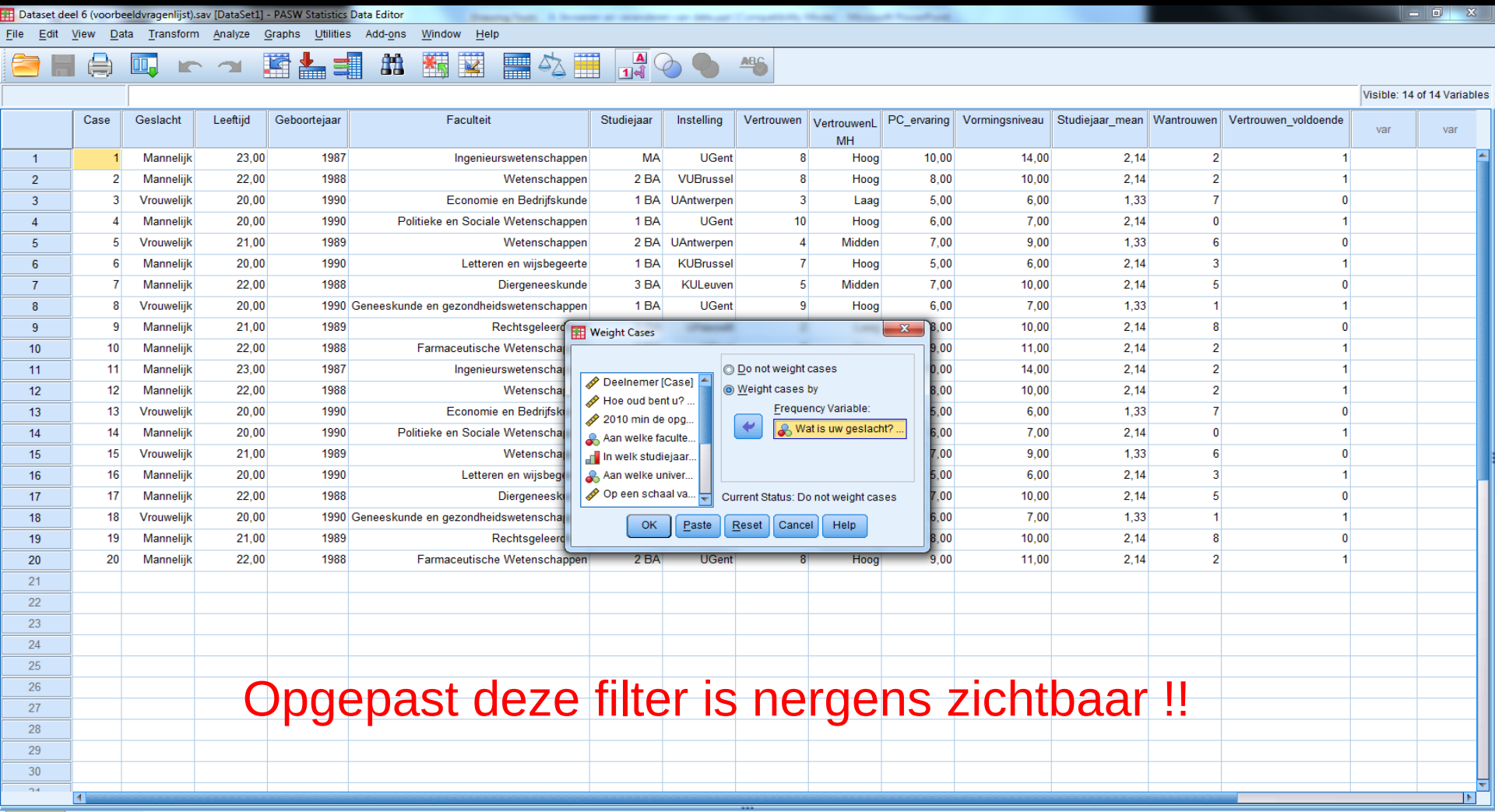
Kijk hiervoor naar de volgende presentatie: “Werken met variabelen”

Data wegen

- Indien er een onevenwicht in de dataset steekt, dat verifieerbaar en corrigeerbaar is
- Vaak op basis van leeftijd, geslacht, regio/afkomst etc… maar ook mogelijk voor inkomensklassen of andere klassen

- Via: **Data** > **Weight**

Weging van de variabelen obv 'Geslacht'



	Case	Geslacht	Leeftijd	Geboortejaar	Faculteit	Studiejaar	Instelling	Vertrouwen	VertrouwenLMH	PC_ervaring	Vormingsniveau	Studiejaar_mean	Wantrouwen	Vertrouwen_voldoende	var	var
1	1	Mannelijk	23,00	1987	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00	14,00	2,14	2	1		
2	2	Mannelijk	22,00	1988	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00	10,00	2,14	2	1		
3	3	Vrouwelijk	20,00	1990	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00	6,00	1,33	7	0		
4	4	Mannelijk	20,00	1990	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00	7,00	2,14	0	1		
5	5	Vrouwelijk	21,00	1989	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00	9,00	1,33	6	0		
6	6	Mannelijk	20,00	1990	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00	6,00	2,14	3	1		
7	7	Mannelijk	22,00	1988	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00	10,00	2,14	5	0		
8	8	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00	7,00	1,33	1	1		
9	9	Mannelijk	21,00	1989	Rechtsgeleerd					8,00	10,00	2,14	8	0		
10	10	Mannelijk	22,00	1988	Farmaceutische Wetenscha					9,00	11,00	2,14	2	1		
11	11	Mannelijk	23,00	1987	Ingenieurswetenscha					0,00	14,00	2,14	2	1		
12	12	Mannelijk	22,00	1988	Wetenscha					8,00	10,00	2,14	2	1		
13	13	Vrouwelijk	20,00	1990	Economie en Bedrijfsk					5,00	6,00	1,33	7	0		
14	14	Mannelijk	20,00	1990	Politieke en Sociale Wetenscha					6,00	7,00	2,14	0	1		
15	15	Vrouwelijk	21,00	1989	Wetenscha					7,00	9,00	1,33	6	0		
16	16	Mannelijk	20,00	1990	Letteren en wijsbeg					5,00	6,00	2,14	3	1		
17	17	Mannelijk	22,00	1988	Diergeneesk					7,00	10,00	2,14	5	0		
18	18	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenscha					6,00	7,00	1,33	1	1		
19	19	Mannelijk	21,00	1989	Rechtsgeleerd					8,00	10,00	2,14	8	0		
20	20	Mannelijk	22,00	1988	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00	11,00	2,14	2	1		

Opgepast deze filter is nergens zichtbaar !!

Weging op basis van geslacht

Ongewogen
30V en 70M

Hoe oud bent u?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20,00	8	40,0	40,0	40,0
	21,00	4	20,0	20,0	60,0
	22,00	6	30,0	30,0	90,0
	23,00	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Gewogen
M = V

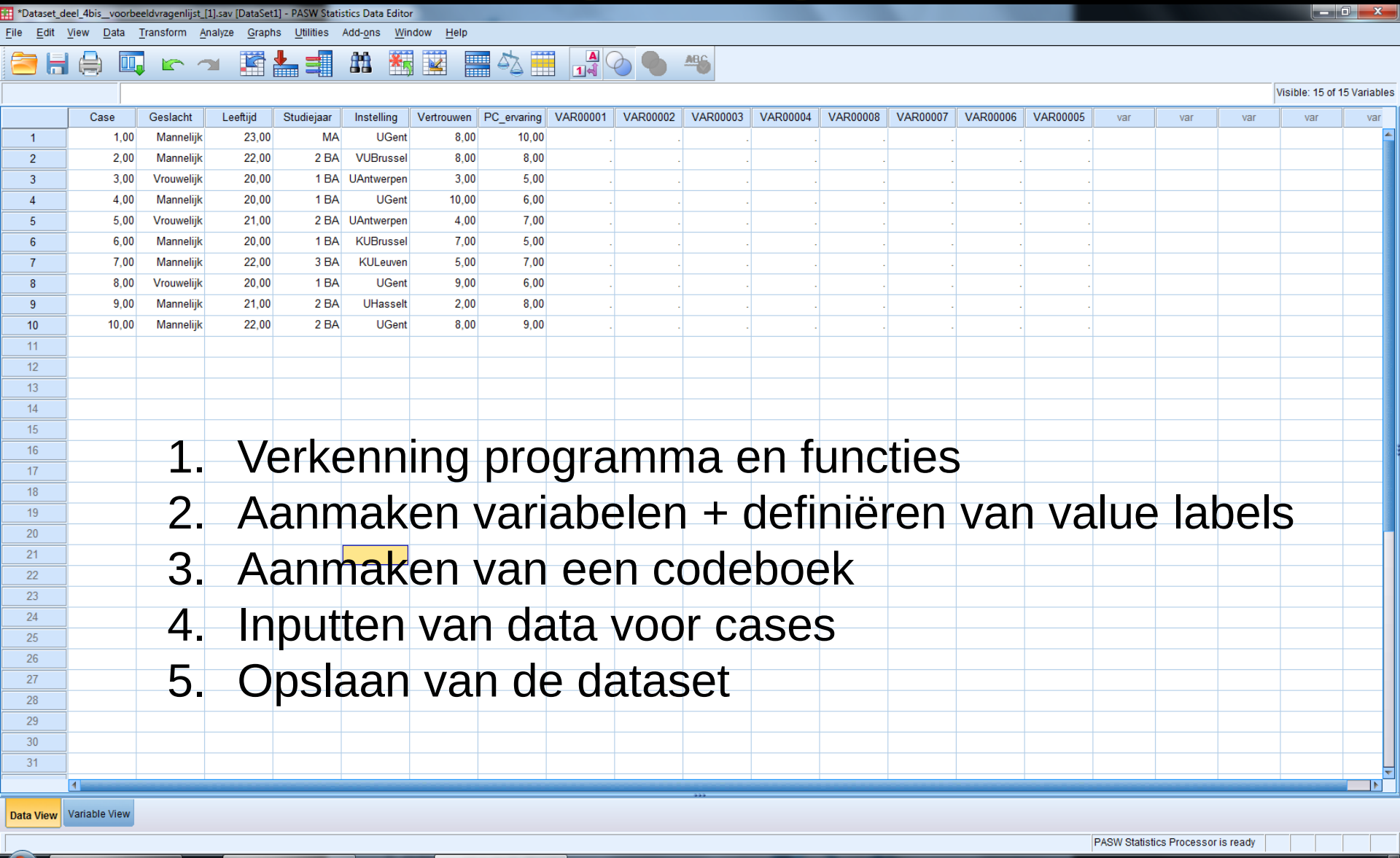
Hoe oud bent u?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20,00	12	46,2	46,2	46,2
	21,00	6	23,1	23,1	69,2
	22,00	6	23,1	23,1	92,3
	23,00	2	7,7	7,7	100,0
	Total	26	100,0	100,0	

Deel 5

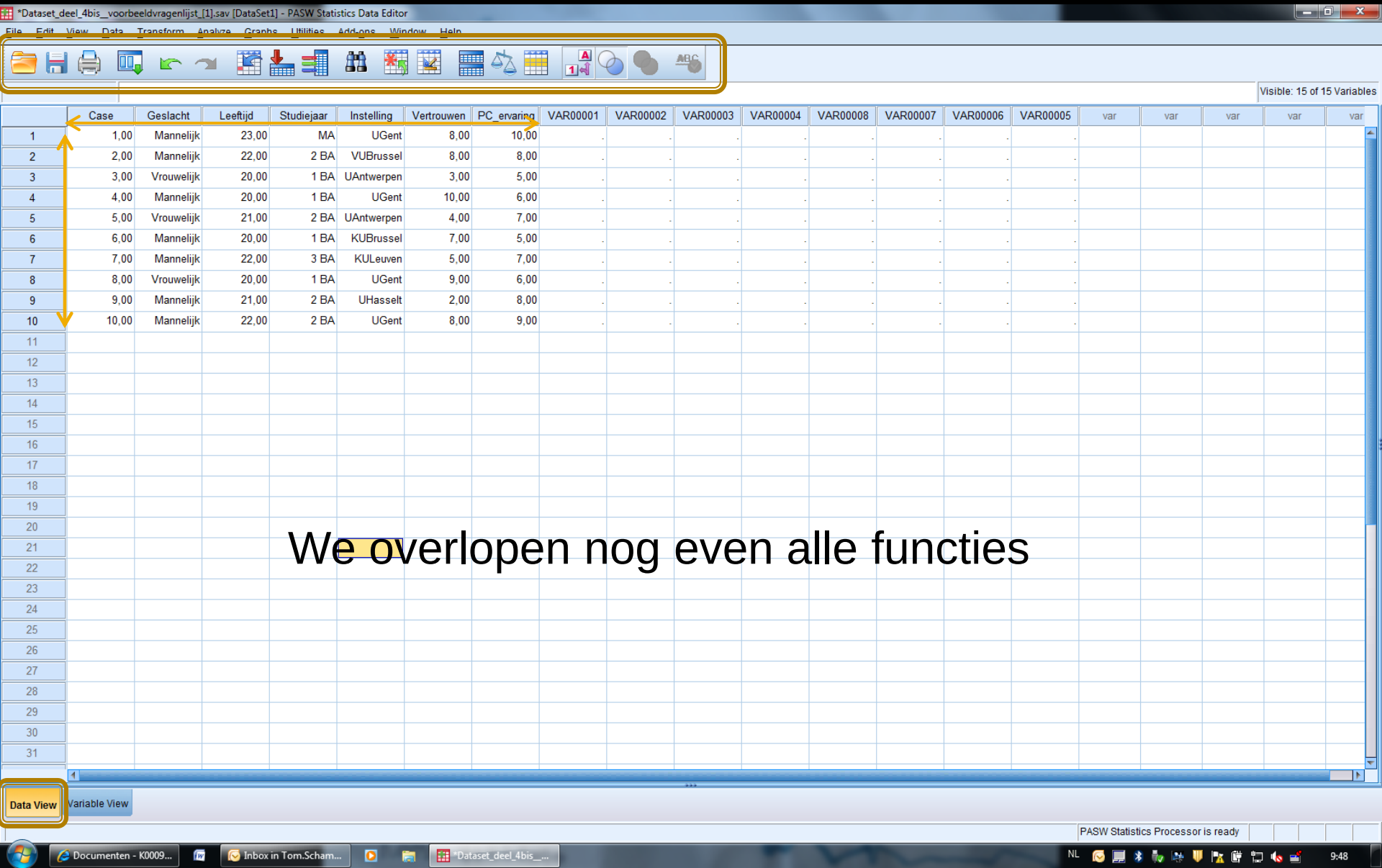
Werken met variabelen

Samenvatting vorige les



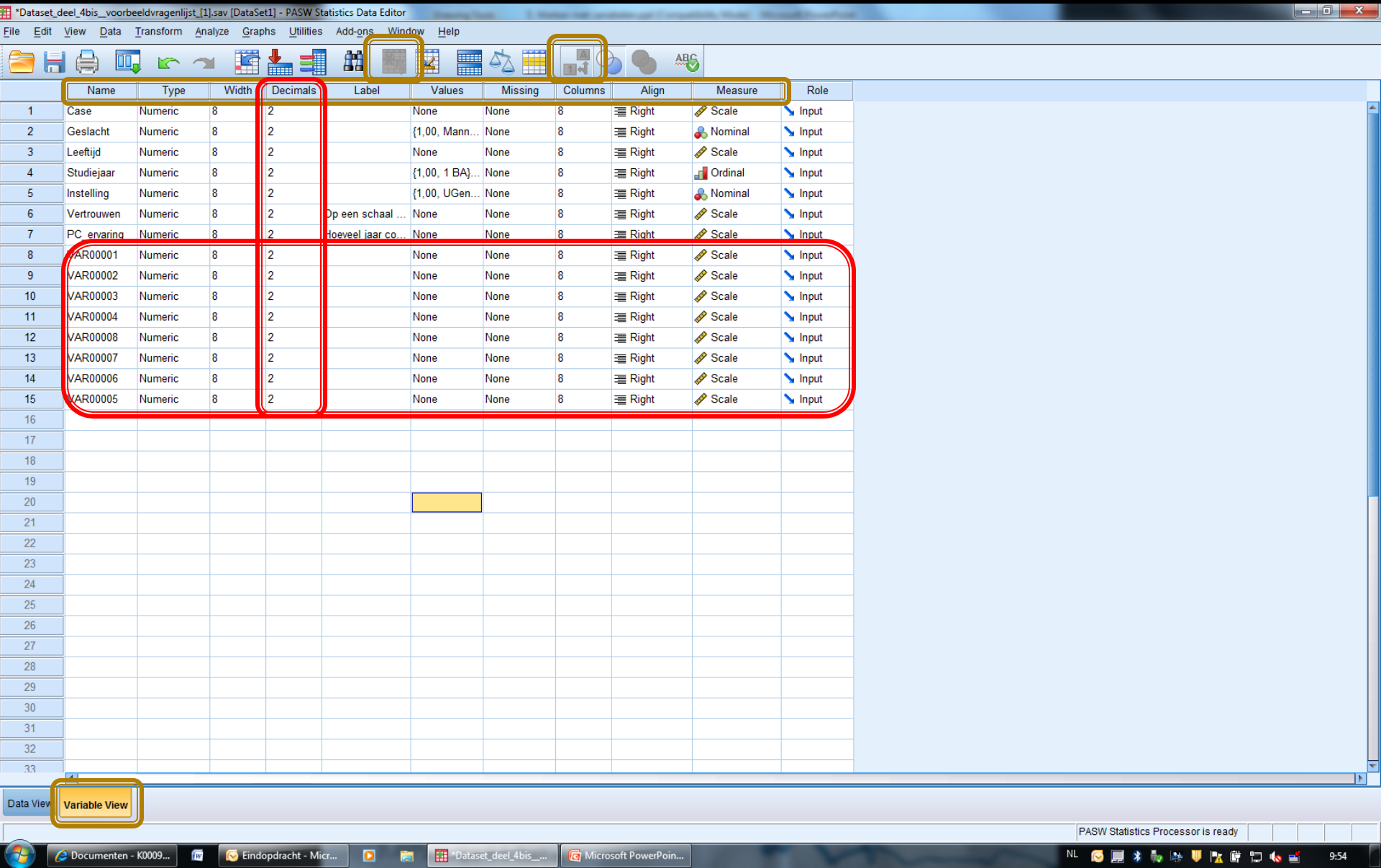
1. Verkenning programma en functies
2. Aanmaken variabelen + definiëren van value labels
3. Aanmaken van een codeboek
4. Inputten van data voor cases
5. Opslaan van de dataset

Data venster / data view



We overlopen nog even alle functies

Variabele venster / variable view



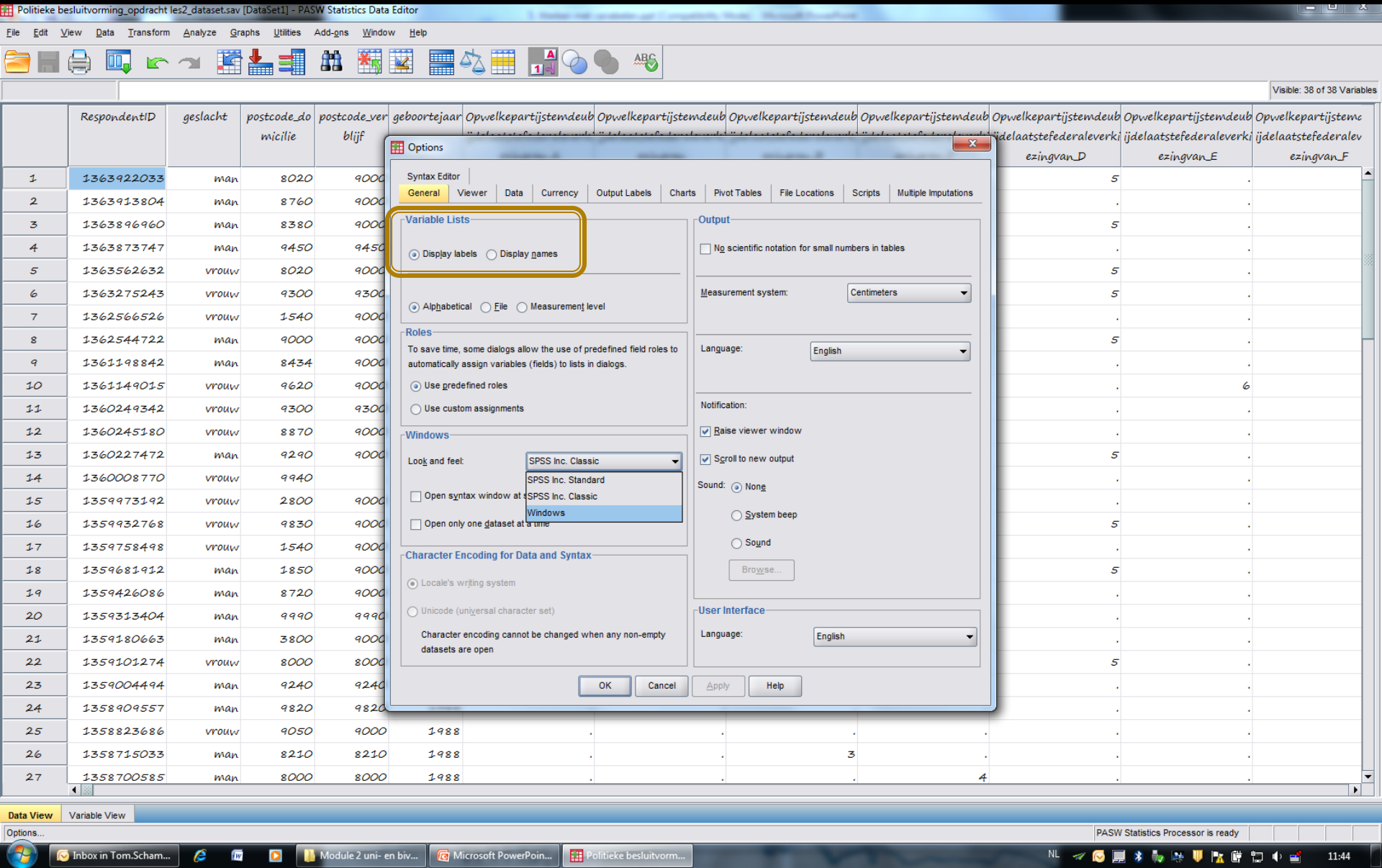
Andere vensters

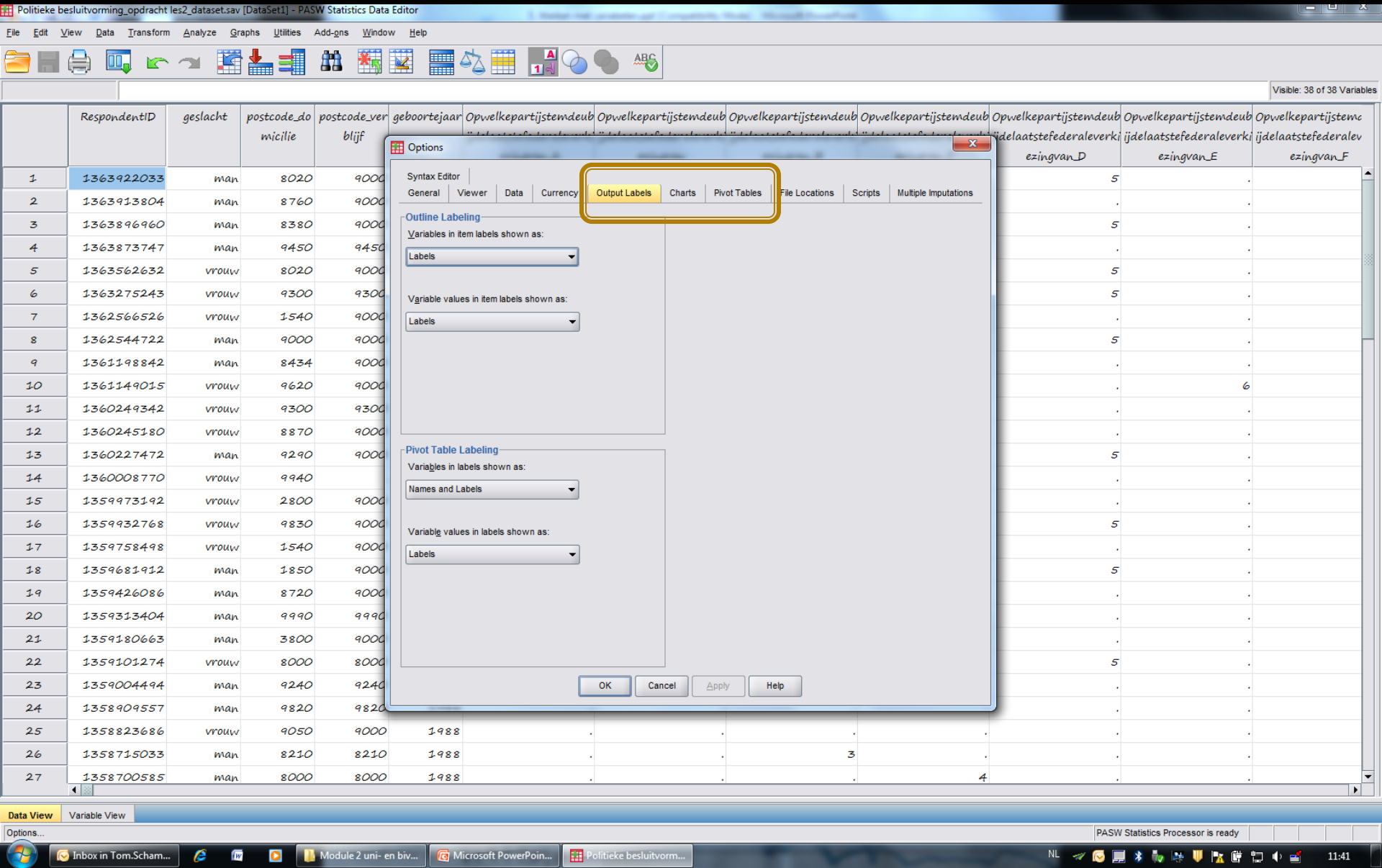
- Het syntaxvenster
  - **File** > **Open** > **Syntax** (bestaande syntax oproepen)
  - **File** > **New** > **Syntax** (nieuwe syntax aanmaken of via copy-paste)
- Het outputvenster
- Dialoogvenster (bij elke operatie: commando, etc.)

SPSS naar je hand zetten

- Via **Edit** > **Options**
  - General (vorm en stijl tekst, lijnen, markers, grafieken, tabellen…)
  - Output labels (of "variabele" naam of "label" naam)
  - Syntax editor (kleurenschema)

- Ontbrekende waarden toevoegen
  - 6 / 66 / 666… ("niet van toepassing")
  - 7 / 77 / 777… (weet niet)
  - 8 / 88 / 888… (geen mening / geen antwoord)
  - 9 / 99 / 999… (missing)


Politieke besluitvorming_opdracht les2_dataset.sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 38 of 38 Variables
RespondentID geslacht postcode_domicilie postcode_verblijf geboortejaar Opwelkepartijstemdeub ezingvan_D ezingvan_E ezingvan_F
1 1363922033 man 8020 9000 5
2 1363913804 man 8760 9000
3 1363896960 man 8380 9000 5
4 1363873747 man 9450 9450
5 1363562632 vrouw 8020 9000 5
6 1363275243 vrouw 9300 9300 5
7 1362566526 vrouw 1540 9000
8 1362544722 man 9000 9000 5
9 1361198842 man 8434 9000
10 1361149015 vrouw 9620 9000 6
11 1360249342 vrouw 9300 9300
12 1360245180 vrouw 8870 9000
13 1360227472 man 9290 9000 5
14 1360008770 vrouw 9940
15 1359973192 vrouw 2800 9000
16 1359932768 vrouw 9830 9000 5
17 1359758498 vrouw 1540 9000
18 1359681912 man 1850 9000 5
19 1359426086 man 8720 9000
20 1359313404 man 9990 9990
21 1359180663 man 3800 9000
22 1359101274 vrouw 8000 8000 5
23 1359004494 man 9240 9240
24 1358909557 man 9820 9820
25 1358823686 vrouw 9050 9000 1988
26 1358715033 man 8210 8210 1988 3
27 1358700585 man 8000 8000 1988 4
Data View Variable View
Options...
PASW Statistics Processor is ready
Options
Syntax Editor
General Viewer Data Currency Output Labels Charts Pivot Tables File Locations Scripts Multiple Imputations
Variable Lists
Display labels Display names
Alphabetical File Measurement level
Roles
To save time, some dialogs allow the use of predefined field roles to automatically assign variables (fields) to lists in dialogs.
Use predefined roles
Use custom assignments
Windows
Look and feel: SPSS Inc. Classic
SPSS Inc. Standard
SPSS Inc. Classic
Windows
Open syntax window at s
Open only one dataset at a time
Character Encoding for Data and Syntax
Locale's writing system
Unicode (universal character set)
Character encoding cannot be changed when any non-empty datasets are open
Output
No scientific notation for small numbers in tables
Measurement system: Centimeters
Language: English
Notification:
Raise viewer window
Scroll to new output
Sound: None
System beep
Sound
Browse...
User Interface
Language: English
OK Cancel Apply Help
Inbox in Tom.Scham...
Module 2 uni- en biv...
Microsoft PowerPoin...
Politieke besluitvorm...
NL
11:44


Politieke besluitvorming_opdracht les2_dataset.sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 38 of 38 Variables
RespondentID geslacht postcode_domicilie postcode_verblijf geboortejaar Opwelkepartijstemdeub Opwelkepartijstemdeub Opwelkepartijstemdeub Opwelkepartijstemdeub Opwelkepartijstemdeub Opwelkepartijstemdeub ijdelaatstefederaleverki ezingvan_D Opwelkepartijstemdeub ijdelaatstefederaleverki ezingvan_E Opwelkepartijstemdeub ijdelaatstefederalev ezingvan_F
1 1363922033 man 8020 9000 5
2 1363913804 man 8760 9000
3 1363896960 man 8380 9000 5
4 1363873747 man 9450 9450
5 1363562632 vrouw 8020 9000 5
6 1363275243 vrouw 9300 9300 5
7 1362566526 vrouw 1540 9000
8 1362544722 man 9000 9000 5
9 1361198842 man 8434 9000
10 1361149015 vrouw 9620 9000 6
11 1360249342 vrouw 9300 9300
12 1360245180 vrouw 8870 9000
13 1360227472 man 9290 9000 5
14 1360008770 vrouw 9940
15 1359973192 vrouw 2800 9000
16 1359932768 vrouw 9830 9000 5
17 1359758498 vrouw 1540 9000
18 1359681912 man 1850 9000 5
19 1359426086 man 8720 9000
20 1359313404 man 9990 9990
21 1359180663 man 3800 9000
22 1359101274 vrouw 8000 8000 5
23 1359004494 man 9240 9240
24 1358909557 man 9820 9820
25 1358823686 vrouw 9050 9000 1988
26 1358715033 man 8210 8210 1988 3
27 1358700585 man 8000 8000 1988 4
Options
Syntax Editor
General Viewer Data Currency Output Labels Charts Pivot Tables File Locations Scripts Multiple Imputations
Outline Labeling
Variables in item labels shown as:
Labels
Variable values in item labels shown as:
Labels
Pivot Table Labeling
Variables in labels shown as:
Names and Labels
Variable values in labels shown as:
Labels
OK Cancel Apply Help
Data View Variable View
Options...
PASW Statistics Processor is ready
Inbox in Tom.Scham... Module 2 uni- en biv... Microsoft PowerPoin... Politieke besluitvorm...
NL 11:41

Wat volgt

- Een variabele verplaatsen
- Een variabele opzoeken (FIND) en tellen (COUNT)
- Variabelen selecteren (GO TO Case / VAR)
- Een variabele toevoegen
- Een variabele aanmaken op basis van reeds bestaande, bv. dummificeren, dichotomiseren… (COMPUTE)
- Een variabele hercoderen, aggregeren, etc. (RECODE)

Een variabele verplaatsen

- Klik bovenaan de kolom (variabele) die je wenst te verplaatsen
- Hou de kolom vast en versleep naar links of rechts en plaats tussen twee andere variabelen of voor de eerste of achter de laatste variabele.

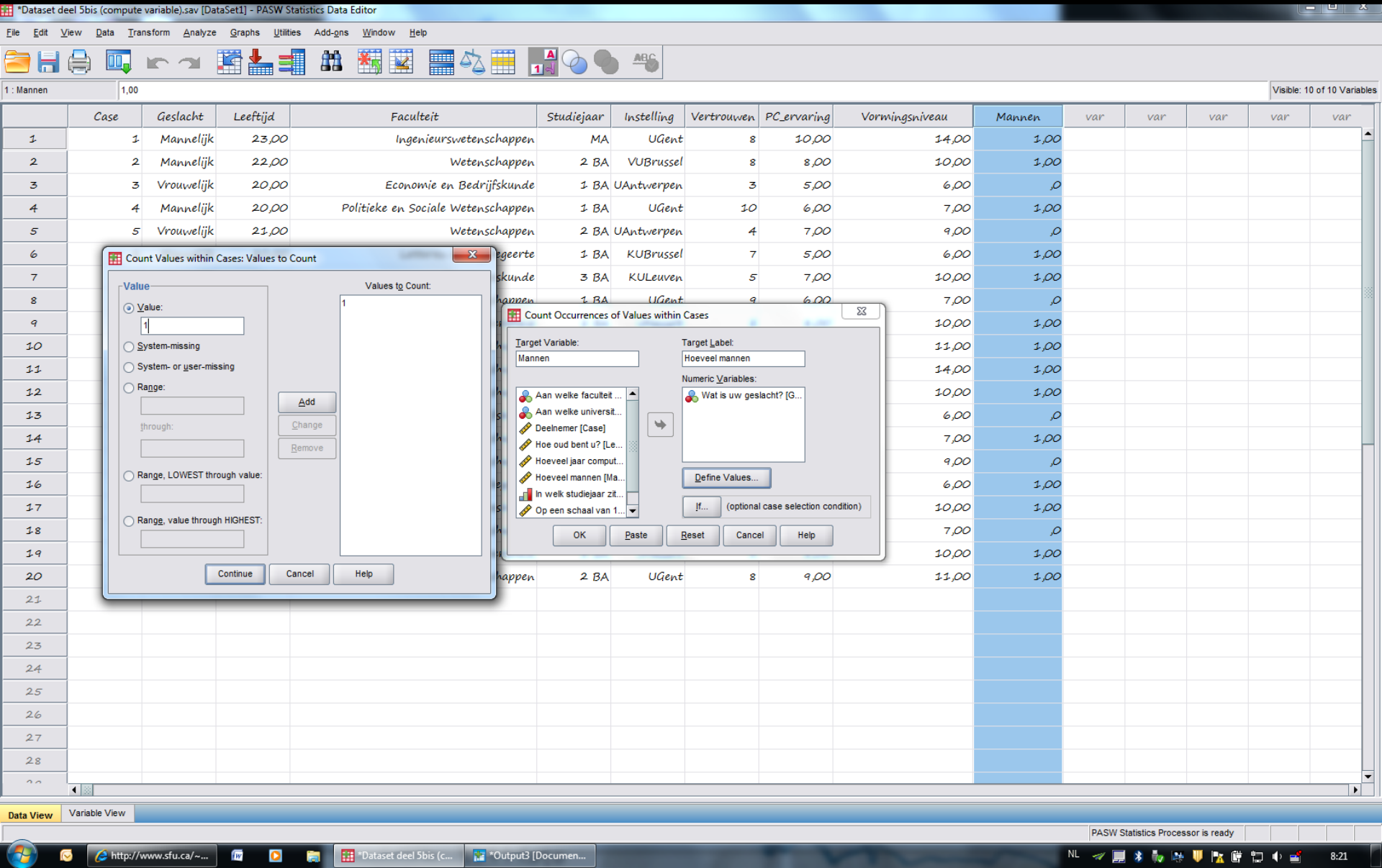
Een variabele zoeken

- Via **Edit** > **Find** (CTRL + F)
- Via **Edit** > **Go to**
- Selecteren van variabelen of cases :
  - Via **Data** > **Select cases**
  - Via Utilities > Use variable sets

Waarden tellen (COUNT)

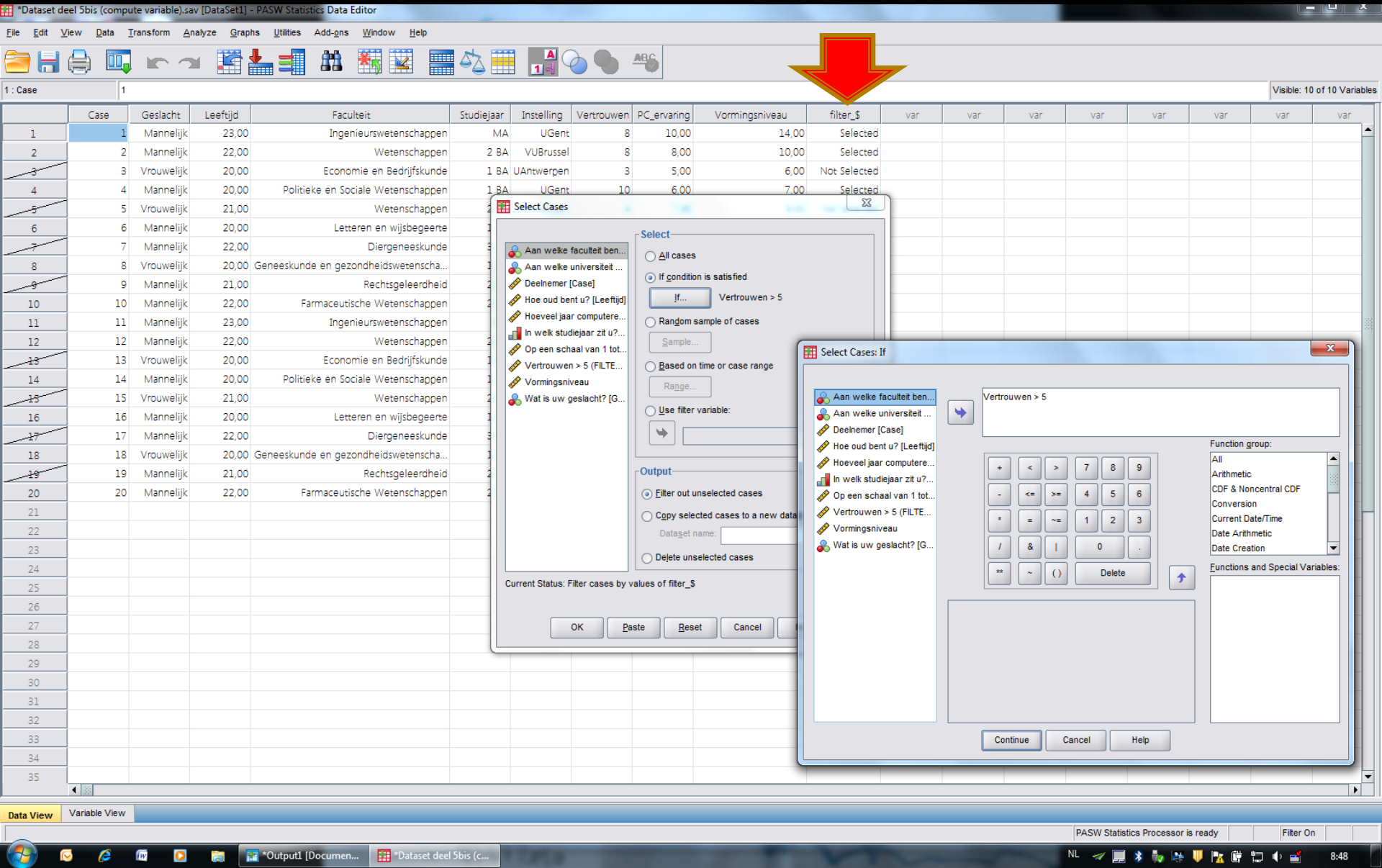
- Op basis van bepaalde criteria opzoeken (tellen) hoeveel respondenten
  - zich in een bepaald waardebereik bevinden
  - een bepaalde waarde hebben

- Meestal om een inzicht te krijgen in hoeveel % missings er zijn
  - 0% = geen missings voor een bepaalde variabele
  - 100% = alleen maar missings


*Dataset deel 5bis (compute variable).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : Mannen
1,00
Visible: 10 of 10 Variables
Case Geslacht Leeftijd Faculteit Studiejaar Instelling Vertrouwen PC_ervaring Vormingsniveau Mannen var var var var var
1 1 Mannelijk 23,00 Ingenieurswetenschappen MA UGent 8 10,00 14,00 1,00
2 2 Mannelijk 22,00 Wetenschappen 2 BA VUBrussel 8 8,00 10,00 1,00
3 3 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3 5,00 6,00 ,0
4 4 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10 6,00 7,00 1,00
5 5 Vrouwelijk 21,00 Wetenschappen 2 BA UAntwerpen 4 7,00 9,00 ,0
6 1 BA KUBrussel 7 5,00 6,00 1,00
7 3 BA KULeuven 5 7,00 10,00 1,00
8 7,00 ,0
9 10,00 1,00
10 11,00 1,00
11 14,00 1,00
12 10,00 1,00
13 6,00 ,0
14 7,00 1,00
15 9,00 ,0
16 6,00 1,00
17 10,00 1,00
18 7,00 ,0
19 10,00 1,00
20 2 BA UGent 8 9,00 11,00 1,00
Count Values within Cases: Values to Count
Value
Value:
1
System-missing
System- or user-missing
Range:
through:
Range, LOWEST through value:
Range, value through HIGHEST:
Values to Count:
1
Add
Change
Remove
Continue
Cancel
Help
Count Occurrences of Values within Cases
Target Variable:
Mannen
Target Label:
Hoeveel mannen
Numeric Variables:
Wat is uw geslacht? [G...
Aan welke faculteit ...
Aan welke universit...
Deelnemer [Case]
Hoe oud bent u? [Le...
Hoeveel jaar comput...
Hoeveel mannen [Ma...
In welk studiejaar zit...
Op een schaal van 1...
Define Values...
If...
(optional case selection condition)
OK
Paste
Reset
Cancel
Help
Data View
Variable View
PASW Statistics Processor is ready
http://www.sfu.ca/~...
*Dataset deel 5bis (c...
*Output3 [Documen...
NL
8:21

Cases selecteren

- Cases selecteren ?
  - **Data** > **Select cases** **(Filter aan / uit)**
    - Selectie op basis van bepaalde voorwaarde(n) (condities)
    - *Random sample of cases* (= selectie ad random uit de steekproef) = willekeurige proportie (van de steekproef)
  - **Goed om weten**: de filter blijft beschikbaar (als VAR) ook al is ie uitgeschakeld

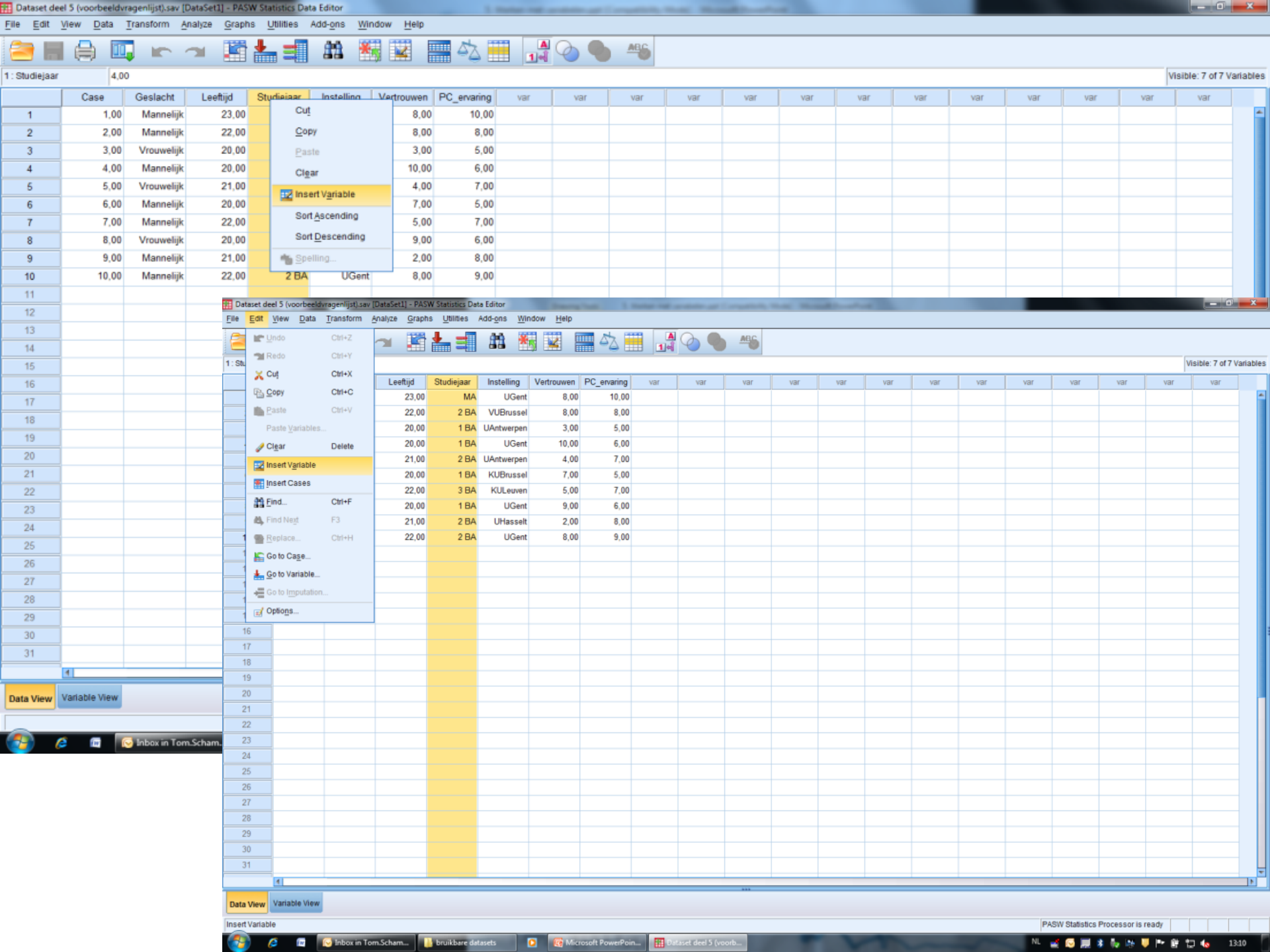

*Dataset deel 5bis (compute variable).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : Case
Visible: 10 of 10 Variables
Case Geslacht Leeftijd Faculteit Studiejaar Instelling Vertrouwen PC_ervaring Vormingsniveau filter_$
1 Mannelijk 23,00 Ingenieurswetenschappen MA UGent 8 10,00 14,00 Selected
2 Mannelijk 22,00 Wetenschappen 2 BA VUBrussel 8 8,00 10,00 Selected
3 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3 5,00 6,00 Not Selected
4 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10 6,00 7,00 Selected
5 Vrouwelijk 21,00 Wetenschappen
6 Mannelijk 20,00 Letteren en wijsbegeerte
7 Mannelijk 22,00 Diergeneeskunde
8 Vrouwelijk 20,00 Geneeskunde en gezondheidswetenscha...
9 Mannelijk 21,00 Rechtsgeleerdheid
10 Mannelijk 22,00 Farmaceutische Wetenschappen
11 Mannelijk 23,00 Ingenieurswetenschappen
12 Mannelijk 22,00 Wetenschappen
13 Vrouwelijk 20,00 Economie en Bedrijfskunde
14 Mannelijk 20,00 Politieke en Sociale Wetenschappen
15 Vrouwelijk 21,00 Wetenschappen
16 Mannelijk 20,00 Letteren en wijsbegeerte
17 Mannelijk 22,00 Diergeneeskunde
18 Vrouwelijk 20,00 Geneeskunde en gezondheidswetenscha...
19 Mannelijk 21,00 Rechtsgeleerdheid
20 Mannelijk 22,00 Farmaceutische Wetenschappen
Select Cases
Aan welke faculteit ben...
Aan welke universiteit ...
Deelnemer [Case]
Hoe oud bent u? [Leeftijd]
Hoeveel jaar computere...
In welk studiejaar zit u?...
Op een schaal van 1 tot...
Vertrouwen > 5 (FILTE...
Vormingsniveau
Wat is uw geslacht? [G...
Select
All cases
If condition is satisfied
If...
Vertrouwen > 5
Random sample of cases
Sample...
Based on time or case range
Range...
Use filter variable:
Output
Filter out unselected cases
Copy selected cases to a new data
Dataset name:
Delete unselected cases
Current Status: Filter cases by values of filter_$
OK Paste Reset Cancel
Select Cases: If
Aan welke faculteit ben...
Aan welke universiteit ...
Deelnemer [Case]
Hoe oud bent u? [Leeftijd]
Hoeveel jaar computere...
In welk studiejaar zit u?...
Op een schaal van 1 tot...
Vertrouwen > 5 (FILTE...
Vormingsniveau
Wat is uw geslacht? [G...
Vertrouwen > 5
+ < > 7 8 9
- <= >= 4 5 6
* = ~= 1 2 3
/ & | 0 .
** ~ ( ) Delete
Function group:
All
Arithmetic
CDF & Noncentral CDF
Conversion
Current Date/Time
Date Arithmetic
Date Creation
Functions and Special Variables:
Continue Cancel Help
Data View Variable View
PASW Statistics Processor is ready
Filter On
*Output1 [Documen...
*Dataset deel 5bis (c...
NL
8:48

Een variabele / case toevoegen

Klik rechts op de balk met de variabelennamen en voeg een variabele toe.

OF: kies linksbovenaan in de takenbalk **Edit > Insert variable**.

Idem: **Insert case**


Dataset deel 5 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : Studiejaar
4,00
Visible: 7 of 7 Variables
Case Geslacht Leeftijd Studiejaar Instelling Vertrouwen PC_ervaring
1,00 Mannelijk 23,00 8,00 10,00
2,00 Mannelijk 22,00 8,00 8,00
3,00 Vrouwelijk 20,00 3,00 5,00
4,00 Mannelijk 20,00 10,00 6,00
5,00 Vrouwelijk 21,00 4,00 7,00
6,00 Mannelijk 20,00 7,00 5,00
7,00 Mannelijk 22,00 5,00 7,00
8,00 Vrouwelijk 20,00 9,00 6,00
9,00 Mannelijk 21,00 2,00 8,00
10,00 Mannelijk 22,00 2 BA UGent 8,00 9,00
Cut
Copy
Paste
Clear
Insert Variable
Sort Ascending
Sort Descending
Spelling...
Data View
Variable View
Dataset deel 5 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Undo Ctrl+Z
Redo Ctrl+Y
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Paste Variables...
Clear Delete
Insert Variable
Insert Cases
Find... Ctrl+F
Find Next F3
Replace... Ctrl+H
Go to Case...
Go to Variable...
Go to Imputation...
Options...
Visible: 7 of 7 Variables
Leeftijd Studiejaar Instelling Vertrouwen PC_ervaring
23,00 MA UGent 8,00 10,00
22,00 2 BA VUBrussel 8,00 8,00
20,00 1 BA UAntwerpen 3,00 5,00
20,00 1 BA UGent 10,00 6,00
21,00 2 BA UAntwerpen 4,00 7,00
20,00 1 BA KUBrussel 7,00 5,00
22,00 3 BA KULeuven 5,00 7,00
20,00 1 BA UGent 9,00 6,00
21,00 2 BA UHasselt 2,00 8,00
22,00 2 BA UGent 8,00 9,00
Data View
Variable View
Insert Variable
PASW Statistics Processor is ready
Inbox in Tom.Scham...
bruikbare datasets
Microsoft PowerPoin...
Dataset deel 5 (voorb...
NL
13:10

Oefening

Selecteer **Variable View**. Double click op *VAR00001* en typ "*Faculteit*" om de nieuwe variabele een naam te geven.

- Ga naar *Values* kolom, klik op de grijze cel.
- Typ "1" in de *Value* box en "Letteren en Wijsbegeerte"; "2" voor "Rechtsgeleerdheid"; "3" voor "Wetenschappen"; "4" voor "Geneeskunde en Gezondheidswetenschappen"; "5" voor "Ingenieurswetenschappen", "6" voor "Economie en Bedrijfskunde"; "7" voor "Diergeneeskunde", "8" voor "Psychologie en Pedagogische Wetenschappen", "9" voor "Bio-ingenieurs- wetenschappen", "10" voor "Farmaceutische Wetenschappen" en "11" voor "Politieke en Sociale Wetenschappen"
- Selecteer telkens **Add.**
- Na de laatste **Add > OK.**

Iedereen voegt deze variabele toe

*Dataset deel 5 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Case	Numeric	8	2		None	None	8	Right	Scale	Input
2	Geslacht	Numeric	8	2		{1,00, Mann...	None	8	Right	Nominal	Input
3	Leeftijd	Numeric	8	2		None	None	8	Right	Scale	Input
4	Richting	Numeric	8	2		{1,00, Letter...	None	8	Right	Scale	Input
5	Studiejaar	Numeric	8	2		{1,00, 1 BA}...	None	8	Right	Nominal	Input
6	Instelling	Numeric	8	2		{1,00, UGen...	None	8	Right	Nominal	Input
7	Vertrouwen	Numeric	8	2	Op een schaal ...	None	None	8	Right	Scale	Input
8	PC_ervaring	Numeric	8	2	Hoeveel jaar co...	None	None	8	Right	Scale	Input



1 Deelnemer
2 Wat is uw geslacht?
3 Hoe oud bent u?
4 Aan welke faculteit bent u ingeschreven?
5 In welk studiejaar zit u?
6 Aan welke universiteit bent u ingeschreven?
7 Op een schaal van 1 tot 10 met 10 het meest, hoeveel vertrouwen heb je erin dat je SPSS zal leren?
8 Hoeveel jaar computerervaring heb je?

Data View | Variable View

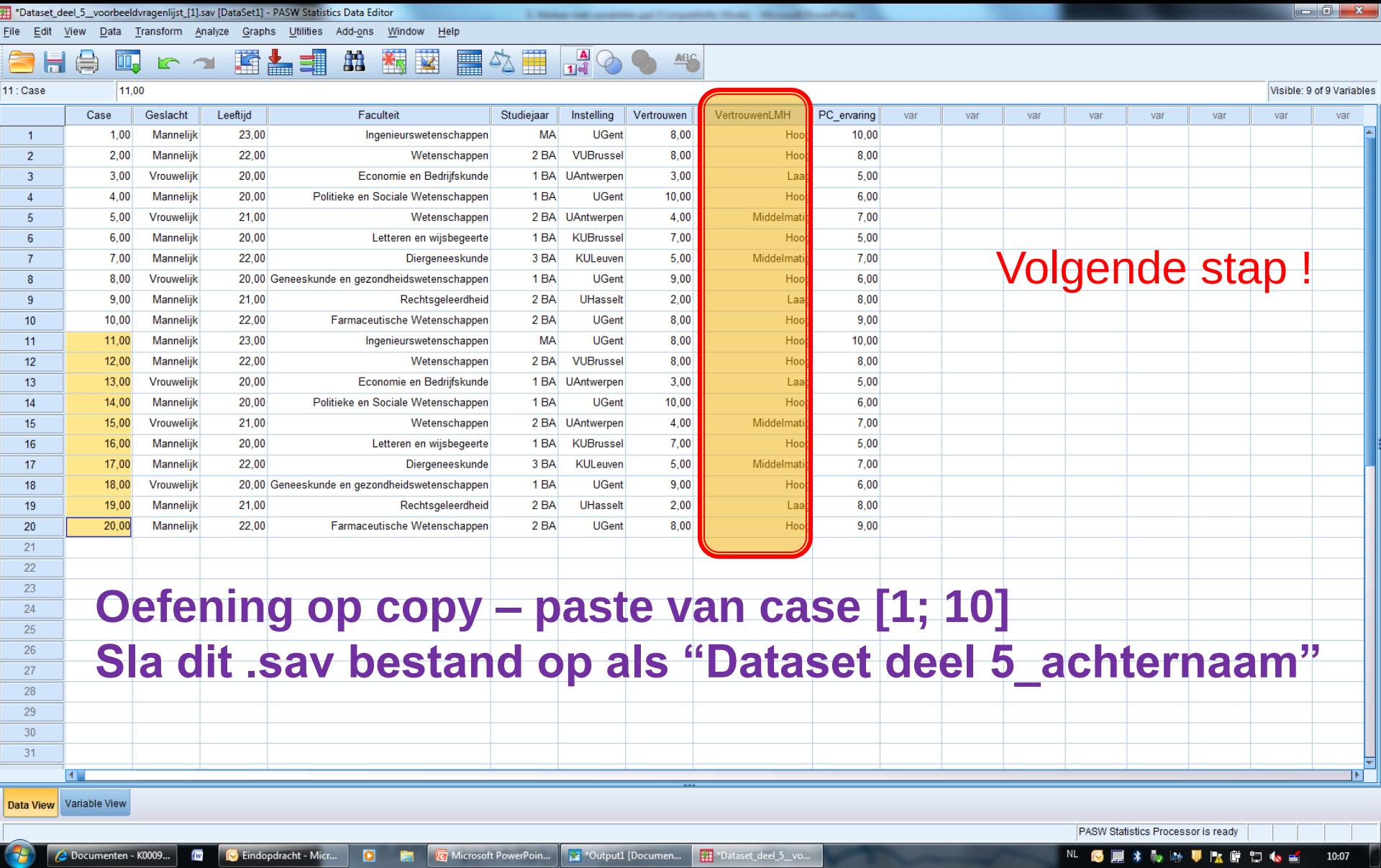
PASW Statistics Processor is ready

Homepage — Univer... | Inbox in Tom.Scham... | bruikbare datasets | Microsoft PowerPoin... | *Dataset deel 5 (voor... | NL | 13:23

Waarden voor alle cases toevoegen

- Selecteer vervolgens : **Data View**.
- Klik op de eerste lijn van de *Richting* column, selecteer de pijl omlaag en selecteer **Ingenieurswetenschappen.**

- Doe dit nu voor alle cases op de volgende manier


*Dataset_deel_5__voorbeeldvragenlijst_[1].sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
11 : Case
11,00
Visible: 9 of 9 Variables
Case Geslacht Leeftijd Faculteit Studiejaar Instelling Vertrouwen VertrouwenLMH PC_ervaring var var var var var var var var
1 1,00 Mannelijk 23,00 Ingenieurswetenschappen MA UGent 8,00 Hoog 10,00
2 2,00 Mannelijk 22,00 Wetenschappen 2 BA VUBrussel 8,00 Hoog 8,00
3 3,00 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3,00 Laag 5,00
4 4,00 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10,00 Hoog 6,00
5 5,00 Vrouwelijk 21,00 Wetenschappen 2 BA UAntwerpen 4,00 Middelmatig 7,00
6 6,00 Mannelijk 20,00 Letteren en wijsbegeerte 1 BA KUBrussel 7,00 Hoog 5,00
7 7,00 Mannelijk 22,00 Diergeneeskunde 3 BA KULeuven 5,00 Middelmatig 7,00
8 8,00 Vrouwelijk 20,00 Geneeskunde en gezondheidswetenschappen 1 BA UGent 9,00 Hoog 6,00
9 9,00 Mannelijk 21,00 Rechtsgeleerdheid 2 BA UHasselt 2,00 Laag 8,00
10 10,00 Mannelijk 22,00 Farmaceutische Wetenschappen 2 BA UGent 8,00 Hoog 9,00
11 11,00 Mannelijk 23,00 Ingenieurswetenschappen MA UGent 8,00 Hoog 10,00
12 12,00 Mannelijk 22,00 Wetenschappen 2 BA VUBrussel 8,00 Hoog 8,00
13 13,00 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3,00 Laag 5,00
14 14,00 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10,00 Hoog 6,00
15 15,00 Vrouwelijk 21,00 Wetenschappen 2 BA UAntwerpen 4,00 Middelmatig 7,00
16 16,00 Mannelijk 20,00 Letteren en wijsbegeerte 1 BA KUBrussel 7,00 Hoog 5,00
17 17,00 Mannelijk 22,00 Diergeneeskunde 3 BA KULeuven 5,00 Middelmatig 7,00
18 18,00 Vrouwelijk 20,00 Geneeskunde en gezondheidswetenschappen 1 BA UGent 9,00 Hoog 6,00
19 19,00 Mannelijk 21,00 Rechtsgeleerdheid 2 BA UHasselt 2,00 Laag 8,00
20 20,00 Mannelijk 22,00 Farmaceutische Wetenschappen 2 BA UGent 8,00 Hoog 9,00
Volgende stap !
Oefening op copy – paste van case [1; 10]
Sla dit .sav bestand op als “Dataset deel 5_achternaam”
Data View
Variable View
PASW Statistics Processor is ready

Variabelen aanmaken op basis van reeds bestaande

Een voorbeeld:

Een nieuwe variabele creëren op basis van reeds bestaande variabelen betekent meestal dat je nood hebt aan **een samengestelde variabele** in je onderzoek. Vooral in de psychologie en sociologie wordt deze methode vaak toegepast. Zo kan je van een persoon zijn individuele scores te weten komen per vraag in een **intelligentietest**. Door middel van het creëren van een samengestelde variabele, tel je de score van alle variabelen op een bepaalde subtest op en bekom je bijvoorbeeld de intelligentiescore op taalgebied. Deze samengestelde variabele kan men dan gebruiken om conclusies te trekken rond het verband tussen intelligentie en andere variabelen (bvb. is er een verschil in intelligentie naar gelang het inkomen of het geslacht).

- Meerdere tests -> 1 totaalscore IQ
- Meerdere vragen -> totaalscore 1 test (numerieke proef, taal, logica etc…)

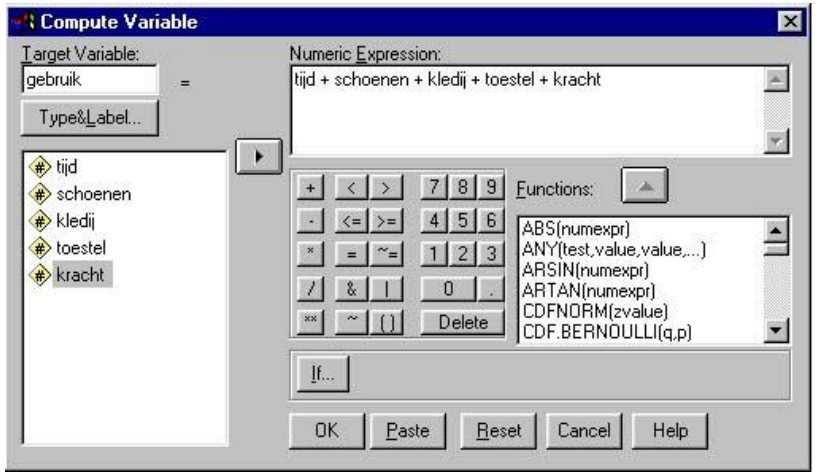
Compute (new) variable

- Om een nieuwe variabele te creëren selecteer je: **Transform > Compute** en het dialoogvenster **'Compute Variable'** komt tevoorschijn.
- Dit dialoogvenster is bijna identiek aan het dialoogvenster voor select cases, if.

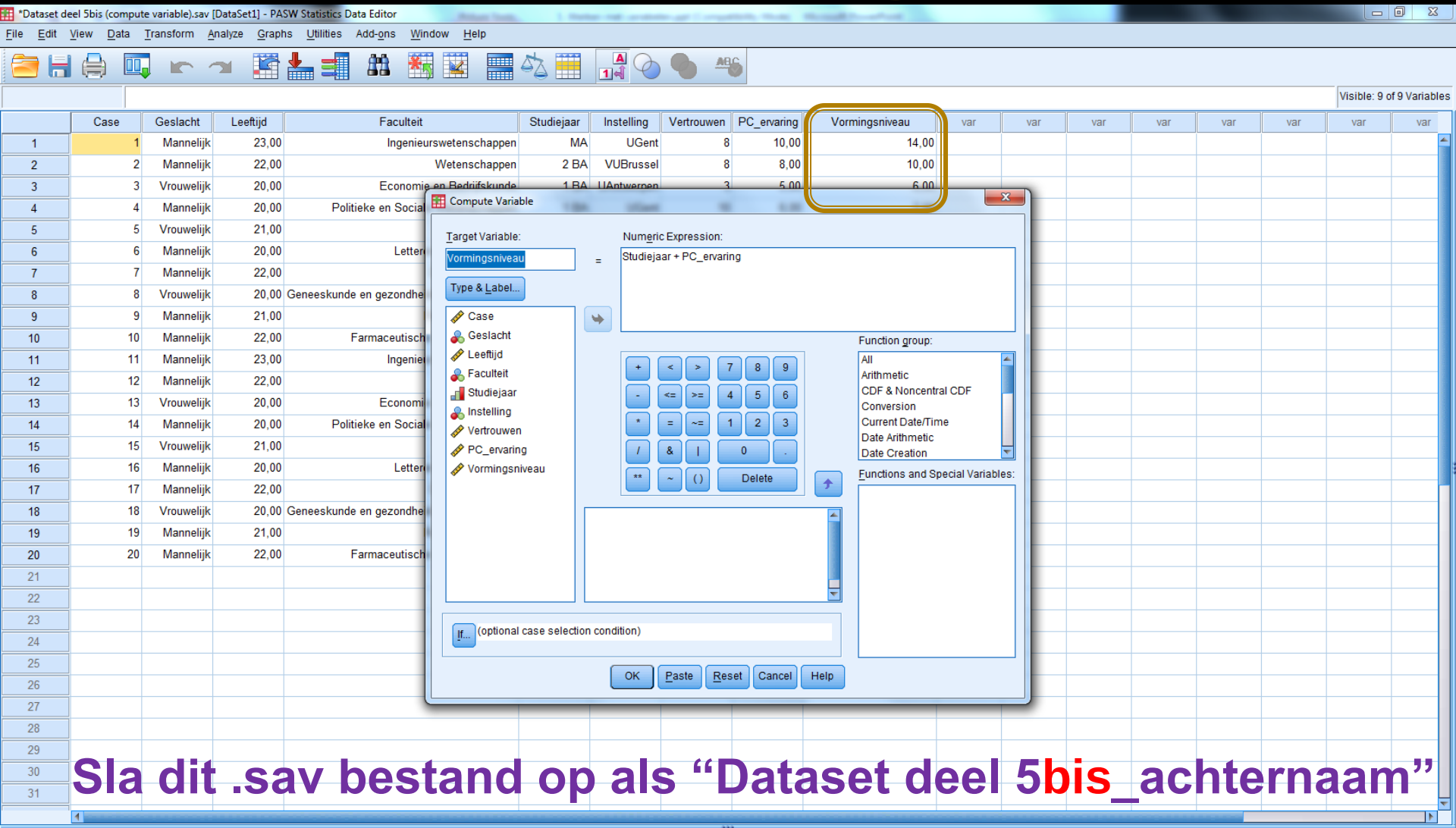
- Om een nieuwe variabele te definiëren:
  - klik in de tekstbox onder 'Target Variable'
  - geef een nieuwe naam voor je samengestelde variabele
  - vorm in de tekstbox onder 'Numeric Expression' de formule in aan de hand waarvan de nieuwe variabele moet worden berekend

Voorbeeld: we doen een onderzoek in een fitness-centrum en vragen aan de sportliefhebbers hoeveel tijd ze gemiddeld spenderen in het centrum, hoe vaak ze nieuwe schoenen/kledij kopen, hoe veel toestellen ze gemiddeld gebruiken en hoelang ze aan krachttraining doen. Stel dat we voor ons onderzoek een globaal idee willen krijgen over de status van onze fitness-gebruikers, meer bepaald is deze gebruiker een occasionele gebruiker, free time gebruiker, liefhebber of profesional. Hiervoor kunnen we de scores op de 5 variabelen samentellen zodat we een eindscore bereiken. Op basis van deze eindscore wordt de gebruiker in 1 van deze status-categoriën gevoegd.

De definitie van onze **samengestelde variabele** (***Target Variable***) zou er dan als volgt uit zien:



Een fictief voorbeeldje



Sla dit .sav bestand op als "Dataset deel 5bis_achternaam"

Oefening

- Via **Transform** > **Compute variable**

  Voorbeeld: bereken op basis van het opgegeven geboortejaar de leefrijd en maak een nieuwe VAR aan: ***Leeftijd***

*Politieke besluitvorming_opdracht les2_oplossing.sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	RespondentID	Numeric	11	0		None	None	11	Right	Scale	Input
2	geslacht	Numeric	8	0	Bent u man of vrouw?	{1, man}...	None	8	Right	Nominal	Input
3	postcode_domicilie	Numeric	8	0	Wat is de postcode van uw domicilie?	None	None	8	Right	Scale	Input
4	postcode_verblijf	Numeric	8	0	Wat is de postcode van uw verblijfplaats?	None	None	8	Right	Scale	Input
5	geboortejaar	Numeric	8	0	In welk jaar bent u geboren?	{28, 1987}...	None	8	Right	Scale	Input
6	leeftijd	Numeric	8	2	Hoe oud is de respondent?	None	None	10	Right	Scale	Input
7	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	8	2	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
8	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	8	2	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
9	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
10	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
11	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
12	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
13	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	8	2	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
14	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
15	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
16	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	Numeric	11	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Right	Nominal	Input
17	Opwelkepartijstemdeubijdelaatstefederaleverkiezi...	String	10	0	Op welke partij stemde u bij de laatste federale verkiezin...	None	None	15	Left	Nominal	Input
18	partijkeuze_2010	Numeric	8	0	Op welke partij stemde u bij de federale Kamerverkiezing...	{1, ongeldig...	None	8	Right	Nominal	Input
19	VRP_uitgenodigd	Numeric	8	0	Bent u ooit uitgenodigd voor een volksraadpleging?	{1, ja}...	None	8	Right	Nominal	Input
20	VRP_deelgenomen	Numeric	8	0	Indien u in de toekomst zou wordenuitgenodigd, zou u de...	{1, ja}...	None	8	Right	Nominal	Input
21	VRP_belangrijkheid	Numeric	8	0	Ik vind volksraadpleging als vorm van directe democratie:	{1, enorm b...	None	8	Right	Scale	Input
22	machtigst_kabinetten_vs_admin	Numeric	8	0	Welke instelling beschouwt u als de machtigste?	{1, veruit k...	None	8	Right	Scale	Input
23	beheersing_cumul	Numeric	8	0	Welke houding moet de federale regering aanhouden ten ...	{1, top-3 p...	None	8	Right	Ordinal	Input
24	beheersing_schijnkandidatuur	Numeric	8	0	Welke houding moet de federale regering aanhouden ten ...	{1, top-3 p...	None	8	Right	Ordinal	Input
25	wetgevend_initiatief	Numeric	8	0	Denkt u dat er een wetgevend initiatief nodig is dat cum...	{1, ja, voor ...	None	8	Right	Nominal	Input
26	Indienjawelkeinitiatieflijktuhetmeestgeschiktmetb...	Numeric	11	0	Indien ja, welke initiatief lijkt u het meest geschikt met b...	None	None	11	Right	Nominal	Input
27	Indienjawelkeinitiatieflijktuhetmeestgeschiktmetb...	Numeric	11	0	Indien ja, welke initiatief lijkt u het meest geschikt met b...	None	None	11	Right	Nominal	Input
28	Indienjawelkeinitiatieflijktuhetmeestgeschiktmetb...	Numeric	11	0	Indien ja, welke initiatief lijkt u het meest geschikt met b...	None	None	11	Right	Nominal	Input
29	Indienjawelkeinitiatieflijktuhetmeestgeschiktmetb...	Numeric	11	0	Indien ja, welke initiatief lijkt u het meest geschikt met b...	None	None	11	Right	Nominal	Input
30	Indienjawelkeinitiatieflijktuhetmeestgeschiktmetb...	Numeric	11	0	Indien ja, welke initiatief lijkt u het meest geschikt met b...	None	None	11	Right	Nominal	Input

Data View | Variable View

PASW Statistics Processor is ready

Inbox in Tom.Scham... | http://webhost.ua.ac... | Politieke besluitvorm... | Microsoft PowerPoin... | *Politieke besluitvor... | NL | 12:13

Dichotomiseren van een variabele

- Via **Transform** > **Compute variable**

- Nieuwe codering: **"0 / 1"** of bijv. **"1 / 2"**
- Op basis van zelf te bepalen grenzen, bereik (ranges), "<=X" of ">=X" etc.

Oefening

Voorbeeld:

- ***Vertrouwen*** transformeren in een nieuwe variabele ***Vertrouwen_LMH*** en daarna bijvoorbeeld naar een opdeling al dan niet ***voldoende*** (0 of 1) op basis van alle waarden <6 -> "0" en waarden >=6 -> "1"

  (zie verder voor uitwerking)

Dummificeren van een variabele

- Het ontbinden van een variabelen met x categorieën naar x-1 variabelen die dichotoom zijn
- Belangrijk voor lineaire regressie waar men een categoriale variabele wil meenemen in een regressievergelijking (module 3)

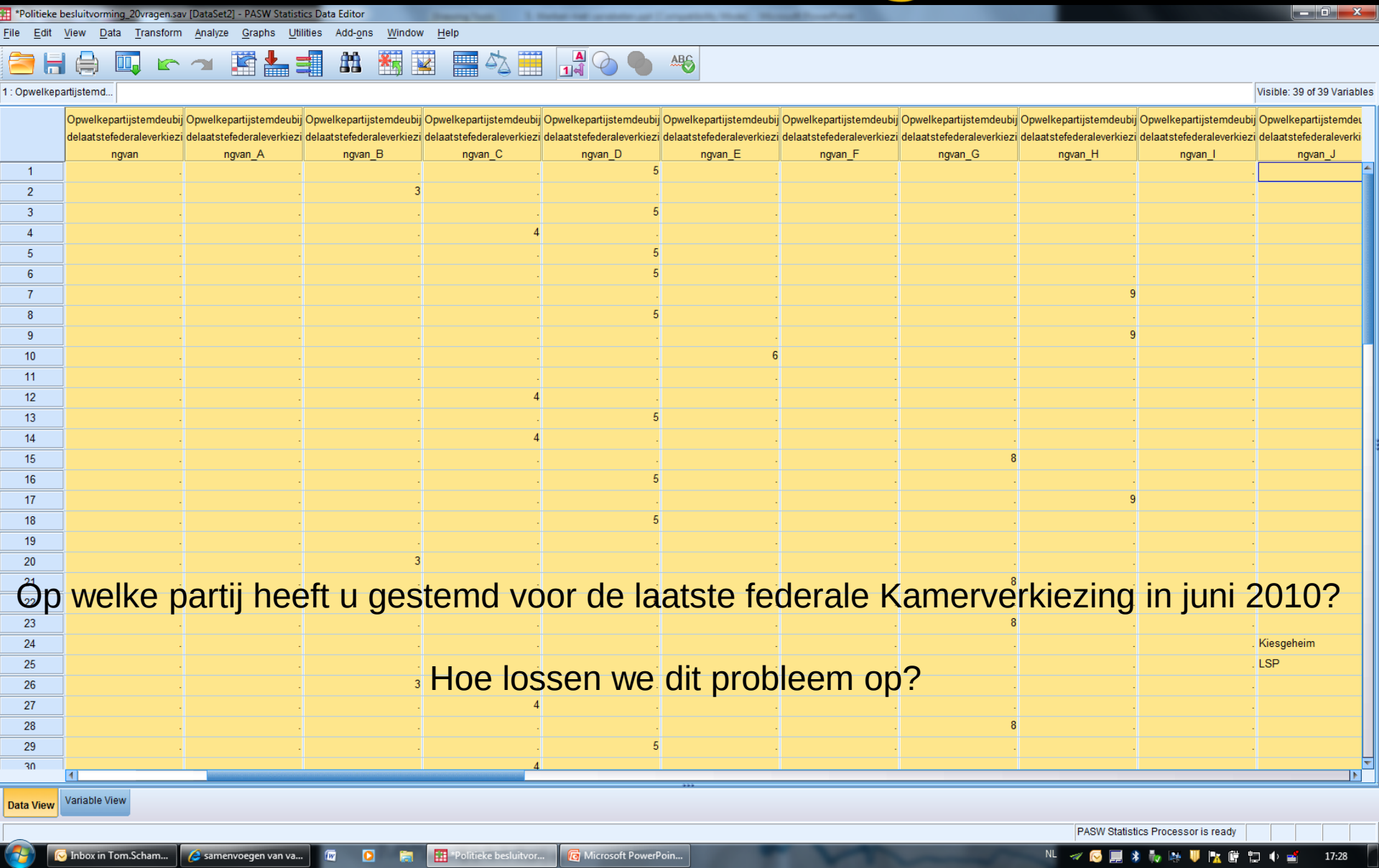
- Voorbeeldje: schaal van 1 -> 5 omzetten naar vier dummyvariabelen

Dummificeren - 2

VAR	DUM_1	DUM_2	DUM_3	DUM_4
Helemaal oneens	0	0	0	0
Eerder oneens	0	0	0	1
Weet niet (onbeslist)	0	0	1	0
Eerder eens	0	1	0	0
Helemaal eens	1	0	0	0

5 categorieën -> 4 dummyvariabelen !!

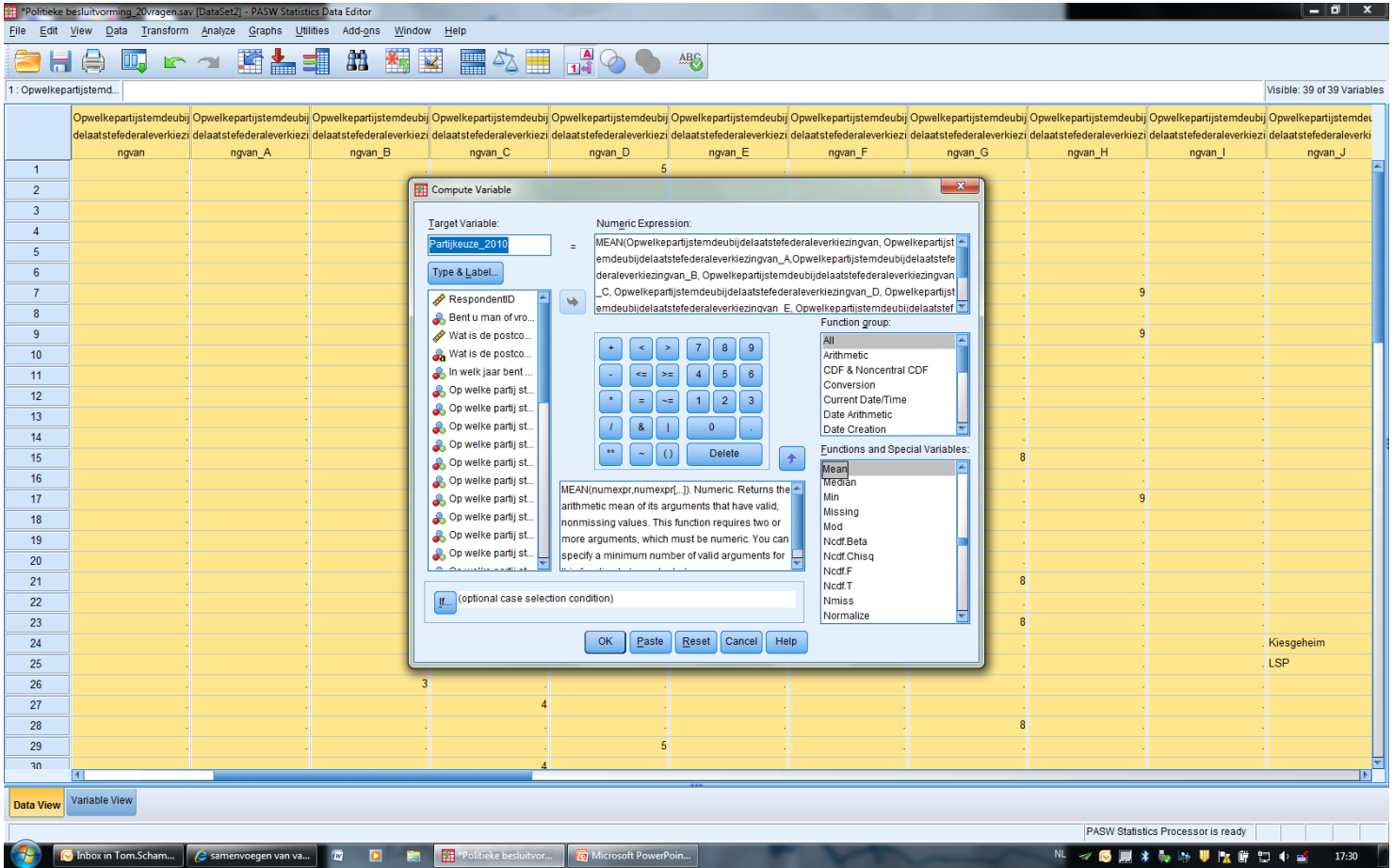
Variabelen samenvoegen



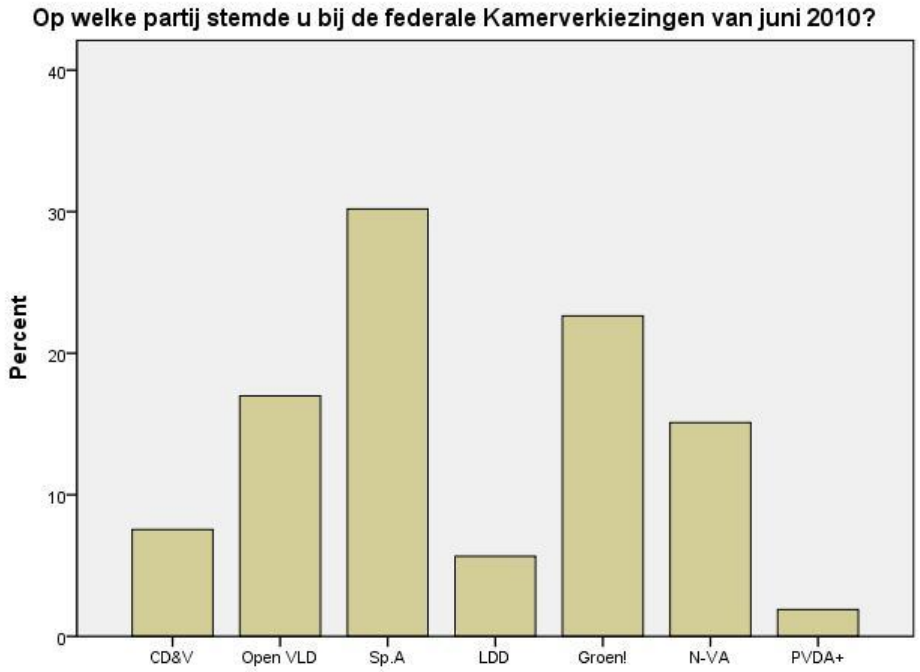
Op welke partij heeft u gestemd voor de laatste federale Kamerverkiezing in juni 2010?

Hoe lossen we dit probleem op?

- Via **Transform > Compute Variable**



Resultaat
VAR Partijkeuze_2010

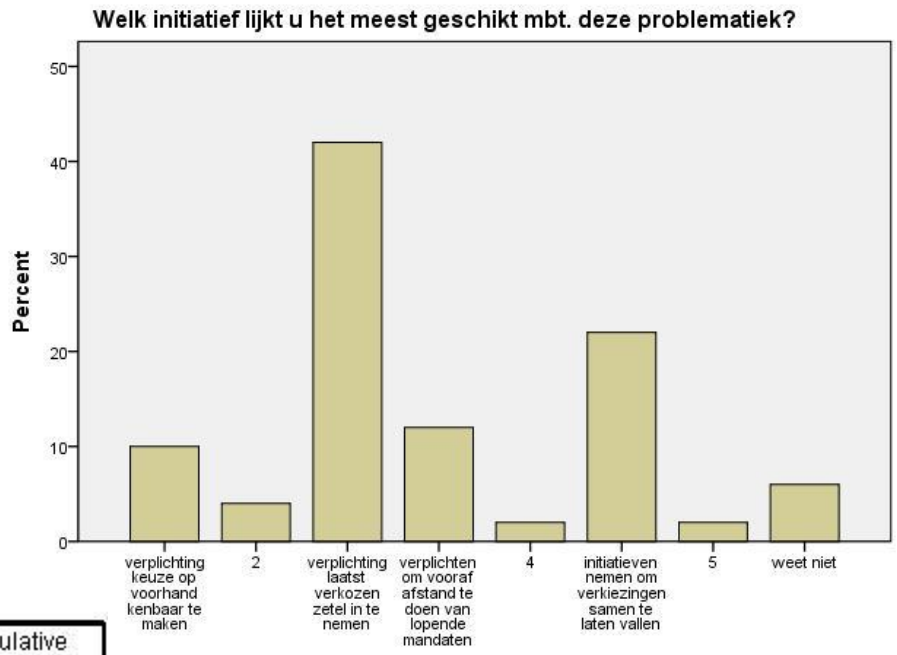


Op welke partij stemde u bij de federale Kamerverkiezingen van juni 2010?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	CD&V	4	6,6	7,5	7,5
	Open VLD	9	14,8	17,0	24,5
	Sp.A	16	26,2	30,2	54,7
	LDD	3	4,9	5,7	60,4
	Groen!	12	19,7	22,6	83,0
	N-VA	8	13,1	15,1	98,1
	PVDA+	1	1,6	1,9	100,0
	Total	53	86,9	100,0	
Missing	System	8	13,1		
Total		61	100,0		

- Idem voor VAR
  - ***welk initiatief om cumulatie en/of schijnkandidatuur tegen te gaan***

VAR welk_initiatief



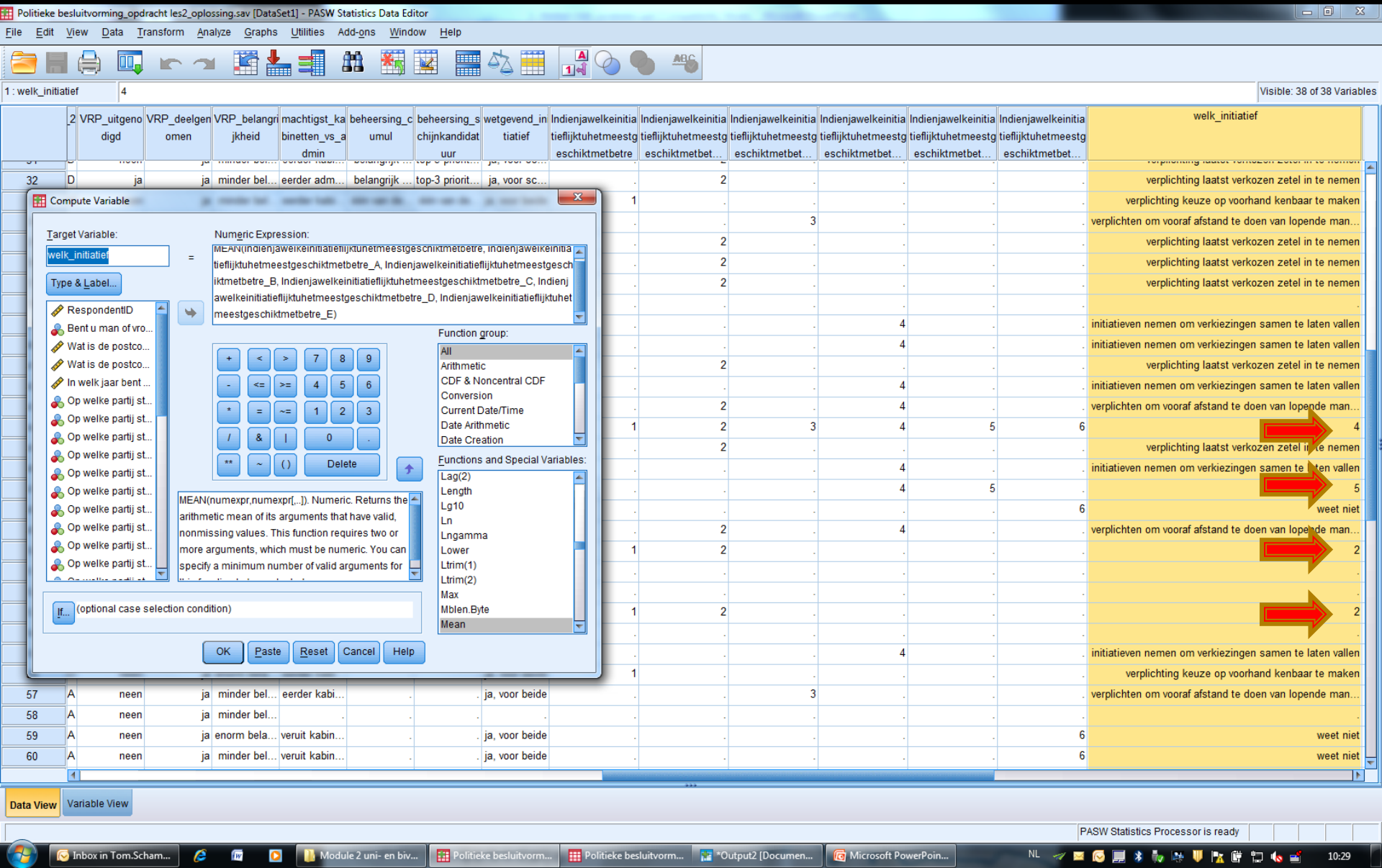
Welk initiatief lijkt u het meest geschikt mbt. deze problematiek?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	verplichting keuze op voorhand kenbaar te maken	5	8,2	10,0	10,0
	2	2	3,3	4,0	14,0
	verplichting laatst verkozen zetel in te nemen	21	34,4	42,0	56,0
	verplichten om vooraf afstand te doen van lopende mandaten	6	9,8	12,0	68,0
	4	1	1,6	2,0	70,0
	initiatieven nemen om verkiezingen samen te laten vallen	11	18,0	22,0	92,0
	5	1	1,6	2,0	94,0
	weet niet	3	4,9	6,0	100,0
	Total	50	82,0	100,0	
Missing	System	11	18,0		
Total		61	100,0		

Welk initiatief lijkt u het meest geschikt mbt. deze problematiek?

HOE KAN DAT ??

VAR welk_initiatief



Variabelen hercoderen

- **Twee vormen:**
  - Hercoderen van categoriale variabelen
  - Hercoderen van continue variabelen
- **Twee resultaten:**
  - In dezelfde variabele
  - In een nieuwe variabele

- Voorbeeld (hercoderen continue variabele)
  - De variabele *Vertrouwen* omvat de data voor een 10-punt Likertschaal (1 – 10).
  - Om de resultaten vergelijkbaar te maken met ander onderzoek waarbij een ander aantal antwoordcategorieën werd gehanteerd (nl. Laag, Middelmatig, Hoog) willen we een groepsvariabele creëren (Ordinal data).

Categoriale variabele hercoderen

- Voorbeeld: variabel '*Studiejaar*' (1Ba, 2Ba, 3Ba en 4Ma)

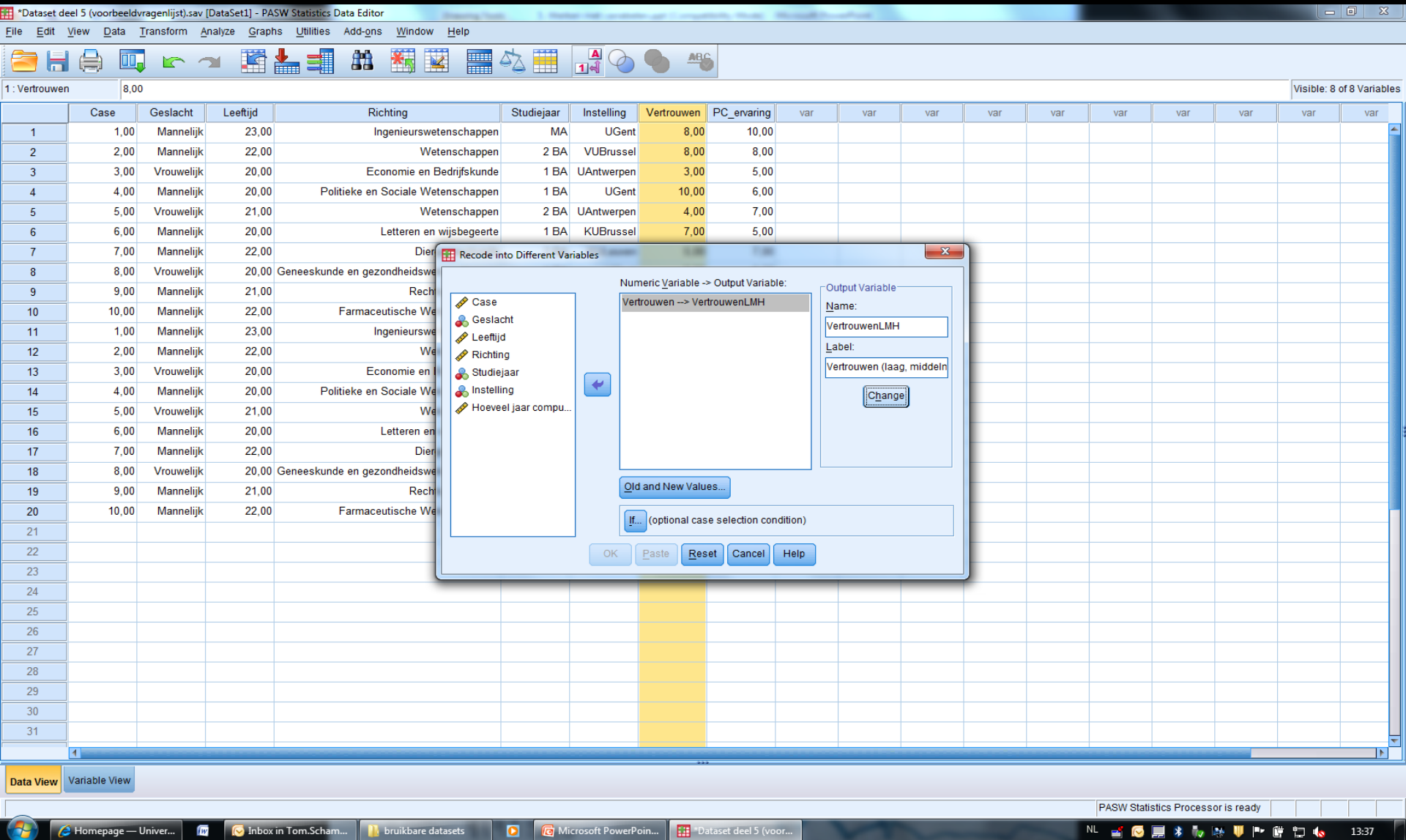
- Hercoderen in een tweeklassevariabele '***Studiejaar_rec***'
  - [Ba (1), Ba (2) en Ba (3)] -> **Bachelor** (value: 1)
  - [Ma (4)] -> **Master** (value: 2)

Continue variabelen hercoderen

Doel: het meetniveau verlagen (van scale -> ordinal)

Selecteer: **Transform > Recode > Into Different Variables**.

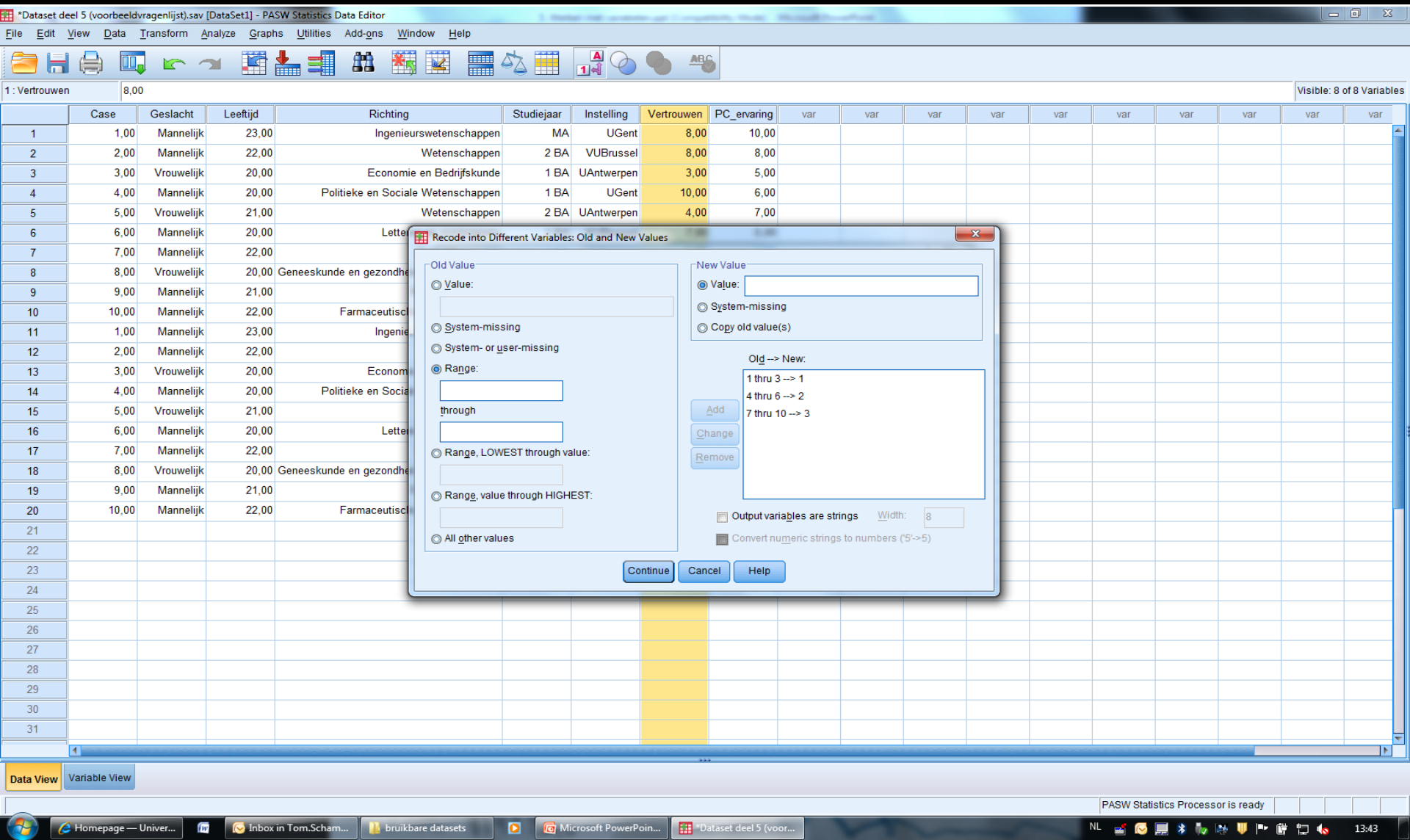
Highlight de "*Vertrouwen*" vraag en klik op de pijl om *Vertrouwen* in de *Input Variable* box te plaatsen.
Typ vervolgens: "**VertrouwenLMH**" in *Output Variable: Name.* Klik op de **Change** button.


*Dataset deel 5 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : Vertrouwen
8,00
Visible: 8 of 8 Variables
Case Geslacht Leeftijd Richting Studiejaar Instelling Vertrouwen PC_ervaring var
1 1,00 Mannelijk 23,00 Ingenieurswetenschappen MA UGent 8,00 10,00
2 2,00 Mannelijk 22,00 Wetenschappen 2 BA VUBrussel 8,00 8,00
3 3,00 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3,00 5,00
4 4,00 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10,00 6,00
5 5,00 Vrouwelijk 21,00 Wetenschappen 2 BA UAntwerpen 4,00 7,00
6 6,00 Mannelijk 20,00 Letteren en wijsbegeerte 1 BA KUBrussel 7,00 5,00
7 7,00 Mannelijk 22,00 Dier
8 8,00 Vrouwelijk 20,00 Geneeskunde en gezondheidswe
9 9,00 Mannelijk 21,00 Rech
10 10,00 Mannelijk 22,00 Farmaceutische We
11 1,00 Mannelijk 23,00 Ingenieurswe
12 2,00 Mannelijk 22,00 We
13 3,00 Vrouwelijk 20,00 Economie en
14 4,00 Mannelijk 20,00 Politieke en Sociale We
15 5,00 Vrouwelijk 21,00 We
16 6,00 Mannelijk 20,00 Letteren en
17 7,00 Mannelijk 22,00 Dier
18 8,00 Vrouwelijk 20,00 Geneeskunde en gezondheidswe
19 9,00 Mannelijk 21,00 Rech
20 10,00 Mannelijk 22,00 Farmaceutische We
Recode into Different Variables
Case
Geslacht
Leeftijd
Richting
Studiejaar
Instelling
Hoeveel jaar compu...
Numeric Variable -> Output Variable:
Vertrouwen --> VertrouwenLMH
Output Variable
Name:
VertrouwenLMH
Label:
Vertrouwen (laag, middeln
Change
Old and New Values...
If... (optional case selection condition)
OK Paste Reset Cancel Help
Data View
Variable View
PASW Statistics Processor is ready
Homepage — Univer...
Inbox in Tom.Scham...
bruikbare datasets
Microsoft PowerPoin...
*Dataset deel 5 (voor...
NL
13:37

- De SPSS gebruiker geeft **7 mogelijkheden om de oude code aan te geven**:
  - **Value**: om één enkele code aan te geven
  - **System-missing**: selecteert alle cases met een system-missing value (dit zijn waarden op een variabele die niet ingevuld zijn, in de datamatrix zijn deze te herkennen door een puntje)
  - **System- or user-missing**: selecteert alle cases met een missing value (zelf gedefinieerd vb. 9 of een puntje)
  - **Range**: geeft een interval aan van oude codes, inclusief de opgegeven grenzen
  - **Range: lowest**: geeft een interval aan vanaf de laagste waarde van de variabele tot en met de opgegeven grens
  - **Range: … through highest**: geeft een interval aan vanaf een opgegeven grens tot en met de hoogste waarde van de variabele
  - **All other values**: selecteert de nog niet aangewezen waarden

Selecteer vervolgens : **Old and New Values.** Onder *Old Value*, selecteer: **Range.**
Typ: "1" in de top box en "3" in de box onder *through*.
Typ: "1" in the *Value* box under *New Value*. Click: **Add.**

Herhaal voor
[4;6] -> "2"
[7; 10] -> "3"


*Dataset deel 5 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : Vertrouwen
8,00
Visible: 8 of 8 Variables
Case Geslacht Leeftijd Richting Studiejaar Instelling Vertrouwen PC_ervaring
1,00 Mannelijk 23,00 Ingenieurswetenschappen MA UGent 8,00 10,00
2,00 Mannelijk 22,00 Wetenschappen 2 BA VUBrussel 8,00 8,00
3,00 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3,00 5,00
4,00 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10,00 6,00
5,00 Vrouwelijk 21,00 Wetenschappen 2 BA UAntwerpen 4,00 7,00
6,00 Mannelijk 20,00
7,00 Mannelijk 22,00
8,00 Vrouwelijk 20,00 Geneeskunde en gezondhe
9,00 Mannelijk 21,00
10,00 Mannelijk 22,00 Farmaceutisch
1,00 Mannelijk 23,00 Ingenie
2,00 Mannelijk 22,00
3,00 Vrouwelijk 20,00 Econom
4,00 Mannelijk 20,00 Politieke en Socia
5,00 Vrouwelijk 21,00
6,00 Mannelijk 20,00 Letter
7,00 Mannelijk 22,00
8,00 Vrouwelijk 20,00 Geneeskunde en gezondhe
9,00 Mannelijk 21,00
10,00 Mannelijk 22,00 Farmaceutisch
Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
through
Range, LOWEST through value:
Range, value through HIGHEST:
All other values
New Value
Value:
System-missing
Copy old value(s)
Old --> New:
1 thru 3 --> 1
4 thru 6 --> 2
7 thru 10 --> 3
Add
Change
Remove
Output variables are strings
Width:
8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
Data View
Variable View
PASW Statistics Processor is ready
Homepage — Univer...
Inbox in Tom.Scham...
bruikbare datasets
Microsoft PowerPoin...
*Dataset deel 5 (voor...
NL
13:43

- Klik **Continue > OK**

```
SYNTAX

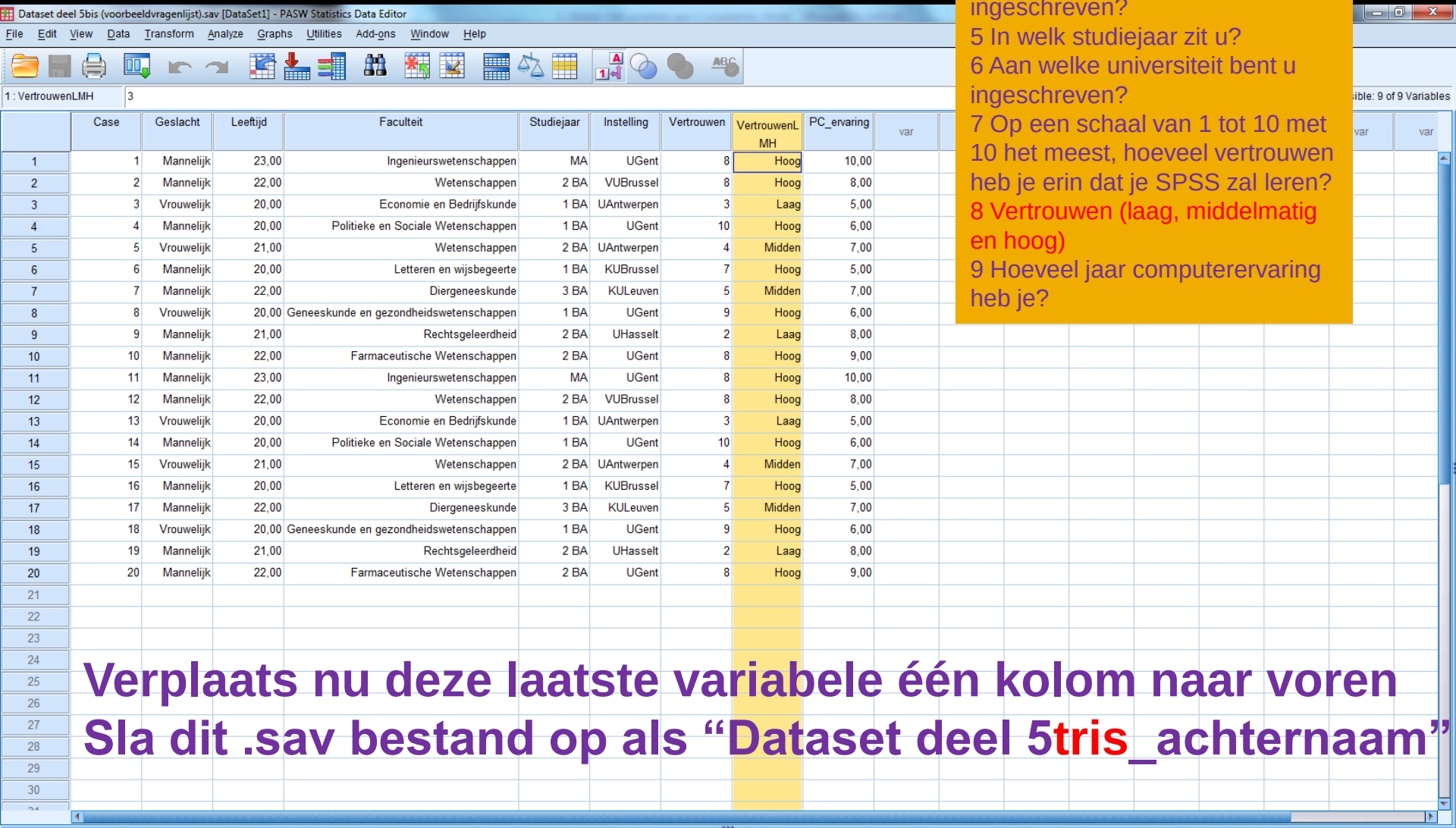
RECODE Vertrouwen (1 thru 3=1) (4 thru 6=2) (7 thru
10=3) INTO VertrouwenLMH.
VARIABLE LABELS  VertrouwenLMH 'Vertrouwen (laag,
middelmatig en hoog)'.
EXECUTE.
```

- De nieuwe variabele verschijnt in een nieuw toegevoegde kolom rechts.

- Ga naar: **Variable View**
- Selecteer op lijn (***VertrouwenLMH***) in de *Values* kolom de grijze cel met de drie dots.

- Definieer de waardes voor elke case:
  - "1" voor "Laag" in de *Label* box + **Add.**
  - "2" voor "Middelmatig" + **Add.**
  - "3" voor "Hoog" + **Add.**

1 Deelnemer
2 Wat is uw geslacht?
3 Hoe oud bent u?
4 Aan welke faculteit bent u ingeschreven?
5 In welk studiejaar zit u?
6 Aan welke universiteit bent u ingeschreven?
7 Op een schaal van 1 tot 10 met 10 het meest, hoeveel vertrouwen heb je erin dat je SPSS zal leren?
8 Vertrouwen (laag, middelmatig en hoog)
9 Hoeveel jaar computerervaring heb je?



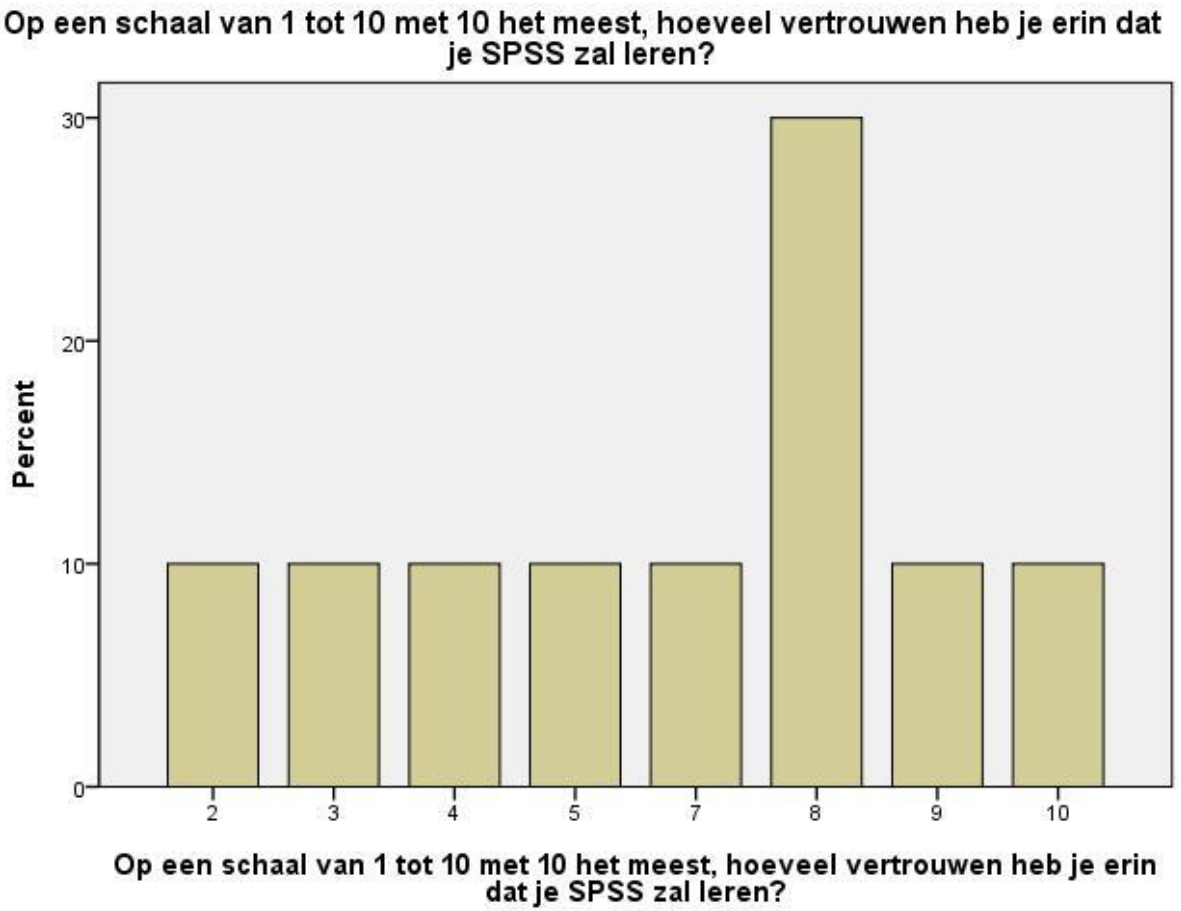
	Case	Geslacht	Leeftijd	Faculteit	Studiejaar	Instelling	Vertrouwen	VertrouwenLMH	PC_ervaring	var	var	var
1	1	Mannelijk	23,00	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00			
2	2	Mannelijk	22,00	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00			
3	3	Vrouwelijk	20,00	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00			
4	4	Mannelijk	20,00	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00			
5	5	Vrouwelijk	21,00	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00			
6	6	Mannelijk	20,00	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00			
7	7	Mannelijk	22,00	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00			
8	8	Vrouwelijk	20,00	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00			
9	9	Mannelijk	21,00	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00			
10	10	Mannelijk	22,00	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00			
11	11	Mannelijk	23,00	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00			
12	12	Mannelijk	22,00	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00			
13	13	Vrouwelijk	20,00	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00			
14	14	Mannelijk	20,00	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00			
15	15	Vrouwelijk	21,00	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00			
16	16	Mannelijk	20,00	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00			
17	17	Mannelijk	22,00	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00			
18	18	Vrouwelijk	20,00	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00			
19	19	Mannelijk	21,00	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00			
20	20	Mannelijk	22,00	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00			
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												

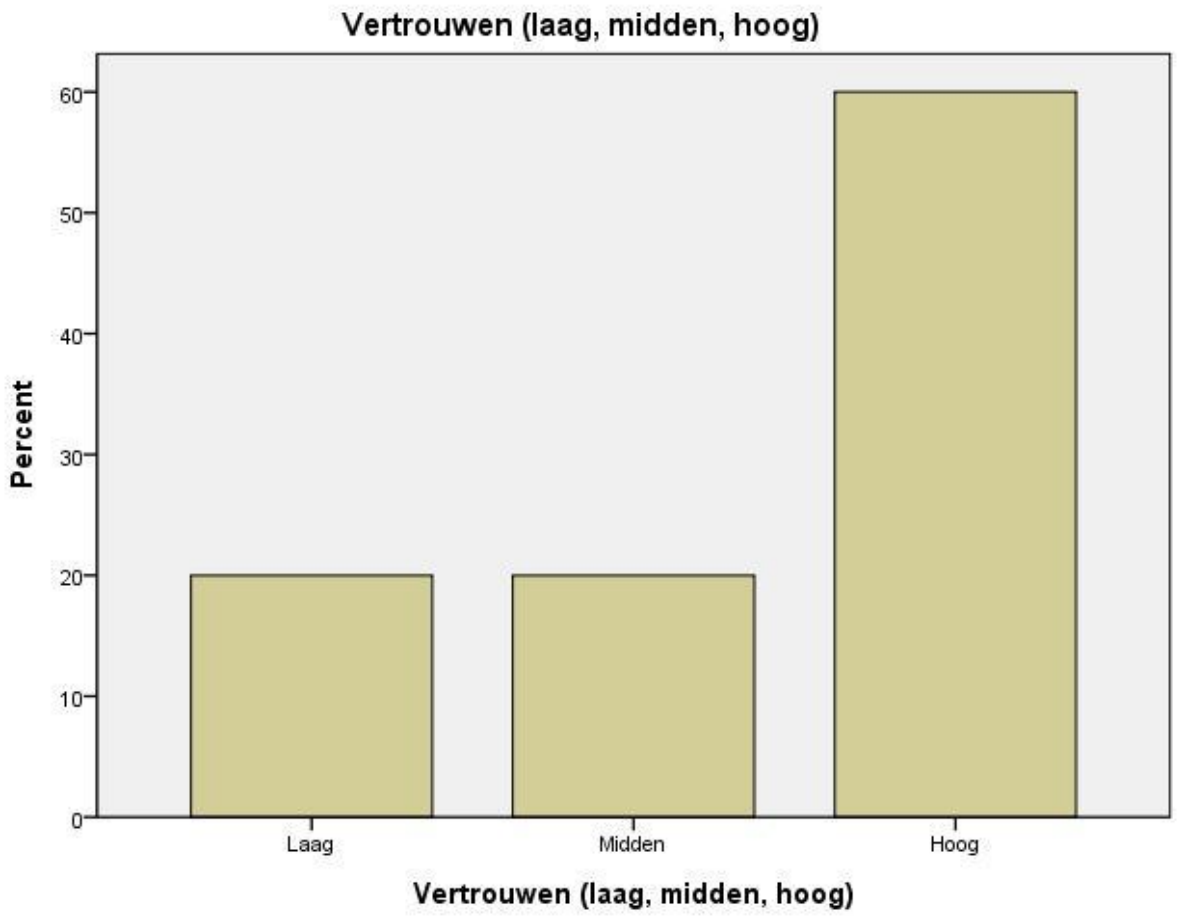
Verplaats nu deze laatste variabele één kolom naar voren
Sla dit .sav bestand op als "Dataset deel 5tris_achternaam"

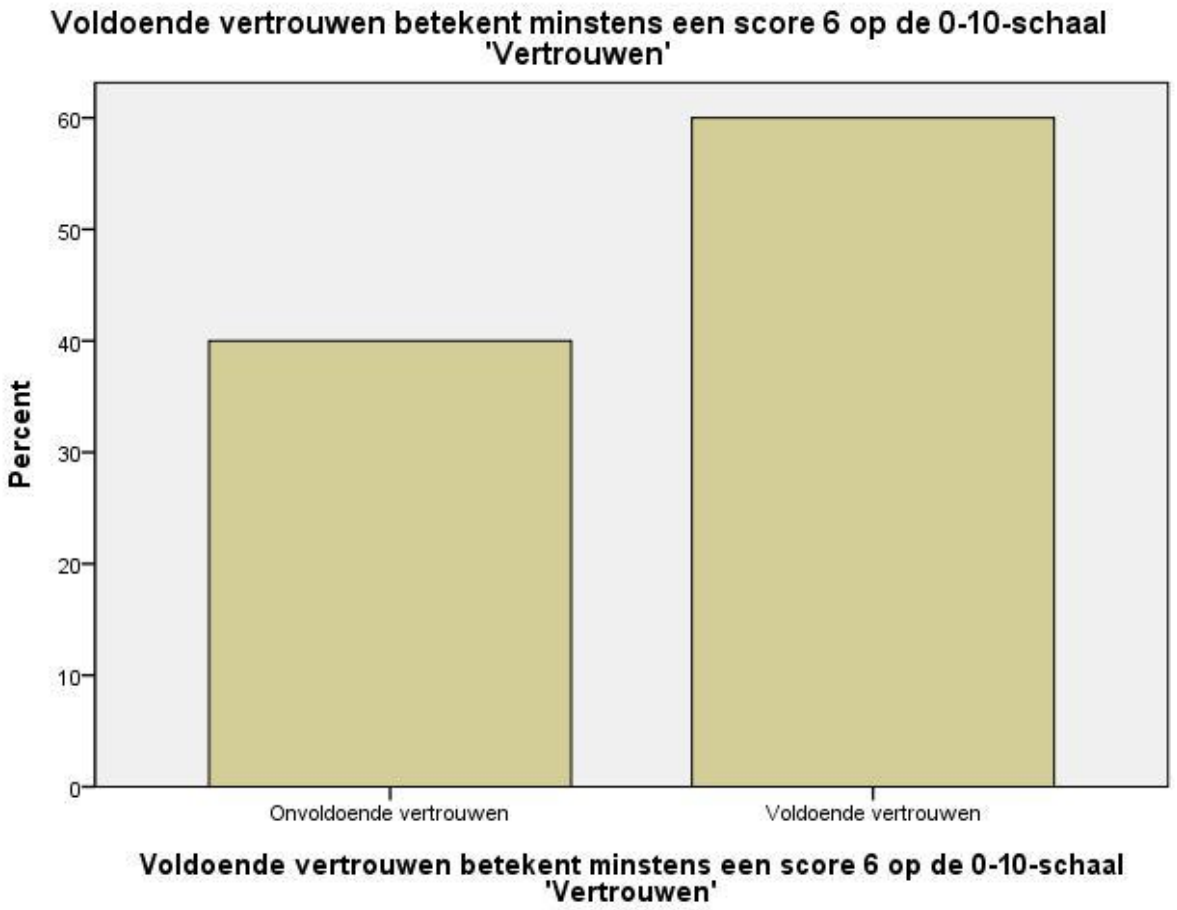
Variatie

- Stap 1: nieuwe variabele aanmaken die enkel de gespecifieerde waarde aanduidt (met een 1)
- Stap 2: frequentie bekijken voor deze nieuwe waarde (0 of 1) -> geeft het aantal/% dat de waarde die aan de voorwaarde voldoet voorkomt
- Voorbeeld: '*Vertrouwen*' (1-10) -> '*VertrouwenLMH*' -> '*Vertrouwen_voldoende*' (voldoende aanwezig of niet)

```
FREQUENCIES VARIABLES=Vertrouwen VertrouwenLMH
Vertrouwen_voldoende
  /BARCHART PERCENT
  /ORDER=ANALYSIS.
```




Vertrouwen (laag, midden, hoog)
Percent
60
50
40
30
20
10
0
Laag
Midden
Hoog
Vertrouwen (laag, midden, hoog)


Voldoende vertrouwen betekent minstens een score 6 op de 0-10-schaal 'Vertrouwen'
Percent
60
50
40
30
20
10
0
Onvoldoende vertrouwen
Voldoende vertrouwen
Voldoende vertrouwen betekent minstens een score 6 op de 0-10-schaal 'Vertrouwen'

Andere voorbeelden van hercoderen van continue variabelen

- Leeftijd -> leeftijdsklassen
- Inkomen -> inkomensklassen
- Etc.

HOE ?

Of: verdelen volgens gelijke klassenbreedtes (cf. hoger) -> meestal op basis van mediaanwaarde

(bv. *Vertrouwen*: 1 en 2; 3 en 4; 5 en 6; 7 en 8; en 9 en 10)

Of: verdelen volgens de nested-means methode

Of: verdelen volgens bestaande voorbeelden, theoretische consensus

Schaalomkering

Het omkeren van de schaal voor de VAR *Vertrouwen* op Likert-schaal (1->10 wordt 10->1) door gebruik te maken van een omkeringsconstante (*11* in dit geval)

= aantal antwoordcategorieën + 1

Nieuwe VAR = omkeringsconstante – om te keren VAR

Nieuwe VAR	omkeringsconstante	Om te keren VAR
1	11	10
2		9
3		8
4		7
5		6
6		5
7		4
8		3
9		2
10		1

Oefening

- Omkeren de schaal voor de volgende VAR:
  - *VRP_belangrijkheid*
  - *effectiviteit_lobbyingkabinetten*

Opmerking: maak steeds een nieuwe VAR aan (geen vervanging van de reeds bestaande!)

Tekst hercoderen (open vragen)

- Stap 1: FREQ van ruwe tekstvariabele

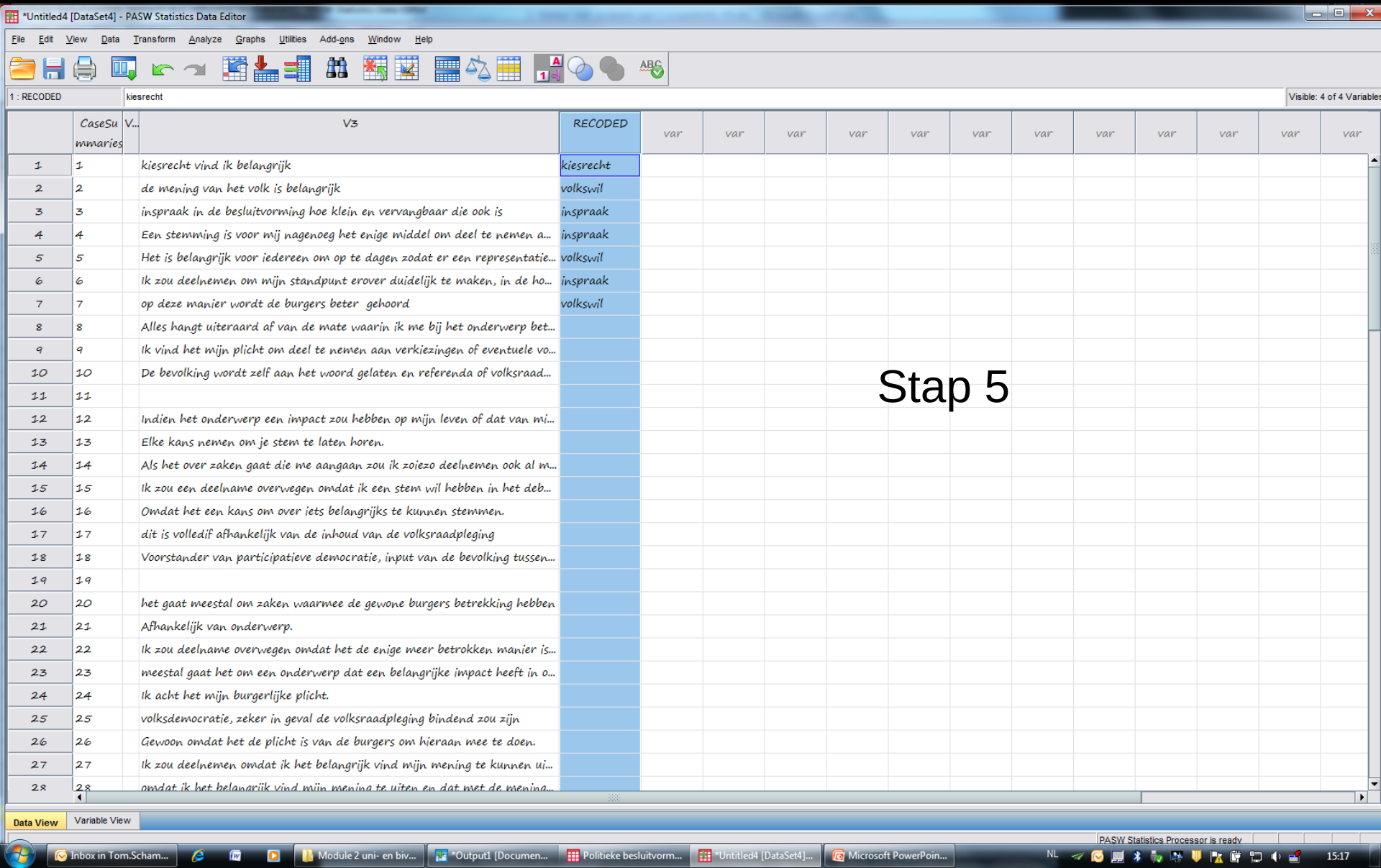
  Via **Analyse** > **Reports** > **Case summaries**
- Stap 2: output opslaan in Excel (exporteren)
- Stap 3: in excel hercoderen (manueel en item per item)
- Stap 4: excel-bestand opnieuw inladen in SPSS (importeren) -> nieuwe SPSS-file (SPSS-bestanden samenvoegen, zie verder)
- Stap 5: FREQ van **gehercodeerde tekst**

Korte demonstratie

Stap 4

Case Summaries[a]

	Case Number	Waarom overweegt u om wel of niet aan een volksraadpleging deel te nemen? (antwoord beperkt tot 3 regels) Ik zou deelname (niet) overwegen om volgende redenen: - Open-Ended Response	VAR_RECODED
1	1	kiesrecht vind ik belangrijk	kiesrecht
2	2	de mening van het volk is belangrijk	volkswil
3	3	inspraak in de besluitvorming hoe klein en vervangbaar die ook is	inspraak
4	4	Een stemming is voor mij nagenoeg het enige middel om deel te nemen aan de democratie.	volkswil
5	5	Het is belangrijk voor iedereen om op te dagen zodat er een representatieve opkomst is.	inspraak
6	6	Ik zou deelnemen om mijn standpunt erover duidelijk te maken, in de hoop dat ik iets kan veranderen.	inspraak
7	7	op deze manier wordt de burgers beter gehoord	volkswil


*Untitled4 [DataSet4] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : RECODED
kiesrecht
Visible: 4 of 4 Variables
CaseSummaries
V...
V3
RECODED
var
1 kiesrecht vind ik belangrijk kiesrecht
2 de mening van het volk is belangrijk volkswil
3 inspraak in de besluitvorming hoe klein en vervangbaar die ook is inspraak
4 Een stemming is voor mij nagenoeg het enige middel om deel te nemen a... inspraak
5 Het is belangrijk voor iedereen om op te dagen zodat er een representatie... volkswil
6 Ik zou deelnemen om mijn standpunt erover duidelijk te maken, in de ho... inspraak
7 op deze manier wordt de burgers beter gehoord volkswil
8 Alles hangt uiteraard af van de mate waarin ik me bij het onderwerp bet...
9 Ik vind het mijn plicht om deel te nemen aan verkiezingen of eventuele vo...
10 De bevolking wordt zelf aan het woord gelaten en referenda of volksraad...
11
12 Indien het onderwerp een impact zou hebben op mijn leven of dat van mi...
13 Elke kans nemen om je stem te laten horen.
14 Als het over zaken gaat die me aangaan zou ik zoiezo deelnemen ook al m...
15 Ik zou een deelname overwegen omdat ik een stem wil hebben in het deb...
16 Omdat het een kans om over iets belangrijks te kunnen stemmen.
17 dit is volledif afhankelijk van de inhoud van de volksraadpleging
18 Voorstander van participatieve democratie, input van de bevolking tussen...
19
20 het gaat meestal om zaken waarmee de gewone burgers betrekking hebben
21 Afhankelijk van onderwerp.
22 Ik zou deelname overwegen omdat het de enige meer betrokken manier is...
23 meestal gaat het om een onderwerp dat een belangrijke impact heeft in o...
24 Ik acht het mijn burgerlijke plicht.
25 volksdemocratie, zeker in geval de volksraadpleging bindend zou zijn
26 Gewoon omdat het de plicht is van de burgers om hieraan mee te doen.
27 Ik zou deelnemen omdat ik het belangrijk vind mijn mening te kunnen ui...
28 omdat ik het belangrijk vind mijn mening te uiten en dat met de mening...
Stap 5
Data View
Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham...
Module 2 uni- en biv...
*Output1 [Documen...
Politieke besluitvorm...
*Untitled4 [DataSet4]...
Microsoft PowerPoin...
NL
15:17

Stap 6

RECODED

		Frequency	Percent	Valid Percent	Cumulative Percent
		55	88,7	88,7	88,7
Valid	inspraak	3	4,8	4,8	93,5
	kiesrecht	1	1,6	1,6	95,2
	volkswil	3	4,8	4,8	100,0
	Total	62	100,0	100,0	

Werken met bestanden

- **Drie technieken**
  - Een bestand toevoegen of bestanden samenvoegen (cases en/of variabelen)
  - Een bestand opdelen
  - Een bestand aggregeren

Cases / variabelen samenvoegen

= 'Concateneren' van bestanden, of invoegen van reeds bestaande data(sets) tot nieuwe grotere

- Via **Data** > **Merge files** > **Add cases**
  > **Add variables**

Een alternatief = "copy-paste" van de betreffende kolom(men)…

Een databestand opsplitsen

- Via **Data** > **Split file**

  Keuze tussen ofwel tussen:

  > **Analyze all cases, do not compare groups**

  > **Compare groups** (cf. 1)

  > **Organize output by groups** (cf. 2)

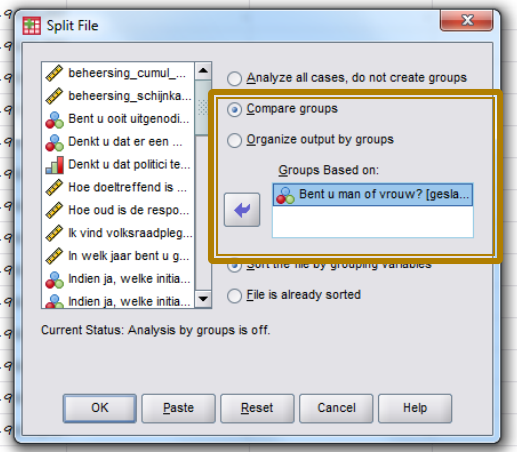
  Voorbeeld: alle *analyses* opsplitsen volgens de categorieën in de VAR ***Geslacht***
  (= twee groepen)

Politieke besluitvorming_opdracht les2_oplossing.sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Visible: 41 of 41 Variables

	RespondentID	geslacht	postcode_domicilie	postcode_verblijf	geboortejaar	leeftijd	Opwelkepartijstemdeubijdelaatstefederaleverkiezingvan	Opwelkepartijstemdeubijdelaatstefederaleverkiezingvan_A	Opwelkepartijstemdeubijdelaatstefederaleverkiezingvan_B	Opwelkepartijstemdeubijdelaatstefederaleverkiezingvan_C	Opwelkepartijstemdeubijdelaatstefederaleverkiezingvan_D	Opwelkepartijstemdeubijdelaatstefederaleverkiezingvan_E	Op ijd
1	1363922033	man	8020	9000	1989	1981,00	.	.	.	.	5	.	
2	1363913804	man	8760	9000	1990	1980,00	.	.	3	.	.	.	
3	1363896960	man	8380	9000	1990	1980,00	.	.	.	.	5	.	
4	1363873747	man	9450	9450	1990	1980,00	.	.	.	4	.	.	
5	1363562632	vrouw	8020	9000	1990	19				.	5	.	
6	1363275243	vrouw	9300	9300	1990	19				.	5	.	
7	1362566526	vrouw	1540	9000	1990	19				.	.	.	
8	1362544722	man	9000	9000	1987	19				.	5	.	
9	1361198842	man	8434	9000	1988	19				.	.	.	
10	1361149015	vrouw	9620	9000	1989	19				.	.	6	
11	1360249342	vrouw	9300	9300	1990	19				.	.	.	
12	1360245180	vrouw	8870	9000	1988	19				4	.	.	
13	1360227472	man	9290	9000	1988	19				.	5	.	
14	1360008770	vrouw	9940	.	1988	19				4	.	.	
15	1359973192	vrouw	2800	9000	1990	19				.	.	.	
16	1359932768	vrouw	9830	9000	1990	19				.	5	.	
17	1359758498	vrouw	1540	9000	1990	19				.	.	.	
18	1359681912	man	1850	9000	1987	19				.	5	.	
19	1359426086	man	8720	9000	1990	1980,00	.	.	.	.	.	.	
20	1359313404	man	9990	9990	1990	1980,00	.	.	3	.	.	.	
21	1359180663	man	3800	9000	1990	1980,00	.	.	.	.	.	.	
22	1359101274	vrouw	8000	8000	1990	1980,00	.	.	.	.	5	.	
23	1359004494	man	9240	9240	1989	1981,00	.	.	.	.	.	.	
24	1358909557	man	9820	9820	1988	1982,00	.	.	.	.	.	.	
25	1358823686	vrouw	9050	9000	1988	1982,00	.	.	.	.	.	.	
26	1358715033	man	8210	8210	1988	1982,00	.	.	3	.	.	.	
27	1358700585	man	8000	8000	1988	1982,00	.	.	.	4	.	.	



Keuze leidt tot twee alternatieve presentaties van dezelfde analyseresultaten

Data View | Variable View

PASW Statistics Processor is ready

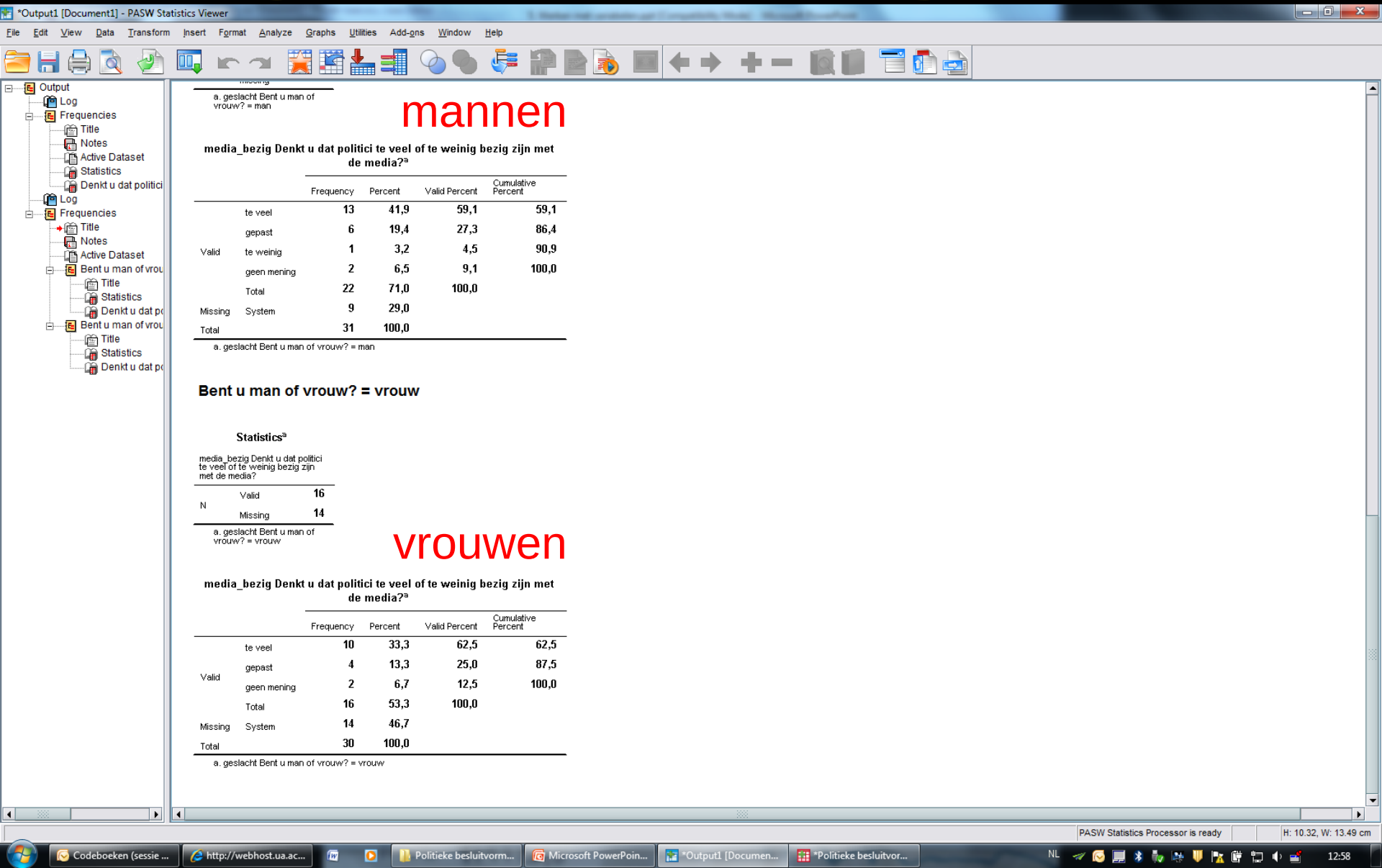
Codeboeken (sessie ... | http://webhost.ua.ac... | Politieke besluitvorm... | Microsoft PowerPoin... | *Output1 [Documen... | Politieke besluitvorm... | NL | 12:52

1. Compare groups

media_bezig Denkt u dat politici te veel of te weinig bezig zijn met de media?

geslacht Bent u man of vrouw?			Frequency	Percent	Valid Percent	Cumulative Percent
man	Valid	te veel	13	41,9	59,1	59,1
		gepast	6	19,4	27,3	86,4
		te weinig	1	3,2	4,5	90,9
		geen mening	2	6,5	9,1	100,0
		Total	22	71,0	100,0	
	Missing	System	9	29,0		
	Total		31	100,0		
vrouw	Valid	te veel	10	33,3	62,5	62,5
		gepast	4	13,3	25,0	87,5
		geen mening	2	6,7	12,5	100,0
		Total	16	53,3	100,0	
	Missing	System	14	46,7		
	Total		30	100,0		

2. Organize output by groups



Bestanden aggregeren

- Via **Data** > **Aggregate**
- Van een set data voor één specifieke variabele een aggregatie (= aggregated variable(s)) maken op basis van een specifiek kenmerk (= **break variable**)

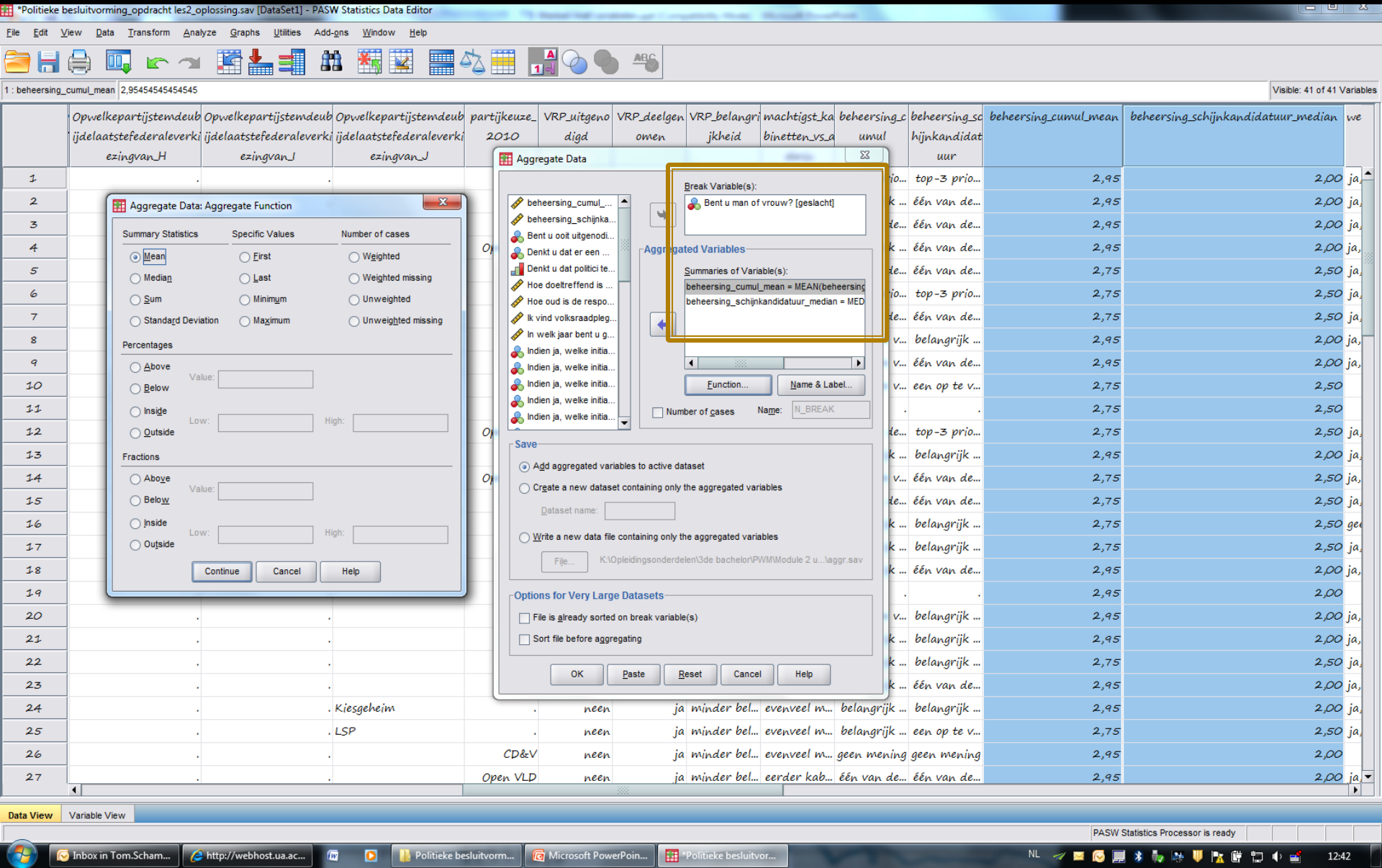
Voorbeeld: geaggregeerd gemiddelde waarde voor de VAR *Opleiding* (*groepsvariabele en – waarde*) voor de VAR *Geslacht* (***groepeervariabele***) -> resultaat 2 waarden (gemiddelde voor alle mannen en voor alle vrouwen)

Oefening

- Doe hetzelfde in de Dataset politieke besluitvorming voor de volgende VAR:

VAR *beheersing_cumul* op basis van het gemiddelde
VAR *beheersing_schijnkandidatuur* obv. de mediaan

Oefening


*Politieke besluitvorming_opdracht les2_oplossing.sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : beheersing_cumul_mean 2,95454545454545
Visible: 41 of 41 Variables
Aggregate Data: Aggregate Function
Summary Statistics
Mean
Median
Sum
Standard Deviation
Specific Values
First
Last
Minimum
Maximum
Number of cases
Weighted
Weighted missing
Unweighted
Unweighted missing
Percentages
Above
Below
Inside
Outside
Value:
Low:
High:
Fractions
Above
Below
Inside
Outside
Value:
Low:
High:
Continue
Cancel
Help
Aggregate Data
Break Variable(s):
Bent u man of vrouw? [geslacht]
Aggregated Variables
Summaries of Variable(s):
beheersing_cumul_mean = MEAN(beheersing
beheersing_schijnkandidatuur_median = MED
Function...
Name & Label...
Number of cases
Name:
N_BREAK
Save
Add aggregated variables to active dataset
Create a new dataset containing only the aggregated variables
Dataset name:
Write a new data file containing only the aggregated variables
File...
K:\Opleidingsonderdelen\3de bachelor\PWM\Module 2 u...\aggr.sav
Options for Very Large Datasets
File is already sorted on break variable(s)
Sort file before aggregating
OK
Paste
Reset
Cancel
Help
Data View
Variable View
PASW Statistics Processor is ready

Opslaan en indienen oefening 2

- Sla het (bewerkte*) databestand op als

" *Politieke besluitvorming_dataset les 2_achternaam.sav* "

En drop in Minerva dropbox, ter attentie van mezelf (tom.schamp@ugent.be)

Ten laatste di 12 april om 12:30pm.
(Laattijdige indieningen worden niet meegenomen voor de evaluatie van het dagelijks werk)

*dat wil zeggen met in begrip van alle hierboven vermelde 'oefeningen'.

Deel 6

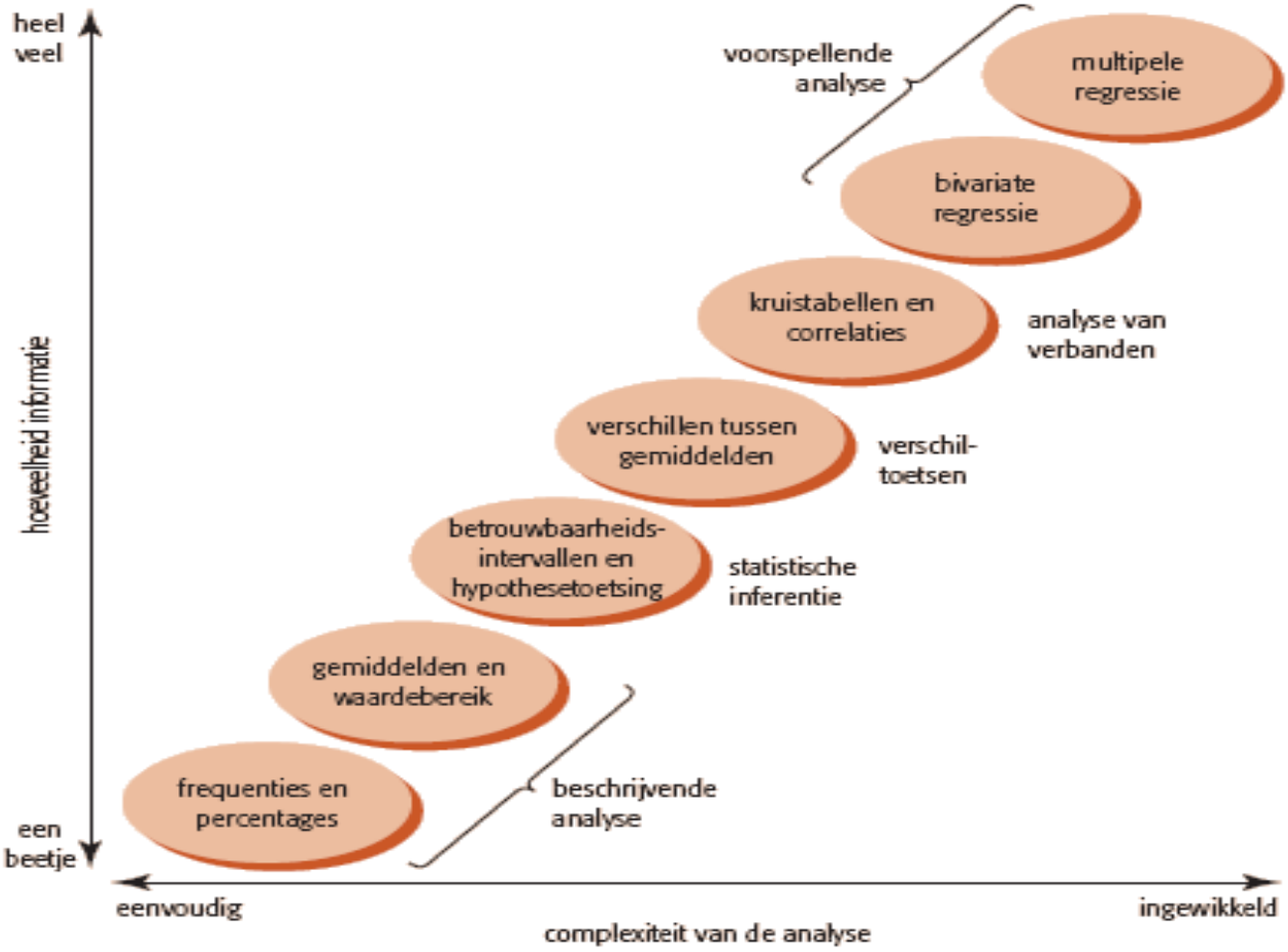
Eenvoudige beschrijvende statistiek

Soorten

- **Datareductie**: het beschrijven van een datamatrix door een klein aantal kenmerkende maten voor de dataverzameling te berekenen.

- **Vier functies van datareductie**:
  - samenvatten van de data
  - toepassen van begrijpelijke conceptualiseringen
  - overbrengen van onderliggende patronen
  - generaliseren van de steekproefuitkomsten naar de populatie (bv. steekproefverdeling)

Indeling


heel veel
hoeveelheid informatie
een beetje
voorspellende analyse
multipele regressie
bivariate regressie
kruistabellen en correlaties
analyse van verbanden
verschillen tussen gemiddelden
verschil-toetsen
betrouwbaarheids-intervallen en hypothesetoetsing
statistische inferentie
gemiddelden en waardebereik
frequenties en percentages
beschrijvende analyse
eenvoudig
ingewikkeld
complexiteit van de analyse

Soorten - 2

- Vijf soorten statistische analyse
  - **Beschrijvende analyse**: gebruikt om de status van de variabelen te beschrijven (hoeveel eenheden, welke waarde voor welke variabelen?)
  - **Analyse van verschillen**: bepalen in welke mate er echte en generaliseerbare verschillen in de populatie bestaan
  - **Analyse van verbanden**: bepalen of er systematische verbanden tussen variabelen bestaan
  - **Inferentiële analyse**: statistische procedures gebruiken om de steekproefuitkomsten te generaliseren
  - **Voorspellende analyse**: hiermee kan de onderzoeker voorspellingen doen over toekomstige gebeurtenissen.

Beschrijvende statistieken

- Drie soorten maten worden gebruikt om de informatie uit een steekproef te beschrijven:

  - **(1) Centrummaten en positiematen**: de datareductie is erop gericht een enkel stuk informatie te rapporteren dat het meest typerende antwoord op een vraag beschrijft.

    - Modus
    - Mediaan
    - Gemiddelde $\bar{x} = \frac{\sum x_i \cdot f_i}{n}$
    - Kwantielen (percentielen, decielen, kwartielen)

Gewogen en niet-gewogen gemiddelden

	scores	totaal	procent	scores	totaal	procent	scores	totaal	procent	scores	totaal	procent
	17	20	0,85	34	40	0,85	34	40	0,85	85	100	0,85
	15	20	0,75	30	40	0,75	30	40	0,75	75	100	0,75
	10	20	0,5	20	40	0,5	15	30	0,5	25	50	0,5
	6	20	0,3	12	40	0,3	6	20	0,3	60	200	0,3
	7	20	0,35	14	40	0,35	7	20	0,35	70	200	0,35
gemiddelden												
niet-gewogen	11	20	0,55	22	40	0,55	18,4	30	0,55	63	130	0,55
gewogen	55	100	0,55	110	200	0,55	92	150	0,61	315	650	0,48

	scores	totaal	procent	**weging**
	85	**100**	0,85	**0,13**
	75	**100**	0,75	**0,12**
	25	**50**	0,5	**0,04**
	60	**200**	0,3	**0,09**
	70	**200**	0,35	**0,11**
gemiddelden				
niet-gewogen	63	130	0,55	
gewogen	315	**650**	0,48	**0,48**





Beschrijvende statistieken - 2

- **(2) Spreidingsmaten**: alle spreidingsmaten (measures of variability) geven het 'typische' verschil weer tussen de waarden in een verzameling.

  - Frequentieverdeling: een optelling van het aantal keren dat elke verschillende waarde in een bepaalde verzameling waarden voorkomt.
  - Waardebereik: geeft de afstand aan tussen de laagste waarde (minimum) en de hoogste waarde (maximum) in een geordende verzameling waarden. (= spread of *range*)

- **De standaardafwijking (voor steekproef / populatie)** (s / σ)
  - een "soort" gemiddelde van de absolute afwijkingen tussen de waarnemingsgetallen en het gemiddelde van die waarnemingsgetallen
  - ook hierbij maken we, in geval van klassen, gebruik van de middens van die klassen.

- **De variantie (voor steekproef / populatie)** (s^2 / σ^2)
  - het gemiddelde van de kwadraten van de waarnemingsgetallen verminderd met het kwadraat van het gemiddelde van die getallen

Statistics

Hoe oud bent u?

N	Valid	20
	Missing	0
Mean		21,1000
Std. Deviation		1,07115
Variance		1,147

Hoe oud bent u?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20,00	8	40,0	40,0	40,0
	21,00	4	20,0	20,0	60,0
	22,00	6	30,0	30,0	90,0
	23,00	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Voorbeeldje: uitwerking van s en s^2 (steekproef n = 20)

- **s** = $\sum$ (20-21,1)^2.8 + (21-21,1)^2.4 + (22-21,1)^2.6 + (23-21,1)^2.2 /20
  = (1,21.8 + 0,01.4 + 0,81.6 + 3,61.2) /20
  = (9,86+0,04+4,86+7,22) /20
  = 1,07

- **s^2** = 1,147

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Beschrijvende statistieken - 3

- Hoe hangt af van **het meetniveau** van elke variabele (Nominal, Ordinal, en Scale (Interval/ Ratio)).
  - Voor **nominale en ordinale variabelen**, zoals *Geslacht* en de gehercodeerde variabele *VertrouwenLMH* such as Gender, kunnen "frequencies" worden weergegeven (% of absolute aantallen mannen, vrouwen, laag vertrouwen etc.). (NIET-METRISCH, CATEGORISCH)
  - Voor **schaalvariabelen**, zoals *Leeftijd* kunnen ook nog Minimum, Maximum, Gemiddelde en de Standaarddeviatie worden opgevraagd. (= METRISCH, CONTINUE)

Beschrijvende statistieken - 2

- **(3) Vormmaten**
  - Normaalverdelingscurve
  - Skwesness (scheefheid)
  - Kurtosis (afplatting)

Samenvatting

	Meetniveau		
	Nominaal	**Ordinaal**	**Interval-Ratio**
Frequentieverdeling	*frequentietabel* *staafdiagram, taartdiagram*	*+ cumulatieve frequentietabel* *+ histogram, lijndiagram*	*Idem*
Centrum- en positiematen	*modus*	*+ mediaan, kwantielen*	*+ rekenkundig gemiddelde* *boxplot*
Spreidingsmaten		*variatie(breedte) en interkwartielafstand*	*+ variantie, standaardafwijking* *boxplot*
Vormmaat		*skewness (scheefheid)* *kurtosis (platheid)*	*skewness (scheefheid)* *kurtosis (platheid)*

De Morgen Economie - Belg verdient gemiddeld 2.936 euro bruto (1168660) - Windows Internet Explorer

http://www.demorgen.be/dm/nl/996/Economie/article/detail/1168660/2010/10/12/Belg-verdient-gemiddeld-2-936-euro-bruto.dhtml

File Edit View Favorites Tools Help

Favorites | De Morgen Economie - Belg verdient gemiddeld ...

Page Safety Tools

ECONOMIE | CONSUMENT

Belg verdient gemiddeld 2.936 euro bruto

12/10/10 14u16



Het gemiddelde maandloon van de Belg bedraagt 2.936 euro bruto (2008). Vijftig procent van de voltijds tewerkgestelde loontrekkende Belgen verdient minder dan 2.595 euro bruto (mediaanloon). Dat meldt de federale overheidsdienst Economie - Algemene Directie Statistiek en Economische Informatie.

Tien procent van de werknemers verdient per maand maximaal 1.911 euro, 20 pct verdient maximaal 2.106 euro. De 10 procent beste verdieners, trekken maandelijks minstens 4.303 euro. Bedrijfsleiders verdienen het best de boterham met gemiddeld 7.870 euro of zo'n 168 pct boven het nationaal gemiddelde.

Horecapersoneel bevindt zich helemaal onderaan de ladder met gemiddeld 1.975 euro of zo'n 33 pct onder het gemiddelde. Het beste loonbriefje valt te rapen in de petrochemie: gemiddeld 4.530 euro of zo'n 54 pct boven het gemiddelde. De nummers twee en drie, onderzoek en ontwikkeling en de chemische nijverheid, volgen op ruime afstand met gemiddeld 3.871 en 3.811 euro. Hotels, restaurants en drankgelegenheden betalen het slechtst (2.129 euro of 27,5 pct onder het gemiddelde).

Loon toegenomen

In de periode 1999-2008 nam het loon gemiddeld met 31,19 pct toe. De grootste stijging is voor kaderleden (+38,15 pct). Qua sector is de stijging (+39,82 pct) het grootst voor de slechtste betaler: hotels, restaurants en drankgelegenheden.

In het arrondissement Brussel worden de hoogste lonen betaald met gemiddeld 3.381. Dinant scoort het laagst: 2.247 euro. De werknemers in Oostende zagen hun loon tussen 1999 en 2008 het sterkst toenemen, zo'n 40 pct. (belga/dea)

EXTRA

ARTIKELS

- Belg verdient gemiddeld 2.936 euro bruto
- Fedis pleit voor afschaffen systeem van maaltijdcheques
- Loon niet belangrijkste reden om trouw te blijven aan bedrijf

MEER OVER Werk

NUTTIGE LINKS

- De Stand van het Land: Schuldenklok
- Website VDAB

15u37 Levenslang hotelverblijf als geschenk na bevalling in lobby

15u30 Vrijspraak voor valse bommeldingen in Kortrijkse dancing

15u29 Rafael Nadal pleit voor langere 'winterstop'

5 MEEST GELEZEN

- Oudste zussenpaar van het land is niet meer
- Meer spoorchaos op komst
- Milquet: "Er valt niets te verduidelijken, alles ligt op tafel"
- Wandelgangen
- Open Vld trekt naar het parlement met wetsvoorstel voor strengere migratie

FOTO-wedstrijd
Succesvol fotograferen

Welke vragen moeten we ons bij deze (voorstelling) van onderzoeksresultaten stellen ?

Nu zaterdag GRATIS

Internet | Protected Mode: On 100%

De Morgen Economi... NL 16:00

REACTIE OP COVERVERHAAL '63.193 M/V OVER HUN SALARIS'

Is Salarisenquête genoeg representatief?

Ik stel me toch wat vragen in verband met de representativiteit van de steekproef. Vooral diegene die menen dat hun salaris aan de lagere kant ligt, zijn volgens mij geïnteresseerd en vullen de enquête in. Het zijn dus niet alleen ceo's die niet antwoorden maar ook heel wat hogere salarissen (vb. vanaf 5.000 euro/maand). Het zou nuttig zijn een korte beschrijving te geven hoe de steekproef tot stand is gekomen. Vrijwillig of gekozen?
Reactie via website

U heeft gelijk dat we met de Salarisenquête de echte toppers niet of slechts beperkt bereiken (deze mensen hebben wellicht geen tijd). Het klopt echter niet dat mensen met een laag salaris oververtegenwoordigd zijn. De afwijkingen ten opzichte van de populatie zijn te wijten aan de methode van dataverzameling: niet via een steekproef maar via zelfselectie (respondenten kiezen zelf om deel te nemen). Het databestand is dan ook vooral een afspiegeling van het lezerspubliek van Vacature. Je moet dus altijd voorzichtig omgaan met een veralgemening van de resultaten. Maar dankzij een zorgvuldige weging kunnen we toch 'indicatieve' uitspraken doen over de gemiddelde lonen van Belgische werknemers.
Gert Theunissen, onderzoeker K.U.Leuven en Luc Sels, decaan Faculty of Business and Economics K.U.Leuven, in opdracht van Vacature

Eerlijke loonsystemen zijn dungezaaid

Hoe komt het toch dat loon nog zo'n taboe is op de arbeidsmarkt? Waarom gooien de werkgevers de lonen niet op het intranet, zodat alle werknemers op de hoogte zijn van elkaars loon? Maar dan heb je natuurlijk een kristalhelder, transparant en bovenal eerlijk loonsysteem nodig waar de wetten voor iedereen gelijk zijn. En niet alleen voor het 'gewone' salaris, maar ook wat extralegale voordelen en bonussen betreft. Koppel die bonussen aan duidelijke doelstellingen, die iedereen kent en aanvaardt. De realiteit is echter dat in de gemiddelde kmo er heel veel mist over de lonen hangt. En dat voedt de speculaties en misverstanden. **Jan De Backer, via e-mail**

WIL U REAGEREN? MAIL NAAR

publiekeopinie@vacature.com

POLL



Schuldig verzu

Terwijl de pedofilieschandalen nieuwe schokgolven rondsturen re lekkere brok op voor de me werk en hun slachtoffers. Sapp de beelden, onthutste commen zo snel? – nu ook al de obliga blijken dat zomaar eventjes 1 gepest voelt. Los van het feit da koesteren bij studies die bijna voorpagina's verschijnen, lijdt he op het werk een onderschat pr met pedofilie, pesten op school het geval is. En dat het fenom verdient, op alle mogelijke vlakk



"Wat gedac vakbondsa die pesten de vingers

FILIP MICHIELS, RED

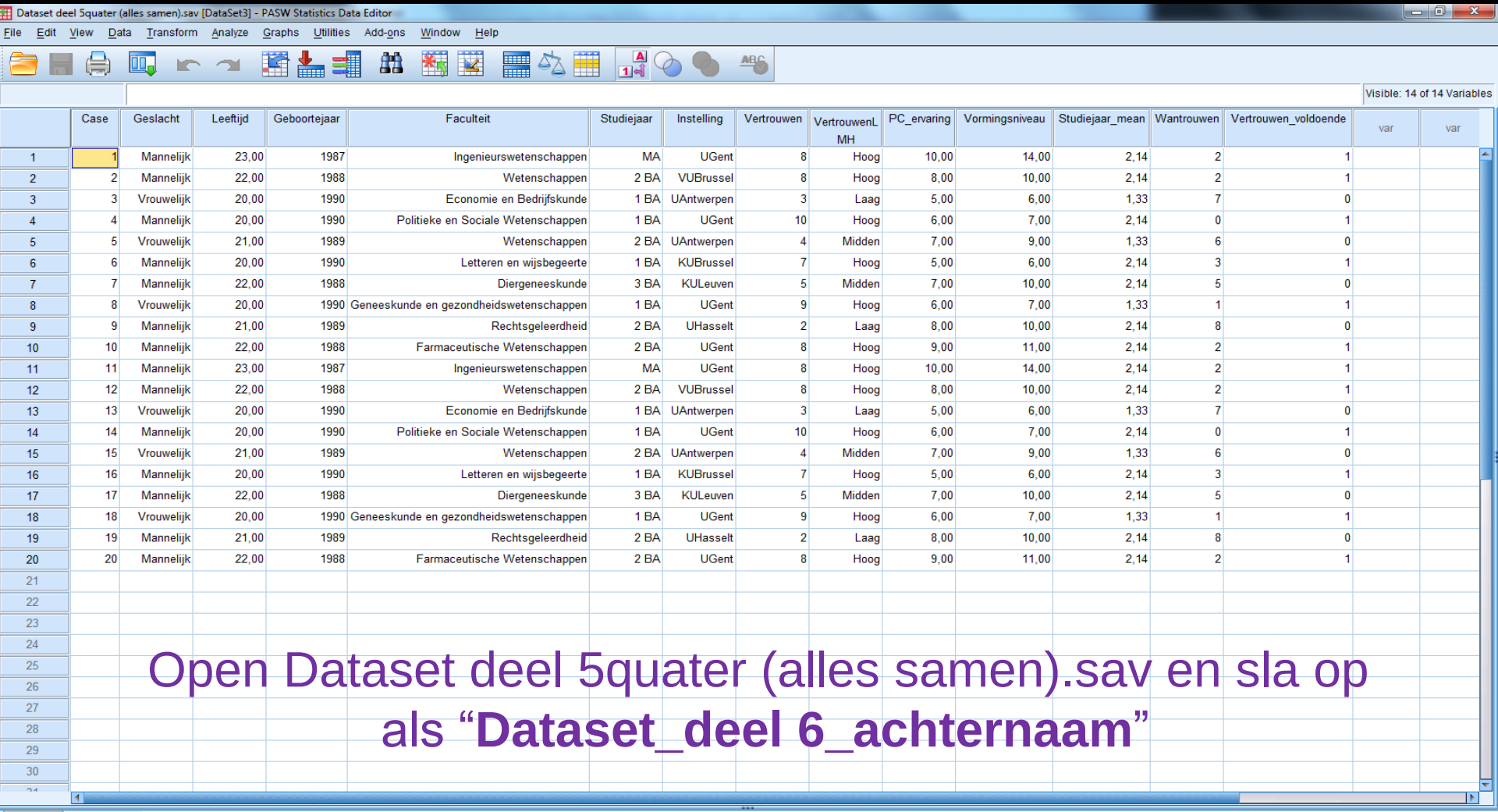
Tot daar evenwel ook het beg heisa van de voorbije dagen. A van uitgaan dat één op tien werk terijen moet ondergaan, dan he allemaal boter op het hoofd. Of in een middelgroot bedrijf van p waarvan er dan gemiddeld 20 g ga's dat niét zouden merken?

Internet | Protected Mode: On
12010 - ...
NL

Indeling

- Univariate beschrijvende analyse
- Bivariate beschrijvende analyse

Laatste aangemaakte dataset



	Case	Geslacht	Leeftijd	Geboortejaar	Faculteit	Studiejaar	Instelling	Vertrouwen	VertrouwenLMH	PC_ervaring	Vormingsniveau	Studiejaar_mean	Wantrouwen	Vertrouwen_voldoende	var	var
1	1	Mannelijk	23,00	1987	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00	14,00	2,14	2	1		
2	2	Mannelijk	22,00	1988	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00	10,00	2,14	2	1		
3	3	Vrouwelijk	20,00	1990	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00	6,00	1,33	7	0		
4	4	Mannelijk	20,00	1990	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00	7,00	2,14	0	1		
5	5	Vrouwelijk	21,00	1989	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00	9,00	1,33	6	0		
6	6	Mannelijk	20,00	1990	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00	6,00	2,14	3	1		
7	7	Mannelijk	22,00	1988	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00	10,00	2,14	5	0		
8	8	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00	7,00	1,33	1	1		
9	9	Mannelijk	21,00	1989	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00	10,00	2,14	8	0		
10	10	Mannelijk	22,00	1988	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00	11,00	2,14	2	1		
11	11	Mannelijk	23,00	1987	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00	14,00	2,14	2	1		
12	12	Mannelijk	22,00	1988	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00	10,00	2,14	2	1		
13	13	Vrouwelijk	20,00	1990	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00	6,00	1,33	7	0		
14	14	Mannelijk	20,00	1990	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00	7,00	2,14	0	1		
15	15	Vrouwelijk	21,00	1989	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00	9,00	1,33	6	0		
16	16	Mannelijk	20,00	1990	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00	6,00	2,14	3	1		
17	17	Mannelijk	22,00	1988	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00	10,00	2,14	5	0		
18	18	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00	7,00	1,33	1	1		
19	19	Mannelijk	21,00	1989	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00	10,00	2,14	8	0		
20	20	Mannelijk	22,00	1988	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00	11,00	2,14	2	1		
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																

Open Dataset deel 5quater (alles samen).sav en sla op als “**Dataset_deel 6_achternaam**”

Frequenties

- Selecteer: **Analyze > Descriptive Statistics > Frequencies**
- Highlight *Geslacht* en verplaatst *Geslacht naar de* Variable(s) box.
- Highlight *Vertrouwen LMH* en verplaats naar de Variable(s) box
- Zorg ervoor dat *Display Frequency Tables* is aangevinkt!
- Selecteer **Charts**
- Bij de *Frequencies: Charts* box, selecteer: **Bar charts** en **Percentages** > **Continue > OK.**

```
SYNTAX

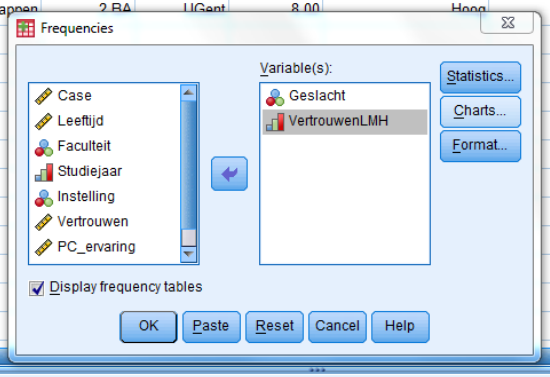
FREQUENCIES VARIABLES=Geslacht VertrouwenLMH
  /BARCHART PERCENT
  /ORDER=ANALYSIS.
```

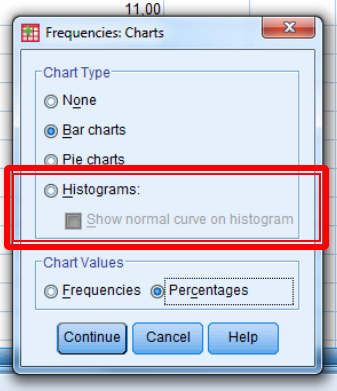
Dataset deel 6 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : Case | 1,00 | Visible: 10 of 10 Variables

	Case	Geslacht	Leeftijd	Faculteit	Studiejaar	Instelling	Vertrouwen	VertrouwenLMH	PC_ervaring	Vormingsniveau	var	var	var	var	var	var
1	1,00	Mannelijk	23,00	Ingenieurswetenschappen	MA	UGent	8,00	Hoog	10,00	14,00						
2	2,00	Mannelijk	22,00	Wetenschappen	2 BA	VUBrussel	8,00	Hoog	8,00	10,00						
3	3,00	Vrouwelijk	20,00	Economie en Bedrijfskunde	1 BA	UAntwerpen	3,00	Laag	5,00	6,00						
4	4,00	Mannelijk	20,00	Politieke en Sociale Wetenschappen	1 BA	UGent	10,00	Hoog	6,00	7,00						
5	5,00	Vrouwelijk	21,00	Wetenschappen	2 BA	UAntwerpen	4,00	Middelmatig	7,00	9,00						
6	6,00	Mannelijk	20,00	Letteren en wijsbegeerte	1 BA	KUBrussel	7,00	Hoog	5,00	6,00						
7	7,00	Mannelijk	22,00	Diergeneeskunde	3 BA	KULeuven	5,00	Middelmatig	7,00	10,00						
8	8,00	Vrouwelijk	20,00	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9,00	Hoog	6,00	7,00						
9	9,00	Mannelijk	21,00	Rechtsgeleerdheid	2 BA	UHasselt	2,00	Laag	8,00	10,00						
10	10,00	Mannelijk	22,00	Farmaceutische Wetenschappen	2 BA	UGent	8,00	Hoog	9,00	11,00						
11	1,00	Mannelijk	23,00	Ingenieurswetenschappen	MA	UGent	8,00	Hoog	10,00	14,00						
12	2,00	Mannelijk	22,00	Wetenschappen	2 BA	VUBrussel	8,00	Hoog	8,00	10,00						
13	3,00	Vrouwelijk	20,00	Economie en Bedrijfskunde	1 BA	UAntwerpen	3,00	Laag	5,00	6,00						
14	4,00	Mannelijk	20,00	Politieke en Sociale Wetenschappen	1 BA	UGent	10,00	Hoog	6,00	7,00						
15	5,00	Vrouwelijk	21,00	Wetenschappen	2 BA	UAntwerpen	4,00	Middelmatig	7,00	9,00						
16	6,00	Mannelijk	20,00	Letteren en wijsbegeerte	1 BA	KUBrussel	7,00	Hoog	5,00	6,00						
17	7,00	Mannelijk	22,00	Diergeneeskunde	3 BA	KULeuven	5,00	Middelmatig	7,00	10,00						
18	8,00	Vrouwelijk	20,00	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9,00	Hoog	6,00	7,00						
19	9,00	Mannelijk	21,00	Rechtsgeleerdheid	2 BA	UHasselt	2,00	Laag	8,00	10,00						
20	10,00	Mannelijk	22,00	Farmaceutische Wetenschappen	2 BA	UGent	8,00	Hoog	9,00	11,00						
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																
31																





Enkel bij ordinaal en scale variabelen

Data View | Variable View

PASW Statistics Processor is ready

Documenten - K0009... | Inbox in Tom.Scham... | Bigham (stap voor st... | 6. Eenvoudige besch... | Dataset deel 6 (voorb...

NL 10:48

Analyse van de Output

- **Frequencies**: aantallen 'valid' en 'missing' data entries voor *Geslacht* en *VertrouwenLMH*
- **Frequency Table**: aantallen voor elke categorie voor *Geslacht* (man en vrouw) en *VertrouwenLMH* (Laag, Middelmatig en Hoog)
- Gevolgd door **Bar Charts**

(Weergave met percentages)

Deel 6_Output1.spv [Document6] - PASW Statistics Viewer

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

- Output
  - Log
  - Frequencies
    - Title
    - Notes
    - Active Dataset
    - Statistics
    - Frequency Table
      - Title
      - Geslacht
      - Vertrouwen (laag, middelmatig en hoog)
    - Bar Chart
      - Title
      - Geslacht
      - Vertrouwen (laag, middelmatig en hoog)

```
FREQUENCIES VARIABLES=Geslacht VertrouwenLMH
  /BARCHART PERCENT
  /ORDER=ANALYSIS.
```

Frequencies

[DataSet1] K:\Opleidingsonderdelen\3de bachelor\Cursusonderdelen\Module 2 uni- en bivariate analyse\bruikbare datasets\Dataset deel 6 (voorbeeldvragen_ijs

Statistics

		Geslacht	Vertrouwen (laag, middelmatig en hoog)
N	Valid	20	20
	Missing	0	0

Frequency Table

Geslacht

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mannelijk	14	70,0	70,0	70,0
	Vrouwelijk	6	30,0	30,0	100,0
	Total	20	100,0	100,0	

Vertrouwen (laag, middelmatig en hoog)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Laag	4	20,0	20,0	20,0
	Middelmatig	4	20,0	20,0	40,0
	Hoog	12	60,0	60,0	100,0
	Total	20	100,0	100,0	

Bar Chart



PASW Statistics Processor is ready

H: 13.34, W: 16.64 cm

Module 2 uni- en biv... | Dataset deel 6 (voorb... | Deel 6_Output1.spv [... | Microsoft PowerPoin...

NL 14:26

Statistics

		Geslacht	Vertrouwen (laag, middelmatig en hoog)
N	Valid	20	20
	Missing	0	0

frequencies

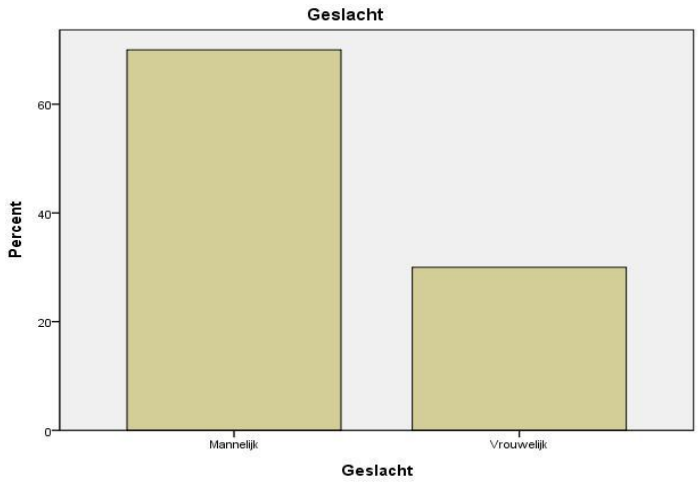
Geslacht

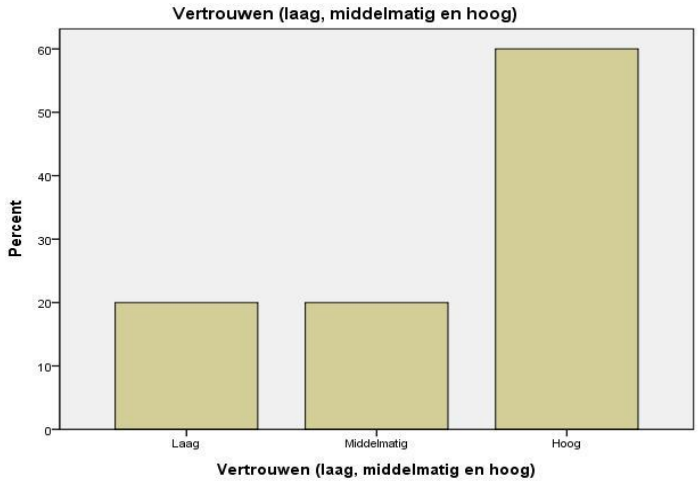
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mannelijk	14	70,0	70,0	70,0
	Vrouwelijk	6	30,0	30,0	100,0
	Total	20	100,0	100,0	

Vertrouwen (laag, middelmatig en hoog)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Laag	4	20,0	20,0	20,0
	Middelmatig	4	20,0	20,0	40,0
	Hoog	12	60,0	60,0	100,0
	Total	20	100,0	100,0	

frequencietabellen %





staafdiagrammen

Opslaan van de Output

- Opgelet:
  - Een tweede file staat nu open, en dat zal het geval zijn voor elke analyse die wordt uitgevoerd
  - Dit is de **output file**
  - Deze file kan worden opgeslagen: je bewaart dan zowel het databestand (.sav) als de outputfile (.spv).
    - Ga naar **File** > **Save As** > (bepaal zelf de locatie)
    - Typ bestandnaam
    - Selecteer **Save**

Hoe eenvoudig ook, er kan veel fout gaan…

Itinera oplossingen voor Brussel_15-bru.pdf - Adobe Reader

File Edit View Document Tools Window Help

1 / 4 71,2% Find

ITINERA INSTITUTE ANALYSE

ONAFHANKELIJKE DENKTANK | KLARE IDEEËN, KLAAR VOOR ACTIE

Oplossingen voor Brussel

2010/15
02|09|2010

MENSEN
WELVAART
BESCHERMING

Het was zo te voorspellen dat de onderhandelingen in het slop gingen raken op het moment dat het "probleem" Brussel wordt aangesneden. Er zijn in dit land geen pasklare oplossingen voor de institutionele problematiek en één van de belangrijkste redenen daarvoor is precies onze hoofdstad. Over het uitgangspunt alleen al geraakt onze politiek het niet eens. Wordt Brussel nu ondergefinancierd of krijgt het net te veel (al dan niet blanco) cheques?

Tabel 1: Ontvangsten van het Brussels Hoofdstedelijk Gewest (2008)

RECETTES TOTALES	2611487	100,00%
Transferts en provenance du Pouvoir fédéral	945551	36,21%
Dotation IPP (y compris intervention de solidarité)	834737	31,96%
Droit de tirage sur le MET	42545	1,63%
Mainmorte	29027	1,11%
Fonds pour l'entretien d'espaces verts	2772	0,11%
Moyens pour compétences issues de Lambermont	3441	0,13%
Moyens aux communes avec échevin flamand	31413	1,20%
Versement destinés à promouvoir le rôle international et la fonction de capital de Bruxelles	0	0,00%
Recettes relatives à l'économie sociale ou plurielle	1599	0,06%
Egalité des chances	17	[illegible]
Impôts régionaux et taxes perçues par la RBC	1305981	50,01%
Impôts régionaux	1198563	45,90%

1

KLARE IDEEËN, KLAAR VOOR ACTIE www.itinerainstitute.org

ITINERA INSTITUTE ANALYSE

Taxes perçues par la RBC	107418	4,11%
Recettes sur Fonds organiques	197672	7,57%
Fonds pour la gestion de la dette	173526	6,64%
Autres Fonds	24146	0,92%
Transfert en provenance de l'Agglo	140886	5,39%
Recette en provenance de la SLRB	5000	[illegible]
Recettes en provenance de l'UE	11013	0,42%
Autres recettes	5384	0,21%

Bron: begroting, CERPE

In elk geval kost koken geld, ook in Brussel. Het Brussels Hoofdstedelijk Gewest is een klein gewest dat wordt geconfronteerd met hoge vaste en moeilijk samendrukbare bestuurskosten die per inwoner hoger liggen dan in de andere gewesten. Bovendien worden in Brussel ook gemeenschapsuitgaven afgewenteld op het gewest, naast de evidente stedelijke uitgaven. Dit wordt uiteraard allemaal nog scherper als rekening gehouden wordt met de realiteit dat het Brussels gewest en de Brusselse gemeenten op budgettair vlak als communicerende vaten werken. De meeruitgaven op het vlak van onderwijs, gezondheidszorg, maar ook justitie en politie zijn echt niet fictief voor iedereen die vertrouwd is met de Brusselse hoofdstedelijke context. Het is vandaag politiek correct om vooral te focussen op het gebrek aan responsabilisering van gemeenschappen en gewesten, maar we zijn vaak blind voor het feit dat op het vlak van de gemeentelijke financiën er ook belangrijke mechanismen bestaan die kunnen aanzetten tot excessieve uitgaven.

Tabel 2: Uitgaven van het Brussels Hoofdstedelijk Gewest (2008)

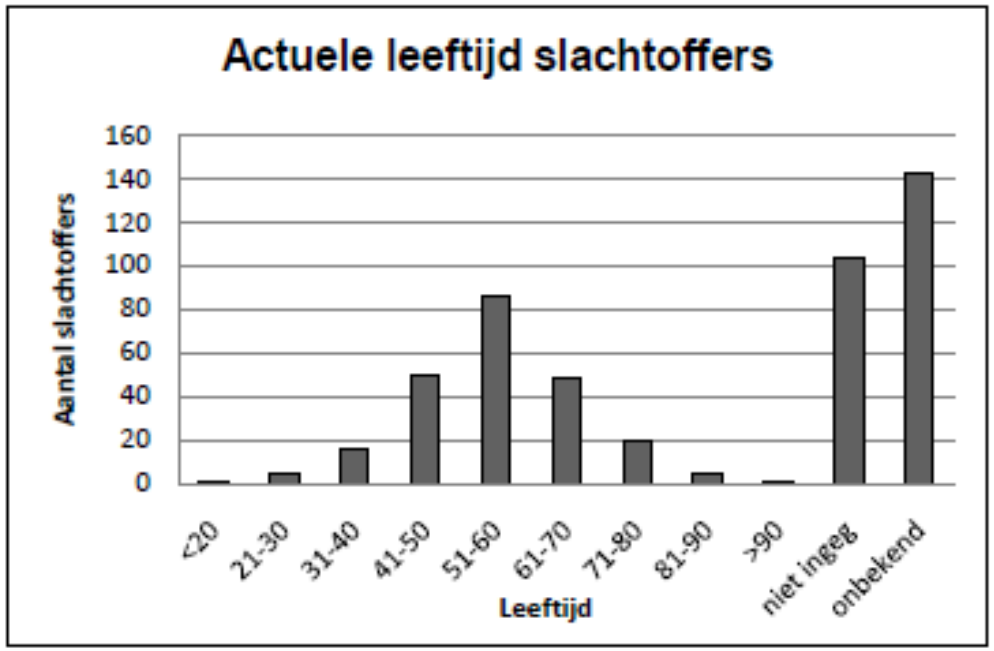
DEPENSES TOTALES	2616324	100,00%
Conseil de la Région de Bruxelles	32697	1,25%
Cabinets ministériels	49743	1,90%
Dépenses générales de l'Administration	392579	15,00%
Développement économique	77050	2,94%
Equipements et déplacements	616885	23,58%
Emploi	205114	7,84%
Pouvoirs locaux	379018	14,49%
Logement	107621	4,11%
Aménagement du territoire	70588	2,70%
Monuments et sites	18955	0,72%
Environnement politique de l'eau et propreté publique	183098	7,00%
Politique de l'énergie	21451	0,82%

2

KLARE IDEEËN, KLAAR VOOR ACTIE www.itinerainstitute.org

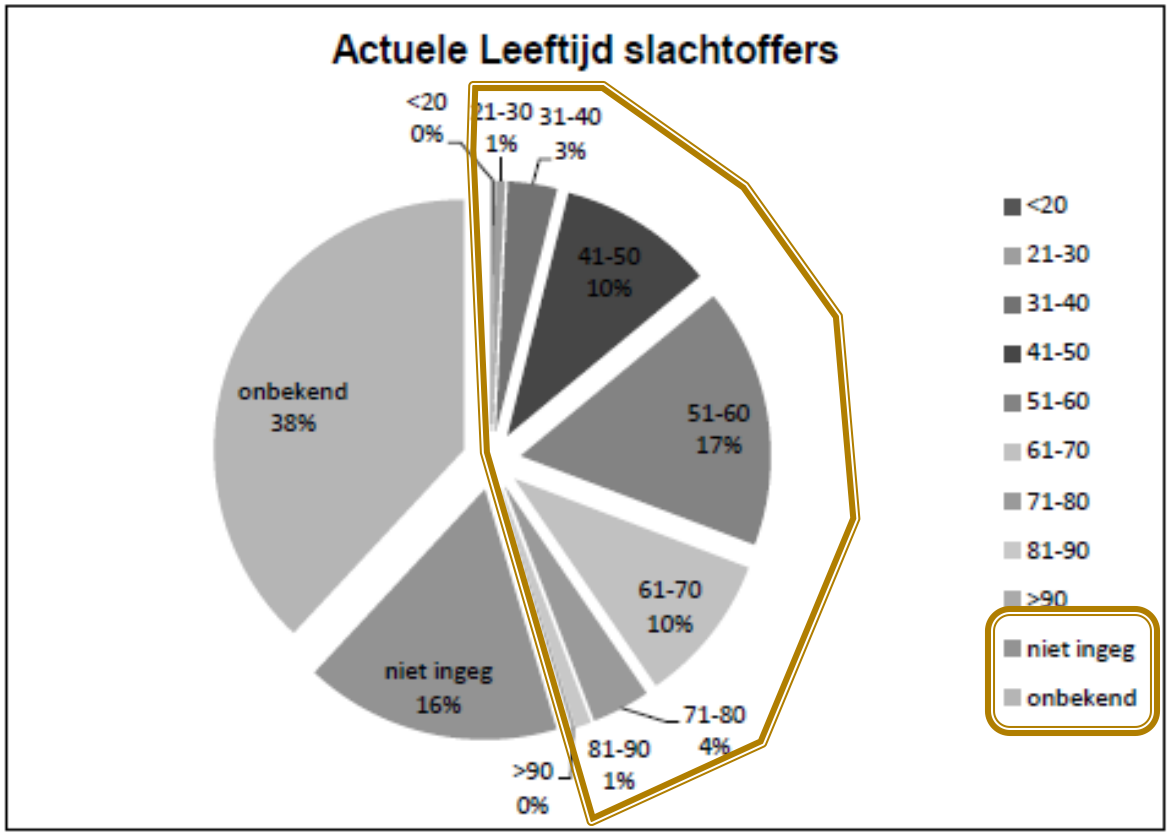
http://www.itinerain... Inbox in Tom.Scham... Module 2 uni- en biv... Itinera oplossingen v... Microsoft PowerPoin... NL 9:04

Bron:
Verslag activiteiten Commissie voor de behandeling van klachten wegens seksueel misbruik in een pastorale relatie
(onafgewerkt wegens inbeslagname op 24 juni 2010)
19 april – 24 juni 2010 10 september 2010
Prof. Peter Adriaenssens, voorzitter, p. 116



leeftijd	aantal
<20	1
21-30	4
31-40	16
41-50	50
51-60	86
61-70	49
71-80	19
81-90	5
>90	1
niet ingeg	103
onbekend	142
totaal	476

%?


Actuele Leeftijd slachtoffers
<20
0%
21-30
1%
31-40
3%
41-50
10%
51-60
17%
61-70
10%
71-80
4%
81-90
1%
>90
0%
niet ingeg
16%
onbekend
38%
<20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
>90
niet ingeg
onbekend

Totaal = 54%

Minimum, Maximum, Gemiddelde, Modus, Mediaan & Standaardafwijking

- Selecteer: **Analyze > Descriptive Statistics > Frequencies** om *Frequencies* te openen.
- Selecteer: **Reset**
- Klik aan *Leeftijd*, *Studiejaar*, *Aantal jaar PC-ervaringen* en plaat ze in de Variable(s) box.
- Is *Display Frequency Tables* aangevinkt?
- Selecteer: **Statistics.** Check **Mean, Mode, Median, Standard Deviation, Minimum** and **Maximum**
- Selecteer: **Continue > OK**

```
FREQUENCIES VARIABLES=Leeftijd Studiejaar PC_ervaring
  /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.
```

Statistics

		Leeftijd	Studiejaar	Hoeveel jaar computerervaring heb je?
N	Valid	20	20	20
	Missing	0	0	0
Mean		21,1000	1,9000	7,1000
Median		21,0000	2,0000	7,0000
Mode		20,00	1,00[a]	5,00[a]
Std. Deviation		1,07115	,96791	1,61897
Minimum		20,00	1,00	5,00
Maximum		23,00	4,00	10,00

a. Multiple modes exist. The smallest value is shown

Leeftijd

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20,00	8	40,0	40,0	40,0
	21,00	4	20,0	20,0	60,0
	22,00	6	30,0	30,0	90,0
	23,00	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Studiejaar

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 BA	8	40,0	40,0	40,0
	2 BA	8	40,0	40,0	80,0
	3 BA	2	10,0	10,0	90,0
	MA	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Hoeveel jaar computerervaring heb je?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5,00	4	20,0	20,0	20,0
	6,00	4	20,0	20,0	40,0
	7,00	4	20,0	20,0	60,0
	8,00	4	20,0	20,0	80,0
	9,00	2	10,0	10,0	90,0
	10,00	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Analyse van de Output

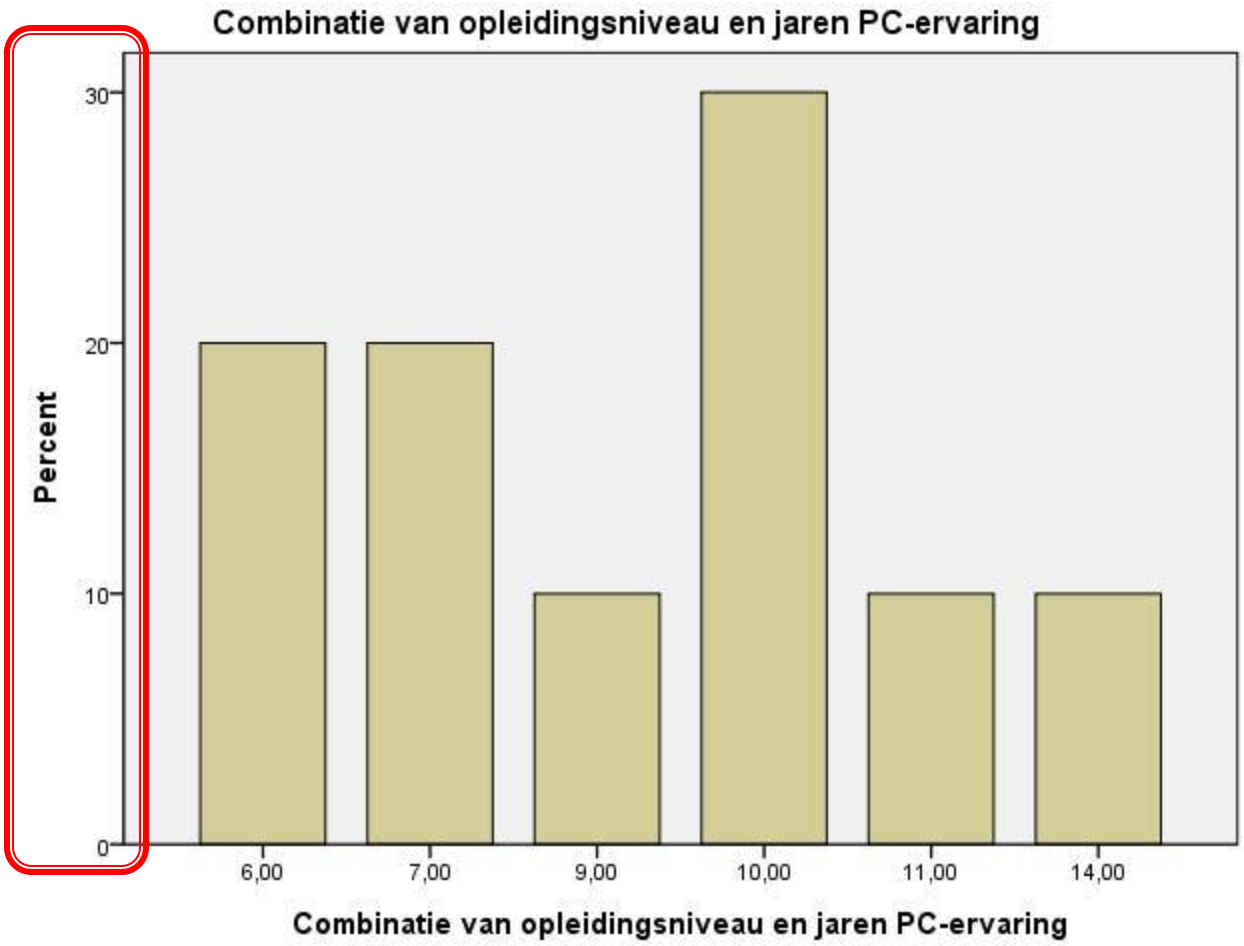
- Frequencies: aantallen 'valid' en 'missing' data entries voor elke variabele m.i.v. **Mean, Median, Mode, Standard Deviation, Minimum, Maximum**.
- Frequency Tables: aantallen voor alle mogelijke antwoordcategorieën voor elk van de geselecteerde variabelen (een tabel voor elke variabele).
  - Bij heel veel antwoordmogelijkheden -> heel wat detail in de tabel
  - Aan de andere kant, de Mean, Median, Mode, Standard Deviation, Minimum, en Maximum zijn op de typische wijze weergegeven.

Oefening 1

- Vraag de **Frequencies** op voor "*Vormingsniveau*"
- Roep volgende **maten van centraliteit en spreiding** op: kwartielen, gemiddelde, mediaan, modus, standaardafwijking en variantie
- Genereer **staafdiagram** met procentuele aanduiding

- **Sla deze output (.spv) op als "Dataset deel 6(vormingsniveau)_achternaam"**

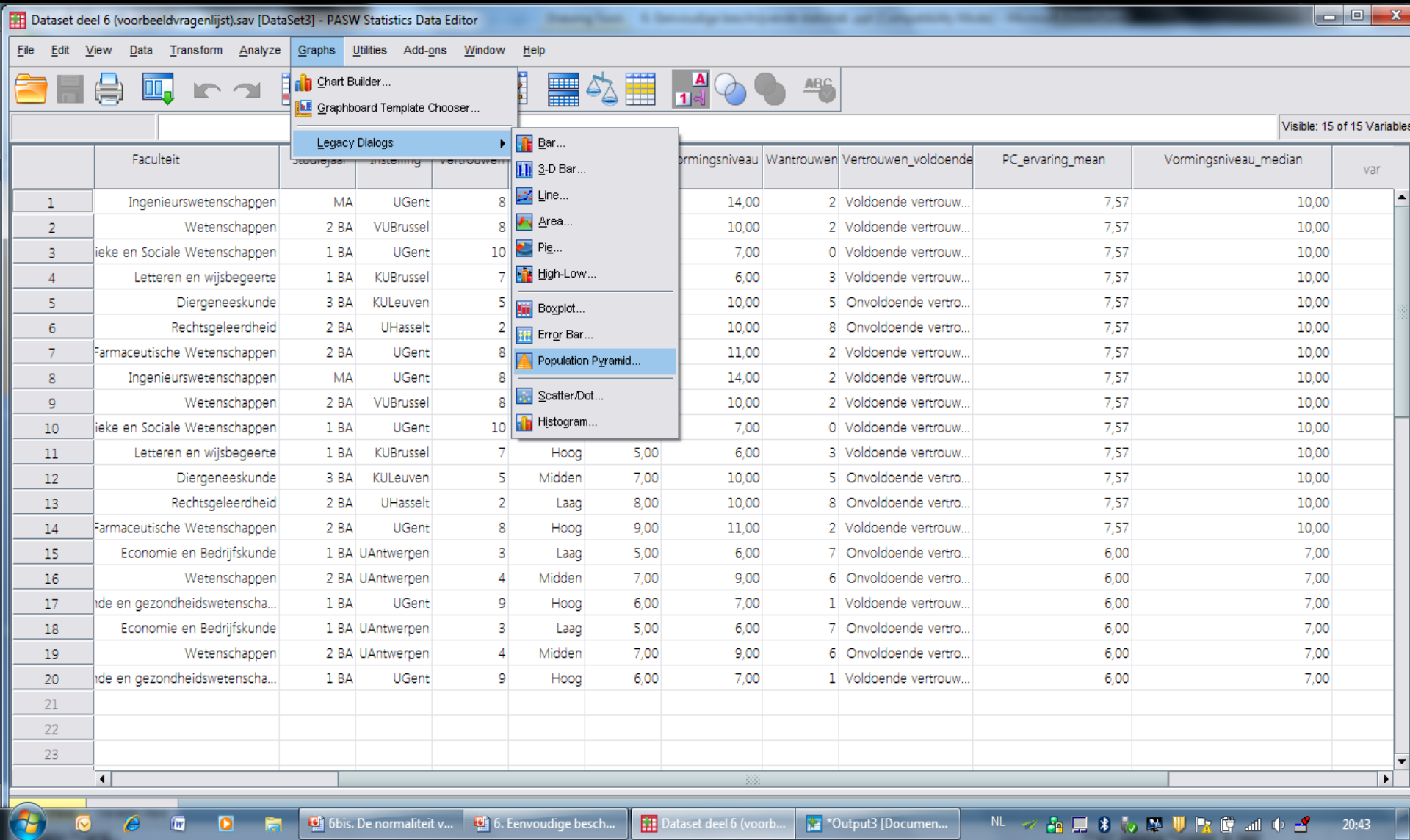
Oplossing


Combinatie van opleidingsniveau en jaren PC-ervaring
Percent
30
20
10
0
6,00
7,00
9,00
10,00
11,00
14,00
Combinatie van opleidingsniveau en jaren PC-ervaring

Snel grafieken maken in spss

- Via: **Graphs** > **Legacy dialogs**

- **Snelkeuze** tussen lijndiagram, staafdiagram, taartgrafiek, scatterplot, boxplot, p etc.
  - Keuze tussen een drietal hoofdtypes van diagrammen of grafieken
  - Keuze van weergave ('*Simple*' en '*Summaries for groups of cases*')
  - Selectie van de variabelen via '**Define**' en verplaats die in het vakje '*Category axis*'
  - Klik op 'OK' voor weergave van de gekozen grafiek

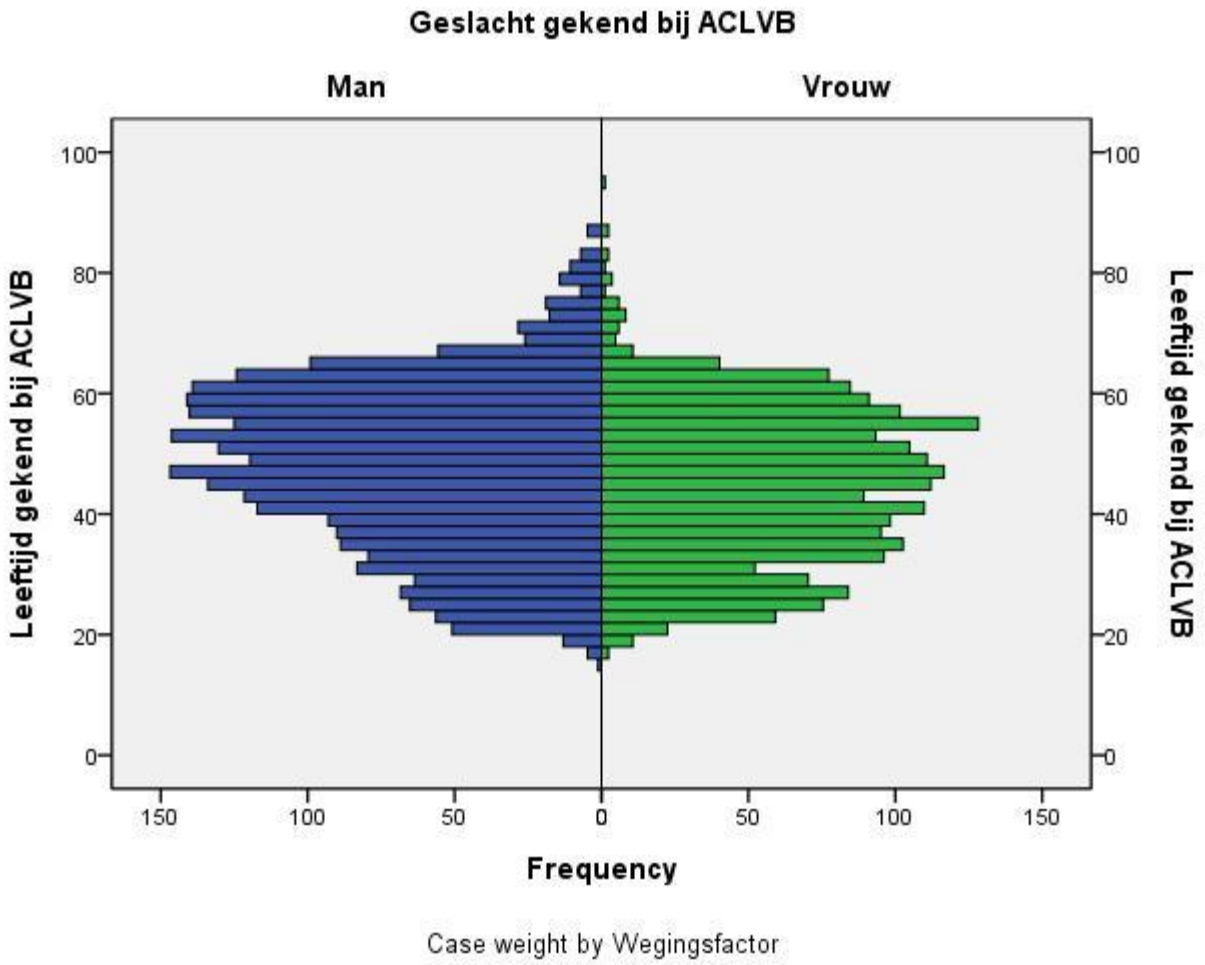

Dataset deel 6 (voorbeeldvragenlijst).sav [DataSet3] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Chart Builder...
Graphboard Template Chooser...
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
Visible: 15 of 15 Variables
Faculteit
Wantrouwen
Vertrouwen_voldoende
PC_ervaring_mean
Vormingsniveau_median
var
Ingenieurswetenschappen MA UGent 8 14,00 2 Voldoende vertrouw... 7,57 10,00
Wetenschappen 2 BA VUBrussel 8 10,00 2 Voldoende vertrouw... 7,57 10,00
ieke en Sociale Wetenschappen 1 BA UGent 10 7,00 0 Voldoende vertrouw... 7,57 10,00
Letteren en wijsbegeerte 1 BA KUBrussel 7 6,00 3 Voldoende vertrouw... 7,57 10,00
Diergeneeskunde 3 BA KULeuven 5 10,00 5 Onvoldoende vertro... 7,57 10,00
Rechtsgeleerdheid 2 BA UHasselt 2 10,00 8 Onvoldoende vertro... 7,57 10,00
Farmaceutische Wetenschappen 2 BA UGent 8 11,00 2 Voldoende vertrouw... 7,57 10,00
Ingenieurswetenschappen MA UGent 8 14,00 2 Voldoende vertrouw... 7,57 10,00
Wetenschappen 2 BA VUBrussel 8 10,00 2 Voldoende vertrouw... 7,57 10,00
ieke en Sociale Wetenschappen 1 BA UGent 10 7,00 0 Voldoende vertrouw... 7,57 10,00
Letteren en wijsbegeerte 1 BA KUBrussel 7 Hoog 5,00 6,00 3 Voldoende vertrouw... 7,57 10,00
Diergeneeskunde 3 BA KULeuven 5 Midden 7,00 10,00 5 Onvoldoende vertro... 7,57 10,00
Rechtsgeleerdheid 2 BA UHasselt 2 Laag 8,00 10,00 8 Onvoldoende vertro... 7,57 10,00
Farmaceutische Wetenschappen 2 BA UGent 8 Hoog 9,00 11,00 2 Voldoende vertrouw... 7,57 10,00
Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 5,00 6,00 7 Onvoldoende vertro... 6,00 7,00
Wetenschappen 2 BA UAntwerpen 4 Midden 7,00 9,00 6 Onvoldoende vertro... 6,00 7,00
de en gezondheidswetenscha... 1 BA UGent 9 Hoog 6,00 7,00 1 Voldoende vertrouw... 6,00 7,00
Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 5,00 6,00 7 Onvoldoende vertro... 6,00 7,00
Wetenschappen 2 BA UAntwerpen 4 Midden 7,00 9,00 6 Onvoldoende vertro... 6,00 7,00
de en gezondheidswetenscha... 1 BA UGent 9 Hoog 6,00 7,00 1 Voldoende vertrouw... 6,00 7,00
6bis. De normaliteit v...
6. Eenvoudige besch...
Dataset deel 6 (voorb...
*Output3 [Documen...
NL
20:43

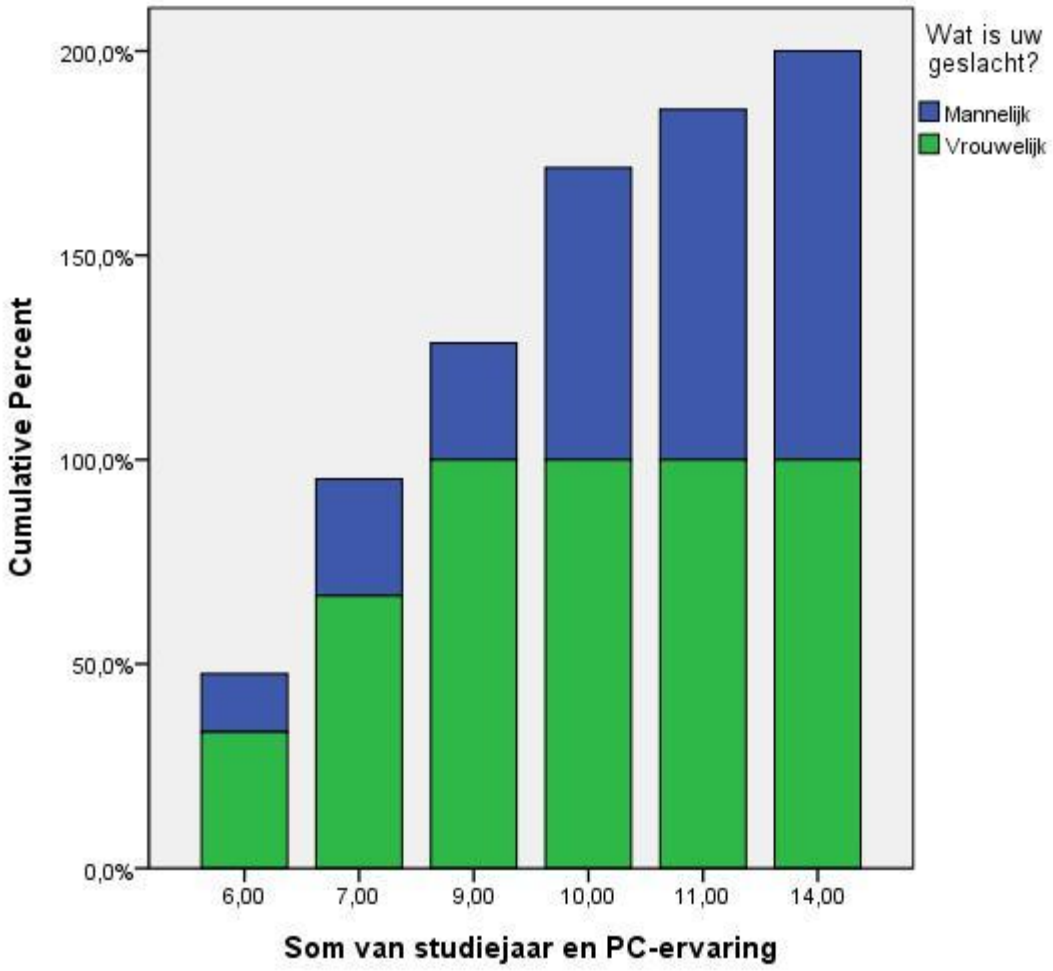
Voorbeeldjes

- Population pyramid (M/V) voor vormingsniveau
- Stacked bar chart (M/V) voor vormingsniveau
  - Hoe interpreteren we deze outputdiagram?

Variant

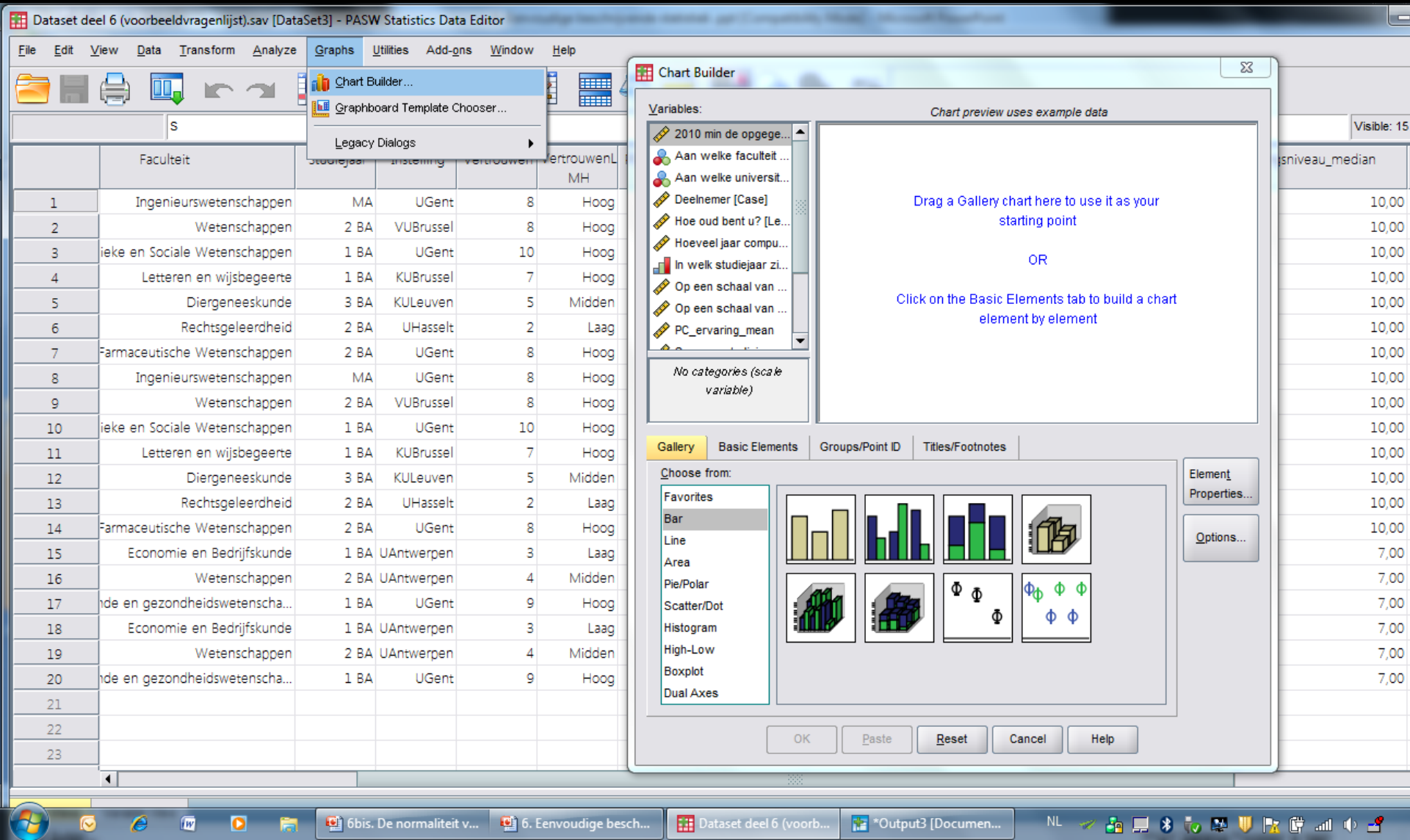
```
XGRAPH CHART=[HISTOBAR] BY Leeftijd_AdmInfo[s]
BY Geslacht_AdmInfo[c]
  /COORDINATE SPLIT=YES
  /BIN START=AUTO SIZE=AUTO.
```




Cumulative Percent
200,0%
150,0%
100,0%
50,0%
0,0%
6,00
7,00
9,00
10,00
11,00
14,00
Som van studiejaar en PC-ervaring
Wat is uw geslacht?
Mannelijk
Vrouwelijk

- Via: **Graphs** > **Chart Builder**

- Voordelen:
  - Meer vormgevingsopties
  - Gebruiksvriendelijk en overzichtelijke opbouw
  - Opgepast: voor aanvang zeker het meetniveau van de variabele goed / correct definiëren!
    - Categoriale variabelen (nominaal, ordinaal)
    - Interval/rationale variabelen (*scale variables*)


Dataset deel 6 (voorbeeldvragenlijst).sav [DataSet3] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Chart Builder...
Graphboard Template Chooser...
Legacy Dialogs
Visible: 15
Faculteit
MH
sniveau_median
1 Ingenieurswetenschappen MA UGent 8 Hoog 10,00
2 Wetenschappen 2 BA VUBrussel 8 Hoog 10,00
3 ieke en Sociale Wetenschappen 1 BA UGent 10 Hoog 10,00
4 Letteren en wijsbegeerte 1 BA KUBrussel 7 Hoog 10,00
5 Diergeneeskunde 3 BA KULeuven 5 Midden 10,00
6 Rechtsgeleerdheid 2 BA UHasselt 2 Laag 10,00
7 Farmaceutische Wetenschappen 2 BA UGent 8 Hoog 10,00
8 Ingenieurswetenschappen MA UGent 8 Hoog 10,00
9 Wetenschappen 2 BA VUBrussel 8 Hoog 10,00
10 ieke en Sociale Wetenschappen 1 BA UGent 10 Hoog 10,00
11 Letteren en wijsbegeerte 1 BA KUBrussel 7 Hoog 10,00
12 Diergeneeskunde 3 BA KULeuven 5 Midden 10,00
13 Rechtsgeleerdheid 2 BA UHasselt 2 Laag 10,00
14 Farmaceutische Wetenschappen 2 BA UGent 8 Hoog 10,00
15 Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 7,00
16 Wetenschappen 2 BA UAntwerpen 4 Midden 7,00
17 de en gezondheidswetenscha... 1 BA UGent 9 Hoog 7,00
18 Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 7,00
19 Wetenschappen 2 BA UAntwerpen 4 Midden 7,00
20 de en gezondheidswetenscha... 1 BA UGent 9 Hoog 7,00
Chart Builder
Variables:
2010 min de opgege...
Aan welke faculteit ...
Aan welke universit...
Deelnemer [Case]
Hoe oud bent u? [Le...
Hoeveel jaar compu...
In welk studiejaar zi...
Op een schaal van ...
Op een schaal van ...
PC_ervaring_mean
No categories (scale variable)
Chart preview uses example data
Drag a Gallery chart here to use it as your starting point
OR
Click on the Basic Elements tab to build a chart element by element
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
Element Properties...
Options...
OK
Paste
Reset
Cancel
Help
6bis. De normaliteit v...
6. Eenvoudige besch...
Dataset deel 6 (voorb...
*Output3 [Documen...
NL

- Verloop:
  - Stap 1: Het grafiektype kiezen (*Gallery*) en slepen naar 'Chart preview pane'
    - Om eigenschappen te zien /tonen: klik venster '*Element properties*' open
  - Stap 2: Kies de variabelen en voeg ze in de grafiek
    - X-as
    - Y-as
  - Stap 3: Verzorg de lay-out van de grafiek

Zelfexploratie

- Hoe grafieken lay-outen? (via **zelf-exploratie**)
  - Ofwel via Chart editor
  - Ofwel via Chart Builder
    - Basic elements
    - Groups / Point ID
    - Title/footnotes
    - Element properties (definiëren van de lijn, het staafje etc., de X-as en de Y-as)
    - Options (bv. opgeven hoe om te gaan met 'missings') Laat '*Listwise deletion*' steeds aangevinkt staan.

Chart editor



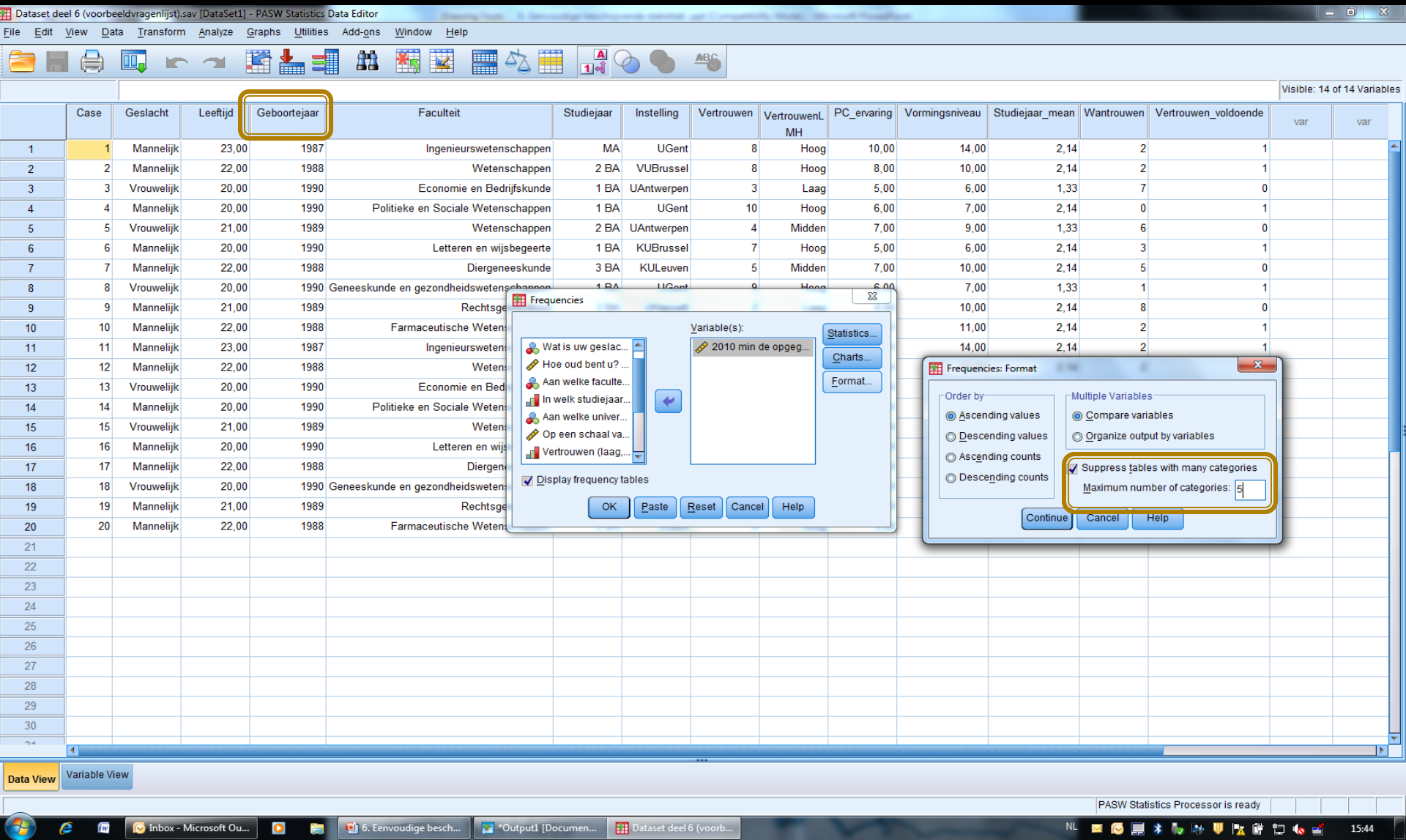
Bereiken door te dubbelklikken op de grafiek of diagram.
Keuzes: bewaren (**Save**), knippen, plakken, assen selecteren, grafieken selecteren, eigenschappen definiëren (**Edit**), werkbalk aanpassen (**View**), lijnen toevoegen, raster toevoegen, etc… (**Options**), labels tonen, markers tonen, curves tekenen (**Elements**), omzetten naar ander grafiektype (**Transform**), en **Help**.

H:375, W:468.75 points
Inbox in Tom.Scham... 6. Eenvoudige besch... *Output1 [Documen... Dataset deel 6 (voorb... Chart Editor NL 14:17

Waarden oplopend of aflopend weergeven

- Frequentietabel opvragen voor de variabele '*Geboortejaar*'
- Via **Analyze** > **Descriptive statistics** > **Frequencies** > **Format** > **Order 'descending values'** aanduiden

  (Opmerking: Optie *'Suppres tables with more than n categories*': er wordt dan enkel een frequentietabel getoond als de categoriën van variabelen kleiner of gelijk zijn aan n)


Dataset deel 6 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 14 of 14 Variables
Case Geslacht Leeftijd Geboortejaar Faculteit Studiejaar Instelling Vertrouwen VertrouwenLMH PC_ervaring Vormingsniveau Studiejaar_mean Wantrouwen Vertrouwen_voldoende var var
1 1 Mannelijk 23,00 1987 Ingenieurswetenschappen MA UGent 8 Hoog 10,00 14,00 2,14 2 1
2 2 Mannelijk 22,00 1988 Wetenschappen 2 BA VUBrussel 8 Hoog 8,00 10,00 2,14 2 1
3 3 Vrouwelijk 20,00 1990 Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 5,00 6,00 1,33 7 0
4 4 Mannelijk 20,00 1990 Politieke en Sociale Wetenschappen 1 BA UGent 10 Hoog 6,00 7,00 2,14 0 1
5 5 Vrouwelijk 21,00 1989 Wetenschappen 2 BA UAntwerpen 4 Midden 7,00 9,00 1,33 6 0
6 6 Mannelijk 20,00 1990 Letteren en wijsbegeerte 1 BA KUBrussel 7 Hoog 5,00 6,00 2,14 3 1
7 7 Mannelijk 22,00 1988 Diergeneeskunde 3 BA KULeuven 5 Midden 7,00 10,00 2,14 5 0
8 8 Vrouwelijk 20,00 1990 Geneeskunde en gezondheidswetenschappen 7,00 1,33 1 1
9 9 Mannelijk 21,00 1989 Rechtsge 10,00 2,14 8 0
10 10 Mannelijk 22,00 1988 Farmaceutische Weten 11,00 2,14 2 1
11 11 Mannelijk 23,00 1987 Ingenieurswetens 14,00 2,14 2 1
12 12 Mannelijk 22,00 1988 Weten
13 13 Vrouwelijk 20,00 1990 Economie en Bed
14 14 Mannelijk 20,00 1990 Politieke en Sociale Weten
15 15 Vrouwelijk 21,00 1989 Weten
16 16 Mannelijk 20,00 1990 Letteren en wij
17 17 Mannelijk 22,00 1988 Diergen
18 18 Vrouwelijk 20,00 1990 Geneeskunde en gezondheidswetens
19 19 Mannelijk 21,00 1989 Rechtsge
20 20 Mannelijk 22,00 1988 Farmaceutische Weten
Frequencies
Wat is uw geslac...
Hoe oud bent u? ...
Aan welke faculte...
In welk studiejaar...
Aan welke univer...
Op een schaal va...
Vertrouwen (laag,...
Variable(s):
2010 min de opgeg...
Statistics...
Charts...
Format...
Display frequency tables
OK Paste Reset Cancel Help
Frequencies: Format
Order by
Ascending values
Descending values
Ascending counts
Descending counts
Multiple Variables
Compare variables
Organize output by variables
Suppress tables with many categories
Maximum number of categories: 5
Continue Cancel Help
Data View
Variable View
PASW Statistics Processor is ready
Inbox - Microsoft Ou...
6. Eenvoudige besch...
*Output1 [Documen...
Dataset deel 6 (voorb...
15:44

Statistics

2010 min de opgegeven leeftijd

N	Valid	20
	Missing	0
Mean		1988,90
Median		1989,00
Mode		1990
Range		3
Minimum		1987
Maximum		1990
Percentiles	25	1988,00
	50	1989,00
	75	1990,00

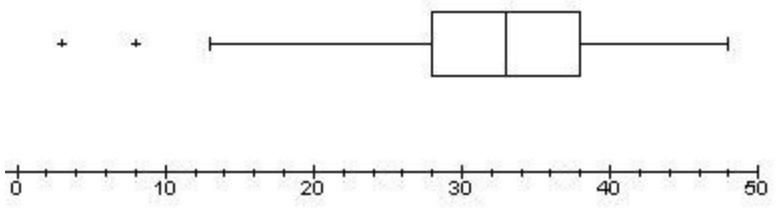
2010 min de opgegeven leeftijd

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1987	2	10,0	10,0	10,0
	1988	6	30,0	30,0	40,0
	1989	4	20,0	20,0	60,0
	1990	8	40,0	40,0	100,0
	Total	20	100,0	100,0	

Cf. presentatie in een boxplot

Boxplot

- De boxplot bestaat dus (in het algemeen) uit twee gedeelten:
  - een getallenlijn en
  - een figuur (de "plot") waarmee de kwartielen en de mediaan worden aangegeven.



Bron: http://www.pandd.demon.nl/statistiek.htm#2

.com/ips5e/content/cat_120/IPS_Splus.pdf - Windows Internet Explorer
/bcs.whfreeman.com/ips5e/content/cat_120/IPS_Splus.pdf
Favorites Help
tp://bcs.whfreeman.com/ips5e/content/cat_12...
38 / 220 151% Find

3. Click on the boxplot button to create the boxplot.

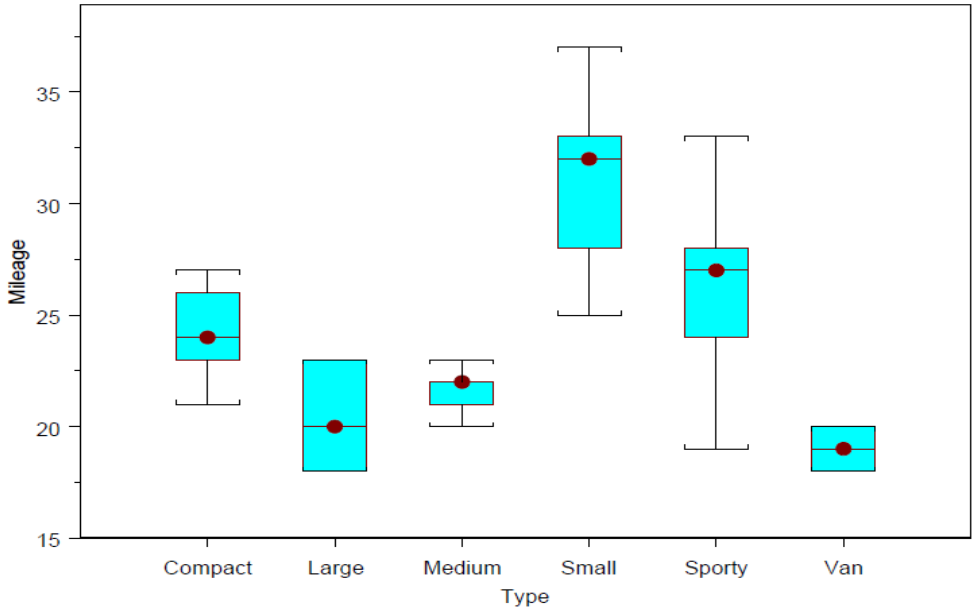


Figure 1.14

Boxplot - 2

- De plot zelf bestaat uit
  - een "centraal" lijnstuk dat de mediaan aangeeft
  - twee lijnstukken die de beide kwartielen aangeven (IKA)
  - twee kleine lijnstukken op een afstand van maximaal 1,5 maal de kwartielafstand (niet verder dan de grootste cq. de kleinste meetwaarde)
  - punten die buiten het bereik van 1, 5 maal de kwartielafstand liggen: uitbijters, uitliggers of outliers.

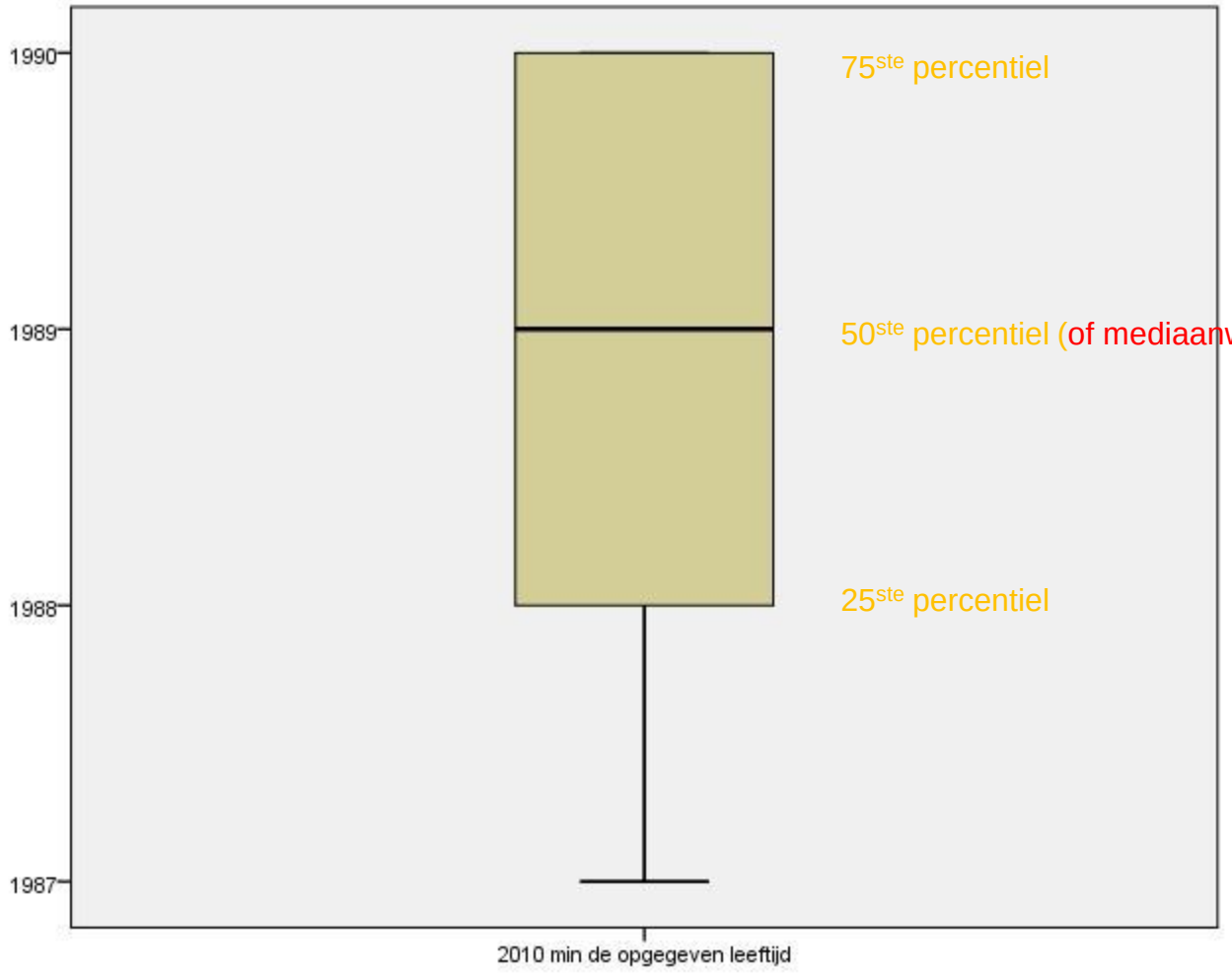
Boxplot - 3

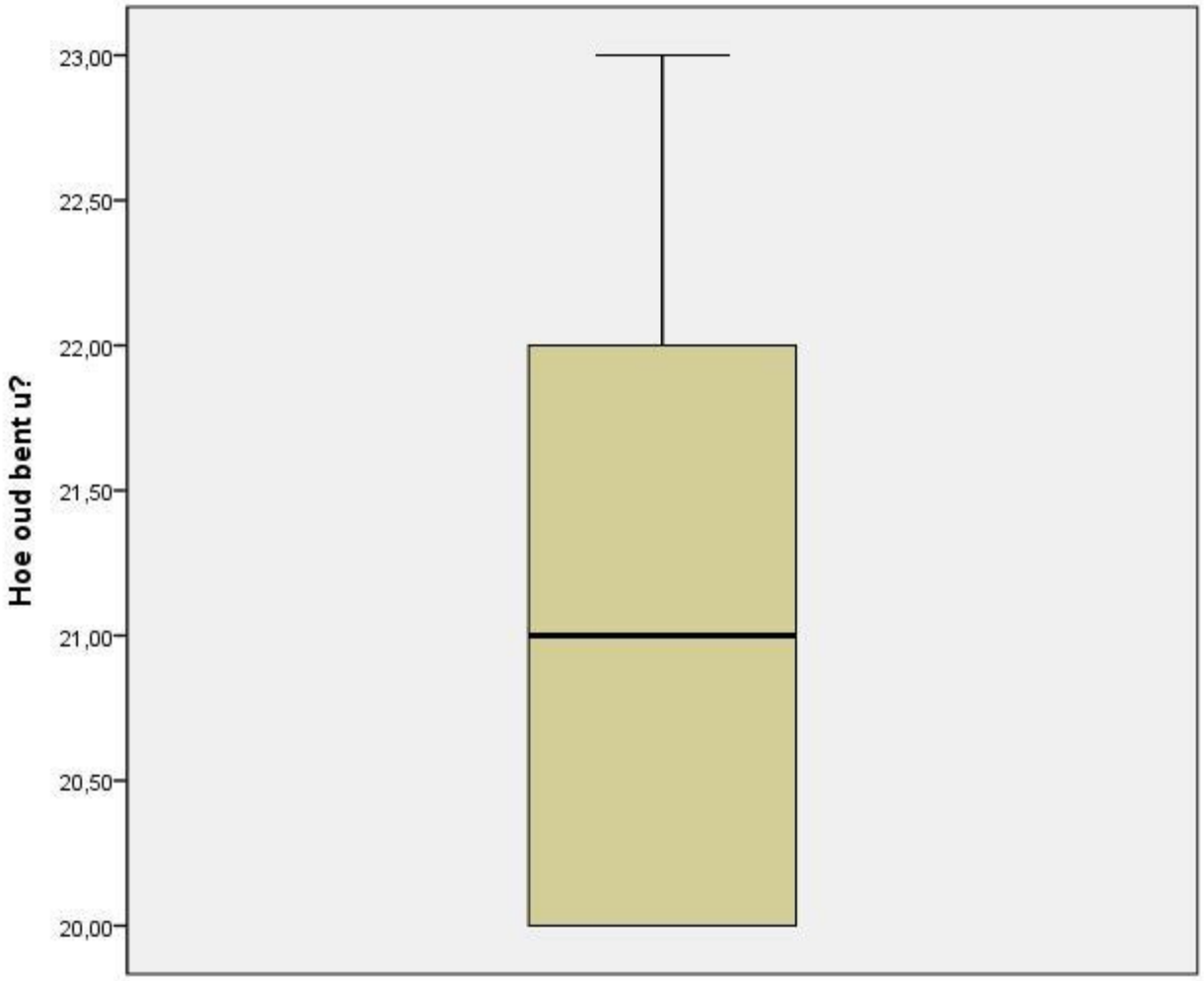
- Voor de verdeling van één variabele bijvoorbeeld '*Geboortejaar*'

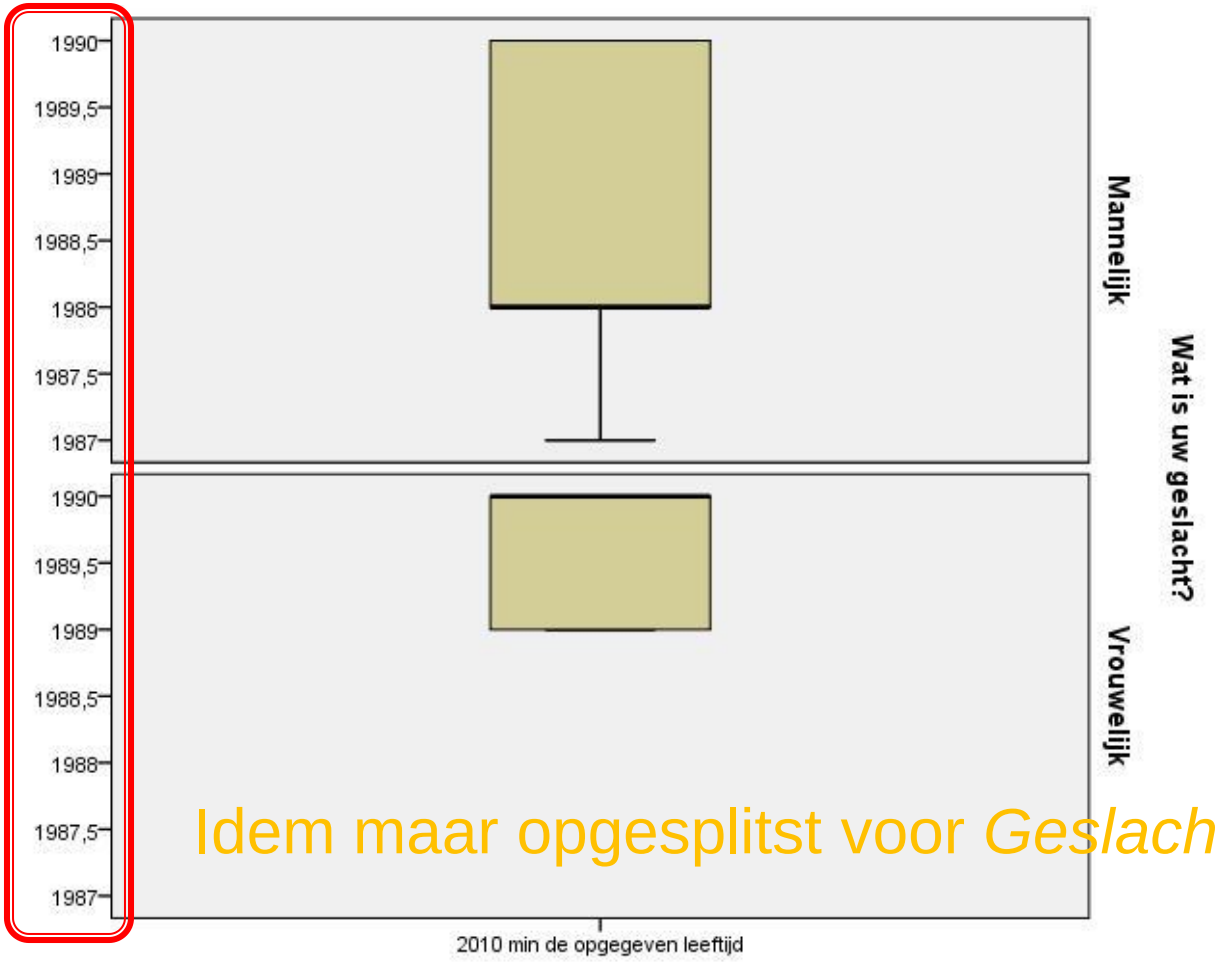
- Via **Graphs** > **Boxplot**
  > optie ***define summaries of separate variables*** aanduiden, variabele kiezen, OK

Bewerkt de plot vervolgens in de *Chart editor*

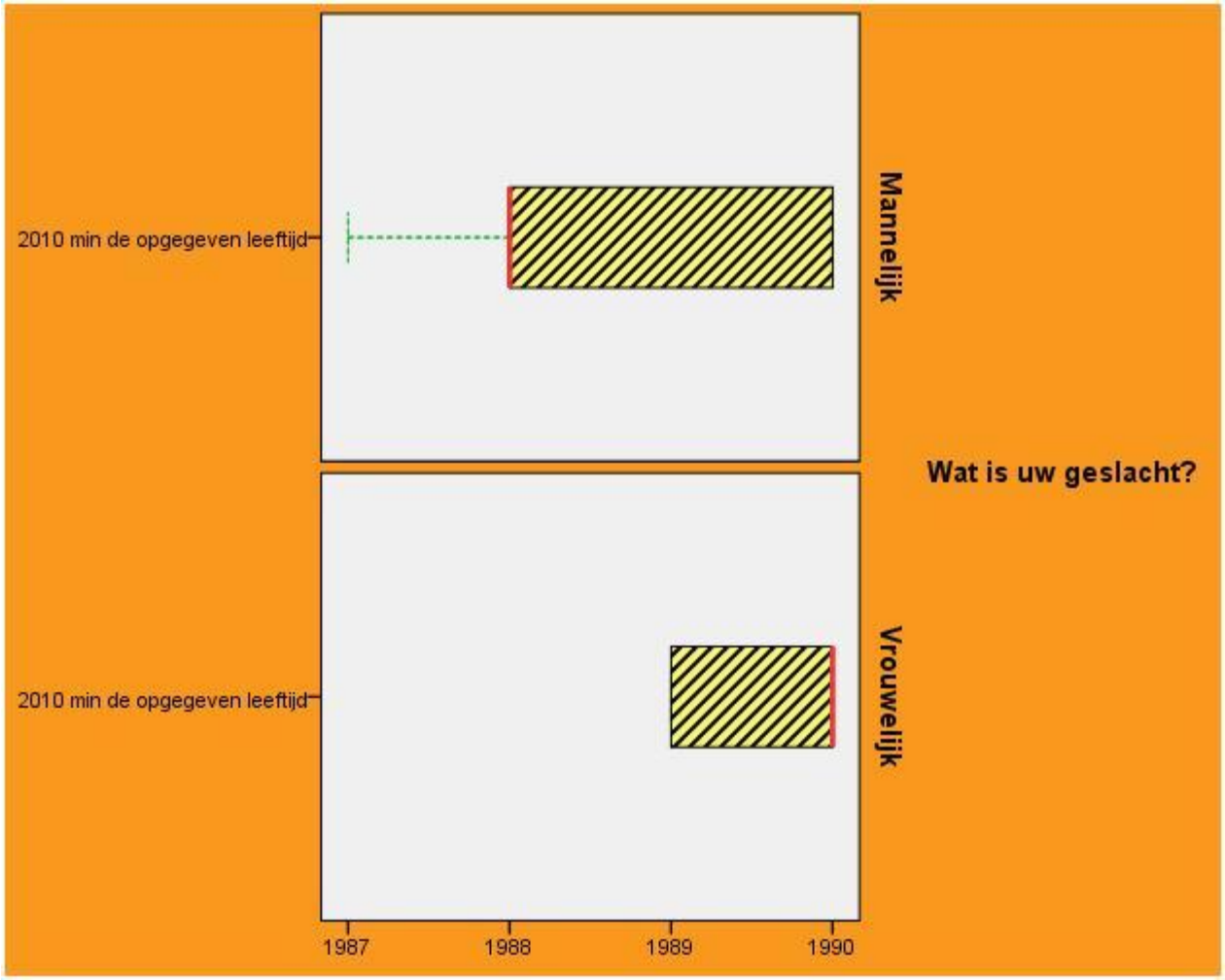
Boxplot *Geboortejaar*




Hoe oud bent u?
23,00
22,50
22,00
21,50
21,00
20,50
20,00

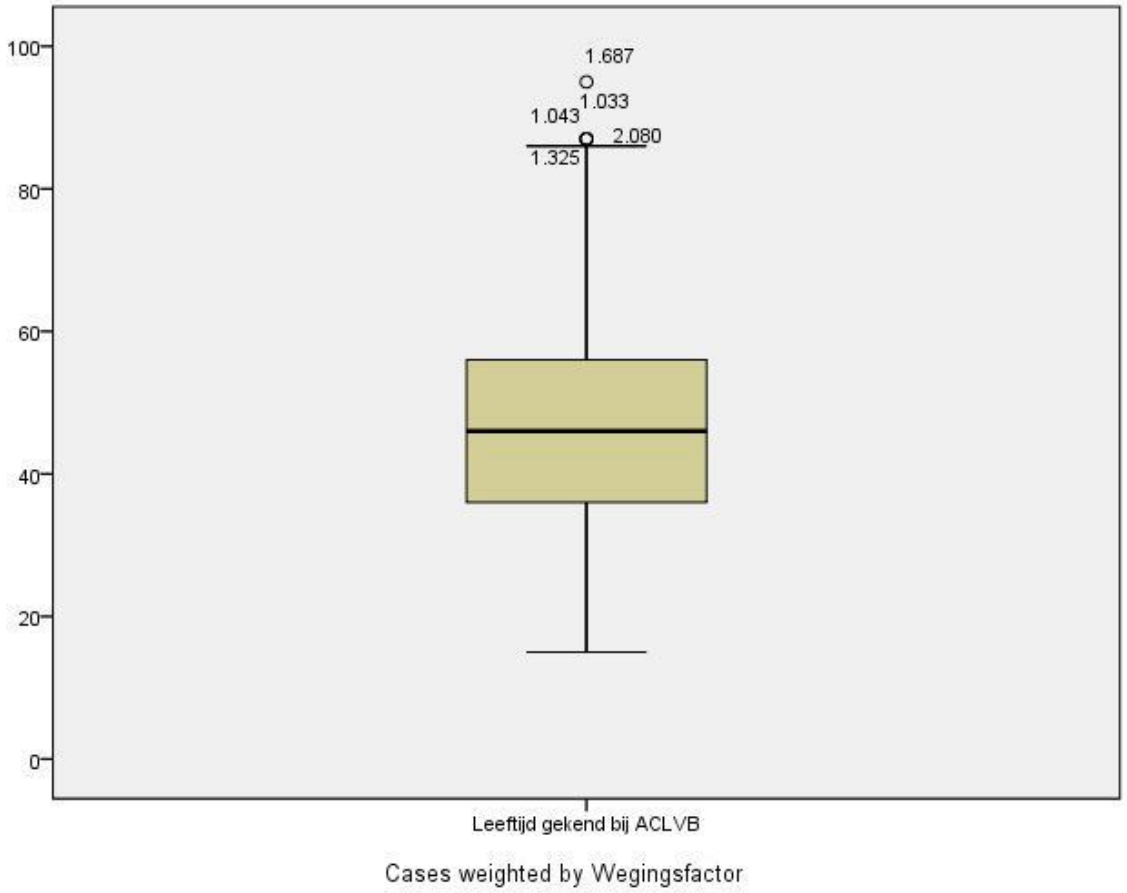


Idem maar opgesplitst voor *Geslacht*


2010 min de opgegeven leeftijd
Mannelijk
Wat is uw geslacht?
2010 min de opgegeven leeftijd
Vrouwelijk
1987
1988
1989
1990

Voorbeeld – aclvb onderzoek

- Interpretatie?



Stam-en-blad plot

Leeftijd gekend bij ACLVB Stem-and-Leaf Plot

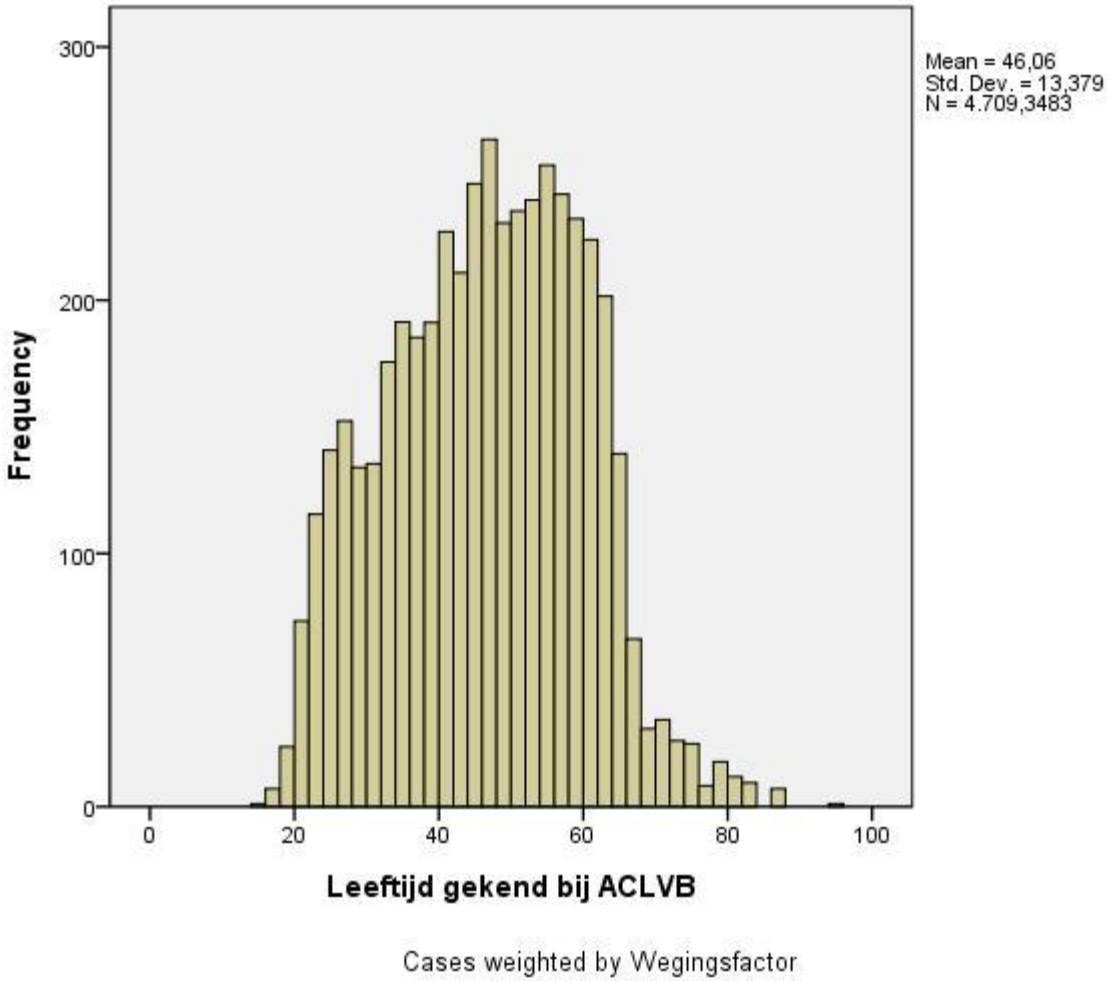
```
 Frequency    Stem &  Leaf

     ,00       1 .
   31,95       1 .  78999&
  253,68       2 .  00001111111222222223333333334444444444
  362,20       2 .  55555555556666666677777777777777888888888999999999999
  399,34       3 .  00000000001111111111222222222222333333333333344444444444444
  479,53       3 .  55555555555555566666666666667777777777777888888888888999999999999999
  550,30       4 .  000000000000000011111111111111112222222222222222333333333333344444444444444444
  627,54       4 .  55555555555555555556666666666666666666666677777777777777778888888888888888899999999999999999
  597,31       5 .  000000000000000001111111111111112222222222222223333333333333333334444444444444444
  604,90       5 .  555555555555555555566666666666666666777777777777777778888888888888889999999999999999999
  508,56       6 .  00000000000000000011111111111111222222222222222333333333333333444444444444
  153,24       6 .  55555555666667777788899
   73,36       7 .  00011223344
   37,86       7 .  556789
   21,30       8 .  0123
    2,37       8 .  &
    5,92 Extremes    (>=87)

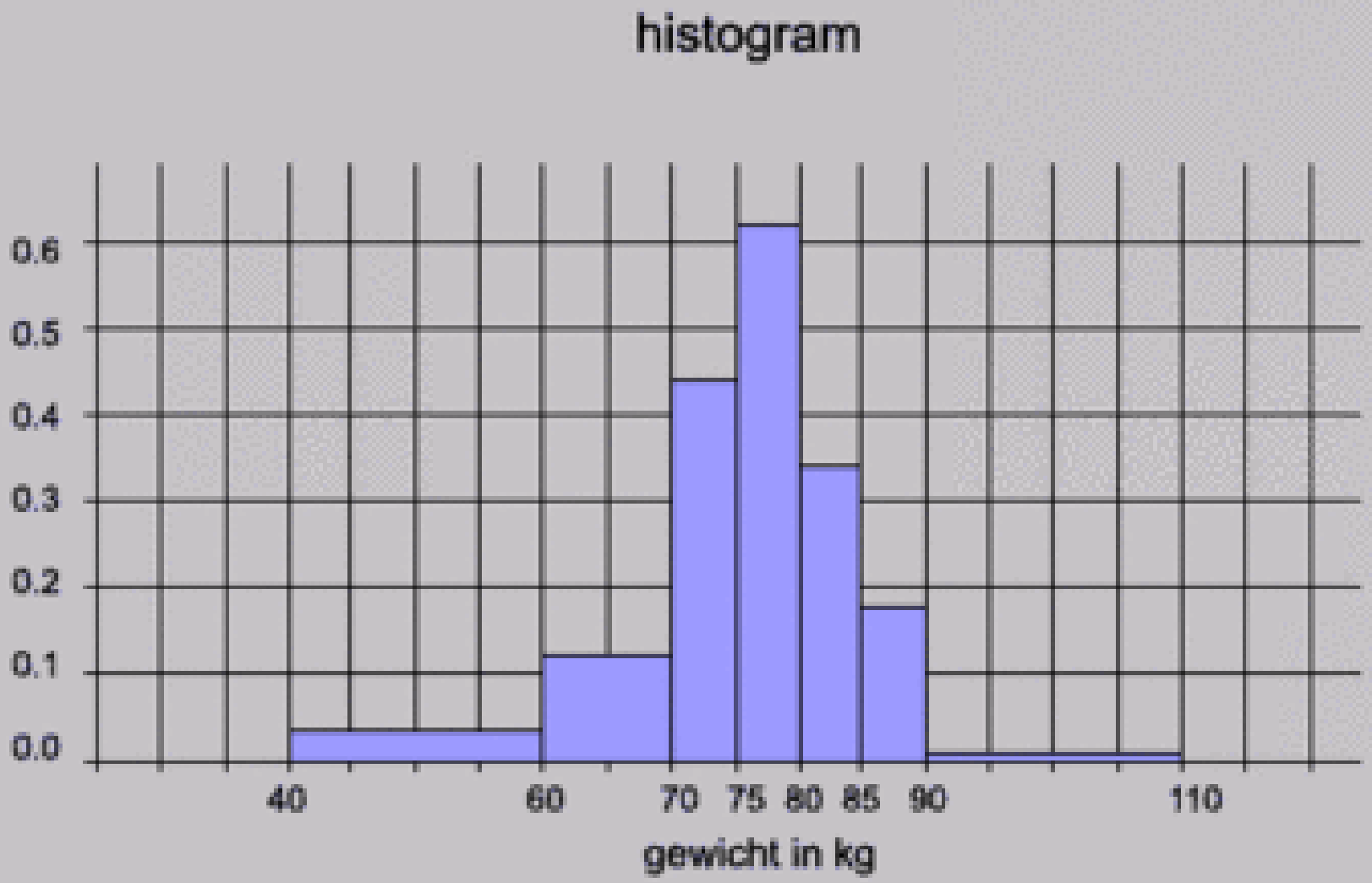
Stem width:      10
Each leaf:       7 case(s)


& denotes fractional leaves.
```

Het histogram (6min)

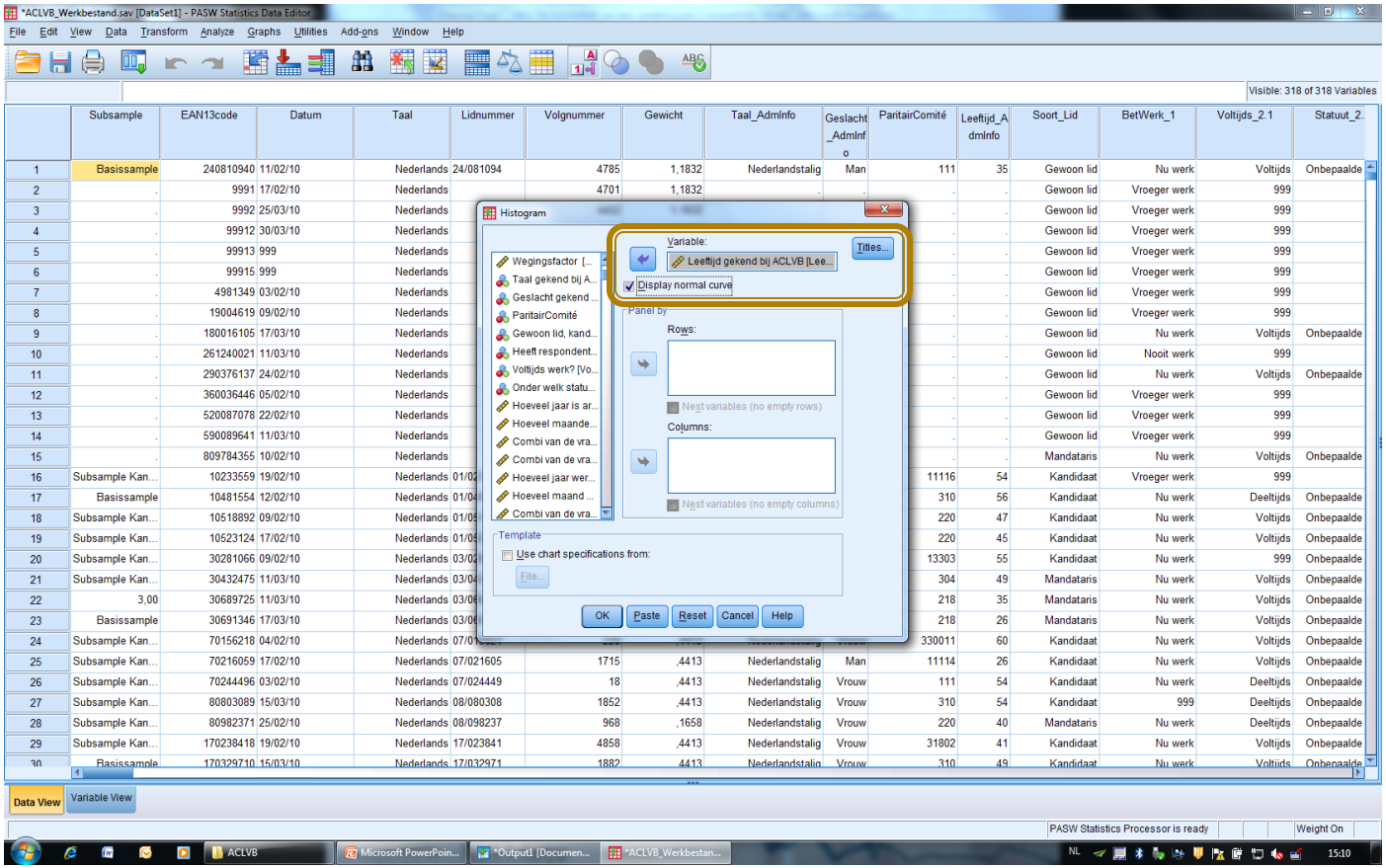


- Definitie en kenmerken (cf. wikipedia)
  - Een **histogram** of **kolommendiagram** is de grafische weergave van de frequentieverdeling van in klassen gegroepeerde data (*ranges*) of enkelvoudige waarden (*buckets*)
  - Dit diagram toont kolommen met oppervlakte ter grootte van de (relatieve) frequenties opgericht boven de klassen.
  - Een histogram geeft *een beeld* van de kansdichtheid waaruit de data afkomstig zijn. (zie ook *kansdichtheidskromme en normaalverdelingscurve*)


histogram
0.6
0.5
0.4
0.3
0.2
0.1
0.0
40
60
70
75
80
85
90
110
gewicht in kg

Hoe opvragen in spss?

- **Graphs > Legacy Dialogs > Histogram**



Oefening 2

- Maak een histogram aan voor de volgende variabelen en interpreteer het resultaat in functie van de specifieke kenmerken van het histogram:
  - *Leeftijd*
  - *Vormingsniveau*

Multiple response analyse

- Wat ?
- Via: **Analyze > > Multiple Reponse > Define variable sets**
- Voorbeeldvragen:
  - *Met welke politieke partijen kunnen de regeringsonderhandeling volgens u slagen?*
  - *Welke initiatieven tegen cumul en/of schrijnkandidatuur?*

*Output2 [Document2] - PASW Statistics Viewer

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

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...

Custom Tables...
Multiple Response Sets...

Output
Log
Graph
Title
Notes
Active Dataset
Scatter of VRP_be
Log
Multiple Response
Title
Notes
Active Dataset
Case Summary
$Welke_initiatieve
Log
Frequencies
Title
Notes
Active Dataset
Statistics
Frequency Table
Title
Welke invloed
Welke invloed
Log
Custom Tables
Title
Notes
Active Dataset
Table 1

Valid			87,0
			90,7
	8,1	9,3	100,0
	87,1	100,0	
Missing	12,9		
Total	100,0		

```
* Custom
CTABLES
  /VLABE
  /TABLE
  /SLABE
```

```
sieke impact_nieuwe DISPLAY=BOTH
AN] + impact_nieuwe [MEAN]
```

Custom

```
[DataSet
```

```
en\3de bachelor\PWM\Module 2 uni- en bivariate analyse\bruikbare datasets\Politieke besluitvorming\Politie
```

	besluitvorming?	besluitvorming?
Mean	3	3

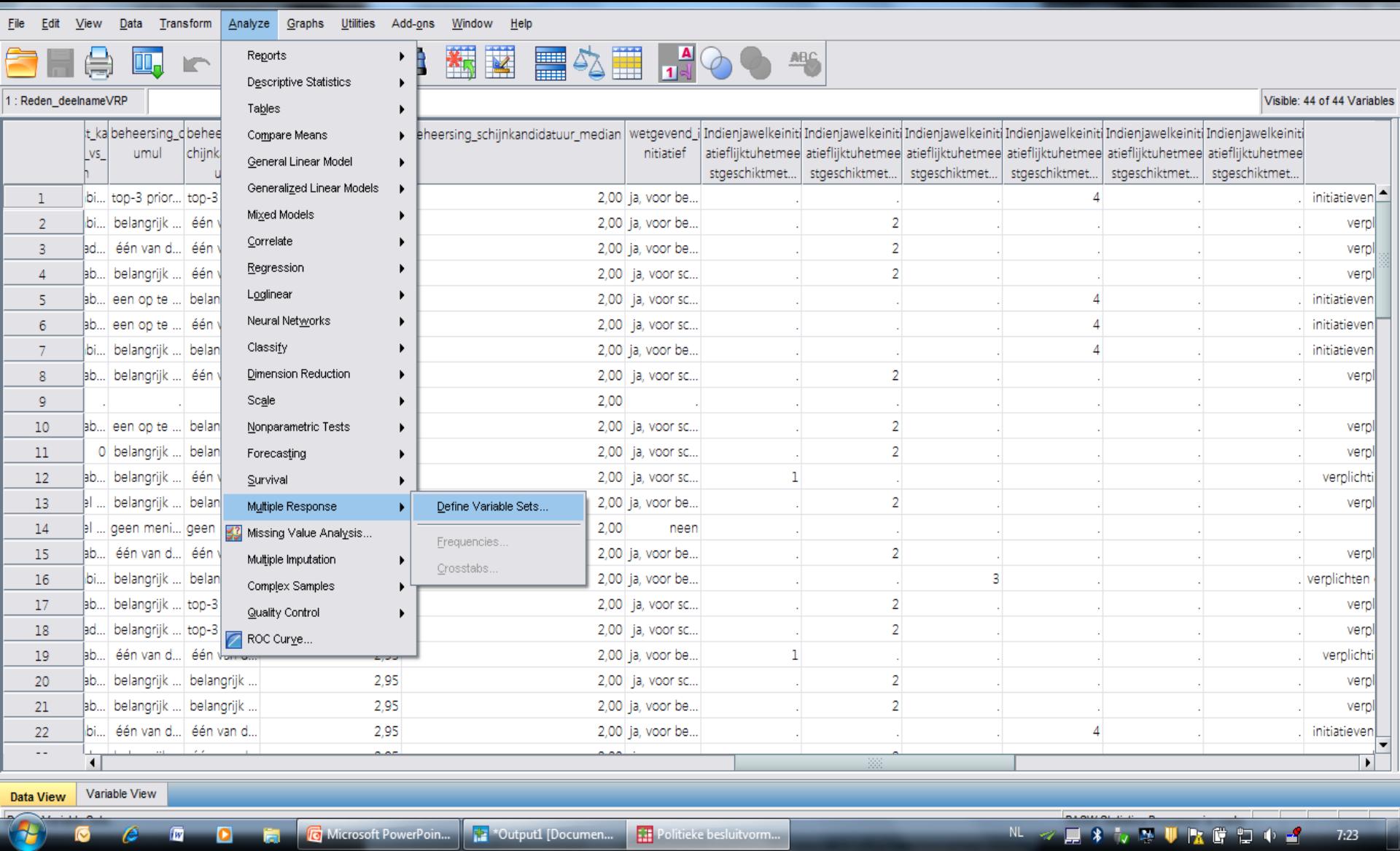
Multiple Response Sets...

PASW Statistics Processor is ready

H: 13.34, W: 16.64 cm

H12 Kiesstelsels (Dev... Microsoft PowerPoin... Politieke besluitvorm... *Output2 [Documen...

NL 8:10

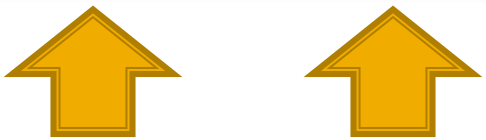

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...
Define Variable Sets...
Frequencies...
Crosstabs...
1 : Reden_deelnameVRP
Visible: 44 of 44 Variables
Data View
Variable View
Microsoft PowerPoin...
*Output1 [Documen...
Politieke besluitvorm...
NL
7:23

Output

$Welke_initiatieven Frequencies

		Responses		
		N	Percent	Percent of Cases
$Welke_initiatieven Welke initiatieven?[a]	1	8	13,3%	16,0%
	2	26	43,3%	52,0%
	3	5	8,3%	10,0%
	4	15	25,0%	30,0%
	5	2	3,3%	4,0%
	6	4	6,7%	8,0%
Total		60	100,0%	120,0%

a. Group



Oefening 3 (thuis)

- Doe dezelfde MRS-analyze voor de VAR:
  - Met welke Vlaamse partijen gaat de formatie lukken? (VAR 23 t.e.m 30 uit ruwe "Politieke besluitvorming_volledige dataset.sav")
  - Met welke Franstalige partijen gaat de formatie niet lukken? (VAR 31 t.e.m. 36)
  - Volg hierbij de stappen op de volgende slide.

- Opdracht :
  - Kopieer deze VAR uit de ruwe dataset.sav Politieke besluitvorming op Minerva
    - **Voor alle zekerheid: sorteer de cases uit beide bestanden eerst op 'RepondentID' (VAR 1; n=60).**
  - Maak aan en bestudeer de frequentietabel voor de MRS-VAR
  - Geef de top-drie Vlaamse en Waalse partijen waarmee de formatie niet zal lukken.

  - Kopieer de outputfile in het reeds geopende WORD-bestand en bespreek de resultaten in functie van de vraag.

$Formatie_partijen_slagen Frequencies

		Responses		
		N	Percent	Percent of Cases
$Formatie_partijen_slagen Met welke Vlaamse en Waalse partijen?[a]	2	39	14,3%	73,6%
	3	26	9,5%	49,1%
	4	33	12,1%	62,3%
	5	1	,4%	1,9%
	6	19	7,0%	35,8%
	7	27	9,9%	50,9%
	8	2	,7%	3,8%
	10	42	15,4%	79,2%
	11	1	,4%	1,9%
	12	31	11,4%	58,5%
	13	31	11,4%	58,5%
	14	21	7,7%	39,6%
Total		273	100,0%	515,1%

a. Group

Enkel nog aanduiden welke nummer voor welke partij staat.
Pas op voor VAR waarvoor geen waarden bestaan !!

- Aangepast via chart editor geeft dit:

$Formatie_partijen_slagen Frequencies

		Responses		
		N	Percent	Percent of Cases
$Formatie_partijen_slagen Met welke Vlaamse en Waalse partijen?[a]	CD&V	39	14,3%	73,6%
	Open VLD	26	9,5%	49,1%
	Sp.a	33	12,1%	62,3%
	Vlaams Belang	1	,4%	1,9%
	N-VA	19	7,0%	35,8%
	Groen!	27	9,9%	50,9%
	PVDA+	2	,7%	3,8%
	PS	42	15,4%	79,2%
	PPopulaire	1	,4%	1,9%
	CDH	31	11,4%	58,5%
	Ecolo	31	11,4%	58,5%
	MR	21	7,7%	39,6%
Total		273	100,0%	515,1%

a. Group

Top-3 Vlaamse partijen

Top-3 Waalse partijen

VAR met dezelfde antwoordcategorieën

- Samen analyseren
- Samenvatten in één enkele FREQ tabel

- Via **Analyze** > **Custom Tables** > Tables of Frequencies

Indeling

- Univariate beschrijvende analyse
- Bivariate beschrijvende analyse

Kruistabellen

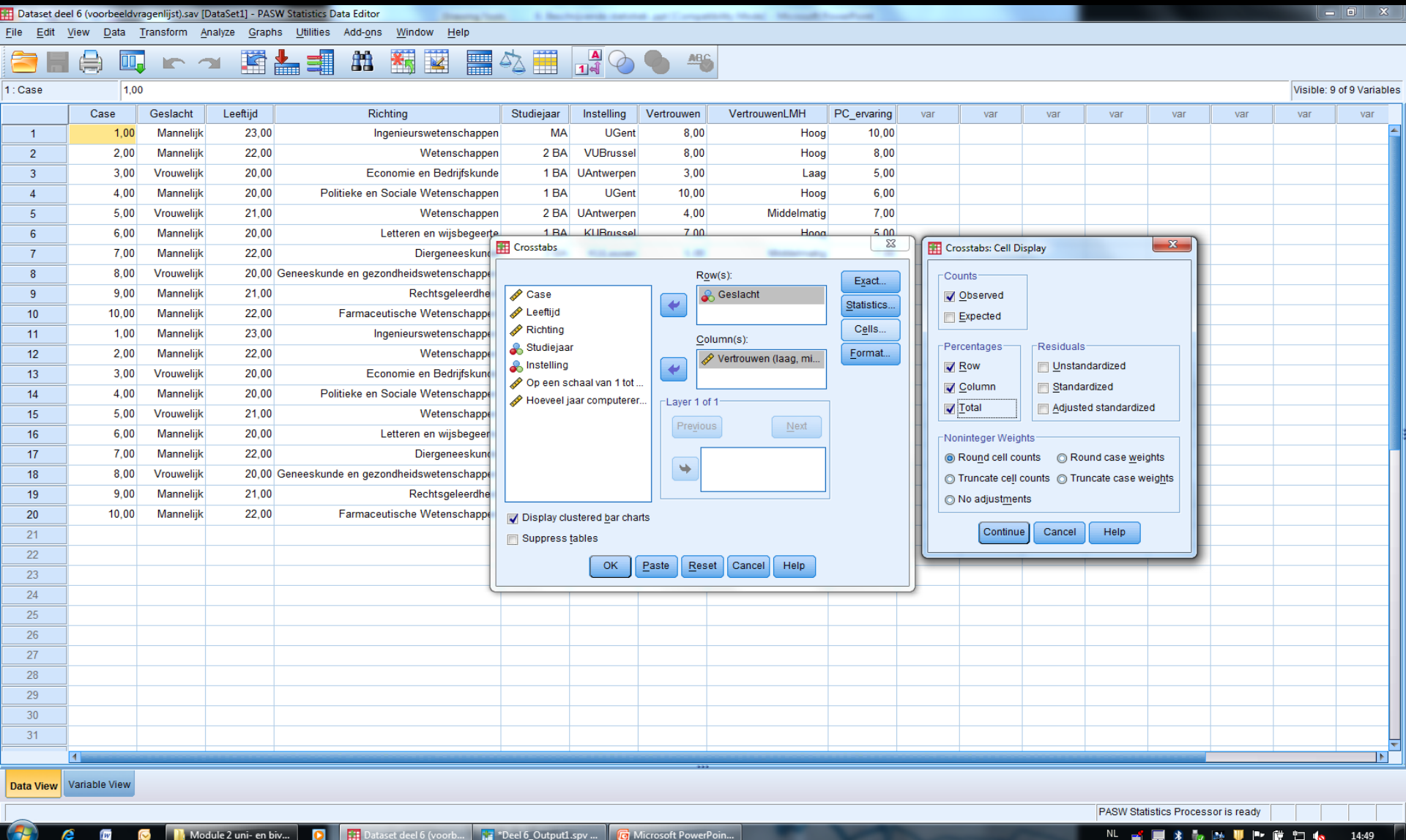
- **Kruistabellen**: een tabel waarin de data (frequentie) voor minstens twee variabelen worden vergeleken door middel van **een indeling in rijen (r) en kolommen (k)**.

- **Frequentietabel**: **vier getallen in elke cel**
  - Frequentie (absolute aantallen)
  - Ruw percentage
  - Tabel met kolompercentages
  - Tabel met rijpercentages

Frequencies voor meerdere variabelen

- Bepaal frequenties voor combinaties van variabelen, bijv. *Geslacht* en *VertrouwenLMH*.
- Selecteer: **Analyze > Descriptive Statistics > Crosstabs**
- Highlight '*Geslacht*' en verplaats naar de Row(s) box.
- Highlight '*VertrouwenLMH*' en verplaatst naar de Column(s) box.
- Check *Display clustered bar charts*
- Selecteer: **Cells**. Check *Percentages* voor ***Row, Column,*** en ***Total***
- Selecteer: **Continue > OK.**

```
CROSSTABS
  /TABLES=Geslacht BY VertrouwenLMH
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT ROW COLUMN TOTAL
  /COUNT ROUND CELL
  /BARCHART.
```



*Deel 6_Output1.spv [Document6] - PASW Statistics Viewer

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

- Output
  - Log
  - Crosstabs
    - Title
    - Notes
    - Active Dataset
    - Case Processing Summary
    - Geslacht * Vertrouwen (laag, mid
    - Bar Chart

```
CROSSTABS
  /TABLES=Geslacht BY VertrouwenLMH
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT ROW COLUMN TOTAL
  /COUNT ROUND CELL
  /BARCHART.
```

Crosstabs

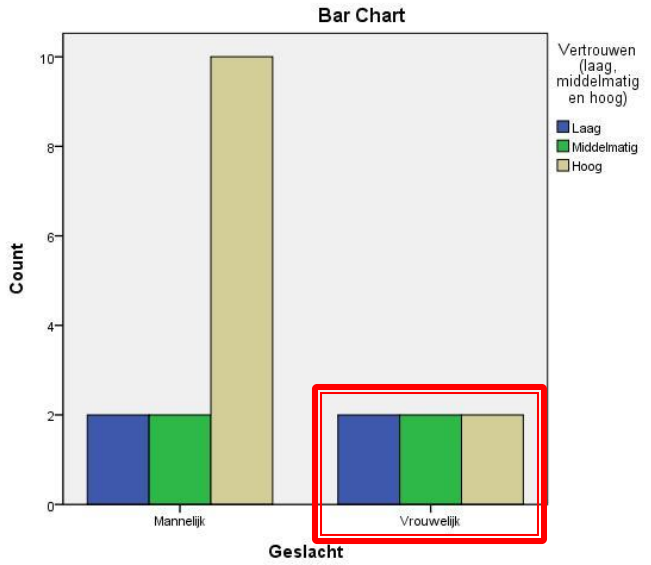
[DataSet1] K:\Opleidingsonderdelen\3de bachelor\Cursusonderdelen\Module 2 uni- en bivariate analyse\bruikbare datasets\Dataset deel 6 (voorbeeldvragenlijst).sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Geslacht * Vertrouwen (laag, middelmatig en hoog)	20	100,0%	0	,0%	20	100,0%

Geslacht * Vertrouwen (laag, middelmatig en hoog) Crosstabulation

			Vertrouwen (laag, middelmatig en hoog)			
			Laag	Middelmatig	Hoog	Total
Geslacht	Mannelijk	Count	2	2	10	14
		% within Geslacht	14,3%	14,3%	71,4%	100,0%
		% within Vertrouwen (laag, middelmatig en hoog)	50,0%	50,0%	83,3%	70,0%
		% of Total	10,0%	10,0%	50,0%	70,0%
	Vrouwelijk	Count	2	2	2	6
		% within Geslacht	33,3%	33,3%	33,3%	100,0%
		% within Vertrouwen (laag, middelmatig en hoog)	50,0%	50,0%	16,7%	30,0%
		% of Total	10,0%	10,0%	10,0%	30,0%
Total		Count	4	4	12	20
		% within Geslacht	20,0%	20,0%	60,0%	100,0%
		% within Vertrouwen (laag, middelmatig en hoog)	100,0%	100,0%	100,0%	100,0%
		% of Total	20,0%	20,0%	60,0%	100,0%



PASW Statistics Processor is ready

Module 2 uni- en biv... | Dataset deel 6 (voorb... | *Deel 6_Output1.spv ... | Microsoft PowerPoin... | NL | 14:53

Analyse van de Output

- **Case Processing Summary**: de aantallen 'valid', 'missing' en 'total' data-counts voor *Geslacht* en *VertrouwenLMH* (dus: deelnemers die beide vragen beantwoordden)
- **Crosstabulation**: de aantallen en percentages van data entries voor elke antwoordcategorie voor beide variabelen samen (= intersecties voor variabele 1 en variabele 2).
- **Bar chart**

- **Dit werkt best voor variabelen die nominaal of ordinaal geschaald zijn.**

Interpretatie van de resultaten

- Van alle respondenten die een hoog vertrouwen stellen in het feit dat ze op het einde van de rit zullen slagen voor de test is 83,3% man.
- Van alle mannen heeft 71,4% een hoog vertrouwen in het feit dat ze zullen slagen voor de test.

- Vraag: Hoeveel % van alle respondenten heeft een laag vertrouwen in de goede afloop?

Oefening 4

- Op basis van de laatste versie van de voorbeelddataset (n = 20) vraag de kruistabel op voor volgende variabelen:

  '*Leeftijd*' en '*Studiejaar*'

  Interpreteer de **kolom- en rijpercentages** en de **clustered bar charts**.

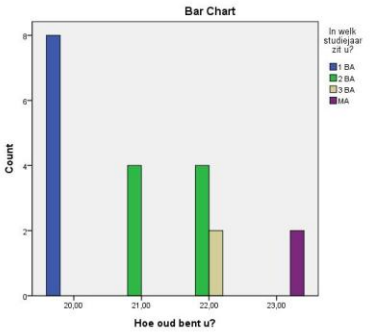
- Copy-paste de output in hetzelfde WORD-bestand waarin je alle vorige outputs hebt geplakt en besproken.

- Kopieer de kruistabel én de clustered bar chart (via en eenvoudige "copy-paste") in een WORD-doc en bespreek beide (5-10 lijnen: **wat valt u op?**)

Oplossing oefening 4

Hoe oud bent u? * In welk studiejaar zit u? Crosstabulation

			In welk studiejaar zit u?				
			1 BA	2 BA	3 BA	MA	Total
Hoe oud bent u?	20,00	Count	8	0	0	0	8
		% within Hoe oud bent u?	100,0%	,0%	,0%	,0%	100,0%
		% within In welk studiejaar zit u?	100,0%	,0%	,0%	,0%	40,0%
		% of Total	40,0%	,0%	,0%	,0%	40,0%
	21,00	Count	0	4	0	0	4
		% within Hoe oud bent u?	,0%	100,0%	,0%	,0%	100,0%
		% within In welk studiejaar zit u?	,0%	50,0%	,0%	,0%	20,0%
		% of Total	,0%	20,0%	,0%	,0%	20,0%
	22,00	Count	0	4	2	0	6
		% within Hoe oud bent u?	,0%	66,7%	33,3%	,0%	100,0%
		% within In welk studiejaar zit u?	,0%	50,0%	100,0%	,0%	30,0%
		% of Total	,0%	20,0%	10,0%	,0%	30,0%
	23,00	Count	0	0	0	2	2
		% within Hoe oud bent u?	,0%	,0%	,0%	100,0%	100,0%
		% within In welk studiejaar zit u?	,0%	,0%	,0%	100,0%	10,0%
		% of Total	,0%	,0%	,0%	10,0%	10,0%
		Count	8	8	2	2	20
		% within Hoe oud bent u?	40,0%	40,0%	10,0%	10,0%	100,0%
		% within In welk studiejaar zit u?	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	40,0%	40,0%	10,0%	10,0%	100,0%



Scatter plot

- Een scatter plot is een gebruikelijke manier om de resultaten van een bivariate analyse weer te geven.
- Een derde en vierde variabele toevoegen kan via *set markers by*.
- Elke as vertegenwoordigt een variabele en de **stippen (bolletjes) vertegenwoordigen de intersectie van de scores** die deelnemers hebben gegeven op beide variabelen.

Bron: http://bcs.whfreeman.com/ips5e/content/cat_120/IPS_Splus.pdf

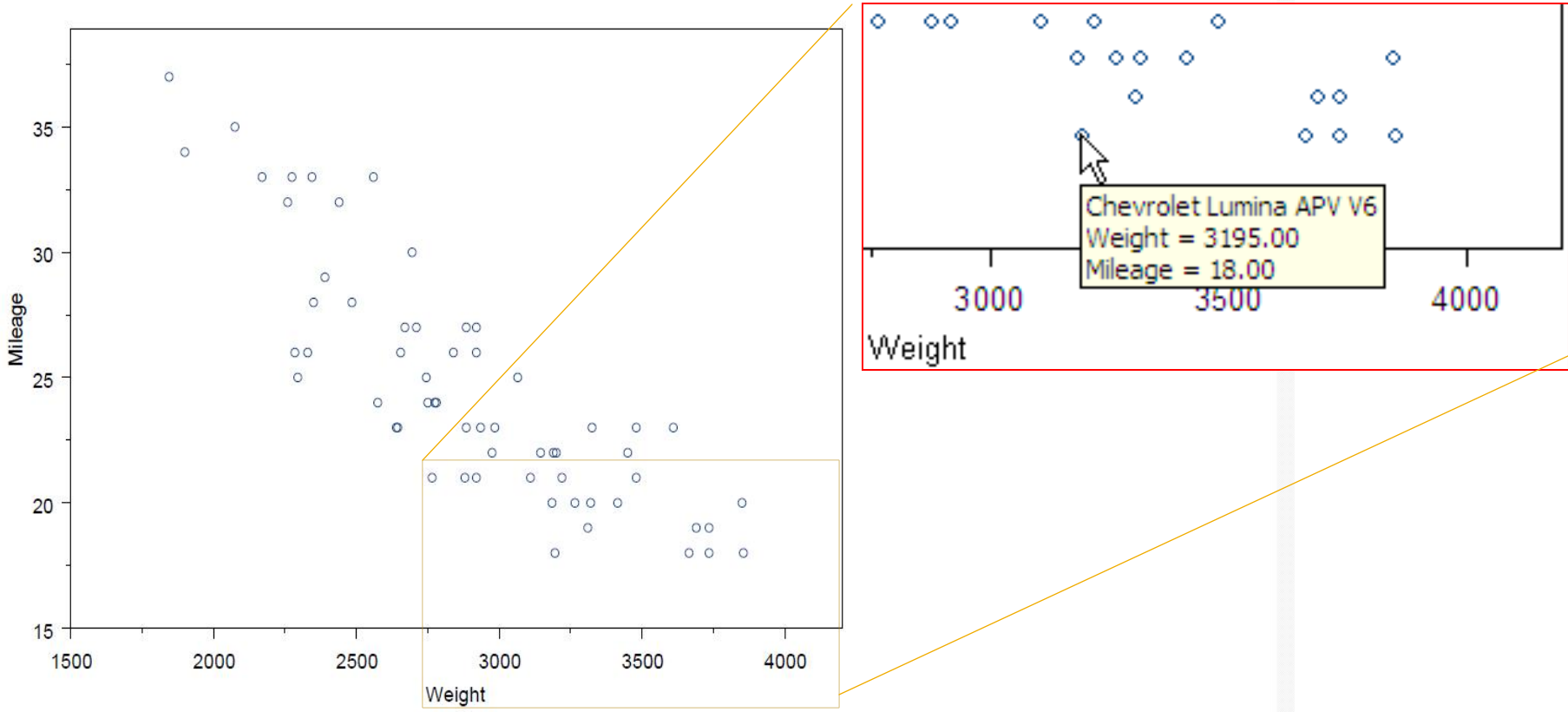
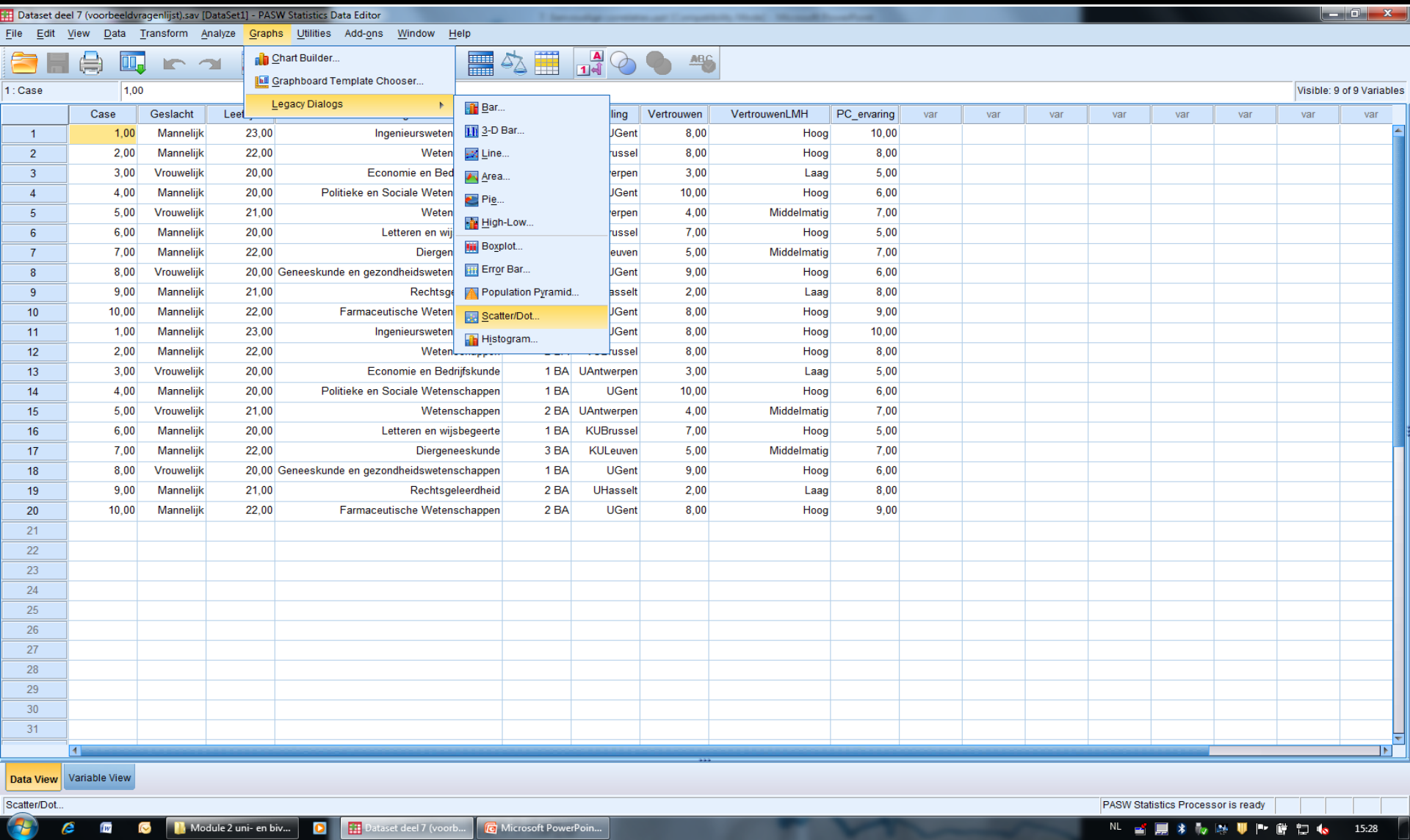


Figure 2.1

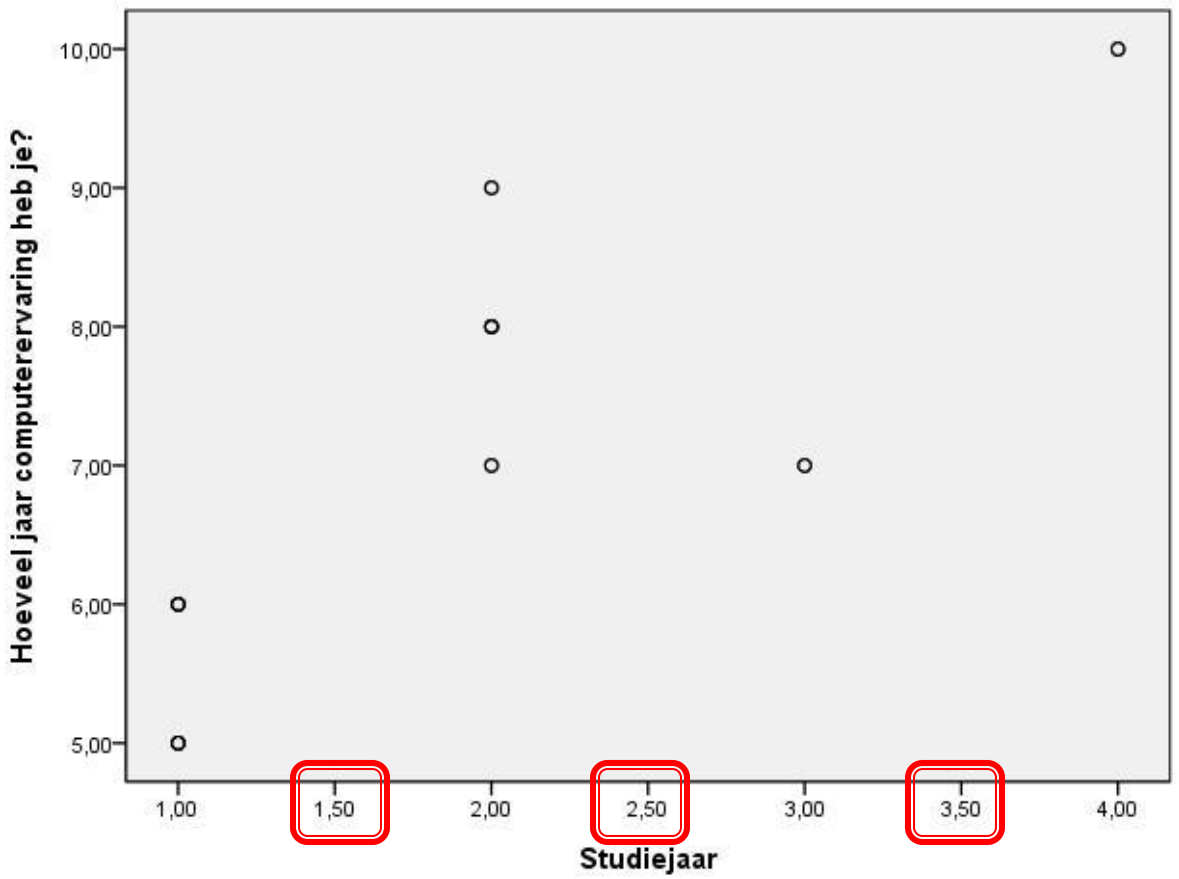
Voorbeeld

- Selecteer: **Graphs > Scatter/Dot > Simple Scatter > Define**
- Highlight *PC-ervaring*. En plaats in de *Y Axis* box.
- Highlight *Studiejaar*. En plaats in de *X Axis* box.


Dataset deel 7 (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Chart Builder...
Graphboard Template Chooser...
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
1 : Case
1,00
Visible: 9 of 9 Variables
Case Geslacht Leef ling Vertrouwen VertrouwenLMH PC_ervaring var var var var var var var var
1 1,00 Mannelijk 23,00 Ingenieursweten UGent 8,00 Hoog 10,00
2 2,00 Mannelijk 22,00 Weten russel 8,00 Hoog 8,00
3 3,00 Vrouwelijk 20,00 Economie en Bed erpen 3,00 Laag 5,00
4 4,00 Mannelijk 20,00 Politieke en Sociale Weten UGent 10,00 Hoog 6,00
5 5,00 Vrouwelijk 21,00 Weten erpen 4,00 Middelmatig 7,00
6 6,00 Mannelijk 20,00 Letteren en wij russel 7,00 Hoog 5,00
7 7,00 Mannelijk 22,00 Diergen euven 5,00 Middelmatig 7,00
8 8,00 Vrouwelijk 20,00 Geneeskunde en gezondheidsweten UGent 9,00 Hoog 6,00
9 9,00 Mannelijk 21,00 Rechtsge asselt 2,00 Laag 8,00
10 10,00 Mannelijk 22,00 Farmaceutische Weten UGent 8,00 Hoog 9,00
11 1,00 Mannelijk 23,00 Ingenieursweten UGent 8,00 Hoog 10,00
12 2,00 Mannelijk 22,00 Weten russel 8,00 Hoog 8,00
13 3,00 Vrouwelijk 20,00 Economie en Bedrijfskunde 1 BA UAntwerpen 3,00 Laag 5,00
14 4,00 Mannelijk 20,00 Politieke en Sociale Wetenschappen 1 BA UGent 10,00 Hoog 6,00
15 5,00 Vrouwelijk 21,00 Wetenschappen 2 BA UAntwerpen 4,00 Middelmatig 7,00
16 6,00 Mannelijk 20,00 Letteren en wijsbegeerte 1 BA KUBrussel 7,00 Hoog 5,00
17 7,00 Mannelijk 22,00 Diergeneeskunde 3 BA KULeuven 5,00 Middelmatig 7,00
18 8,00 Vrouwelijk 20,00 Geneeskunde en gezondheidswetenschappen 1 BA UGent 9,00 Hoog 6,00
19 9,00 Mannelijk 21,00 Rechtsgeleerdheid 2 BA UHasselt 2,00 Laag 8,00
20 10,00 Mannelijk 22,00 Farmaceutische Wetenschappen 2 BA UGent 8,00 Hoog 9,00
21
22
23
24
25
26
27
28
29
30
31
Data View
Variable View
Scatter/Dot...
PASW Statistics Processor is ready
Module 2 uni- en biv...
Dataset deel 7 (voorb...
Microsoft PowerPoin...
NL
15:28

- Inlassen van **een voetnoot of een commentaar ("subscript")**:
  - Selecteer: **Titles**.
  - In de *Footnote Line 1* box, typ "Figuur 1. Er is een significant positief verband tussen de variabelen Studiejaar en PC-ervaring."
  - Selecteer: **Continue > OK**
- Resultaat zie volgende slide

```
GRAPH
  /SCATTERPLOT(BIVAR)=Studiejaar WITH PC_ervaring
  /MISSING=LISTWISE.
```



Figuur 1. Er is een significant positief verband tussen de variabelen Studiejaar en PC-ervaring.

- Om de schaal van de X-as aan te passen
  - Double klik op de plot
  - In de *Chart editor* selecteer de X-as (blauw)
  - Pas de schaal aan in de tab 'Scale'
  - Klik **Apply** en **Close**
  - Klik buiten de box in de SPSS Viewer om de Chart Editor af te sluiten en de aanpassingen zullen zichtbaar worden.

- Rechtermuisklik op de plot**+ Copy / Paste** om de plot in een document te plakken.

*Output15 [Document16] - PASW Statistics Viewer

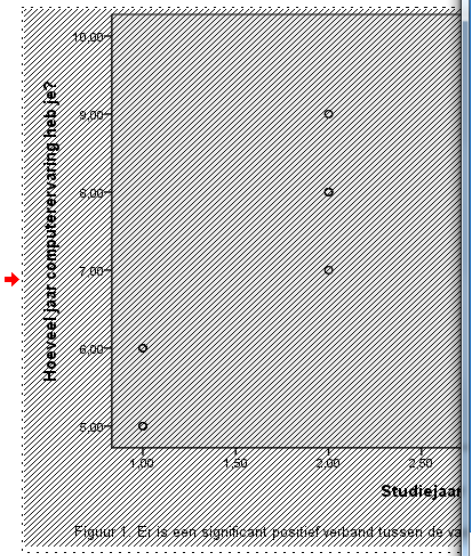
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

- Title
- Notes
- Active Dataset
- Scatter of PC_ervaring Studieja

```
GRAPH
  /SCATTERPLOT(BIVAR)=Studiejaar WITH PC_ervaring
  /MISSING=LISTWISE
  /FOOTNOTE='Figuur 1. Er is een significant po
```

Graph

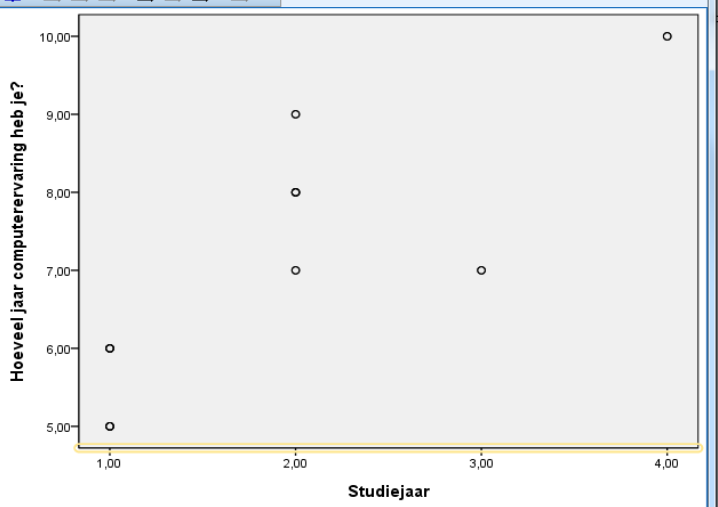
[DataSet1] K:\Opleidingsonderdelen\3de bachelor



Figuur 1. Er is een significant positief verband tussen de va

Chart Editor

File Edit View Options Elements Help



Figuur 1. Er is een significant positief verband tussen de variabelen Studiejaar en PC-ervaring.

H:375, W:468.75 points

Properties

Labels & Ticks | Number Format | Variables

Chart Size | Scale | Lines

Range

	Auto	Custom	Data
Minimum	☑	1	1
Maximum	☑	4	4
Major Increment	☐	1	
Origin	☑	0	

☐ Display line at origin

Type

- ◉ Linear
- ○ Logarithmic — Base: 10 ☑ Safe
- ○ Power — Exponent: 0,5 ☑ Safe

Lower margin (%): 5 Upper margin (%): 5

Apply Close Help

PASW Statistics Processor is ready H: 13.34, W: 16.64 cm

Module 2 uni- en biv... | Dataset deel 7 (voorb... | *Output15 [Docume... | Chart Editor | Microsoft PowerPoin...

NL 15:47

Oefening 5 (thuis)

- Dataset "Politieke besluitvorming.sav" laatste versie:
  - Hercodeer de VAR VRP_belangrijkheid in een nieuwe VAR VRP_belangrijkheid_REC met 'oplopende schaal' (dus: 1 = helemaal niet belangrijk, 2 = neutraal, 3 = eerder belangrijk, 4 = helemaal wel belangrijk)

- Maak een kruistabel (+ bar chart) en een scatterplot aan voor de VAR *Leeftijd* en deze VAR *VRP_belangrijkheid_REC*

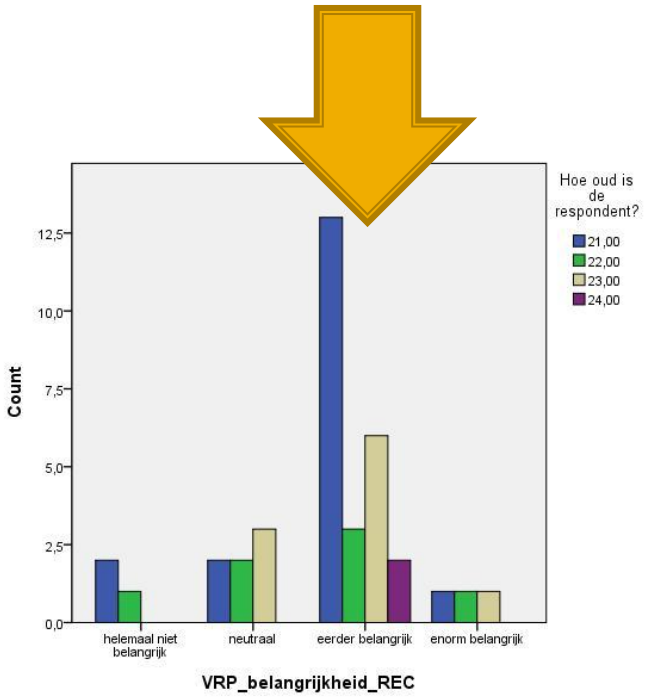
- Pas de indeling van de X en Y-as van de plot aan indien nodig. Copy-paste de plot in WORD en interpreteer het resultaat.
- Drop het WORD-bestand met alle aangemaakte tabellen én de bespreking in **Minerva dropbox (t.a.v. tom.schamp@ugent.be)**

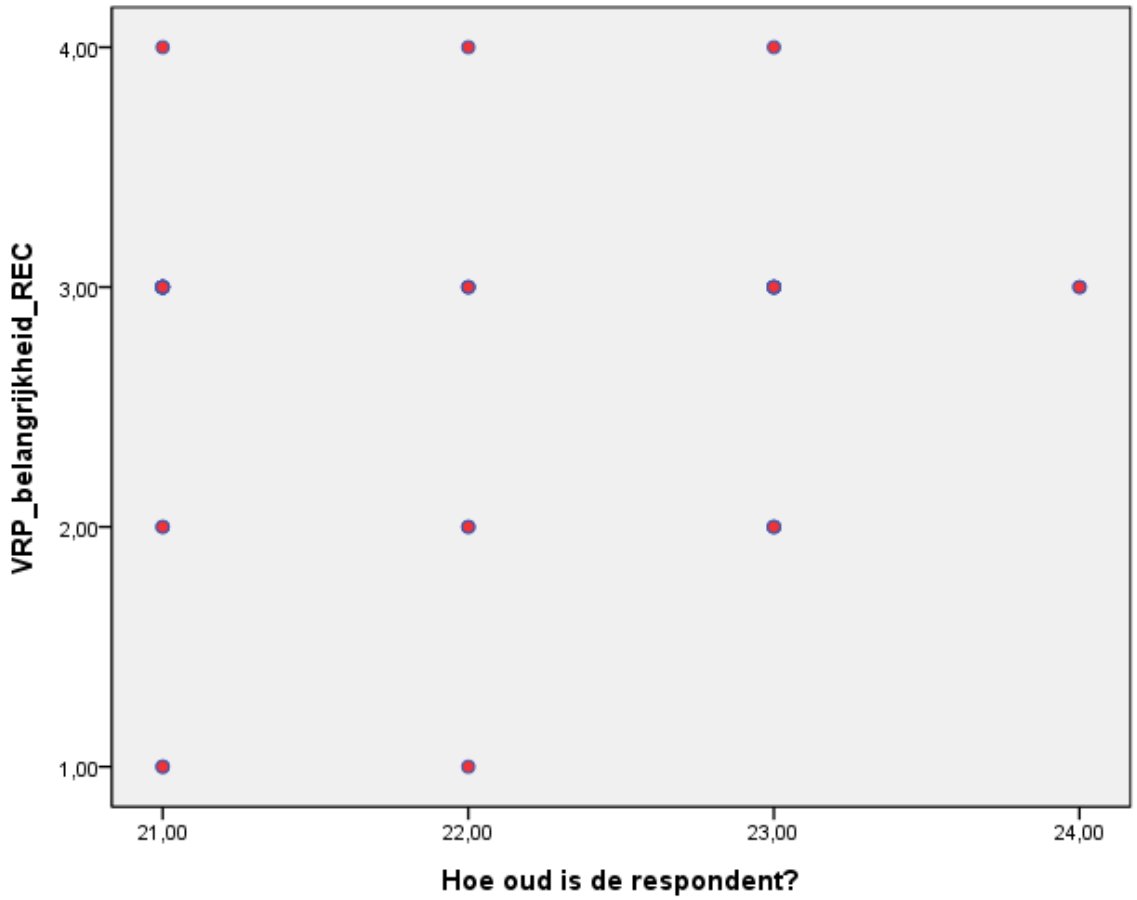
- Deadline: **wo 27 april a.s. 12:30pm** !

Oplossing oefening 5

VRP_belangrijkheid_REC * leeftijd Hoe oud is de respondent? Crosstabulation

			leeftijd Hoe oud is de respondent?				
			21,00	22,00	23,00	24,00	Total
VRP_belangrijkheid_REC	helemaal niet belangrijk	Count	2	1	0	0	3
		% within VRP_belangrijkheid_REC	66,7%	33,3%	,0%	,0%	100,0%
		% within leeftijd Hoe oud is de respondent?	11,1%	14,3%	,0%	,0%	8,1%
		% of Total	5,4%	2,7%	,0%	,0%	8,1%
	neutraal	Count	2	2	3	0	7
		% within VRP_belangrijkheid_REC	28,6%	28,6%	42,9%	,0%	100,0%
		% within leeftijd Hoe oud is de respondent?	11,1%	28,6%	30,0%	,0%	18,9%
		% of Total	5,4%	5,4%	8,1%	,0%	18,9%
	eerder belangrijk	Count	13	3	6	2	24
		% within VRP_belangrijkheid_REC	54,2%	12,5%	25,0%	8,3%	100,0%
		% within leeftijd Hoe oud is de respondent?	72,2%	42,9%	60,0%	100,0%	64,9%
		% of Total	35,1%	8,1%	16,2%	5,4%	64,9%
	enorm belangrijk	Count	1	1	1	0	3
		% within VRP_belangrijkheid_REC	33,3%	33,3%	33,3%	,0%	100,0%
		% within leeftijd Hoe oud is de respondent?	5,6%	14,3%	10,0%	,0%	8,1%
		% of Total	2,7%	2,7%	2,7%	,0%	8,1%
		Count	18	7	10	2	37
		% within VRP_belangrijkheid_REC	48,6%	18,9%	27,0%	5,4%	100,0%
		% within leeftijd Hoe oud is de respondent?	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	48,6%	18,9%	27,0%	5,4%	100,0%




VRP_belangrijkheid_REC
4,00
3,00
2,00
1,00
21,00
22,00
23,00
24,00
Hoe oud is de respondent?

Wat hebben we geleerd ?

- **Frequentietabel opvragen**: Analyze, Descriptive statisctics, Frequencies, variabele(n) overbrengen naar rechterkolom, OK

- **Frequentietabel opvragen en univariate statistische parameters berekenen**: Analyze, Descriptive statisctics, Frequencies, variabele(n) overbrengen naar rechterkolom, Statistics, aanduiden van parameters die je wil berekenen (vb.mean = rekenkundig gemiddelde), Continue, OK

- **Univariate statistische parameters berekenen zonder frequentietabel**: Analyze, Descriptive statisctics, Frequencies, variabele(n) overbrengen naar rechterkolom, optie 'Display frequencies tables' uitzetten, Statistics, aanduiden van parameters die je wil berekenen (vb.mean=rekenkundig gemiddelde), Continue, OK

Wat geleerd ? - 2

- **Frequentietabel opvragen en frequentieverdeling in grafische vorm weergeven**: Analyze, Descriptive statisctics, Frequencies, variabele(n) overbrengen naar rechterkolom, Charts, aanduiden van gewenste grafiek, Continue, OK

- **Frequentieverdeling in grafische vorm weergegeven** (zonder frequentieverdeling): 2 mogelijkheden:
  - Analyze, Descriptive statisctics, Frequencies, variabele(n) overbrengen naar rechterkolom, optie 'Display frequencies tables' uitzetten, Charts, aanduiden van gewenste grafiek, Continue, OK
  - Graphs, gewenste grafiek aanduiden (vb. Bar chart, pie chart, …), variabele(n) overbrengen, OK

    Opmerking: Graphs, Gallery : hierin geeft men je uitleg over de verschillende mogelijke grafieken

Wat geleerd ? - 3

- **Frequentietabel opvragen en waarden veranderen van oplopend naar aflopend**: Analyze, Descriptive statistics, Frequencies, variabele(n) overbrengen naar rechterkolom, Format, bij optie Order 'descending values' aanduiden, Continue, OK (Opmerking: Optie *'Suppres tables with more than n categories*': er wordt dan enkel een frequentietabel getoond als de categoriën van variabelen kleiner of gelijk zijn aan n.)

- **Kruistabel opvragen**: Analyze, Descriptive statistics, Crosstabs, variabelen overbrengen naar rechterkolommen, OK

- **Opvragen van een boxplot voor de verdeling van 1 variabele**: Graphs, Boxplot, optie *define summaries of separate variables* aanduiden, variabele kiezen, OK
- **Scatterplot opvragen**: Graphs, Scatter, optie Simple Define kiezen, variabelen overbrengen naar rechterkolommen, OK

Deel 6bis

De normaliteit van een verdeling

De normaalverdeling

- Een andere manier om de verdeling van **een metrische variabele** te beschrijven (*naast: min/max, range, variantie, standaardafwijking (cf. link Khanacademy.org)*).

- **De normaalverdelingsfunctie is een type kansdichtheidsfunctie**
  - De bijbehorende kansdichtheid is **hoog in het midden**, en wordt naar lage en hoge waarden steeds kleiner zonder ooit echt nul te worden.
  - Ook wel Gauss-verdeling genaamd (vernoemd naar de Duitse wiskundige Carl Friedrich Gauss)

- Normaalverdelingen geven ons een idee van een mogelijke theoretische verdeling
  - Ze verschillen sterk van elkaar in **gemiddelde** en **standaardafwijking**
  - Ze zijn identiek in **de symmetrische klokvorm**

- Centrale begrippen en concepten:
  - Standaardafwijking (13min)
  - Betrouwbaarheidsinterval (14min)
  - Centrale limietstelling (10min): voor n steekproeven

    ∑ van de steekproevenverdelingen // verdeling van de populatie
  - Kans (p) en overschrijdingskans (α): 5%, 1% of 0.1%…

- Wiskundige formule v/d **normaalverdelingskromme**:
  - Twee parameters, μ en σ, komen voor
  - Dus: een normaalverdeling heeft een **verwachtingswaarde μ** en **standaardafwijking σ**. (of variantie, σ^2)

- De normale verdeling is perfect symmetrisch :
  - de verwachtingswaarde **μ** van de verdeling is **het 'middelpunt'** van de grafiek van de verdelingsfunctie of het (geschatte) steekproefgemiddelde
  - de 'breedte' van de grafiek (van de kansdichtheid) wordt gekarakteriseerd door de standaarddeviatie **σ** (of de variantie **σ^2**): **schatters voor de steekproefvariantie**

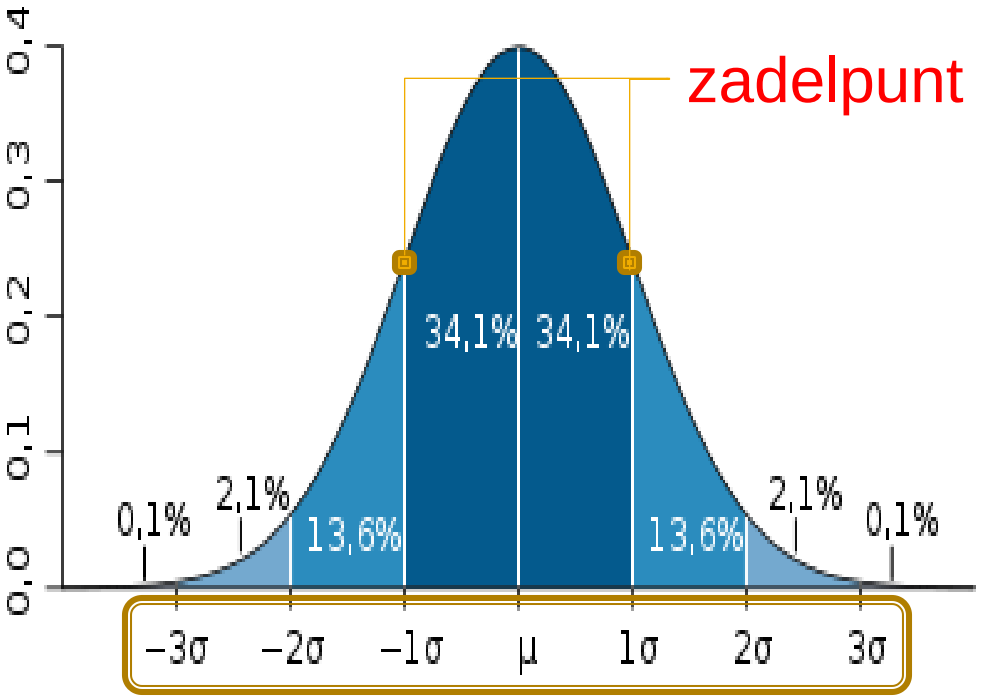
- De functie is "een theoretische benadering van de verdeling"… (een geïdealiseerd beeld)
- Voor heel veel natuurlijk voorkomende verschijnselen is een normale verdeling een goede beschrijving van de frequentie waarmee bepaalde meetwaarden kunnen voorkomen; daarom wordt vaak een normale verdeling verondersteld voor de **onderliggende (populatie)verdeling**.

- Voorbeeld: Zo kan men een schatting maken van bv. de ***gemiddelde lengte*** en ***de standaardafwijking*** van de verdeling van (alle) Nederlandse mannen door een steekproef van een honderdtal mannen te nemen en daarvan de lengte te meten.

- Andere voorbeelden: schoenmaat, gewicht, dagopbrengst zonnepanelen….

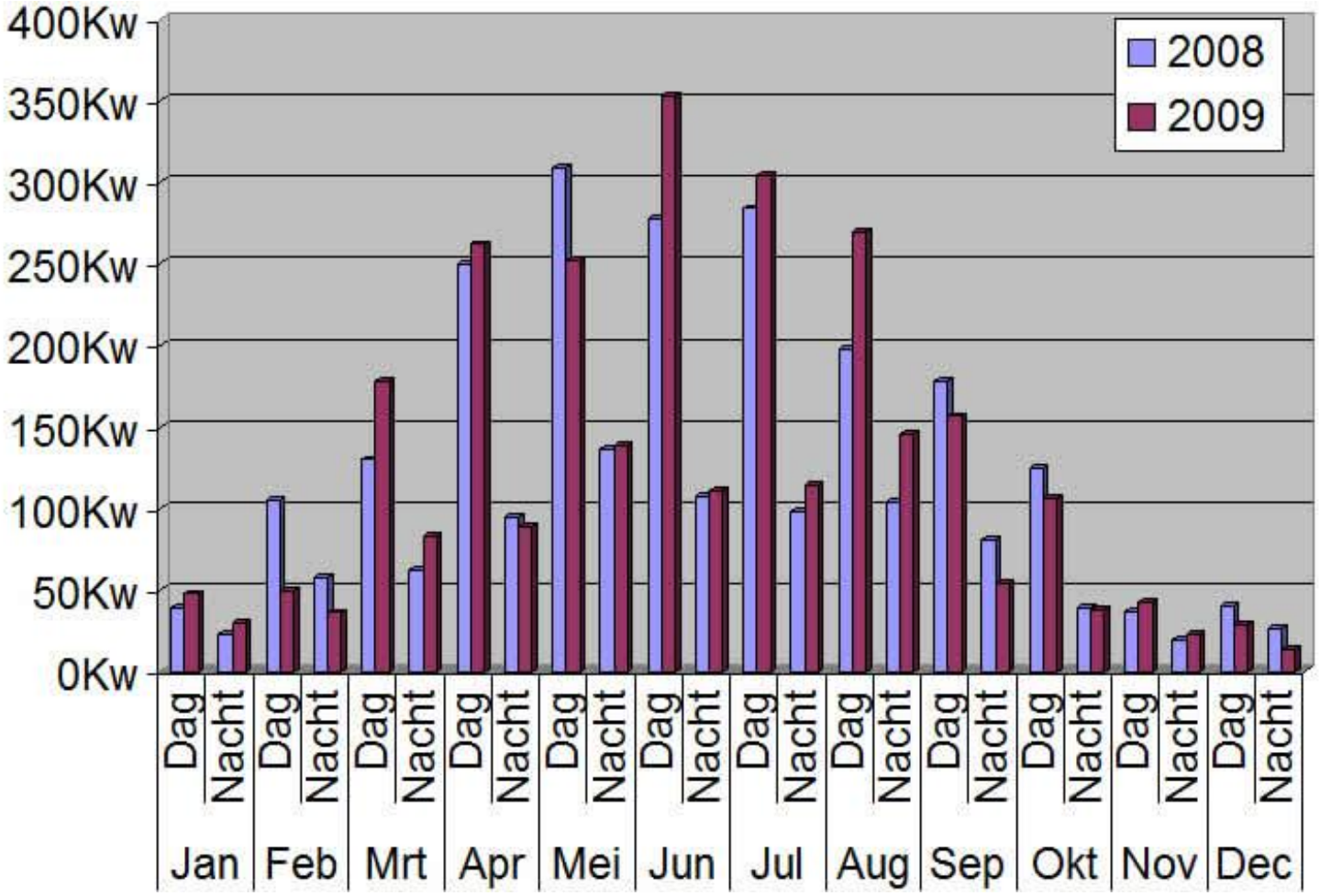
Normaalverdelingskromme

- **Functieverloop**
  - Minimum op – oneindig
  - μ – (3, 2, 1) σ
  - Gemiddelde (μ)
  - μ + (1, 2, 3) σ
  - Maximum op + oneindig



Uitgewerkt excel-voorbeeld via Khanacademy.org

Voorbeeld


400Kw
350Kw
300Kw
250Kw
200Kw
150Kw
100Kw
50Kw
0Kw
2008
2009
Dag
Nacht
Jan
Feb
Mrt
Apr
Mei
Jun
Jul
Aug
Sep
Okt
Nov
Dec

Functieverloop

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Binnen een afstand van één standaardafwijking (**s**) van de verwachtingswaarde ligt ongeveer **68,2%** van het oppervlak onder de grafiek van de kansdichtheid van de normale verdeling
- Binnen de afstand van twee standaardafwijkingen (**2s**) van de verwachtingswaarde: ongeveer **95,4%**
- De curve gaat daarna vrij snel naar nul: ongeveer **99,6%** van het oppervlak ligt binnen drie standaardafwijkingen (**3s**) afstand van het midden.
- En **99,8%** binnen vier standaardafwijkingen (**4s**) van het midden. Afwijkingen van meer dan vier keer de standaardafwijking zijn dus zeer zeldzaam.

Centrale limietstelling

- **De "willekeurige verdeling"**

  Dankzij de centrale limietstelling weten we dat voor veel verdelingen de verdeling van de som en dus ook de verdeling van het gemiddelde van een groot aantal onafhankelijke waarnemingen (steekproeven) daaruit, bij voldoende veel metingen bij benadering de vorm van een normale verdeling heeft.
  Bijgevolg gelden de bovengenoemde percentages niet alleen voor de normale verdeling, maar bij benadering ook voor het gemiddelde van een groot aantal onafhankelijke waarnemingen uit veel onbekende verdelingen.

Standaardnormale verdeling

- = gestandaardiseerde normale verdeling

  Z-scores (8min) $z = (x - \mu) / \sigma$

  - Bijzondere normaalverdeling met **gemiddelde = 0** en **σ = 1**
  - Ook wel 'de moeder van alle normaalverdelingen' genoemd

  - Via **Transform** > **Compute**
  - Of via **Analyze > Descriptive statistics > Descriptives**
    + vink "Save standardized values as variables"

1. Neem een willekeurige dataset of maak er een aan.

2. Enter 720 and 820 into the first column.

3. At the menu, select Data > Distribution Functions....

4. In the Distribution Functions dialog box, select the name of the data set for Data Set:.

5. For Source Column: select the column that you just created.

6. The output is sent to a new column. Designate the name of this new column in the Target Column: field (the default name is `Probability`).

7. Under Distribution Parameters, For Mean: enter 1026, and for Std. Deviation: enter 209.

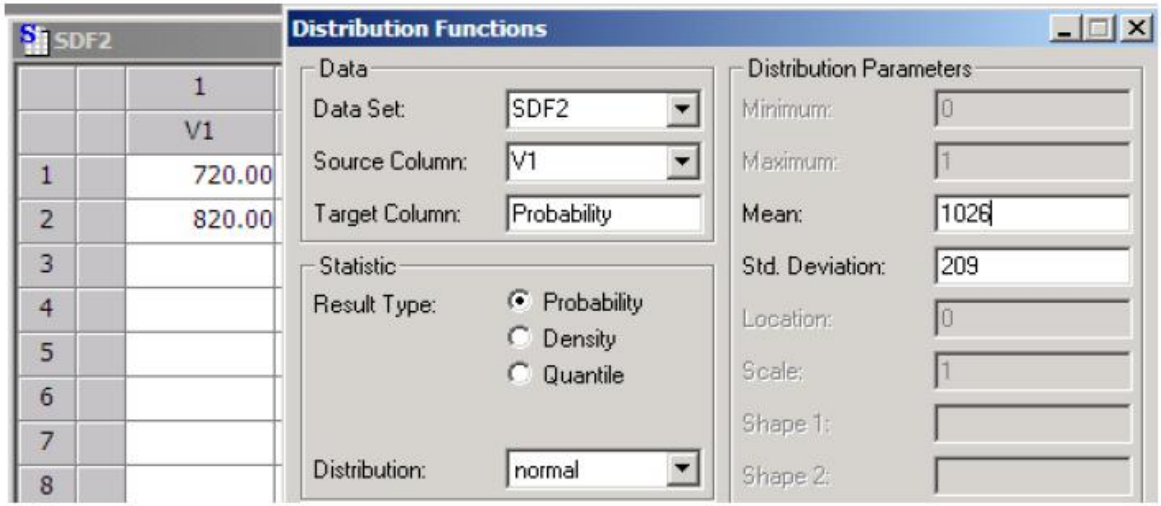


Figure 1.16

Bron: http://bcs.whfreeman.com/ips5e/content/cat_120/IPS_Splus.pdf

8. Click OK.

9. Select the new column and click on the Increase precision button twice to view more significant digits.

If we let X denote the SAT score, then the S-PLUS output tells us that $P(X \leq 720) = 0.0716$ and $P(X \leq 820) = 0.1622$. Thus, we want $P(720 \leq X \leq 820) = 0.1622 - 0.0716 = 0.0906$.

End of Example 1.8

Unknown Zone | Protecte
http://bcs.whfreema... Module 2 uni- en biv... 00. Voorkennis Statis... 6. Eenvoudige besch... NL

Scheefheid (skewness)

- *"Skewness is a measure of symmetry, or more precisely, the lack of symmetry (cf. situatie 4). A distribution, or data set, is symmetric if it looks the same to the left and right of the center point."*

- [-2; 2] (of [-1; 1]) = symmetrisch

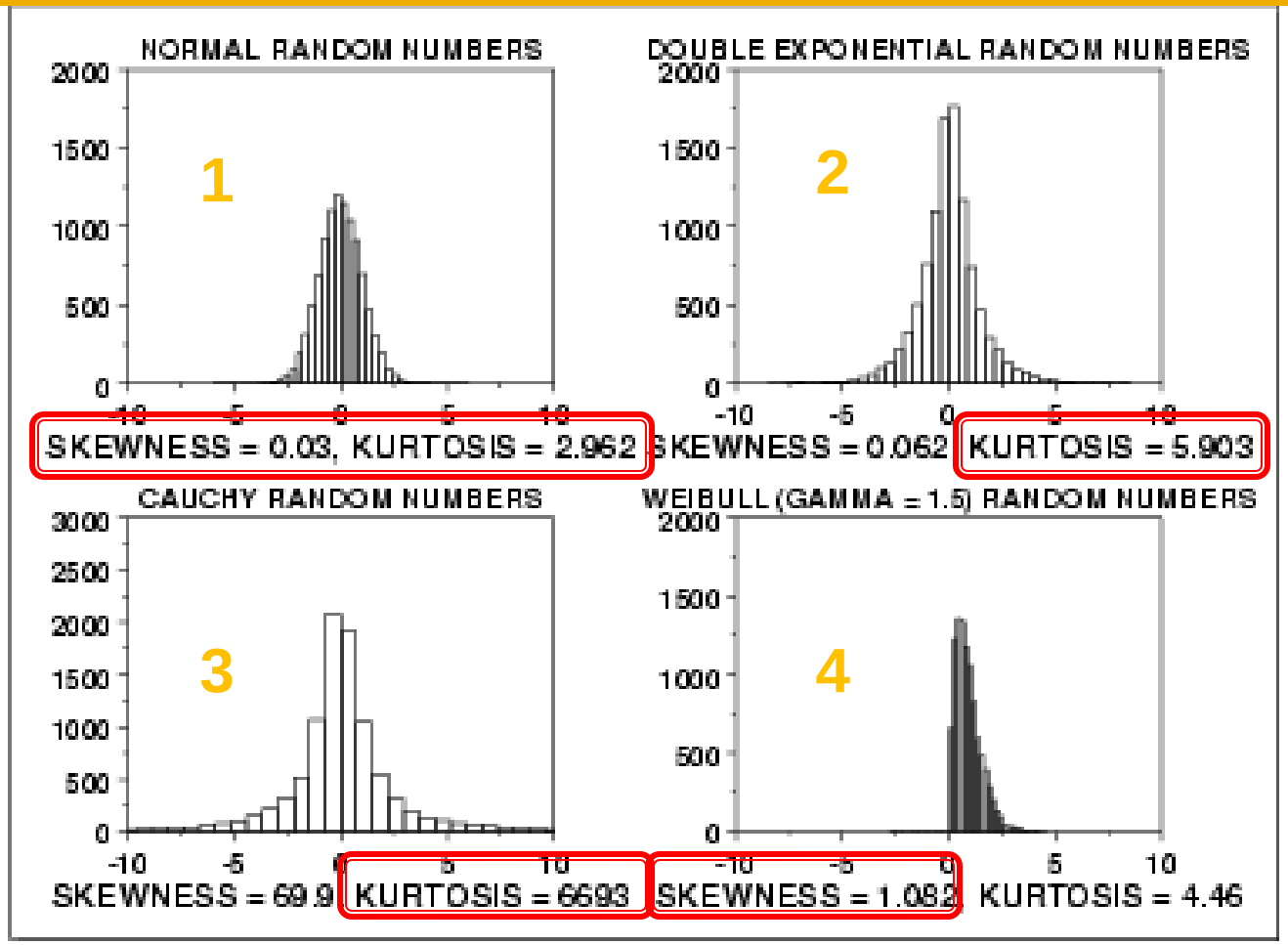
- **Toepassing vier gevallen** (Khanacademy.org)

Afplatting (kurtosis)

- *"Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. (situatie 3) Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case."*

- **Leptokurtisch** (+) vs. **platokurtisch** (-)
- [-2; 2] = niet significant afwijkend van de normaalverdelingskromme

"The histogram is an effective graphical technique for showing both the skewness and kurtosis of data set."



Onderzoek van de normaliteit

Via **Analyze** > **Descriptive statistics** > **Explore**

Opvragen van :

- Plots
- Boxplot
- Stemleaf
- Histogram
- Nplot

*Output1 [Document1] - PASW Statistics Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Monospaced 10,5

Output
- Log
- Explore
  - Title
  - Notes
  - Active Dataset
  - Case Processing Summar
  - Descriptives
  - Tests of Normality
  - Leeftijd gekend bij ACLVB
    - Title
    - Histogram
    - Stem-and-Leaf Plot
    - Normal Q-Q Plot
    - Detrended Normal Q-C
    - Boxplot

```
GET
  FILE='K:\Opleidingsonderdelen\3de bachelor\PWM\Module 2 uni- en bivariate analyse\bruikbare datasets\ACLVB\ACLVB_Werkbestand.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
EXAMINE VARIABLES=Leeftijd_AdmInfo
  /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Explore

[DataSet1] K:\Opleidingsonderdelen\3de bachelo

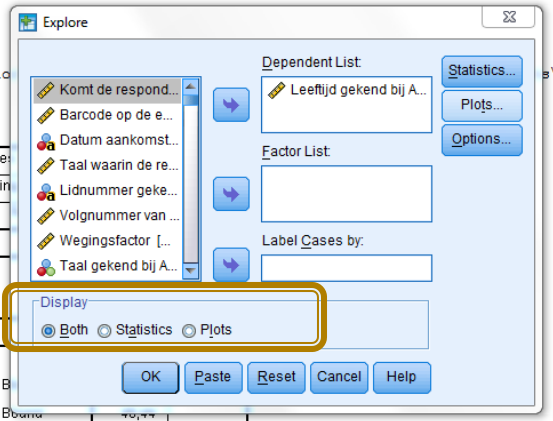
Case Processing Summary

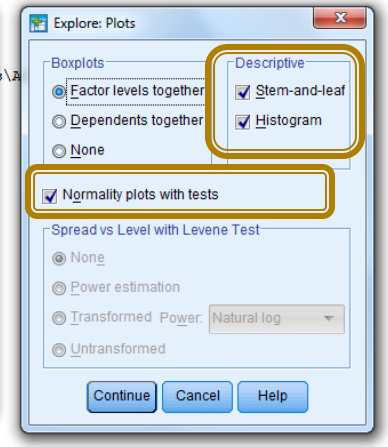
	Case			
	Valid		Missin	
	N	Percent	N	
Leeftijd gekend bij ACLVB	4709	99,0%	50	

Descriptives

Leeftijd gekend bij ACLVB	Mean		
	95% Confidence Interval for Mean	Lower B	
		Upper Bound	
	5% Trimmed Mean	45,97	
	Median	46,00	
	Variance	178,987	
	Std. Deviation	13,379	
	Minimum	15	
	Maximum	95	
	Range	80	
	Interquartile Range	20	
	Skewness	,005	,036
	Kurtosis	-,565	,071

Tests of Normality



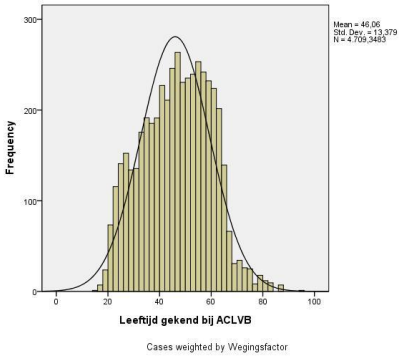


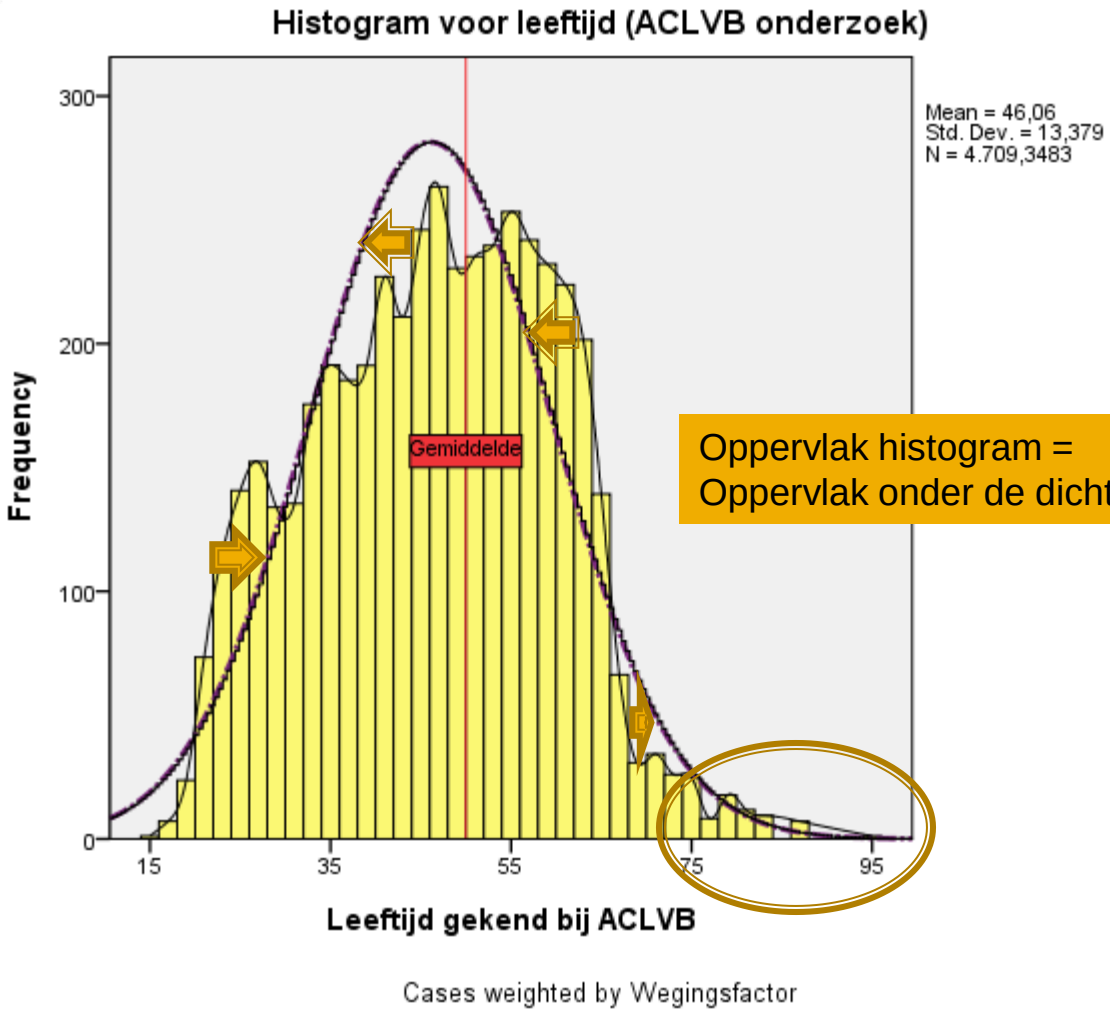
PASW Statistics Processor is ready

H: 5.87, W: 37.33 cm

Inbox in Tom.Scham... 6bis. De normaliteit v... 6. Eenvoudige besch... *Output1 [Documen... *ACLVB_Werkbestan... NL 11:31

Voorbeeld - aclvb





Oppervlak histogram =
Oppervlak onder de dichtheidskromme f(x)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Leeftijd gekend bij ACLVB	4709	99,0%	50	1,0%	4759	100,0%

Descriptives

			Statistic	Std. Error
Leeftijd gekend bij ACLVB	Mean		46,06	,195
	95% Confidence Interval for Mean	Lower Bound	45,67	
		Upper Bound	46,44	
	5% Trimmed Mean		45,97	
	Median		46,00	
	Variance		178,987	
	Std. Deviation		13,379	
	Minimum		15	
	Maximum		95	
	Range		80	
	Interquartile Range		20	
	Skewness		,005	,036
	Kurtosis		-,565	,071

Kolmogorov-Smirnov toets

- De ***Kolmogorov-Smirnov goodness of fit*** **toets** is gebaseerd op een maat voor het verschil in twee verdelingen.
- In de vorm voor één steekproef, is het *een aanpassingstoets*, waarmee onderzocht wordt of de waargenomen verdeling f(o) afwijkt van een bekende theoretische verdeling zoals de normale verdeling (f(e))

- **De toetsingsgrootheid (het kengetal)** is de grootste afstand (uitgedrukt in %) tussen de waargenomen verdelingsfunctie en de theoretische verdelingsfunctie van de in het geding zijnde bekende verdeling

- Zie ook: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

Tests of Normality

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Leeftijd gekend bij ACLVB	,050	4709	,000

a. Lilliefors Significance Correction



- (a) Statistisch gezien benadert f(o) de f(e) (*normaalverdeling*)
- (b) Significant op een p <.001 niveau (α < 1%)

- Conclusie: **VAR leeftijd is normaal verdeeld** !

- K-S-test hanteren bij onderzoek van volgende distributies:
  - een normale verdeling
  - een log-normal distributie
  - een Weibull distributie
  - een exponentiële distributie
  - een logistische verdeling

- Opgepast:
  - Moeilijke interpretatie van het kengetal (de maximale afwijking) , steeds samen met significantie
  - K-S-test best alleen gebruiken bij grote steekproeven (n > 100)

Q-Q-plot

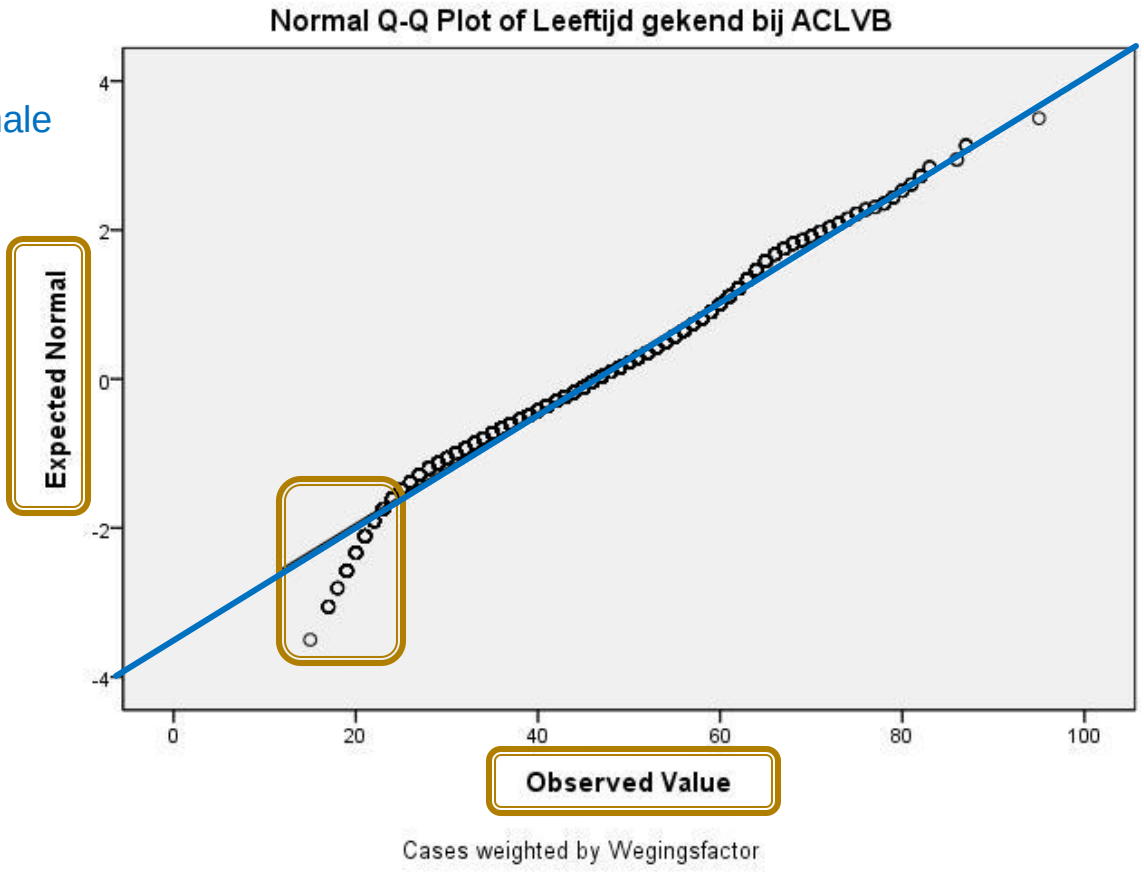
- Een grafiek die toont hoe goed de waargenomen data (verdeling) aansluiten bij de normale verdeling die wort voorgesteld als een rechte in plaats van als een curve
- Afwijkingen = afwijkingen ten opzichte van de normale verdeling
  - Hoe groter de afwijking , hoe minder normaal verdeeld
  - Hoe dichter bij de rechte, hoe beter normaal verdeeld

- Via **Analyze** > **Despriptive statistics** > **Q-Q plots**

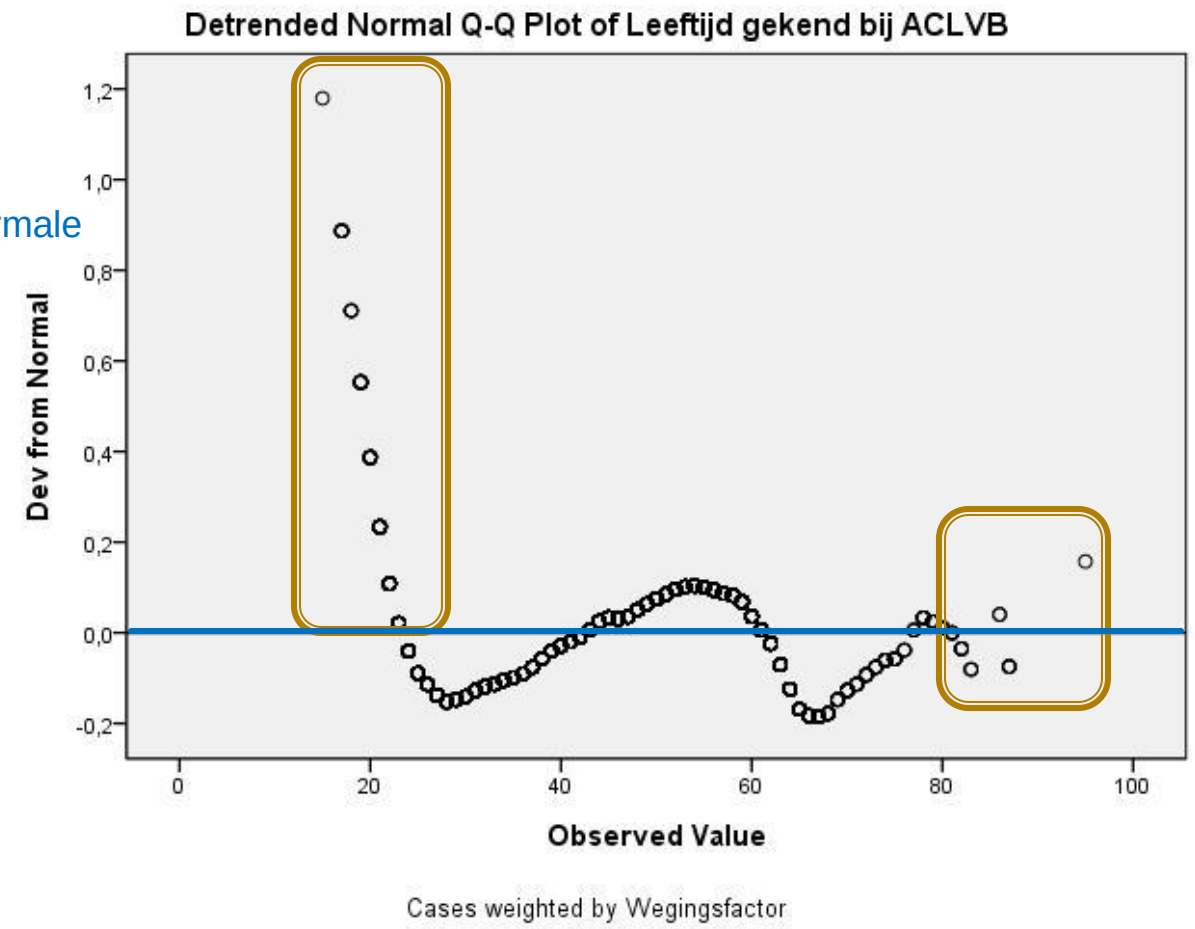
Analyze > Descriptive Statistics

- **Q-Q plot**

Y-as:
de standaardnormale verdeling



Y-as:
de standaardnormale verdeling



Evaluatie van de output

- **Is de VAR *Leeftijd* normaal verdeeld ??**
  - Histogram + normaalverdelingscurve
  - Normal Q-Q plot (bogen, concentraties en zwevende punten) -> ***zijn er zaken die opvallen ?***
  - Kolmogorov-Smirnov kengetal

Bron: http://bcs.whfreeman.com/ips5e/content/cat_120/IPS_Splus.pdf

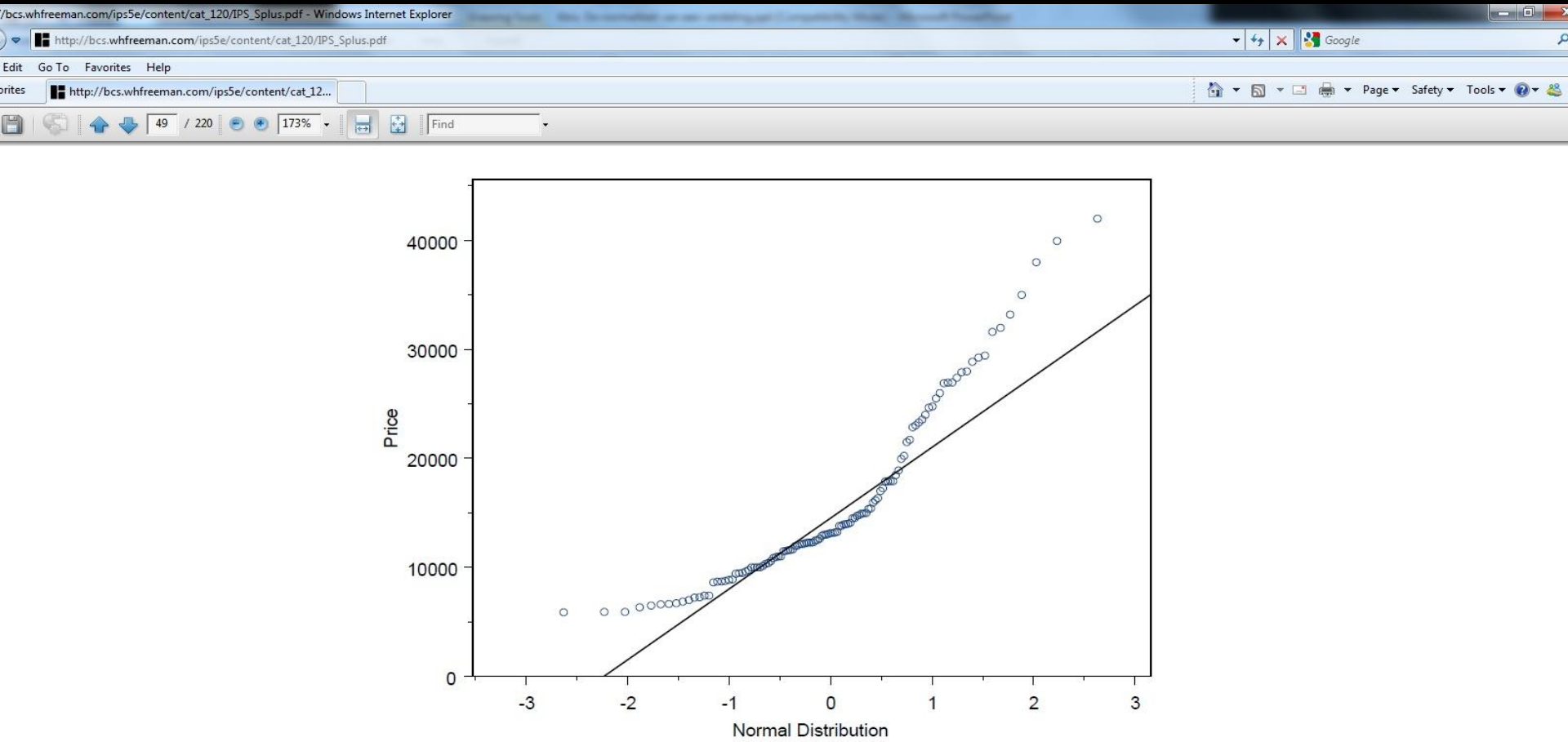


Figure 1.25

It is clear that `Price` does not follow a normal distribution; the largest values in `Price` are much greater than we would expect from a normal distribution.

- **Shapiro-Wilk normaliteitstest**
  - Idem als Kolmogorov-Smirnov-test
  - Enkel **voor kleine steekproeven** (tussen 3 en 50 cases)
  - De interpretatie van de kengetallen is voor de rest exact dezelfde
- OPMERKING ! *Grote steekproeven* vereisen een hoger p-niveau ($p < .01$ of zelfs $p < .001$); *kleine steekproeven* een lager p-niveau ($p < .05$)

Herhalingsoefening (tijd : 20')

- Lees het excel-bestand "*Dataset_survey_16112010_gedefinieerd[1].xlsx*" in SPSS in
- Onderzoek en interpreteer in een nieuw WORD-bestand "**Oefening sessie 4_normaliteit_uw achternaam**" de normaliteit van de verdeling van de nieuw aan te maken VAR ***Leeftijd*** aan de hand van
  - boxplot
  - Histogram + normaalverdelingskromme
  - Q-Q-plot

Deel 7

Eenvoudige Correlaties

Wat is correlatie ?

- Definitie = een consistent en systematisch verband, samenhang of relatie tussen *de niveaus* of *benamingen* van twee variabelen (onafhankelijke VAR (X) en/of afhankelijke VAR (Y)). [met *X op de x-as*; *Y op de y-as*]

Aanwezigheid van samenhang =/= **afhankelijkheid, volgtijdelijkheid of oorzakelijkheid**

- Samenhang tussen variabelen wordt gemeten aan de hand van "**associatiematen**".

- **Wijziging VAR 1 (X-as) <-> Wijziging VAR 2 (Y-as)**

Wat is correlatie ? - 2

- Men spreekt van correlatie als er tussen de beide grootheden een zekere lineaire samenhang is, in die zin dat waarden van de beide grootheden dezelfde of een tegengestelde tendens vertonen.
  - Zo is er sprake van correlatie tussen de grootheden lengte en gewicht bij (volwassen) mensen. Met toenemende lengte neemt gemiddeld gezien ook het gewicht toe.
- Correlatie kan, maar hoeft niet op een causaal verband te duiden.
  - Zo is er correlatie tussen sociale klasse en inkomen, berustend op de genoten opleiding. Ook is er de zgn. **schijn-** of **nonsenscorrelatie**, zoals tussen de schade bij een brand en het aantal ingezette brandweerlieden of het voorkomen van ooievaars en geboortes
- Afhankelijk van de soort grootheden zijn er **verschillende correlatiecoëfficiënten** om de correlatie te meten. Ook dient weer goed onderscheid gemaakt te worden tussen populatie en steekproef.

Soorten verbanden

- **Vier soorten verbanden**

  - **Niet-monotoon**: bij een niet-monotoon verband wordt de aanwezigheid (of afwezigheid) van de ene variabele systematisch gerelateerd aan de aanwezigheid (of afwezigheid) van een andere variabele.

Voorbeeld: McDonald's weet bijvoorbeeld uit ervaring dat ochtendklanten meestal koffie kopen terwijl middagklanten meestal frisdrank kopen.
Maar, dat verband is op geen enkele manier exclusief. Er is geen garantie dat een ochtendklant altijd koffie bestelt of dat een middagklant altijd frisdrank koopt.

Soorten verbanden - 2

- **Monotoon**: een monotoon verband is een verband waarbij de onderzoeker slechts de algemene richting van de samenhang tussen de twee variabelen kan aanduiden (ofwel stijgend ofwel dalende tendens)

Voorbeeld: De eigenaar van een schoenwinkel weet dat oudere kinderen doorgaans grotere schoenen nodig hebben dan jongere kinderen.
Maar, er is geen manier om de leeftijd van kinderen te koppelen aan de juiste schoenmaat. Er bestaat geen universele regel voor de snelheid waarmee de voet van een kind groeit, noch voor de uiteindelijke schoenmaat die het kind krijgt. Er bestaat echter een monotoon stijgend verband tussen de leeftijd van het kind en zijn schoenmaat.

Soorten verbanden - 3

- **Lineair**: een rechtlijnig verband tussen twee variabelen.
- Hier levert kennis van de omvang van de ene variabele *automatisch* kennis op over de omvang van de andere variabele.
- Zie ook: regressievergelijking (module 3)

formule voor de rechte lijn $y = a + bx$

- Voorbeelden: sommige niet zo evident of eenduidig…

  - Opleidingsniveau – gezondheid (wederkerig?)
  - Werkloosheid (X) – armoede (Y)
  - Verstedelijking (X) – individualisme (Y)
  - Politieke voorkeur (X) – partijkeuze (Y)
  - Locatie woonst (X) – aantal sociale contacten (Y)
  - Geslacht (X) – intelligentie (IQ versus EQ) (Y)
  - Aanwezigheid werklozen – criminaliteit (niet eenduidig)

De Standaard Online - Hoe meer werklozen hoe meer criminaliteit - Windows Internet Explorer
http://www.standaard.be/artikel/detail.aspx?artikelid=KK3390PT
File Edit View Favorites Tools Help
De Standaard Online - Hoe meer werklozen hoe ...
Page Safety Tools

De Standaard

HOME « MAANDAG 6 DECEMBER 2010



Kleintjes met ambitie

'Klumpies' is een nieuwe animatiereeks op Ketnet.

BINNENLAND

Hoe meer werklozen hoe meer criminaliteit

Geen verband tussen misdaad en vreemdelingen in gemeente

maandag 06 december 2010 Pieter Lesaffer

AANRADEN 61



BRUSSEL - Niet het aantal allochtonen bepaalt de criminaliteit in een gemeente, wel de werkloosheid. En in Brussel ligt de criminaliteit lager dan men zou kunnen verwachten.

In de steden is er meer criminaliteit dan op het platteland. Dat is geen verrassing.

MEEST RECENT Meest aangeraden Meest gelezen

1. Guy Quaden noemt markten 'irrationeel'
2. Tramverkeer in Gent verstoord door afgerukte bovenleiding
3. Nationale Bank trekt groeiraming op
4. Onderhoud tussen Leterme en Kagame afgelast
5. Ochtendspits NMBS verliep traag

Wat schrijft **Wikileaks** over België?



Ontdek jouw Oranje kant !
Kijk snel op het aanbod
2/4

Internet | Protected Mode: On 150%
De Standaard Online ... Inbox in Tom.Scham... 7. Eenvoudige correl... Dataset deel 7 (voorb... *ACLVB_Werkbestan...
NL 14:51

- ‘We hebben alle mogelijke factoren bekeken, en het verband tussen criminaliteit en werkloosheid bleek het sterkst', zegt **Marc Hooghe**, politicoloog aan de Leuvense universiteit. ‘Dit is ons antwoord op de studie van Marion van San. Zij stelde begin de jaren 2000 vast dat de criminaliteit onder allochtone jongeren hoger lag dan bij anderen. Maar Van San hield geen rekening met sociaal-economische factoren.'

- ‘In onze studie zochten we ook naar een verband tussen vreemdelingen en criminaliteit, maar we hebben er geen gevonden. We hebben per gemeente het aandeel niet-Europese inwoners onderzocht, het aandeel Turken, Marokkanen… Het waren geen bepalende factoren. Van allochtonen met de Belgische nationaliteit hebben we geen cijfers, maar we weten dat dit aandeel overeenkomt met het aandeel inwoners met een vreemde nationaliteit per gemeente. Onze conclusie gaat dus ook op voor allochtonen.’

Soorten verbanden - 4

- **Curvilineaire verbanden**: die verbanden waarbij de ene variabele samenhangt met de andere variabele, maar het verband eerder de vorm van een kromme heeft dan van een rechte lijn.

Een voorbeeld van een curvilineair verband is de kromme voor de levenscyclus van een product die het verkooppatroon beschrijft van een nieuw product dat langzaam groeit tijdens de introductie, vervolgens in de groeifase snel in opwaartse richting spurt en ten slotte een plateau bereikt of aanzienlijk langzamer groeit als de markt verzadigd raakt.

Andere voorbeelden: de Laffer-curve (economie), de "opwinding-resultaat"-curve (psychometrie) e.a.

Verbanden karakteriseren !!!

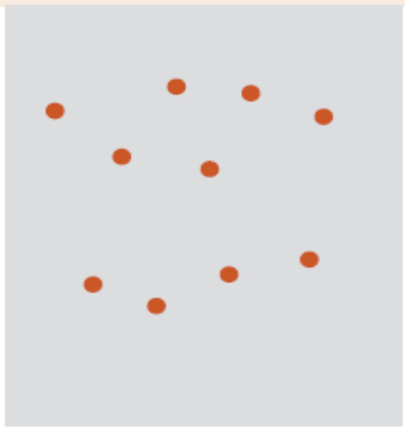
- **Aanwezigheid**: verwijst naar de vraag of er een systematisch verband bestaat tussen twee variabelen.
- **Richting**: een monotoon verband kan stijgen of dalen.
- **Sterkte**: afhankelijk van het soort verband dat wordt onderzocht kan de sterkte van de samenhang tussen twee variabelen worden voorgesteld als sterk, gematigd, zwak of niet-bestaand.

Let op: de volgorde is van belang !!

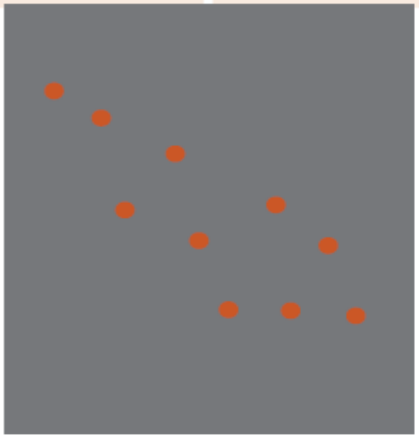
eerst de aanwezigheid van samenhang onderzoeken ;
dan pas de richting ;
en dan de sterkte

Sterkte van het verband

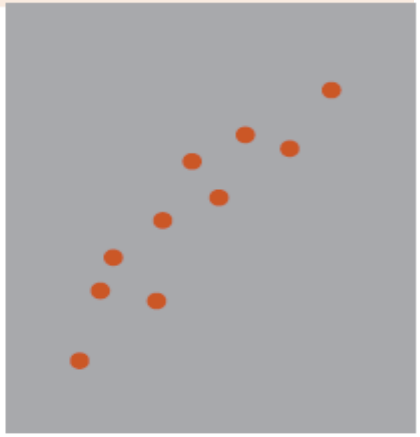
Waardebereik van de coëfficiënt	Sterkte van het verband
± 0,81 tot ± 1,00	sterk
± 0,61 tot ± 0,80	gematigd
± 0,41 tot ± 0,60	zwak
± 0,21 tot ± 0,40	zeer zwak
± 0,00 tot ± 0,20	geen verband



(a) geen verband



(b) negatief verband



(c) positief verband

Verbanden karakteriseren - 2

- Samenhang tussen meerdere variabelen controleren in eenzelfde analyse kan dus.
- Resultaat = een matrix van richting (+ of -) en de **sterkte van de samenhang** (dicht bij 0 of bij (-)1).

- ***Grosso modo drie soorten*** *(cf. overzichtstabel)*
  - *Bij 2 nominale variabelen: gebruik de Pearson* χ^2*-analyse*
  - *Bij 2 ordinaal geschaalde variabelen: gebruik Spearman Rho (zelfs voor niet normaal-verdeelde variabelen) en Kendall's tau*
  - *Bij 2 scale-variabelen: gebruik de Pearson correlation (analyse van samenhang tussen twee of meer parametrische, lineaire variabelen)*

Overzicht

	Meetniveau variabele 2			
Meetniveau variabele 1		Nominaal	Ordinaal	Interval/ratio
	Nominaal	**Chi², Phi, Cramer's V** *geclusterd staafdiagram, gestapeld staafdiagram*	**Chi², Phi, Cramer's V** *idem*	**T-toets** **+ Anova** *+ boxplot voor meerdere variabelen*
	Ordinaal		**+ Spearman Rho, Kendall's tau** *idem*	**T-toets** **+ Anova** *+ boxplot voor meerdere variabelen*
	Interval/Ratio			**+ Correlatie (Pearson R)** *+ scatterplot*

Samenhang bestuderen tussen twee nominale variabelen

- **Chi-kwadraat**

  - Bij een chi-kwadraatanalyse (χ^2-analyse) bestudeer je de frequenties van **twee nominale variabelen** in **een kruistabel**.

  - In deze situatie, twee variabelen met een nominale schaal, testen we voor **een niet-monotone verband**.

Geen eenvoudige interpretatie van de aan- of afwezigheid van samenhang !

Chi-kwadraat (χ^2)

- Wanneer de χ^2-toets hanteren ?
  - Als je wil weten of er een (significant) verband bestaat tussen twee (categorische) variabelen… (onafhankelijkheidstoets), bijvoorbeeld
    - Relatie "ontbijt-lunch en melk-frisdrank drinken", of "man-vrouw en pils-wijn drinken" …
  - Als je wil toetsen of de gevonden data passen bij een veronderstelde verdeling (aanpassingstoets), bijvoorbeeld
    - Onderzoek of verdeling op basis van geslacht overeenkomt met de populatie (50/50), of of de categorieën waarop de steekproef wordt onderverdeeld ook overeenstemmen met deze in de populatie: taalrol, geografie (postcode) enzovoorts
  - Als je wil toetsen of verschillende steekproeven uit éénzelfde verdeling afkomstig zijn (= homogeniteitstoets)

De Chi²-toets

- Essentie van deze toets :
  1. de toets berekent een theoretische verdeling (geeft het beeld van een puur toevallige verdeling van de waarden voor de twee variabelen in de veronderstelling dat die onafhankelijk zijn van elkaar) (f(e))
  2. de toets vergelijkt die theoretische verdeling met de feitelijk verdeling (f(o))
  3. als e =/=o dan betekent dit dat de geobserveerde waarden in de tabel niet toevallig zijn => wijst op een samenhang

Chi²-analyse – wat ?

- **Vergelijking van**
  - **Waargenomen frequenties**: de feitelijke tellingen in de kruistabel. (*f(o)*)
  - **Verwachte frequenties**: *theoretische frequenties* die worden afgeleid van deze hypothese van geen verband tussen de twee variabelen. (*f(*e); louter toevallige verdeling)

Berekende χ^2 -waarde (= toetsingsgrootheid)

Doel: interpretatie van de toetsingsgrootheid χ^2
als *een maat voor de afwijking van de nulhypothese*.

chi-kwadraat formule

$$\chi^2 = \sum_{i-1}^{n} \frac{(\text{waargenomen}_i - \text{verwachte}_i)^2}{\text{verwachte}_i}$$

waarbij

waargenomen_i = waargenomen frequentie in cel i

verwachte_i = verwachte frequentie in cel i

n = aantal cellen

Verschillen worden gekwadrateerd omdat
anders negatieve en positieve waarden elkaar kunnen opheffen

- Illustratie 1 (evenredige verdeling) - $\chi^2 = 0$, $p = .25$

	Man	Vrouw
Lengte >1.70m	50	50
Lengte <1,70m	50	50

VERDELING OBV TOEVAL

- Illustratie 2 (niet-evenredige verdeling) - $\chi^2 = ?$, $p = ?$

	Ontbijt	Lunch
Koffie	95 (a)	5 (b)
Prik	5 (c)	95 (d)

χ^2 berekenen !!!

Basis van Chi-kwadraattoets

- **Berekening van de verwachte frequenties**
  - S1:Berekening van de verwachte frequenties e (bij onafhankelijkheid van de variabelen moet er een gelijkmatige verdeling zijn, omwille van Hyp-0)
  - S2: **Deel het rijtotaal voor elke cel door het totaal aantal waarnemingen in de tabel (*f*)**
  - S3: **Deel het kolomtotaal voor elke cel door het totaal aantal waarnemingen in de tabel (*f*)**
  - S4: **Vermenigvuldig beide verhouding** (= de kans van een waarneming om in een bepaalde cel te belanden)
  - S5: **Vermenigvuldig de kans met het totaal aantal waarnemingen** in te tabel; het resultaat is e voor die cel

Uitwerking van het voorbeeld

Situatie 2

- S2, 3 en 4: berekening verwachte waarden
  - (100/200) * (100/200) = .25
  - .25 *200 = 50 (verwachte waarde in cel (a), a')
  - Verwachte waarde in cel (b), b' = 100- (a) = 50 etc…
- Stap 2: berekening chi-kwadraat (toets)
  - $\chi^2 = (a-a')^2/a' + (b-b')^2/b' + (d-d')^2/d' + (d-d')^2/d'$

    $= 40,5 + 40,5 + \ldots = 162$

- De χ^2-toetsingsgrootheid is dus **een som**
  - van alle verschillen tussen verwachte (e) en waargenomen frequenties (*f*) over alle mogelijkheden (cellen)

- Maar een χ^2 van 162 *wat leert ons dit ???*

  -> in verband met de aanwezigheid van samenhang tussen VAR X en VAR Y ??

Interpretatie van χ^2-uitkomst

- **Uit de χ^2-analyse komt naar voren:**
  - **De grootte van χ^2:**
    - χ^2 wordt groter naarmate het verschil tussen de waarden (de verdeling) van *f* en o groter wordt (hier veronderstellen we geen verband), met een kans p = 0,000 : er is samenhang tussen de twee onderzochte variabelen; puur toeval speelt dus niet in de waargenomen verdeling (of bijna niet)
    - Als de verdeling van de waargenomen waarden gelijk is aan de verdeling van de verwachte waarden dan is χ^2 = 0 en is de kans p = 1,000 (100% kans) dat we de nulhypothese weerhouden : er is geen samenhang tussen beide variabelen; samenhang die er is is 100% toevallig
  - **De kans (p) dat de onderzoeker bewijs vindt voor de nulhypothese** (dat er geen verband is tussen *f e* en o) en de verschillen aan toeval te wijten zijn [0,000 ; 1,000] => zie tabel !

Significantieniveau bepalen

Een verschil met een waarschijnlijkheid tussen 0,01 en 0,05 --> significant
Een verschil met een waarschijnlijkheid < 0.01 --> zeer significant
Hoe groter de uitkomst van de X^2, hoe kleiner de kans op toeval.

Betrouwbaarheids-niveau van de conclusie:

<90%
(a)[90-94,99%[
(b)[95-98,99%[
(c)[99-99,99%[
(d)>99,99%

Sign. Heel sign.

kans op toeval / vrijheidsgraden	0,10 (a)	0,05 (b)	0,01 (c)	0,001 (d)
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.815	11.345	16.266
4	7.779	9.488	13.277	18.467
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.322
8	13.362	15.507	20.090	26.125
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	29.141	36.123
14	22.307	24.996	30.578	37.697

X^2

Het aantal vrijheidsgraden is het aantal groepen = (2-1)*(2-1) = 1

Hoe kleiner de p (kans) hoe meer zekerheid dat de verschillen voor zover ze zich voordoen niet aan toeval kunnen worden toegeschreven.

χ^2-waarde 162 uit ons voorbeeld met vrijheidsgraad = 1 geeft ons een p-waarde < .0001

Tegen een p-niveau <.001 is de kans dat de onderzoeker bewijs vindt voor **de nulhypothese** (dat er geen verband is tussen f en o) en dat verschillen aan toeval te wijten zijn, miniem. Dus Ho **wordt verworpen – er is een verband tussen VAR Ontbijt en VAR drank**

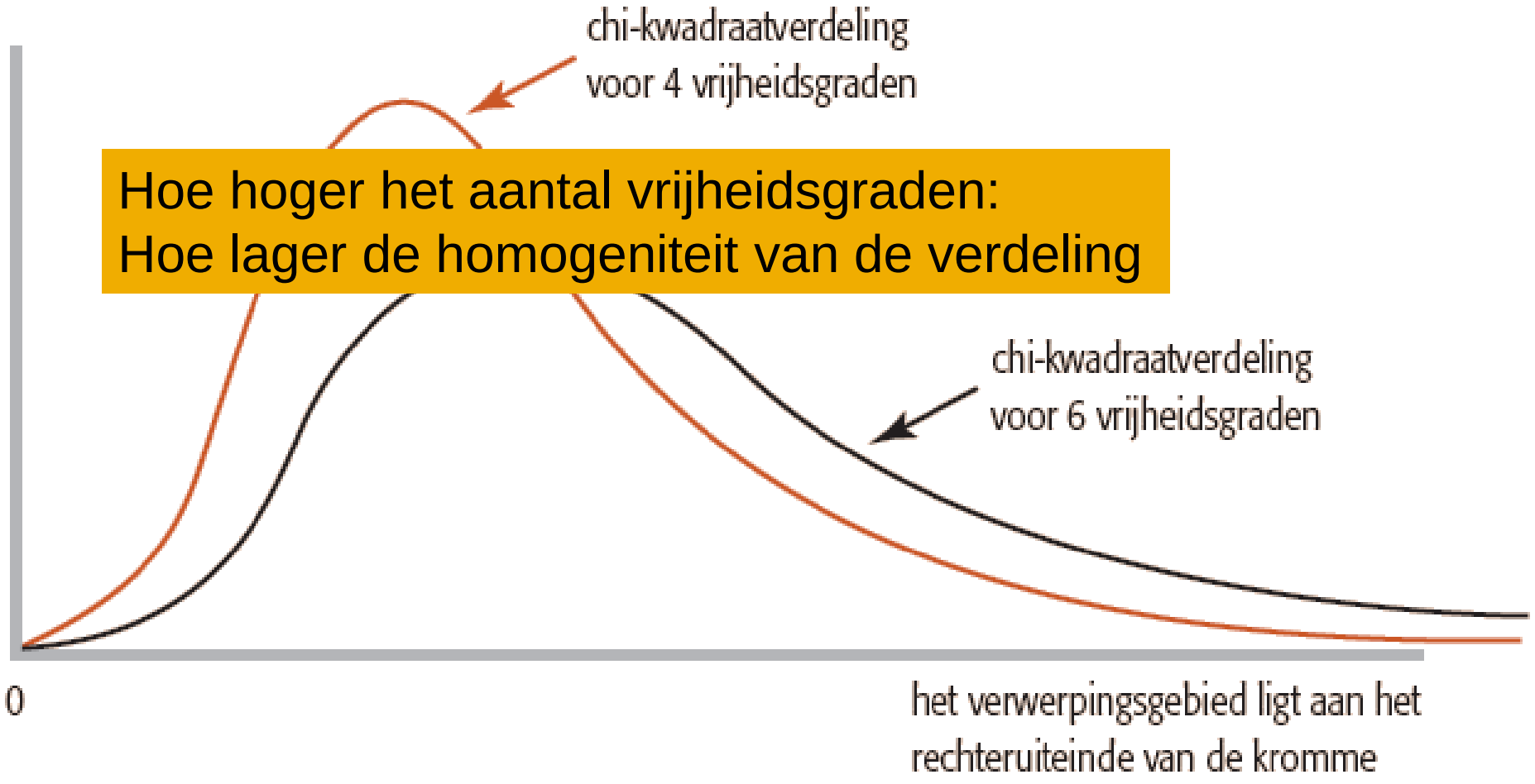
"Respons alleen zegt nog niets over de representativiteit, i.e. de mate waarin de kenmerken van de populatie (alle ACLVB-leden) weerspiegeld worden in de steekproef."

- Voorbeeld: relatie tussen de waargenomen aantallen in een steekproef en de geplande (verwachte) aantallen

		Geplande Steekproef		Gerealiseerde Steekproef	
		N	%	N	%
Geslacht	*Man*	6903	59,50%	2760	58,60%
	Vrouw	4697	40,50%	1942	41,40%
χ2=1,25 df=1 p=0,263					
Leeftijd	*-30j*	2062	17,80%	580	12,30%
	30-60j	8360	72,10%	3547	75,30%
	+60j	1178	10,20%	585	12,40%
χ2=109,07 df=2 p=0,000					
Taalrol	*Ned.*	3845	66,90%	3214	68,20%
	Fr.	7755	33,10%	1498	31,80%
χ2=142,83 df=1 p=0,000					

?

n (= aantal cellen) bepaalt het verloop van de χ^2-verdeling



Beperkingen van χ^2

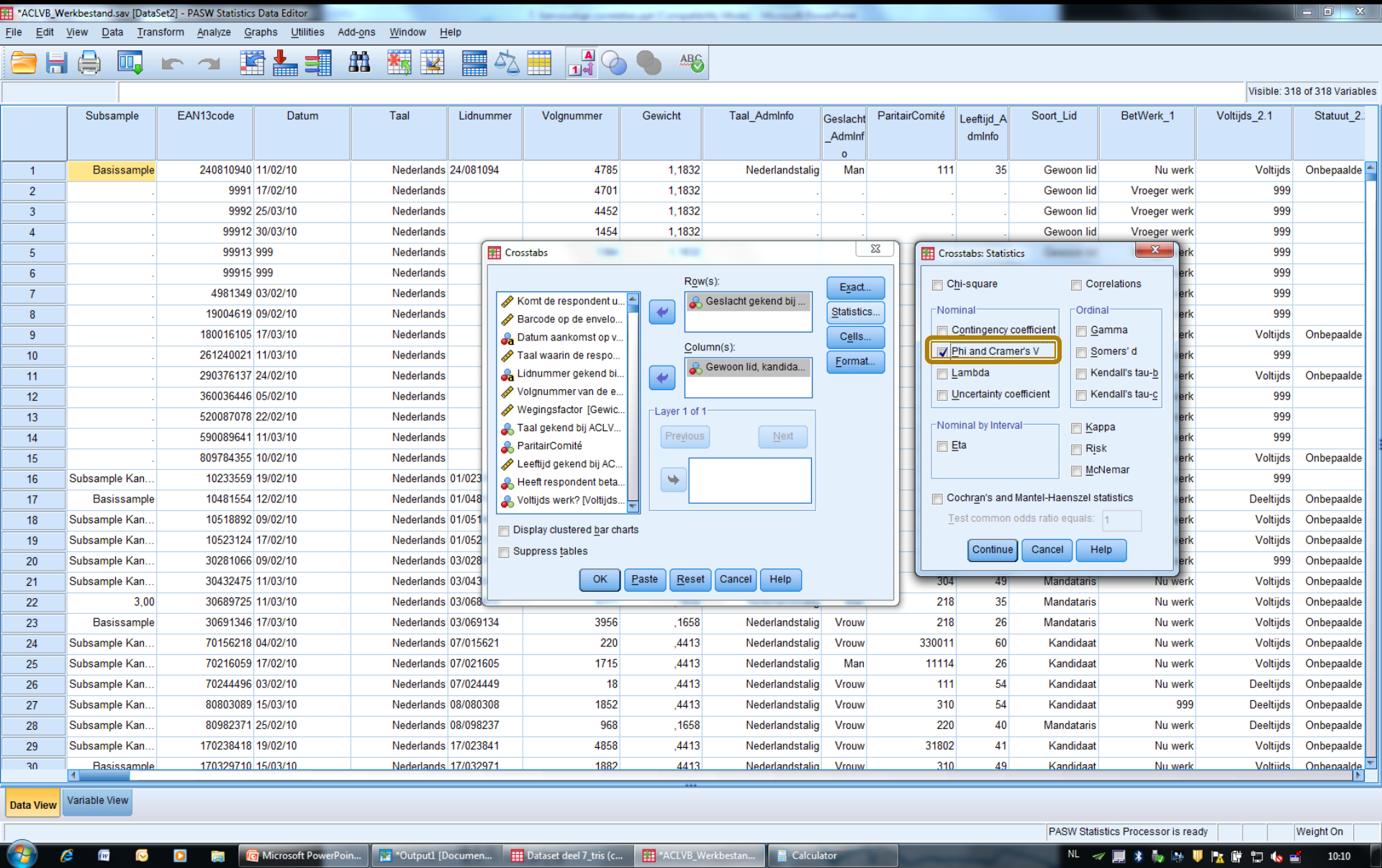
- Ten eerste,
  - **Niet toepassen !!**
    - op een tabel met minder dan 50 observaties per rij of kolom
    - op een tabel met minder dan 5 observaties per cel
  - Oplossing: Bij N<50 of cellen <5 waarnemingen voert SPSS *automatisch* de ***Fisher's exact test*** uit. (= analoge interpretatie als χ^2) (soms ook niet !!)

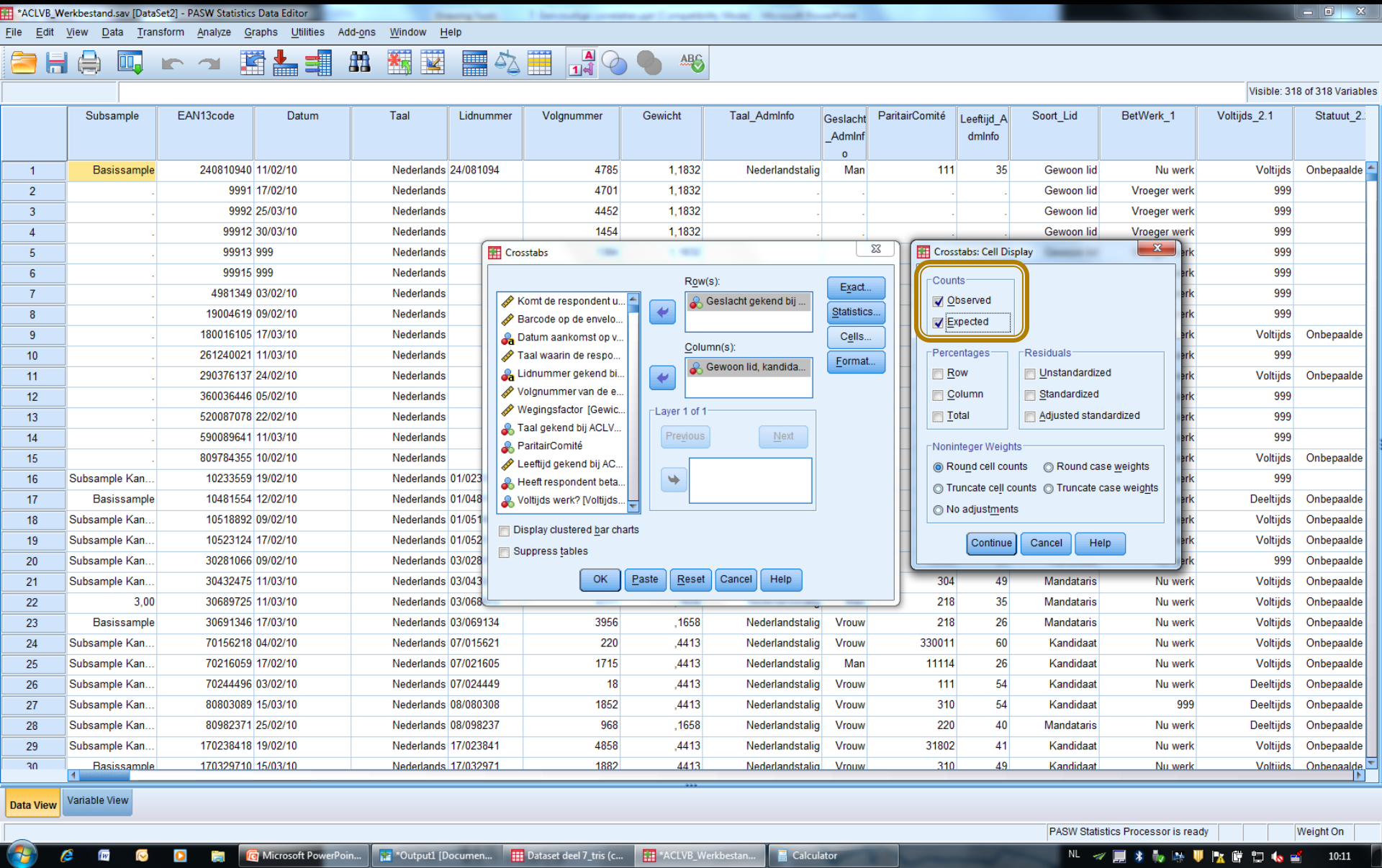
- Anderzijds, χ^2 **zegt niets over…**
  - …de richting van de samenhang
  - …de sterkte van het verband (ook al is het significant)
    - Cf. Phi en Cramer's V (verder!!)
  - …eventuele aannames ivm. de gemiddelden voor een populatie (-> parametervrije of niet-parametrische toetsen: z, T, F)

Andere toetsen voor associatie tussen twee nominale variabelen

- **Phi**: enkel voor 2*2 tabellen
  - $\Phi = \sqrt{X^2 / n}$ (waarbij n= 1)
  - ϕ = [-1;1] met 0 = onafhankelijkheid en -1 of 1 voor perfecte verbanden
- **Cramer's V** voor r*k tabellen
  - Correctie op Chi^2 voor aantal cellen en voor aantal r en k
  - $V = \sqrt{X^2 / n(k-1)}$ (waarbij k = min(r, k))
  - V = [-1;1] met analoge interpretatie als Phi

Data aclvb – voorbeeld 1




*ACLVB_Werkbestand.sav [DataSet2] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 318 of 318 Variables
Subsample EAN13code Datum Taal Lidnummer Volgnummer Gewicht Taal_AdmInfo Geslacht_AdmInfo ParitairComité Leeftijd_AdmInfo Soort_Lid BetWerk_1 Voltijds_2.1 Statuut_2
1 Basissample 240810940 11/02/10 Nederlands 24/081094 4785 1,1832 Nederlandstalig Man 111 35 Gewoon lid Nu werk Voltijds Onbepaalde
2 . 9991 17/02/10 Nederlands 4701 1,1832 . . . . Gewoon lid Vroeger werk 999
3 . 9992 25/03/10 Nederlands 4452 1,1832 . . . . Gewoon lid Vroeger werk 999
4 . 99912 30/03/10 Nederlands 1454 1,1832 . . . . Gewoon lid Vroeger werk 999
5 . 99913 999 Nederlands 999
6 . 99915 999 Nederlands 999
7 . 4981349 03/02/10 Nederlands 999
8 . 19004619 09/02/10 Nederlands 999
9 . 180016105 17/03/10 Nederlands Voltijds Onbepaalde
10 . 261240021 11/03/10 Nederlands 999
11 . 290376137 24/02/10 Nederlands Voltijds Onbepaalde
12 . 360036446 05/02/10 Nederlands 999
13 . 520087078 22/02/10 Nederlands 999
14 . 590089641 11/03/10 Nederlands 999
15 . 809784355 10/02/10 Nederlands Voltijds Onbepaalde
16 Subsample Kan... 10233559 19/02/10 Nederlands 01/023 999
17 Basissample 10481554 12/02/10 Nederlands 01/048 Deeltijds Onbepaalde
18 Subsample Kan... 10518892 09/02/10 Nederlands 01/051 Voltijds Onbepaalde
19 Subsample Kan... 10523124 17/02/10 Nederlands 01/052 Voltijds Onbepaalde
20 Subsample Kan... 30281066 09/02/10 Nederlands 03/028 999 Onbepaalde
21 Subsample Kan... 30432475 11/03/10 Nederlands 03/043 304 49 Mandataris Nu werk Voltijds Onbepaalde
22 3,00 30689725 11/03/10 Nederlands 03/068 218 35 Mandataris Nu werk Voltijds Onbepaalde
23 Basissample 30691346 17/03/10 Nederlands 03/069134 3956 ,1658 Nederlandstalig Vrouw 218 26 Mandataris Nu werk Voltijds Onbepaalde
24 Subsample Kan... 70156218 04/02/10 Nederlands 07/015621 220 ,4413 Nederlandstalig Vrouw 330011 60 Kandidaat Nu werk Voltijds Onbepaalde
25 Subsample Kan... 70216059 17/02/10 Nederlands 07/021605 1715 ,4413 Nederlandstalig Man 11114 26 Kandidaat Nu werk Voltijds Onbepaalde
26 Subsample Kan... 70244496 03/02/10 Nederlands 07/024449 18 ,4413 Nederlandstalig Vrouw 111 54 Kandidaat Nu werk Deeltijds Onbepaalde
27 Subsample Kan... 80803089 15/03/10 Nederlands 08/080308 1852 ,4413 Nederlandstalig Vrouw 310 54 Kandidaat 999 Deeltijds Onbepaalde
28 Subsample Kan... 80982371 25/02/10 Nederlands 08/098237 968 ,1658 Nederlandstalig Vrouw 220 40 Mandataris Nu werk Deeltijds Onbepaalde
29 Subsample Kan... 170238418 19/02/10 Nederlands 17/023841 4858 ,4413 Nederlandstalig Vrouw 31802 41 Kandidaat Nu werk Voltijds Onbepaalde
30 Basissample 170329710 15/03/10 Nederlands 17/032971 1882 ,4413 Nederlandstalig Vrouw 310 49 Kandidaat Nu werk Voltijds Onbepaalde
Crosstabs
Komt de respondent u...
Barcode op de envelo...
Datum aankomst op v...
Taal waarin de respo...
Lidnummer gekend bi...
Volgnummer van de e...
Wegingsfactor [Gewic...
Taal gekend bij ACLV...
ParitairComité
Leeftijd gekend bij AC...
Heeft respondent beta...
Voltijds werk? [Voltijds...
Row(s):
Geslacht gekend bij ...
Column(s):
Gewoon lid, kandida...
Layer 1 of 1
Previous
Next
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...
OK Paste Reset Cancel Help
Crosstabs: Cell Display
Counts
Observed
Expected
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue Cancel Help
Data View
Variable View
PASW Statistics Processor is ready
Weight On
Microsoft PowerPoin...
*Output1 [Documen...
Dataset deel 7_tris (c...
*ACLVB_Werkbestan...
Calculator
NL
10:11

Analyse van de output

VAR Y (afhankelijke)

Mogelijke aanpassing :

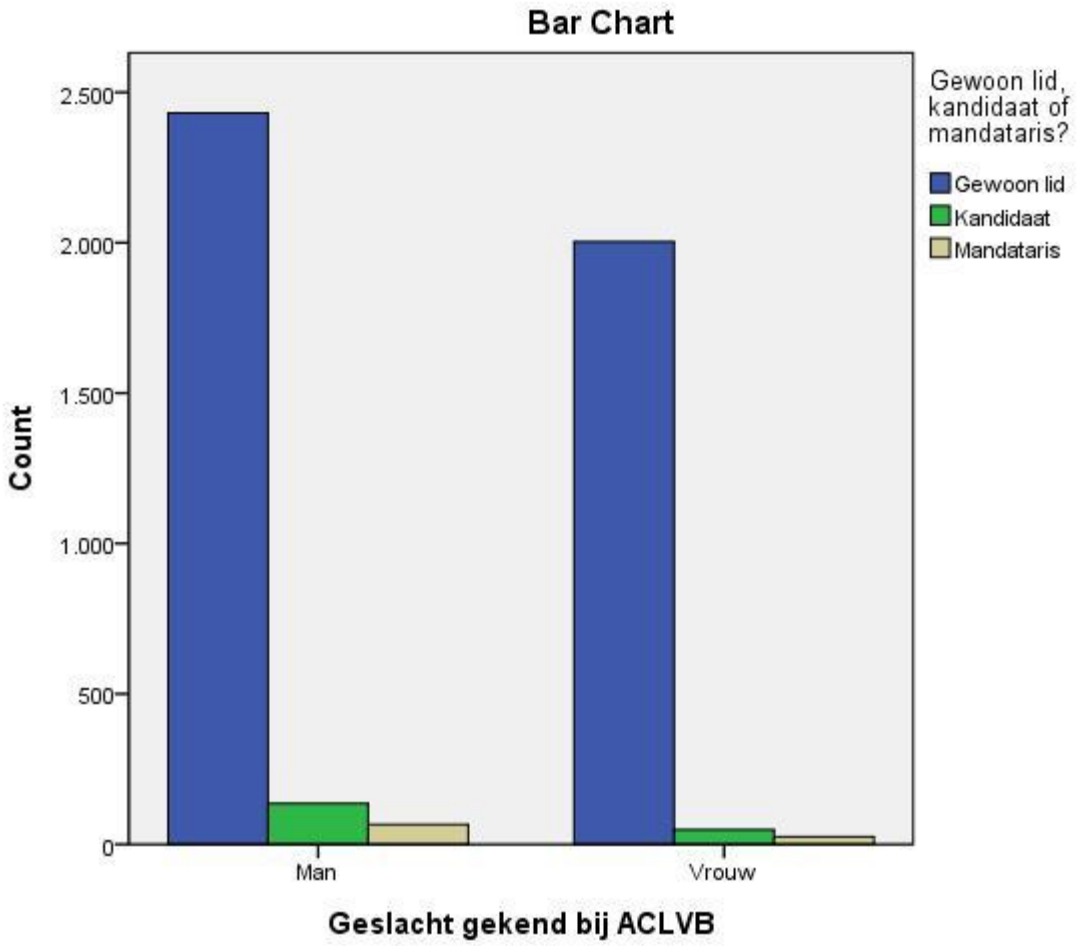
VAR X (onafhankelijke)

Geslacht gekend bij ACLVB * Gewoon lid, kandidaat of mandataris? Crosstabulation

			Gewoon lid, kandidaat of mandataris?			
			Gewoon lid	Kandidaat	Mandataris	Total
Geslacht gekend bij ACLVB	Man	Count	2431	136	66	2633
		Expected Count	2479,2	102,9	50,9	2633,0
		% within Geslacht gekend bij ACLVB	92,3%	5,2%	2,5%	100,0%
		% within Gewoon lid, kandidaat of mandataris?	54,8%	73,9%	72,5%	55,9%
		% of Total	51,6%	2,9%	1,4%	55,9%
	Vrouw	Count	2003	48	25	2076
		Expected Count	1954,8	81,1	40,1	2076,0
		% within Geslacht gekend bij ACLVB	96,5%	2,3%	1,2%	100,0%
		% within Gewoon lid, kandidaat of mandataris?	45,2%	26,1%	27,5%	44,1%
		% of Total	42,5%	1,0%	,5%	44,1%
Total		Count	4434	184	91	4709
		Expected Count	4434,0	184,0	91,0	4709,0
		% within Geslacht gekend bij ACLVB	94,2%	3,9%	1,9%	100,0%
		% within Gewoon lid, kandidaat of mandataris?	100,0%	100,0%	100,0%	100,0%
		% of Total	94,2%	3,9%	1,9%	100,0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,088	,000
	Cramer's V	,088	,000
N of Valid Cases		4709	


Bar Chart
Gewoon lid, kandidaat of mandataris?
Gewoon lid
Kandidaat
Mandataris
Count
2.500
2.000
1.500
1.000
500
0
Man
Vrouw
Geslacht gekend bij ACLVB

Interpretatie van de samenhang voor Phi en Cramer's V

- Sterk verband — value ≥.60
- Matig — value = [.30;.60[
- Zwak tot heel zwak — value <.30
- Geen verband — value = 0

Bron: cursus statistiek John Lievens

Oefening 2 (10')

- Open **Dataset "*Politieke besluitvorming_opdracht les3_oplossing.sav*"**

- Ga de samenhang na tussen de volgende variabelen: '***Geslacht***' (kolom) en '***VRP belangrijkheid_REC***' (rij)
  - Anders geformuleerd*: Is er een verschil in de verdeling van de VAR VRP_belangrijkheid_REC (5 categorieën) voor de VAR Geslacht ?*

Hoe ?

Via: **Analyze** > **Descriptive statistics** > **Crosstabs**

> **Statistics** en selecteer "**Chi-square**"
> Vink '**Display clustered bar chart**" aan
> Vink bij cellen (*Crosstabs Cell Display*) ook de '**Expected count**' aan.

```
GET
  FILE='K:\Opleidingsonderdelen\3de bachelor\PWM\Module 2 uni- en bivariate analyse\bruikbare datasets\Politieke besluitvorming\Politieke besluitvorming_opdracht les3_oplossing.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
CROSSTABS
  /TABLES=VRP_belangrijkheid_REC BY geslacht
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED COLUMN TOTAL
  /COUNT ROUND CELL
  /BARCHART.
```

Interpretatie van de output

- **Kruistabel**
  - Counts
  - Expected counts
  - Percentages etc…
- **Maten voor symmetrie**
  - **Grootte van de Pearson Chi-square** (kengetal), Phi of Cramer's V
  - p –niveau (is samenhang betekenisvol of niet?)
  - Significantiewaarde die wordt bekeken in het licht van het gestelde **significantieniveau α** (**.05, .01 of .001**) - voor grote datasets hanteren we het significantieniveau .001, voor kleinere datasets is sign. niveau van .05 OK (= relatief lage betrouwbaarheid)
- **Bar charts (clustered)**

Interpretatie van de output

Een verschil met een waarschijnlijkheid tussen 0,01 en 0,05 --> significant
Een verschil met een waarschijnlijkheid < 0.01 --> zeer significant
Hoe groter de uitkomst van de X^2, hoe kleiner de kans op toeval en hoe groter de samenhang.

Betrouwbaarheid:
<90%
(a)[90-94,99%[
(b)[95-98,99%[
(c)[99-99,99%[
(d)>99,99%

		Sign.		Heel sign.
kans op toeval / vrijheidsgraden	0,10 (a)	0,05 (b)	0,01 (c)	0,001 (d)
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.815	11.345	16.266
4	7.779	9.488	13.277	18.467
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.322
8	13.362	15.507	20.090	26.125
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	29.141	36.123
14	22.307	24.996	30.578	37.697

X^2

Het aantal vrijheidsgraden is het aantal groepen = (2-1)*(5-1) = **3**

Uitspraak over grootte Chi² ?
Is er samenhang tussen beide variabelen.

Uitspraak over significantie van de samenhang tussen VAR X en VAR Y ?

Wordt Ho weerhouden ? Speelt toeval ?

Oplossing

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,567[a]	3	,463
Likelihood Ratio	2,693	3	,441
Linear-by-Linear Association	2,477	1	,115
N of Valid Cases	55		

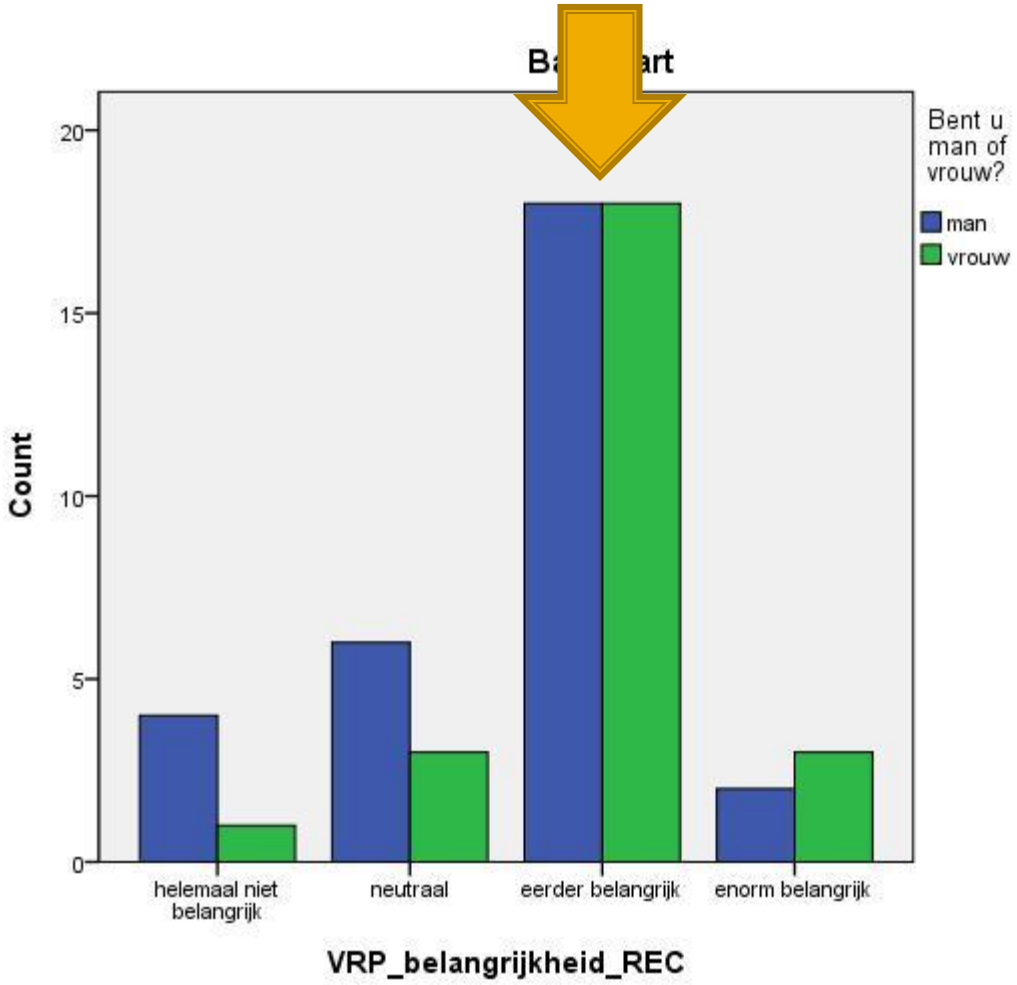
a. 6 cells (75,0%) have expected count less than 5. The minimum expected count is 2,27.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,216	,463
	Cramer's V	,216	,463
N of Valid Cases		55	

VRP_belangrijkheid_REC * geslacht Bent u man of vrouw? Crosstabulation

VRP_belangrijkheid_REC		geslacht Bent u man of vrouw? man	vrouw	Total
helemaal niet belangrijk	Count	4	1	5
	Expected Count	2,7	2,3	5,0
	% within geslacht Bent u man of vrouw?	13,3%	4,0%	9,1%
	% of Total	7,3%	1,8%	9,1%
neutraal	Count	6	3	9
	Expected Count	4,9	4,1	9,0
	% within geslacht Bent u man of vrouw?	20,0%	12,0%	16,4%
	% of Total	10,9%	5,5%	16,4%
eerder belangrijk	Count	18	18	36
	Expected Count	19,6	16,4	36,0
	% within geslacht Bent u man of vrouw?	60,0%	72,0%	65,5%
	% of Total	32,7%	32,7%	65,5%
enorm belangrijk	Count	2	3	5
	Expected Count	2,7	2,3	5,0
	% within geslacht Bent u man of vrouw?	6,7%	12,0%	9,1%
	% of Total	3,6%	5,5%	9,1%
Total	Count	30	25	55
	Expected Count	30,0	25,0	55,0
	% within geslacht Bent u man of vrouw?	100,0%	100,0%	100,0%
	% of Total	54,5%	45,5%	100,0%


Ba art
Bent u man of vrouw?
man
vrouw
Count
20
15
10
5
0
helemaal niet belangrijk
neutraal
eerder belangrijk
enorm belangrijk
VRP_belangrijkheid_REC

Tot slot

- Aangeven dat een verband 'zwak' of 'sterk' is is ontoereikend als resultaat
- Hoe vertalen in zinvolle conclusies?
  - Interpretatie kruistabel
  - Interpretatie clustered bar charts
  - Interpretatie in functie van een bestaande theorie of een bestaand model
    - Bv. intermediaire variabelen etc…
  - Dan nog kunnen we niet pretenderen dat de conclusie correct en volledig is !

Associatiematen die rechtstreeks gebaseerd zijn op de waargenomen frequenties

- **Oddsratio**
  - http://nl.wikipedia.org/wiki/Odds_ratio

- **Yules Q**
  - http://nl.wikipedia.org/wiki/Yule's_Q

Deel 7.bis

Eenvoudige correlaties - Ordinale en Scale variabelen

Vooraf

- **Stijgt het meetniveau, dan stijgt ook de statistische informatie die we kunnen opvragen per variabele (zie overzichtstabel: technieken die van toepassing zijn op een lager meetniveau mag men steeds toepassen op een hoger, maar niet omgekleerd)**
- Het verschil tussen *ordinale* en *continue* variabelen:
  - *Vaste meeteenheid* nodig om lineaire vergelijkingen te hanteren bij de studie van samenhang tussen VAR1 en VAR2
- Samenhang wordt uitgedrukt aan de hand van een correlatiecoëfficiënt

Samenhang bestuderen tussen twee ordinale variabelen

- Hoe kunnen we het verband tussen twee variabelen van ordinaal niveau meten?
  - In welke mate gaan lage waarden van VARx met lage waarden VAR y samen en hoge met hoge?
  - In dit geval proberen we de aanwezigheid, richting en de sterke van **een monotoon verband** te bepalen:
    - Algemene verband: is er samenhang of niet?
    - Plus **de richting** (-/+)

Correlatiecoëfficiënt (r)

- **Symmetrische maat**
  - *Effect van VAR1 op VAR2* = effect van *VAR2 op VAR1 (wederkerigheid) (2-tailed)*

- R drukt enkel uit hoe sterk twee variabelen bij elkaar horen (= aanwezigheid van samenhang)
- Zegt niets over de invloed of het effect van VAR1 op VAR2 of omgekeerd (*=/= causaliteit*)
  - Ook al weten we dat bepaalde variabelen wel een effect hebben (bv. 'geslacht_vrouwelijk' op 'zwanger zijn')

Welke statististische controle

- We worden bij het onderzoek van verbanden tussen twee ordinale variabelen geholpen door drie statistische controles (parameters):

  - **Kendall tau-b**
  - **Somer's d** (niet verder besproken)
  - **Spearman Rho**

Kendall's tau

- **Symmetrische associatiematen** !! (net als Cramer's V en Phi, ook *symmetric measures*) dus geen aanames over welke VAR welke WAR beïnvloedt !!

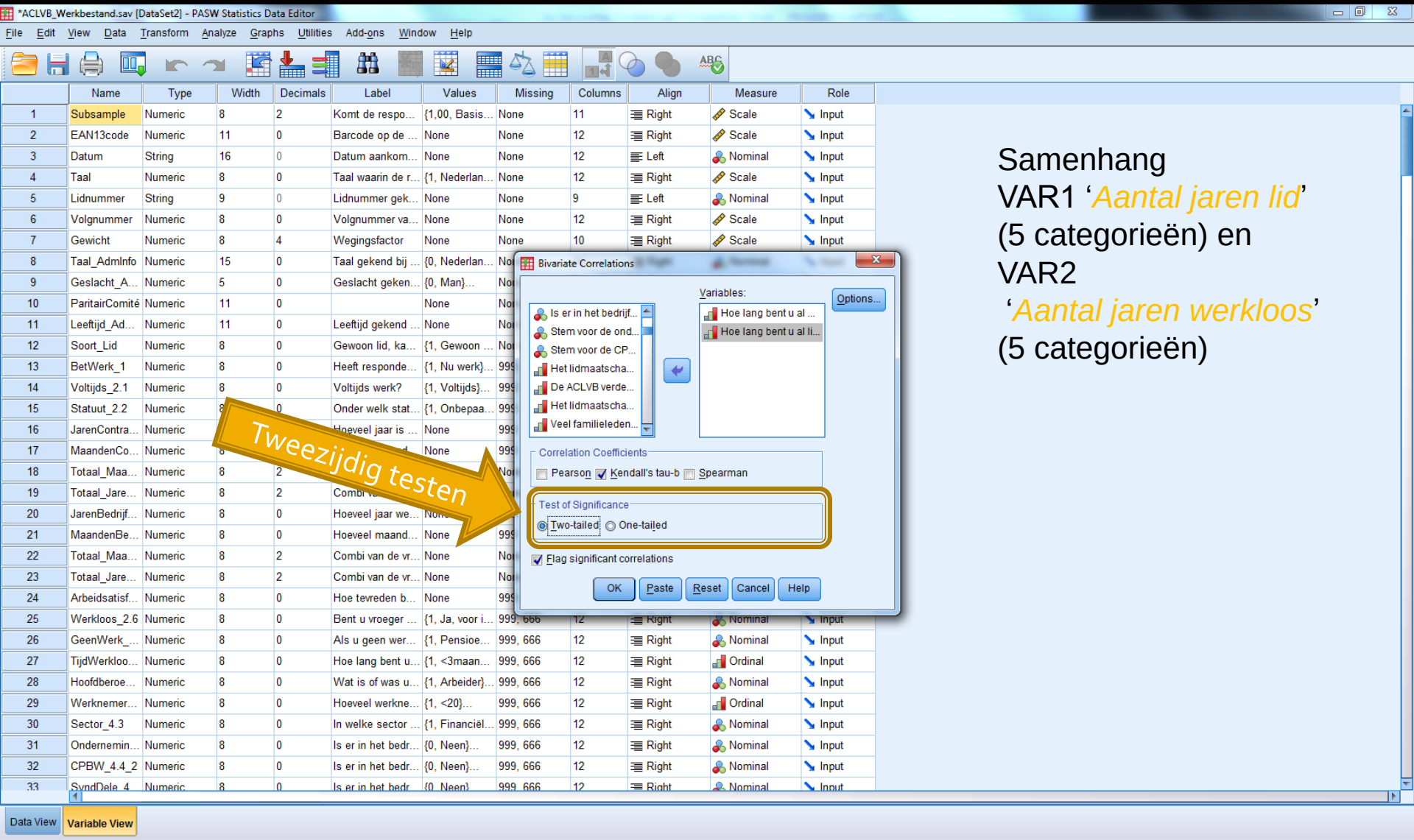
- **Coëfficiënt = [-1; 1]**
  - Bij volledige overeenstemming tussen de beide ordinale volgordes, dus als in elk paar beide grootheden hetzelfde rangnummer hebben (*concordantie*), dan is de coëfficiënt 1. (perfect + verband)

    C = n(n-1)/2
  - In het tegenovergestelde geval, dus als de volgordes volledig tegengesteld verlopen en alle paren *discordant* zijn, is de coëfficiënt met de waarde -1. (perfect – verband)

    C = 0
  - Indien de volgordes onafhankelijk zijn, heeft de coëfficiënt de waarde 0. (geen verband)

Waar in SPSS?

```
NONPAR CORR
  /VARIABLES=TijdWerkloos_3.2 TijdLid_6
  /PRINT=KENDALL TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```



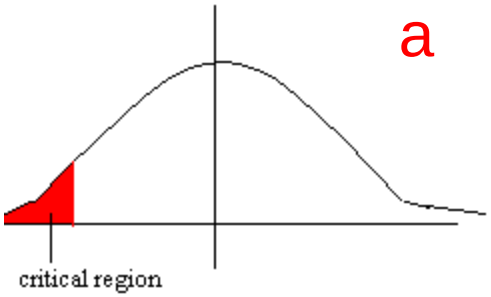
Samenhang
VAR1 '*Aantal jaren lid*'
(5 categorieën) en
VAR2
'*Aantal jaren werkloos*'
(5 categorieën)

Twee significantietoetsen voor de Ho

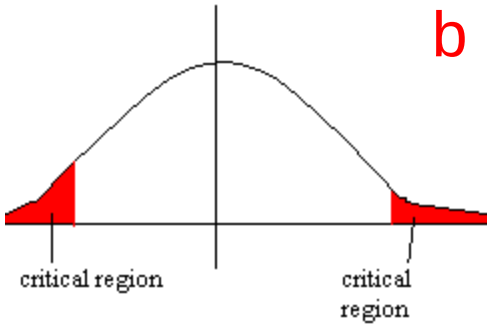
One-tailed: test van ongelijkheid, dus kijkt naar stijging of daling van de parameter in de kritische regio (bepaald door significantieniveau*) (a)
Two-tailed: test van gelijkheid, dus kijkt naar alle veranderingen van de parameter (stijging en daling) (b)

*Gebruikelijke p-niveaus: .05, .01 of .001 (kans om verkeerdelijk de Ho te verwerpen)

a



b



Analyse output

Interpretatie: **R = .361****
De samenhang is positief en significant (tot op een .01-grens)

*Vanwaar **n = 891** ?*
Belangrijke info !!
Wat met de missings ??

MAAR idee van mogelijke causaliteit ?
Bv. lid zijn -> werkloosheid of omgekeerd?? Meer info via een scatterplot bv.

Correlations

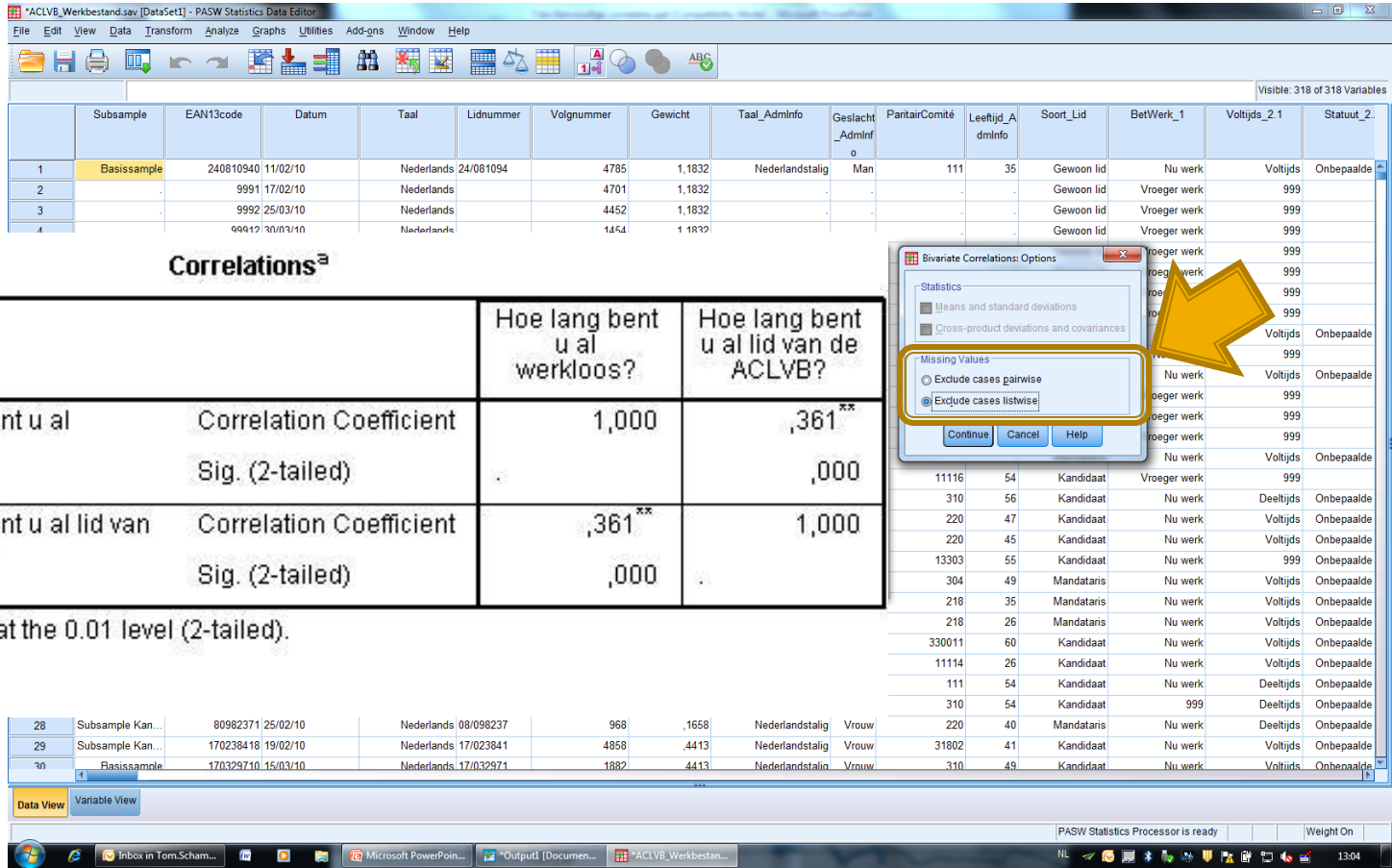
			Hoe lang bent u al werkloos?	Hoe lang bent u al lid van de ACLVB?
Kendall's tau_b	Hoe lang bent u al werkloos?	Correlation Coefficient	1,000	,361**
		Sig. (2-tailed)	.	,000
		N	911	891
	Hoe lang bent u al lid van de ACLVB?	Correlation Coefficient	,361**	1,000
		Sig. (2-tailed)	,000	.
		N	891	3661

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

- Wanneer '*Exclude cases pairwise*' ?
  - **De standaardinstelling !**
  - Cases waarvoor per correlatie *tweevoudige missings* bestaan, worden weggelaten uit de analyse
  - Als er te weinig cases zouden overblijven na deleting van door missings besmette cases

- Wanneer '*Exclude cases listwise*' ?
  - Betekent het weglaten van elke case van zodra ze een missing bevat (met andere woorden hier bestaat de kans dat veel meer cases wegvallen)
  - Als dat risico op teveel weggelaten cases niet bestaat, is er geen probleem

- Zelfde correlatie maar nu '*Exclude cases listwise*'



Correlations[a]

			Hoe lang bent u al werkloos?	Hoe lang bent u al lid van de ACLVB?
Kendall's tau_b	Hoe lang bent u al werkloos?	Correlation Coefficient	1,000	,361**
		Sig. (2-tailed)	.	,000
	Hoe lang bent u al lid van de ACLVB?	Correlation Coefficient	,361**	1,000
		Sig. (2-tailed)	,000	.

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

a. Listwise N = 891

Alternatieve weg via crosstabs

TijdLid_6 Hoe lang bent u al lid van de ACLVB? * TijdWerkloos_3.2 Hoe lang bent u al werkloos? Crosstabulation

			TijdWerkloos_3.2 Hoe lang bent u al werkloos?					
			<3maanden	3-6maanden	6maanden-1jaar	1-2jaar	>2jaar	Total
TijdLid_6 Hoe lang bent u al lid van de ACLVB?	<6maanden	Count	15	13	4	1	2	35
		% within TijdWerkloos_3.2 Hoe lang bent u al werkloos?	12,4%	13,3%	3,1%	,7%	,3%	3,3%
		% of Total	1,4%	1,2%	,4%	,1%	,2%	3,3%
	6maanden-1jaar	Count	15	11	22	7	2	57
		% within TijdWerkloos_3.2 Hoe lang bent u al werkloos?	12,4%	11,2%	16,9%	5,1%	,3%	5,4%
		% of Total	1,4%	1,0%	2,1%	,7%	,2%	5,4%
	1-5jaar	Count	32	25	34	59	79	229
		% within TijdWerkloos_3.2 Hoe lang bent u al werkloos?	26,4%	25,5%	26,2%	43,1%	13,8%	21,6%
		% of Total	3,0%	2,4%	3,2%	5,6%	7,5%	21,6%
	5-10jaar	Count	24	22	39	23	116	224
		% within TijdWerkloos_3.2 Hoe lang bent u al werkloos?	19,8%	22,4%	30,0%	16,8%	20,2%	21,1%
		% of Total	2,3%	2,1%	3,7%	2,2%	10,9%	21,1%
	>10jaar	Count	35	27	31	47	375	515
		% within TijdWerkloos_3.2 Hoe lang bent u al werkloos?	28,9%	27,6%	23,8%	34,3%	65,3%	48,6%
		% of Total	3,3%	2,5%	2,9%	4,4%	35,4%	48,6%
Total		Count	121	98	130	137	574	1060
		% within TijdWerkloos_3.2 Hoe lang bent u al werkloos?	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	11,4%	9,2%	12,3%	12,9%	54,2%	100,0%

Opvragen van correlaties in Crosstabs geeft ons sowieso

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Phi	,517			,000
	Cramer's V	,259			,000
Ordinal by Ordinal	Kendall's tau-b	,361	,024	14,639	,000
	Spearman Correlation	,415	,028	14,843	,000[c]
Interval by Interval	Pearson's R	,411	,028	14,668	,000[c]
N of Valid Cases		1060			

a. Not assuming the null hypothesis.

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

c. Based on normal approximation.

Conclusie: sterk positieve associatie tussen VAR T_WH en VAR T_lid aclvb

Berekening van r

- Berekening van r door "*een combinatie te nemen van de afwijkingen van de waarden op beide variabelen ten opzichte van hun gemiddelde*"

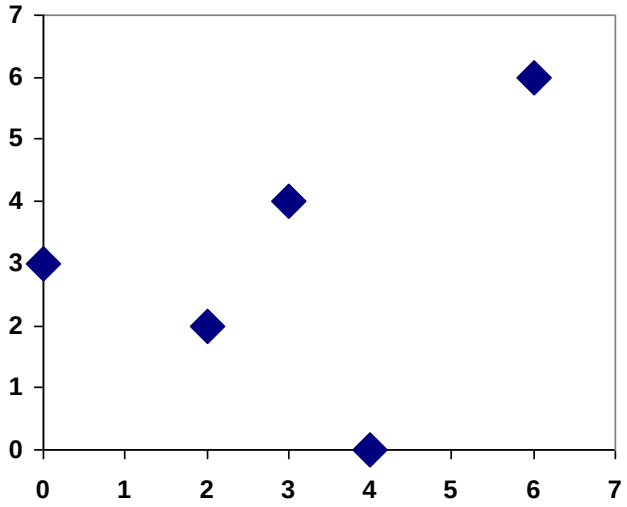
- Gaat uit van **een lineair verband**
  - Studie van de scatterplot
  - Studie van de covariantie
    - Als X ↑ ↓ en Y ↑ ↓ : cov (x,y) = positief
    - Als X ↑ ↓ en Y ↓ ↑ : cov (x,y) = negatief
    - Bij niet-constante relatie: cov (x,y) = 0

$$\operatorname{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

  - Studie van de lineaire regressie (***f(y) = ax + b***)
    - *Verklaren en voorspellen*

Covariantie

Scatterplot



Covariantiematrix

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
0	3	-3	0	0
2	2	-1	-1	1
3	4	0	1	0
4	0	1	-3	-3
6	6	3	3	9
$\bar{x} = 3$	$\bar{y} = 3$			$\sum = 7$

$$\mathrm{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}))}{n-1} = \frac{7}{4} = 1.75$$

Wat betekent dit??

Covariantie - 2

- *Covariantie zegt ons eigenlijk niet zo heel veel….*
- Reden: Cov (x, y) is afhankelijk van de grootte van de variantie van de x- en y-waarden (“*range*”)

(101-51)*(100-50)
(81-51)*(80-50)
Etc…

	High variance data				Low variance data		
Subject	x	y	x error * y error		x	y	X error * y error
1	101	100	2500		54	53	9
2	81	80	900		53	52	4
3	61	60	100		52	51	1
4	51	50	0		51	50	0
5	41	40	100		50	49	1
6	21	20	900		49	48	4
7	1	0	2500		48	47	9
Mean	51	50			51	50	
Sum of x error * y error :			7000		Sum of x error * y error :		28
Covariance:			1166.67	>	Covariance:		4.67

Als X ↑ ↓ en Y ↑ ↓ : cov (x,y) = positief

Correlations

		TijdLid_6 Hoe lang bent u al lid van de ACLVB?	TijdWerkloos_3.2 Hoe lang bent u al werkloos?
TijdLid_6 Hoe lang bent u al lid van de ACLVB?	Pearson Correlation	1	,408**
	Sig. (2-tailed)		,000
	Sum of Squares and Cross-products	5493,550	683,394
	Covariance	1,194	,645
	N	4603	1061
TijdWerkloos_3.2 Hoe lang bent u al werkloos?	Pearson Correlation	,408**	1
	Sig. (2-tailed)	,000	
	Sum of Squares and Cross-products	683,394	2240,199
	Covariance	,645	2,066
	N	1061	1085

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

Somer's D

Somer's D heeft de volgende eigenschappen:

- Voor a-symmetrische maten !! (*directional measure*): meet samenhang tussen VAR1 en VAR2 maar veronderstelt de samenhang tussen VAR2 en VAR1 niet.

- De interpretatie van de kengetallen is idem als bij Kendall's tau-b
- Coëfficiënt = [-1; 1]
  - Bij volledige overeenstemming tussen de beide volgordes VAR x en VAR Y de coëfficiënt met de waarde 1.
  - Met in het tegenovergestelde geval de coëfficiënt met de waarde -1.
  - Met indien de volgordes totaal onafhankelijk zijn de coëfficiënt met waarde 0.
  - **En met een waarde tussen [-1 ; 1] in alle andere gevallen.**

Spearman R

- Gebaseerd op de rangnummers van de waarden ipv. op de data zelf (*verschil in ordening van de waarden*)

IQ, X_i	Hours of TV per week, Y_i	rank x_i	rank y_i	d_i	d_{i^2}
86	0	1	1	0	0
97	20	2	6	−4	16
99	28	3	8	−5	25
100	27	4	7	−3	9
101	50	5	10	−5	25
103	29	6	9	−3	9
106	7	7	3	4	16
110	17	8	5	3	9
112	6	9	2	7	49
113	12	10	4	6	36

Aantal unieke waardes = 6

$\Sigma dj^2 = 194$

- Berekening van Spearman R
  - 10 waarnemingsparen (= N-1)
  - n = aantal waargenomen paren
  - d = het verschil tussen de rangnummers van het betrokken paar waarden van X en Y

$$\sum d_i^2 = 194$$

$$\rho = 1 - \frac{6 \times 194}{10(10^2 - 1)}$$ ⇨ $\rho = -0.175757575$

  - De lage R-waarde wijst op een lage correlatie tussen VAR1 en VAR2. Zwak negatief verband.

Spearman – interpretatie R

- **Aanwezigheid van het verband.** Bepaalt of er een significant verband bestaat. De p-waarde van de coëfficiënt moet eerst onderzocht worden. Als die >0,05, dan is er "geen verband". Als p ≤ 0,05, dan is er wel een verband. Significantie stijgt als p-level kleiner is (<0,01 bv)

- **Richting van het verband.** Kijk naar de coëfficiënt. Is het verband positief (+) of negatief (-)?

- **Sterkte van het verband.** Dit getal ligt altijd tussen **[-1 ; 1]**. Correlatiecoëfficiënten die dichter bij +1 of -1 liggen, geven een sterker verband aan. o geeft geen verband aan.

Opgepast ! In het geval dat de originele waardes voor X en Y reeds samenhang of koppelingen ('*ties*') vertonen mag Spearman Rho niet gebruikt worden maar moet de ***Pearson correlation coefficient*** worden berekend.

Voorbeeld aclvb

Correlations

			TijdLid_6 Hoe lang bent u al lid van de ACLVB?	TijdWerkloos_3.2 Hoe lang bent u al werkloos?
Kendall's tau_b	TijdLid_6 Hoe lang bent u al lid van de ACLVB?	Correlation Coefficient	1,000	,361**
		Sig. (2-tailed)	.	,000
		N	3661	891
	TijdWerkloos_3.2 Hoe lang bent u al werkloos?	Correlation Coefficient	,361**	1,000
		Sig. (2-tailed)	,000	.
		N	891	911
Spearman's rho	TijdLid_6 Hoe lang bent u al lid van de ACLVB?	Correlation Coefficient	1,000	,415**
		Sig. (2-tailed)	.	,000
		N	3661	891
	TijdWerkloos_3.2 Hoe lang bent u al werkloos?	Correlation Coefficient	,415**	1,000
		Sig. (2-tailed)	,000	.
		N	891	911

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

Interpretatie van R (idem)

- **R** varieert van **[-1; 1]**, voor perfecte – of + verbanden, en **0 = geen verband**
  - [-.10;0] of [0;.10] = Zeer zwak of geen verband
  - [-.25;-.10[of [.11; .25] = Zwak verband
  - [-.50;-.25[of [.26;.50] = Sterk verband
  - [-1.0;-.50[of [.51; 1.0] = Heel sterk verband

Bron: SPSS in onderzoek, p. 243

Significantie

- Naast samenhang kijken naar significantie:
  - Prob <.05 sign. tot op 0.05 niveau (*)
  - Prob <.01 sign. tot op 0.01 niveau (**)
  - Prob <.001 sign. tot op 0.001 niveau (***)

Bijv. **Spearman Correlation** van .415** met "**Sig. 2-tailed**" van .000 : sterk positief verband is *niet* aan toeval te wijten

Oefening

- Open SPSS bestand **Politieke besluitvorming_opdracht les5.sav**
- Bestudeer de samenhang tussen ***Leeftijd*** en ***VRP_belangrijkheid***, en vertrek bij het formuleren van de nulhypothese.

- Hypothesis (Ho): …
- Variabelen en meetniveau:
  - Variable 1: *Leeftijd* (Scale) (onafhankelijke -> kolom)
  - Variable 2: *VRP_belangrijkheid* (Ordinaal) (afhankelijke)

- Selecteer: **Analyze > Correlate > Bivariate**
- Plaats *VAR* naar de *Variables* box.
- Check *Spearman*, *Two-tailed*, en *Flag significant correlations*.
- Selecteer: **Options**. Check *Means and standard deviations.*
- Selecteer: **Continue > OK**

```
NONPAR CORR
  /VARIABLES=VRP_belangrijkheid geboortejaar
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Analyse van de output

- Correlations
  - Significantie?

Correlations

			VRP_belangrijkheid Ik vind volksraadpleging als vorm van directe democratie:	geboortejaar In welk jaar bent u geboren?
Spearman's rho	VRP_belangrijkheid Ik vind volksraadpleging als vorm van directe democratie:	Correlation Coefficient	1,000	-,037
		Sig. (1-tailed)	.	,290
		N	224	224
	geboortejaar In welk jaar bent u geboren?	Correlation Coefficient	-,037	1,000
		Sig. (1-tailed)	,290	.
		N	224	244

Alternatief (crosstabs)

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Interval by Interval	Pearson's R	,034	,061	,505	,614[c]
Ordinal by Ordinal	Spearman Correlation	-,037	,064	-,556	,579[c]
N of Valid Cases		224			

a. Not assuming the null hypothesis.

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

c. Based on normal approximation.

Samenhang bestuderen tussen twee scale-variabelen

- Hoe kunnen we het verband tussen twee variabelen van **interval- en rationiveau** meten? (*metrisch* niveau)
  - In dit geval proberen we de aanwezigheid, richting en de sterke **van een monotoon verband** te bepalen (idem twee ordinale variabelen dus)
  - Tussen **twee stochastische variabelen** (onbekend verloop)
  - We worden hierbij geholpen door **de product-momentcorrelatiecoëfficiënt van Pearson Rho**

Pearson's R (rho)

- **Standaardisatie** van deze kengetallen voor **cov (x, y)**

- Pearson R verdeelt de covariantie door de multipele-standaarddeviaties van x en y (zie verder)

Covariantie -> Pearson's R

$$\mathrm{cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

formule voor de product-moment-correlatiecoëfficiënt van Pearson

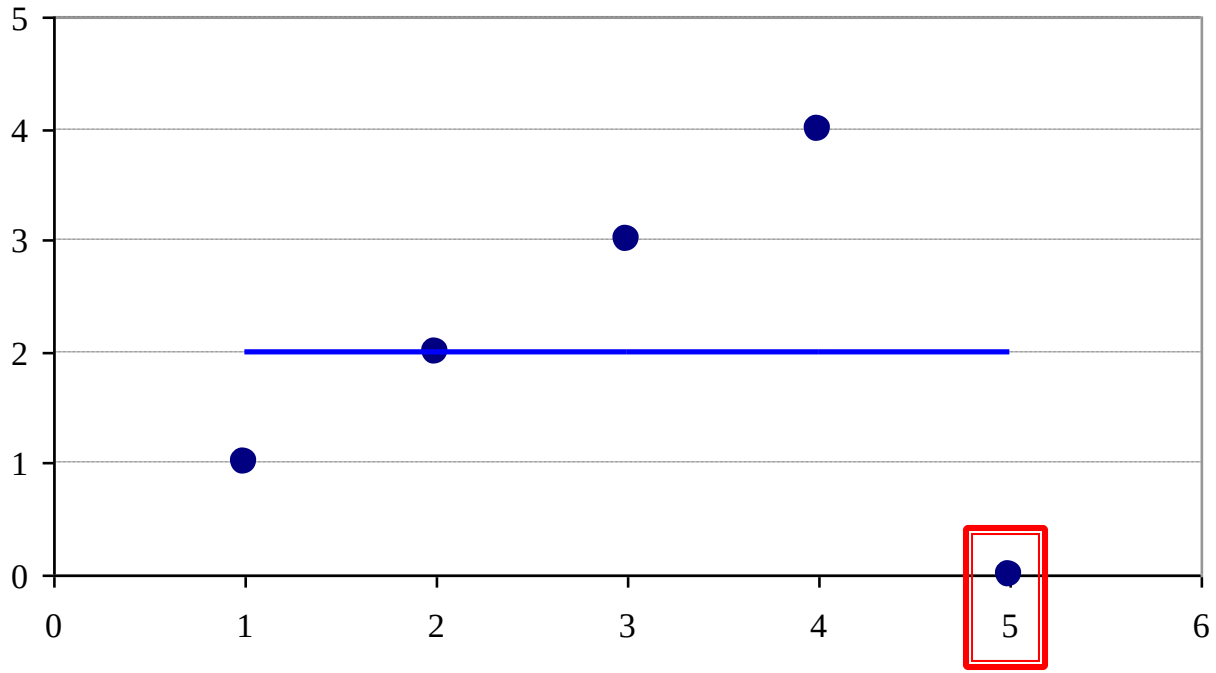
$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y}$$

waarbij

x_i	=	elke *x*-waarde
$\bar{x}$	=	gemiddelde van de *x*-waarden
y_i	=	elke *y*-waarde
$\bar{y}$	=	gemiddelde van de *y*-waarden
n	=	aantal gepaarde gevallen
s_x, s_y	=	standaardafwijking van *x* en *y*

$$r_{xy} = \frac{\sum_{i=1}^{n} Z_{x_i} * Z_{y_i}}{n-1}$$

Opgelet !!
r is zeer gevoelig voor extreme waarden



Kenmerken

- **Aanwezigheid van het verband.** Bepaalt of er een significant verband bestaat. De p-waarde van de coëfficiënt moet eerst onderzocht worden. Als Prob >0.05, dan is er "geen verband". Als p ≤ 0,05, dan is er wel een verband.

- **Richting van het verband.** Kijk naar de coëfficiënt. Is het verband positief (+) of negatief (-)?

- **Sterkte van het verband.** Dit getal ligt altijd tussen **[-1 ; 1]**. Correlatiecoëfficiënten die dichter bij +1 of -1 liggen, geven een sterker verband aan. Met als waarde **0 = geen verband**. [= effect van de standardisatie !!]

Interpretatie van R (idem)

- **R** varieert van **[-1; 1]**, voor perfecte – of + verbanden, en **0 = geen verband**
  - [-.10;0] of [0;.10] = Zeer zwak of geen verband
  - [-.25;-.10[of [.11; .25] = Zwak verband
  - [-.50;-.25[of [.26;.50] = Sterk verband
  - [-1.0;-.50[of [.51; 1.0] = Heel sterk verband

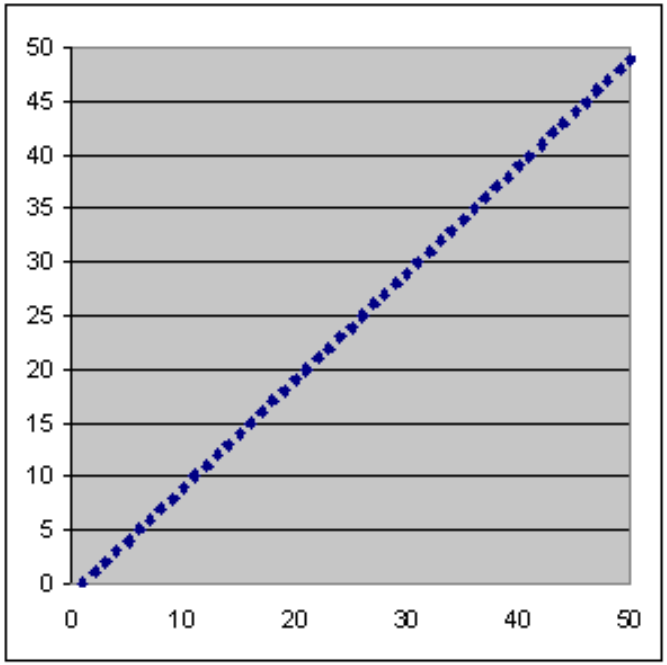
Bron: SPSS in onderzoek, p. 243

Significantie

- Naast samenhang kijken naar significantie:
  - Prob <.05 sign. tot op 0.05 niveau (*)
  - Prob <.01 sign. tot op 0.01 niveau (**)
  - Prob <.001 sign. tot op 0.001 niveau (***)

Bijv. **Pearson Correlation** van -.150 met "**Sig. 2-tailed**" van .004 : zwak negatief verband is *niet* aan toeval te wijten

Pearson R - 2

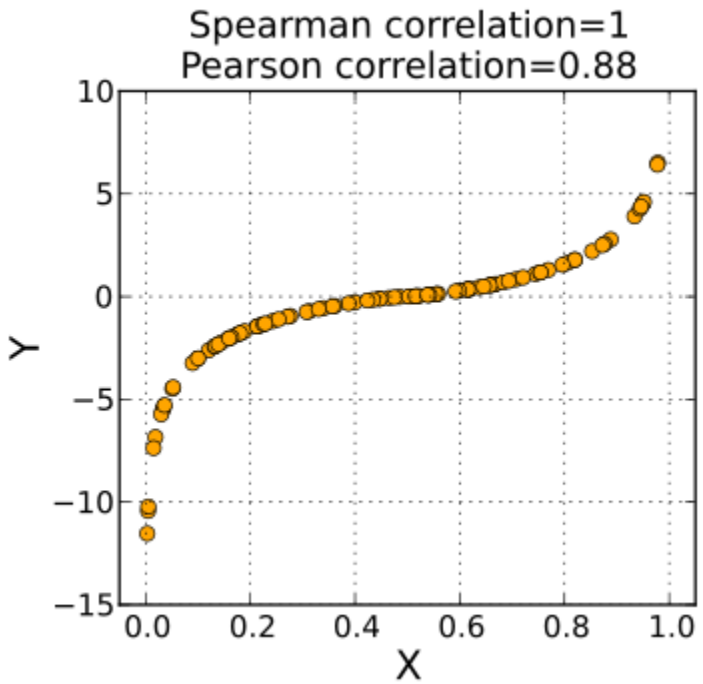


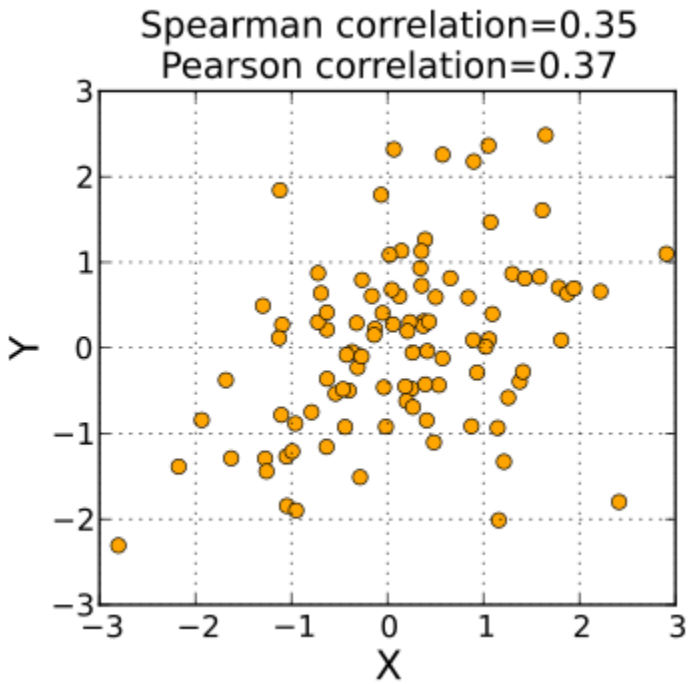
- Als **r = 1** of **r = -1** :
  - *Voorspelbaarheid* van waarde VARy op basis van waarde VARx
- r is feitelijk $\hat{r}$
  - r = werkelijke r van de gehele populatie
  - $\hat{r}$ = een schatting van r op basis van de gegevens

Pearson R - 3

- Speciale overwegingen bij procedures voor **lineaire correlaties** en hun interpretatie:
  - Bij de correlatiecoëfficiënt wordt er alleen naar het verband tussen twee variabelen gekeken. Er wordt niet gekeken naar interacties met andere variabelen.
  - Er wordt bij de correlatiecoëfficiënt expliciet niet uitgegaan van een oorzaak-gevolg-verband
  - Onderzoekers zullen altijd **de nul-hypothese** (geen verband) testen.

Spearman R versus Pearson R





A Spearman correlation of 1 results when the two variables being compared are monotonically related, even if their relationship is not linear. In contrast, this does not give a perfect Pearson correlation.

Bron:
http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient

Oefening

- Bestudeer de samenhang tussen ***Beheersing_cumul*** en ***Beheersing_schijnkandidatuur***, en vertrek bij het formuleren van de nulhypothese.

- Herschaal de VAR naar volgende categorieën:
  - 1: helemaal niet prioritair
  - 2: eerder niet prioritair
  - 3: eerder wel prioritair
  - 4: helemaal wel prioritair

  (missings of geen antwoord ->: missings)

- Wat is de nulhypothese?: …

- Variabelen en meetniveau:
  - Variable 1: *Beheersing_cumul* (Scale)
  - Variable 2: *Beheersing_schijnkandidatuur* (Scale)

- Selecteer: **Analyze > Correlate > Bivariate**
- Plaats *VAR* naar de *Variables* box.
- Check *Pearson*, *Two-tailed*, en *Flag significant correlations*.
- Selecteer: **Options**. Check *Means and standard deviations.*
- Selecteer: **Continue > OK**

Analyse van de Output

- Beschrijvende statistieken: Means, Standard Deviations, en N voor *Beheersing_cumul* en *Beheersing_schijnkandidatuur*.

Descriptive Statistics

	Mean	Std. Deviation	N
Beheersing_cumul Hoe prioritair is de beheersing van cumul?	2,5091	,93384	220
Beheersing_ schijnkandidatuur Hoe prioritair is de beheersing van schijnkandidatuur?	2,3091	,78508	220

- Correlaties:
  - De variabelen worden opgelijst en snijden elkaar (intersect) in de grid/matrix
  - In elke intersectie staat
    - Een mate van samenhang (correlatiecoëfficiënt)
    - het significantieniveau en
    - het aantal (N).

Interpretatie ?

Correlations

		Beheersing_cumul Hoe prioritair is de beheersing van cumul?	Beheersing_schijnkandidatuur Hoe prioritair is de beheersing van schijnkandidatuur?
Beheersing_cumul Hoe prioritair is de beheersing van cumul?	Pearson Correlation	1	-,041
	Sig. (2-tailed)		,543
	Sum of Squares and Cross-products	190,982	-6,618
	Covariance	,872	-,030
	N	220	220
Beheersing_schijnkandidatuur Hoe prioritair is de beheersing van schijnkandidatuur?	Pearson Correlation	-,041	1
	Sig. (2-tailed)	,543	
	Sum of Squares and Cross-products	-6,618	134,982
	Covariance	-,030	,616
	N	220	220

Opdracht

- Open de laatste bewaarde **Dataset deel 7bis (voorbeeldbestand).sav**
- Maak een nieuwe variabele '***Vormingsniveau_rec***' aan o.b.v. de bestaande variabele '*Vormingsniveau*'
  - System missing (alle ontbrekende waarden)
  - 1 = laag (voor waarden [1;5]),
  - 2 = middelmatig (voor waarden [6;10]),
  - 3 = hoog (voor waarden [11;15]) en
  - 4 = zeer hoog [voor alle waarden >15)
- Onderzoek samenhang (test voor correlatie) tussen:
  - ***Vormingsniveau_rec*** en ***VertrouwenLMH*** *(opdracht A)*
  - ***Leeftijd*** *en* ***PC_ervaring*** *en* ***Wantrouwen*** *(opdracht B) + puntenwolk voor Leeftijd en PC-ervaring*

- Sla de oplossing (.spv outputfile) en de interpretatie van (A) en (B) op in WORD-bestand met naam "**Dataset deel 7bis (voorbeeldbestand)_achternaam**"

Centraal bij de analyse van de output:

Hoe interpreteren we de gevonden verbanden ?
*** bestaan**
*** richting**
*** sterkte**
Ofwel via Crosstabs + statistics
Ofwel via correlations (juiste maat kiezen)

Oplossing (A) – Spearman Rho

- Statistics
- Correlations

Correlations

			Vormingsniveau in 5 klassen	Vertrouwen (laag, midden, hoog)
Spearman's rho	Vormingsniveau in 5 klassen	Correlation Coefficient	1,000	,395**
		Sig. (2-tailed)	.	,000
		N	160	160
	Vertrouwen (laag, midden, hoog)	Correlation Coefficient	,395**	1,000
		Sig. (2-tailed)	,000	.
		N	160	160

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

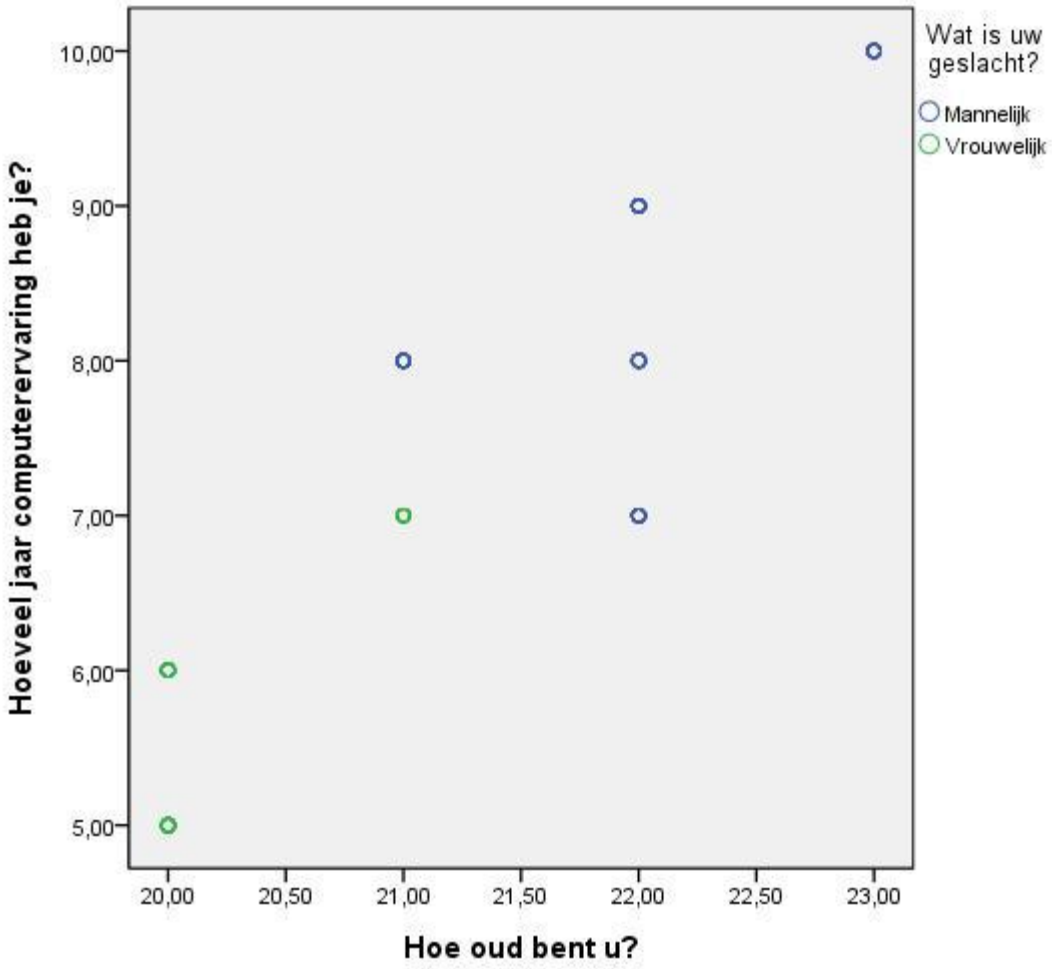
Oplossing (B) – Pearson R

- Statistics
- Correlations

Correlations

		Hoeveel jaar computerervaring heb je?	Op een schaal van 0 tot 10 met 10 als hoogste niveau van wantrouwen, hoeveel wantrouwen heeft u dat u geen SPSS zal leren?	Hoe oud bent u?
Hoeveel jaar computerervaring heb je?	Pearson Correlation	1	-,113	,904**
	Sig. (2-tailed)		,154	,000
	N	160	160	160
Op een schaal van 0 tot 10 met 10 als hoogste niveau van wantrouwen, hoeveel wantrouwen heeft u dat u geen SPSS zal leren?	Pearson Correlation	-,113	1	-,059
	Sig. (2-tailed)	,154		,455
	N	160	160	160
Hoe oud bent u?	Pearson Correlation	,904**	-,059	1
	Sig. (2-tailed)	,000	,455	
	N	160	160	160

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


Hoeveel jaar computerervaring heb je?
10,00
9,00
8,00
7,00
6,00
5,00
20,00
20,50
21,00
21,50
22,00
22,50
23,00
Hoe oud bent u?
Wat is uw geslacht?
Mannelijk
Vrouwelijk

Voorbeeld
gemengde meetniveaus

```
NONPAR CORR
  /VARIABLES=Studiejaar Vormingsniveau_rec
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

			In welk studiejaar zit u?	Vormingsniveau in 5 klassen
Spearman's rho	In welk studiejaar zit u?	Correlation Coefficient	1,000	,511**
		Sig. (2-tailed)	.	,000
		N	160	160
	Vormingsniveau in 5 klassen	Correlation Coefficient	,511**	1,000
		Sig. (2-tailed)	,000	.
		N	160	160

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

Significantieniveau p < .01 (= hoog)
.513** correlatiecoëfficiënt wijst op een duidelijk sterk en positief verband
Betrouwbaarheidsinterval = 99% (slechts 1% kans op foute analyse)
Conclusie: Hyp0 ('geen verband') = verworpen.
MAAR !!! VARx en VARy zijn niet onafhankelijk van elkaar !!! ->> Pearson R berekenen (= maat van lineaire associatie)

Tot slot - Pearson R^2 (regressie)

- Regressieanalyse veronderstelt **causaliteit** tussen de variabelen, dus **een effect van VARx op VARy**
- Bij enkelvoudige regressieanalyse stelt de onderzoeker eerst vast *welke VAR afhankelijk* is en *welke VAR onafhankelijk*.
- Bij regressierekening omschrijven we de vorm en richting van het systematische verband tussen een afhankelijke en onafhankelijke variabelen.
- We gaan ervan uit dat **het verband rechtlijnig** (lineair) is.

Least square regression (LQR)

- Deze kan algebraïsch worden weergegeven als een regressiefunctie

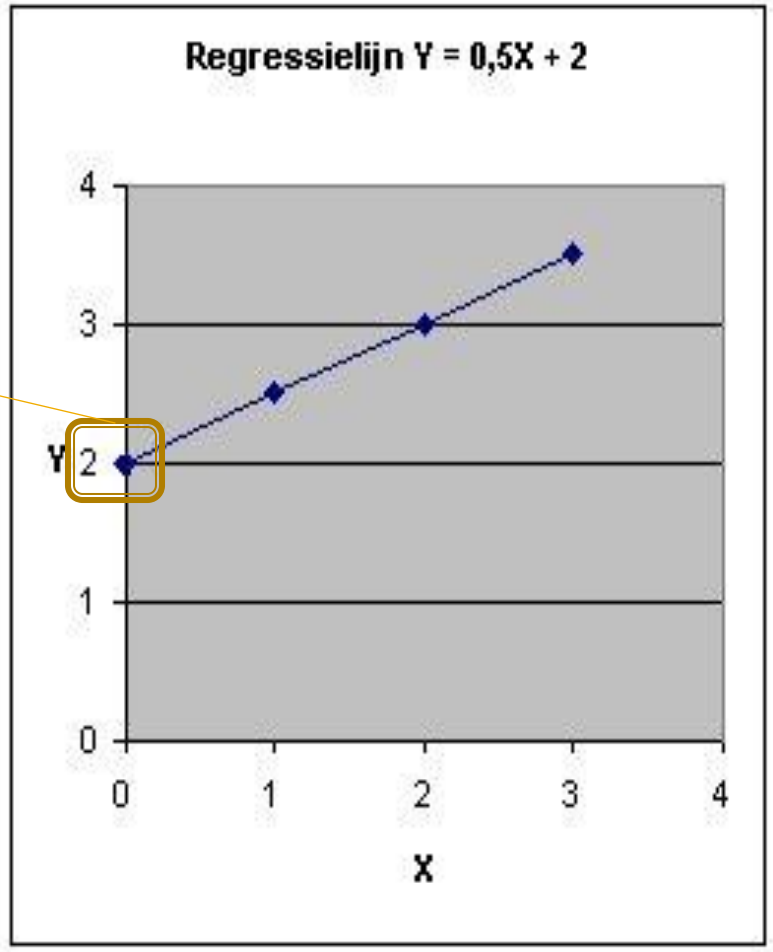
  y = a + bx

  Waarbij (= enkelvoudig lineair model)

  - X = de onafhankelijke variabele ('oorzaak')
  - Y = de afhankelijke variabele ('gevolg')
  - a = **de constante**, die het snijpunt (intercept) met de Y-as vormt
  - b = **de hellingscoëfficiënt** (of richtingscoëfficiënt, RICO)

- Voorbeeldje:

$$Y = 2 + 0.5X$$



- De regressielijn snijdt de Y-as bij het punt (X=0, Y=2): de constante (a) 2 vormt de intercept.
- De richtingscoëfficiënt (b) is 0,5: voor iedere eenheid in de toename van X, neemt Y met 50% daarvan toe (of: de toename van Y is de helft van die van X).

R^2 - determinatiecoëfficiënt

- Er is geen consensus over de exacte definitie van R^2.
- De **determinatiecoëfficiënt R^2** is een maat voor het deel van **de variabiliteit** dat wordt verklaard door het statistisch model.

- Enkel in het geval van lineaire regressie zijn alle definities *equivalent*.
  - In geval **van enkelvoudige lineaire regressie** is R^2 simpelweg gelijk aan het kwadraat van een correlatiecoëfficiënt r.
  - In het geval van **multipele lineaire regressie** valt R^2 te definiëren als de proportie verklaarde variantie van het regressiemodel

Interpretatie R^2

R	R^2 (afgerond) (zie regressie)	Verklaarde variantie	Interpretatie kracht van het verband
< 0,3	< 0,1	< 10%	zeer zwak
0,3 - 0,5	0,1 - 0,25	10 - 25%	zwak
0,5 - 0,7	0,25 - 0,5	25 - 50%	matig
0,7 - 0,85	0,5 - 0,75	50 - 75%	sterk
0,85 - 0,95	0,75 - 0,9	75 - 90%	zeer sterk
> 0,95	> 0,9	> 90%	uitzonderlijk sterk (suspect!)

Hoe lineaire verbanden nagaan in SPSS?

- Als we in de schermen van de module Linear Regression de afhankelijke en onafhankelijke variabelen opgeven dan zullen de volgende tabellen worden gegenereerd:
  - Model Summary
  - ANOVA
  - Coeficients.

Via **Analyze** > **Regression** > **Linear regression**

Voorbeeld aclvb

Model 1: VARy = T_lid

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,408[a]	,166	,166	1,007

a. Predictors: (Constant), TijdWerkloos_3.2 Hoe lang bent u al werkloos?

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	214,485	1	214,485	211,500	,000[a]
	Residual	1074,076	1059	1,014		
	Total	1288,561	1060			

a. Predictors: (Constant), TijdWerkloos_3.2 Hoe lang bent u al werkloos?

b. Dependent Variable: TijdLid_6 Hoe lang bent u al lid van de ACLVB?

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2,840	,090		31,728	,000
	TijdWerkloos_3.2 Hoe lang bent u al werkloos?	,314	,022	,408	14,543	,000

a. Dependent Variable: TijdLid_6 Hoe lang bent u al lid van de ACLVB?

Model 2: VARy = T_werkloos

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,408[a]	,166	,166	1,309

a. Predictors: (Constant), TijdLid_6 Hoe lang bent u al lid van de ACLVB?

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	362,441	1	362,441	211,500	,000[a]
	Residual	1814,997	1059	1,714		
	Total	2177,439	1060			

a. Predictors: (Constant), TijdLid_6 Hoe lang bent u al lid van de ACLVB?

b. Dependent Variable: TijdWerkloos_3.2 Hoe lang bent u al werkloos?

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1,738	,153		11,326	,000
	TijdLid_6 Hoe lang bent u al lid van de ACLVB?	,530	,036	,408	14,543	,000

a. Dependent Variable: TijdWerkloos_3.2 Hoe lang bent u al werkloos?

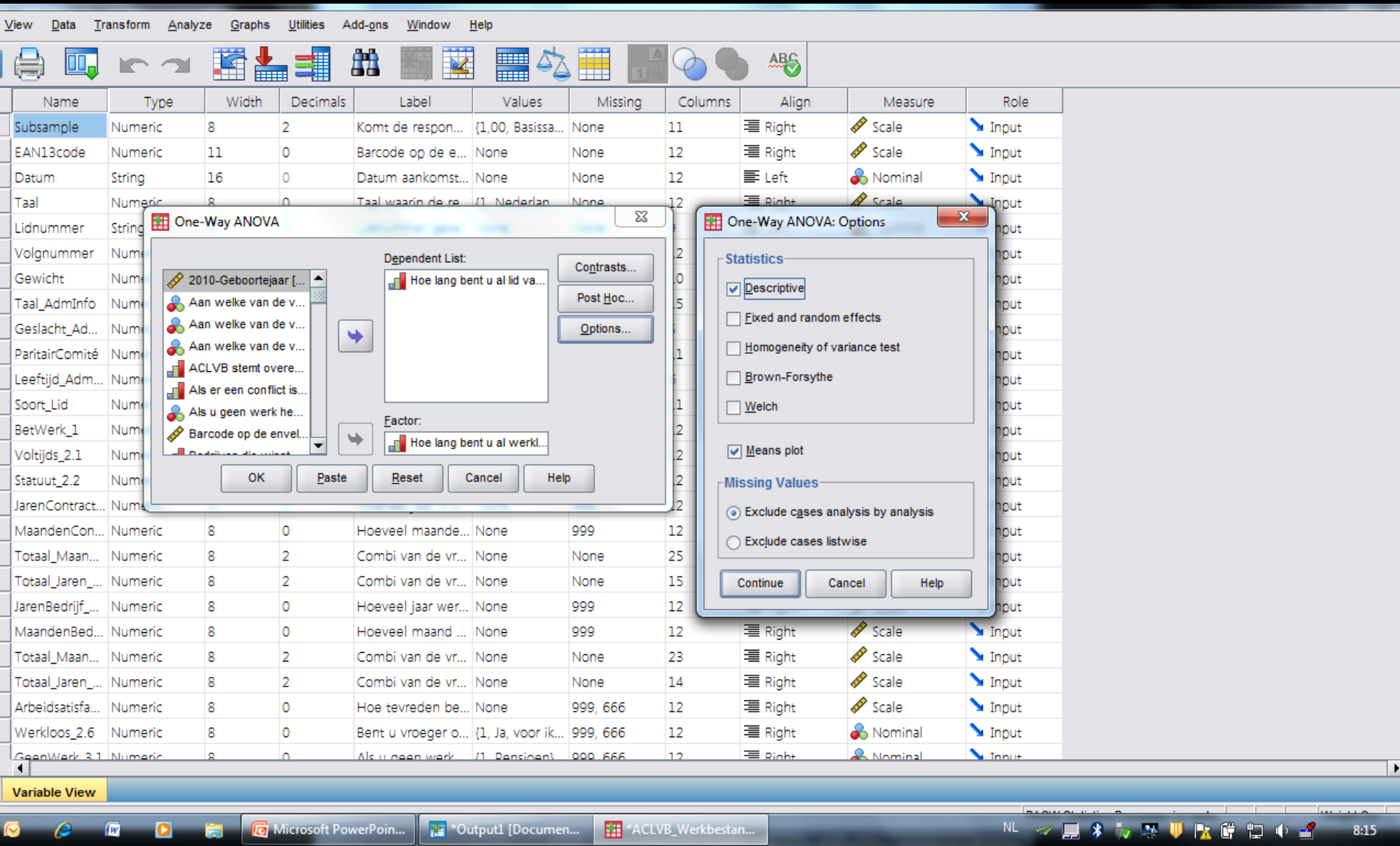
Interpretatie – model summary

- De tabel Model Summary geeft de **Pearson's R** correlatie weer tussen de variabelen meestal in een percentage 'verklaarde variantie' uitgedrukt.
- De kolom **'R square'** is een zogenaamde *goodness-of-fit* maat. In een enkelvoudige lineaire regressie is dat het kwadraat van de correlatie. En omwille van *equivalentie* geeft deze maat uitdrukking aan hoe goed de geobserveerde data clusteren rond de geschatte regressielijn.
- In dit voorbeeld betekent R square dus dat de totale variatie van VARy voor **slechts 17%** verklaard kan worden door de lineaire regressie c.q. de verschillen in VARx.

Interpretatie - ANOVA

- De **F-test** in de ANOVA tabel geeft ons een indicatie van de samenhang tussen de variantie voor VARx en VARy.
- F toetst ook de nulhypothese: **b = o** (dus 'er is geen samenhang in **de puntenwolk**' en dus de regressielijn loopt horizontaal - zonder helling).
- Of via **Analyze** > **Compare means** > **One-way ANOVA**

- Uit de tabel kunnen we opmaken dat bij de studie van de variantie van VARx en VARy er tussen VARx en VARy een significant verband is en dat VARx een voorspeller is van VARy (regressiecoëfficiënt b ≠ o) en omgekeerd


View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
Subsample Numeric 8 2 Komt de respon... {1,00, Basissa... None 11 Right Scale Input
EAN13code Numeric 11 0 Barcode op de e... None None 12 Right Scale Input
Datum String 16 0 Datum aankomst... None None 12 Left Nominal Input
Taal Numeric 8 0 Taal waarin de re... {1, Nederlan... None 12 Right Scale Input
Lidnummer String 8 0 Lidnummer None None 12 Left Nominal Input
Volgnummer Numeric 8 0 Volgnummer None None 12 Right Scale Input
Gewicht Numeric 8 0 Gewicht None None 10 Right Scale Input
Taal_AdmInfo Numeric 8 0 Taal waarin de ... {1, Nederlan... None 5 Right Nominal Input
Geslacht_Ad... Numeric 8 0 Geslacht None None 5 Right Nominal Input
ParitairComité Numeric 8 0 Paritair Comité None None 11 Right Nominal Input
Leeftijd_Adm... Numeric 8 0 Leeftijd None None 5 Right Scale Input
Soort_Lid Numeric 8 0 Soort Lid None None 11 Right Nominal Input
BetWerk_1 Numeric 8 0 Betrekking Werk 1 None None 12 Right Nominal Input
Voltijds_2.1 Numeric 8 0 Voltijds 2.1 None None 12 Right Nominal Input
Statuut_2.2 Numeric 8 0 Statuut 2.2 None None 12 Right Nominal Input
JarenContract... Numeric 8 0 Jaren Contract... None None 12 Right Scale Input
MaandenCon... Numeric 8 0 Hoeveel maande... None 999 12 Right Scale Input
Totaal_Maan... Numeric 8 2 Combi van de vr... None None 25 Right Scale Input
Totaal_Jaren_... Numeric 8 2 Combi van de vr... None None 15 Right Scale Input
JarenBedrijf_... Numeric 8 0 Hoeveel jaar wer... None 999 12 Right Scale Input
MaandenBed... Numeric 8 0 Hoeveel maand ... None 999 12 Right Scale Input
Totaal_Maan... Numeric 8 2 Combi van de vr... None None 23 Right Scale Input
Totaal_Jaren_... Numeric 8 2 Combi van de vr... None None 14 Right Scale Input
Arbeidsatisfa... Numeric 8 0 Hoe tevreden be... None 999, 666 12 Right Scale Input
Werkloos_2.6 Numeric 8 0 Bent u vroeger o... {1, Ja, voor ik... 999, 666 12 Right Nominal Input
GeenWerk_3.1 Numeric 8 0 Als u geen werk ... {1, Pensioen} 999, 666 12 Right Nominal Input
Variable View
One-Way ANOVA
2010-Geboortejaar [...
Aan welke van de v...
Aan welke van de v...
Aan welke van de v...
ACLVB stemt overe...
Als er een conflict is...
Als u geen werk he...
Barcode op de envel...
Dependent List:
Hoe lang bent u al lid va...
Factor:
Hoe lang bent u al werkl...
Contrasts...
Post Hoc...
Options...
OK Paste Reset Cancel Help
One-Way ANOVA: Options
Statistics
Descriptive
Fixed and random effects
Homogeneity of variance test
Brown-Forsythe
Welch
Means plot
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Continue Cancel Help
Microsoft PowerPoin...
*Output1 [Documen...
*ACLVB_Werkbestan...
NL
8:15

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
TijdLid_6 Hoe lang bent u al lid van de ACLVB? * TijdWerkloos_3.2 Hoe lang bent u al werkloos?	Between Groups	(Combined)	246,925	4	61,731	62,590	,000
		Linearity	214,485	1	214,485	217,469	,000
		Deviation from Linearity	32,441	3	10,814	10,964	,000
	Within Groups		1041,636	1056	,986		
	Total		1288,561	1060			

Hoe groter F hoe groter de samenhang tussen de co-variantie x, y

Maar waar het sterkst ???

Multiple Comparisons

TijdLid_6 Hoe lang bent u al lid van de ACLVB?
Bonferroni

(I) TijdWerkloos_3.2 Hoe lang bent u al werkloos?	(J) TijdWerkloos_3.2 Hoe lang bent u al werkloos?	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
<3maanden	3-6maanden	,003	,135	1,000	-,38	,38
	6maanden-1jaar	-,148	,125	1,000	-,50	,20
	1-2jaar	-,377*	,124	,024	-,73	-,03
	>2jaar	-1,089*	,099	,000	-1,37	-,81
3-6maanden	<3maanden	-,003	,135	1,000	-,38	,38
	6maanden-1jaar	-,150	,133	1,000	-,52	,22
	1-2jaar	-,380*	,132	,040	-,75	-,01
	>2jaar	-1,092*	,109	,000	-1,40	-,79
6maanden-1jaar	<3maanden	,148	,125	1,000	-,20	,50
	3-6maanden	,150	,133	1,000	-,22	,52
	1-2jaar	-,230	,122	,593	-,57	,11
	>2jaar	-,942*	,096	,000	-1,21	-,67
1-2jaar	<3maanden	,377*	,124	,024	,03	,73
	3-6maanden	,380*	,132	,040	,01	,75
	6maanden-1jaar	,230	,122	,593	-,11	,57
	>2jaar	-,712*	,095	,000	-,98	-,45
>2jaar	<3maanden	1,089*	,099	,000	,81	1,37
	3-6maanden	1,092*	,109	,000	,79	1,40
	maanden-1jaar	,942*	,096	,000	,67	1,21
	1-2jaar	,712*	,095	,000	,45	,98

*. The mean difference is significant at the 0.05 level.

Groep 3 van >=2jaar WH verschilt significant van de andere twee groepen in gemiddeld aantal jaren lid van ACLVB
De andere groepen verschillen niet significant van elkaar

TijdLid_6 Hoe lang bent u al lid van de ACLVB?

	TijdWerkloos_3.2 Hoe lang bent u al werkloos?	N	Subset for alpha = 0.05		
			1	2	3
Tukey HSD[a,b]	3-6maanden	97	3,40		
	<3maanden	121	3,40		
	6maanden-1jaar	130	3,55	3,55	
	1-2jaar	136		3,78	
	>2jaar	575			4,49
	Sig.		,707	,293	1,000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 141,888.

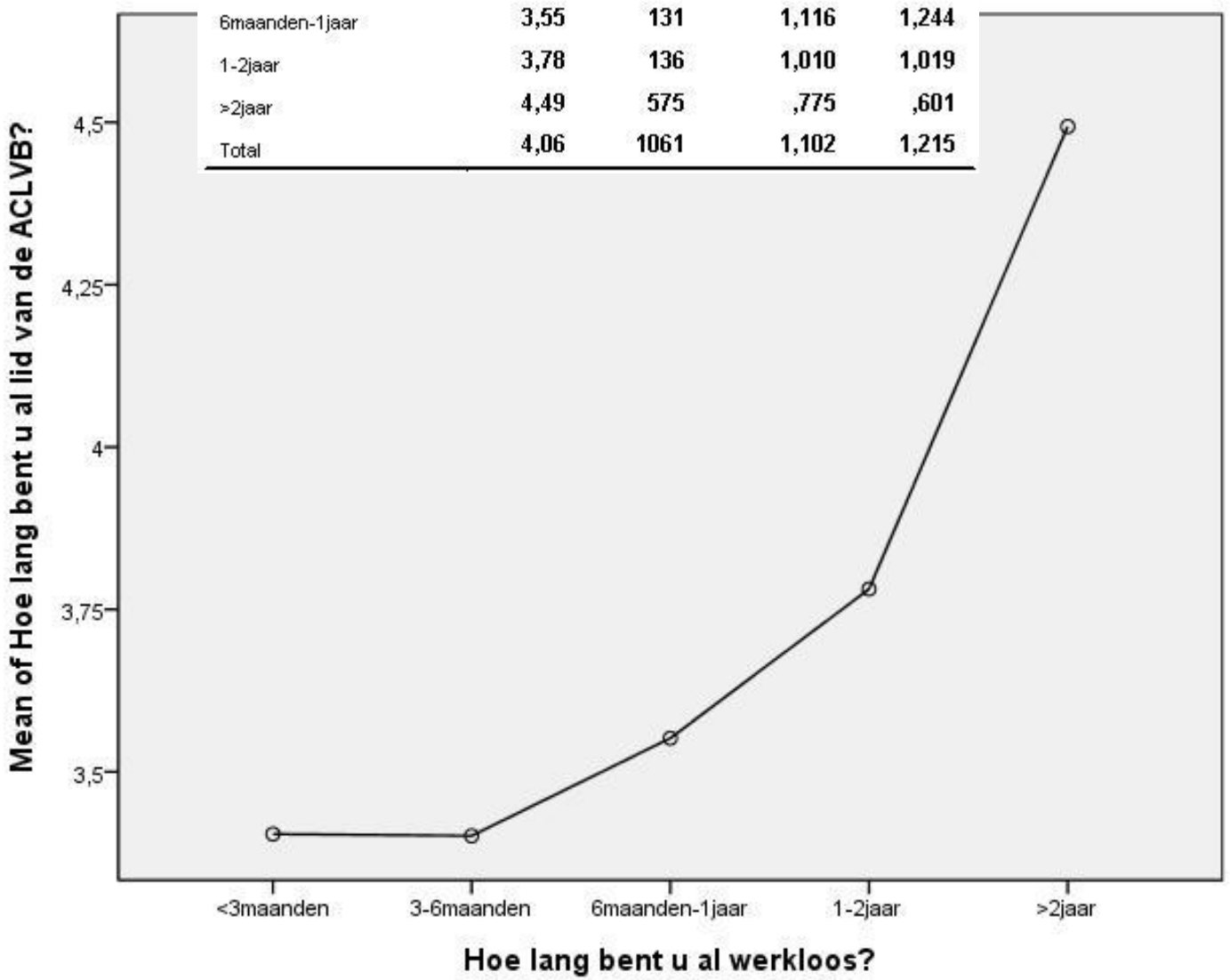
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Screeplot

Report

TijdLid_6 Hoe lang bent u al lid van de ACLVB?

TijdWerkloos_3.2 Hoe la···	Mean	N	Std. Deviation	Variance
<3maanden	3,40	122	1,355	1,837
3-6maanden	3,40	97	1,352	1,828
6maanden-1jaar	3,55	131	1,116	1,244
1-2jaar	3,78	136	1,010	1,019
>2jaar	4,49	575	,775	,601
Total	4,06	1061	1,102	1,215



N=1060

Interpretatie - Coefficients

- De **t-test** in de derde tabel vergelijkt de gemiddelden van de VARx en VARy
- t geeft niet meer informatie dan hetgeen we al wisten uit de F-test in de ANOVA tabel, namelijk dat VARx **een significante voorspeller** is van VARy én omgekeerd.

- Of via **Analyze** > **Compare means** > **Independent Samples t-test**

- In deze tabel wordt de lineaire regressievergelijking gegeven.
  - **De constante a** zien we (enigszins verwarrend in SPSS) staan onder de kolom B en heeft de waarde 2,84 in model 1 en 1,74 in model 2.
  - **De hellingscoëfficiënt b (rico)** wordt ook onder de kolom B vermeld en heeft de waarde 0,408 in beide modellen.

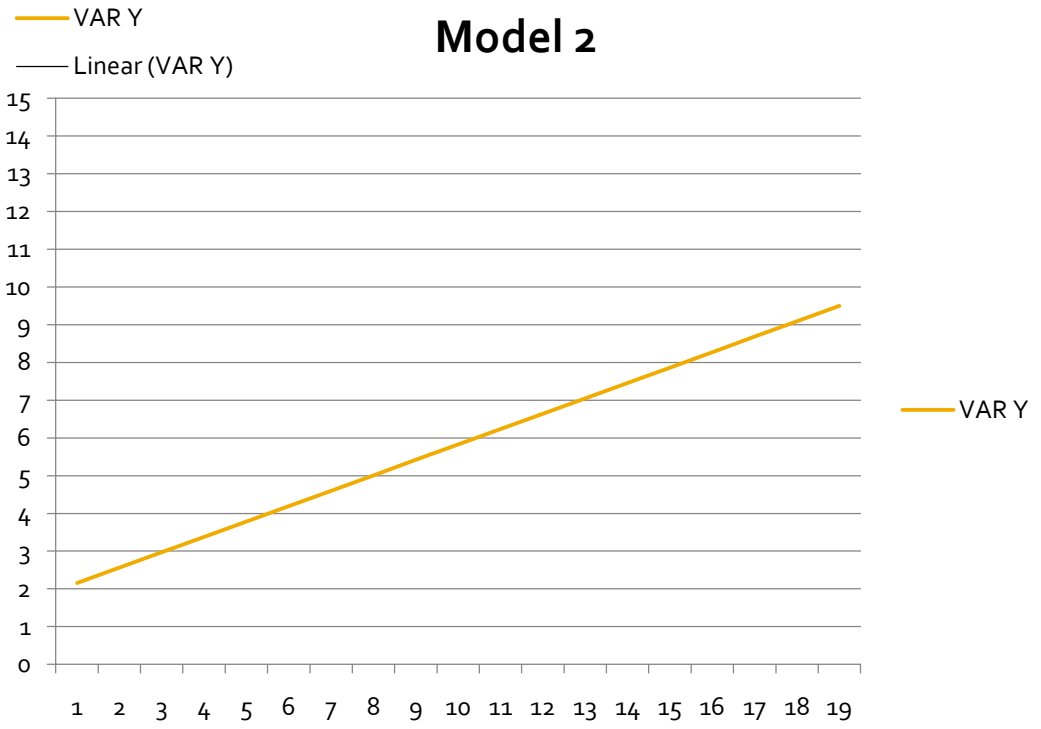
- De regressievergelijking voor model 1:

  *Tijd_lid (Y) = **2,84** + (0,408*Tijd_werkloos)*
- De regressievergelijking voor model 2:

  *Tijd_werkloos (Y) = 1,744 + (0,408*Tijd_lid)*

Simulatie van effect via excel:





- **Wanneer regressieanalyse** uitvoeren?
  - Het berekenen van *trends bij tijdreeksen*
  - Het berekenen en *toetsen van causale verbanden*
  - Het *interpoleren* van ontbrekende waarden
  - Het doen van *conditionele voorspellingen* (indien de trend zich voortzet, dan…)

Wat hebben we geleerd ?

Wanneer we twee variabelen van nominaal of ordinaal meetniveau met elkaar in verband brengen dan kunnen we **de Crosstabs-procedure** gebruiken om kruistabellen en bijbehorende statistische parameters te berekenen.

Om de relatie tussen twee variabelen op interval of ratio meetniveau na te gaan is het produceren van **een scatterplot** (grafische weergave van de relatie) aangeraden. Voor de sterkte van het verband tussen deze twee variabelen voer je dan **een correlatie-analyse** uit. (opmerking: deze lijst is gebaseerd op de menustructuur van SPSS10.0.)

Wat geleerd ? - 2

- **Samenhang nagaan voor twee variabelen (nominaal, ordinaal en interval/ratio)**
- **Correlatiecoëfficiënt berekenen**: Analyze, Correlate, Bivariate, aanduiden van variabelen, aanduiden welke correlatiecëfficiënt je wil berekenen (vb. Pearson), OK
- **Aanzet tot regressie en berekenen van regressiecoëfficiënten** en R-coefficiënten kunnen interpreteren. (Pearson R en R^2), OK

Deel 8

Inferentie, generaliseren en hypothesen toetsen

Inferentiële statistiek

- **Inferentie**: een vorm van logica waarbij je een generalisatie (= veralgemenende uitspraak) doet over een hele klasse (lees: populatie) op basis van wat je hebt waargenomen bij een kleine deelverzameling (klein aantal leden) van die klasse

- **Statistische inferentie**: een verzameling procedures waarbij de steekproefomvang en steekproefgrootheden worden gebruikt om de populatieparameters *zo correct mogelijk* te schatten

Parameters schatten

- Bij parameterschatting gebruik je steekproefinformatie om een interval te berekenen dat het bereik van een parameter, bijvoorbeeld het populatiegemiddelde (μ) of populatiepercentage (π), beschrijft

- Hierbij heb je drie waarden nodig:
  - Steekproefgrootheid
  - Standaardfout
  - Betrouwbaarheidsintervallen

- Steekproefgrootheid: Steekproefgrootheden worden gegenereerd van de steekproefdata en worden gebruikt om populatieparameters te schatten

- Standaardfout: een **maat voor de variabiliteit in de steekproefverdeling** gebaseerd op wat er in theorie zou gebeuren als we een zeer groot aantal onafhankelijke steekproeven uit dezelfde populatie zouden trekken

- Betrouwbaarheidsintervallen: geven **de mate van nauwkeurigheid** aan die de onderzoeker wenst en wordt als betrouwbaarheidsniveau vastlegt in de vorm van *een percentage*.
  - B.I. van 95% is heel courant voor kleine tot middelgrote steekproeven
  - B.I. van >95% voor grote steekproeven

Standaardfout

formule voor de standaardfout van het gemiddelde

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

waarbij

$s_{\bar{x}}$ = standaardfout van het gemiddelde
s = standaardafwijking
n = steekproefomvang

Standaardfout - 2

formule voor de standaardfout van het percentage

$$s_p = \sqrt{\frac{p \times q}{n}}$$

waarbij

s_p = standaardfout van het percentage
p = het steekproefpercentage
q = $(100 - p)$
n = steekproefomvang

Betrouwbaarheidsinterval

- **De berekening van een betrouwbaarheidsinterval voor een gemiddelde of een percentage**:
  - Stap 1: Neem eerst de steekproefgrootheid, hetzij het gemiddelde of het percentage, p.
  - S2: Bepaal de hoeveelheid variabiliteit in de steekproef in de vorm van de standaardfout van het gemiddelde sx-gem of de standaardfout van het percentage sp.
  - S3: Geef aan hoe groot de steekproef n is.
  - S4: Stel het gewenste betrouwbaarheidsniveau vast om de waarde voor z te bepalen.
  - S5: Bereken het (95%-)betrouwbaarheidsinterval.

Hypothese-toetsen

- **Hypothese toetsen**: een statistische procedure die wordt gebruikt om **de hypothese op basis van de steekproefinformatie te 'accepteren' of te 'verwerpen'**.

  - **Intuïtieve hypothesetoetsing**: dit gebeurt als iemand iets dat hij heeft waargenomen gebruikt om te kijken of het overeenstemt met zijn idee over dat onderwerp of er juist tegen ingaat.
    - Basis: heuristiek, anekdotiek, experiment etc..;

Hypothese-toetsen - 2

- **Statistische hypothesen toetsen**:
  - Begin met een bewering over een populatiekenmerk.
  - Trek **een aselecte steekproef** en bepaal **de steekproefgrootheid**.
  - Vergelijk deze steekproefuitkomst met de parameter in je hypothese (cf. info over het populatiekenmerk).
  - Beslis of de steekproef de oorspronkelijke hypothese ondersteunt.
  - Als de steekproef de hypothese niet ondersteunt, herzie dan je hypothese, zodat zij in overeenstemming is met de steekproefgrootheid.

Hypothese-toetsen - 3

- **Niet-gerichte hypothese**: geeft niet de richting aan waarin de populatieparameter wordt verondersteld te vallen vergeleken met een gemiddelde of percentage.

- **Richtingshypothese**: geeft de richting aan waarin de populatieparameter wordt verondersteld te vallen vergeleken met een gemiddelde of percentage.

Toetsen van de nulhypothese

- Ho = test van gelijkheid aan
  - een bepaalde waarde (x)
  - een bepaald gemiddelde van een groep ($\bar{x} = \frac{\sum x_i \cdot f_i}{n}$)
  - een normaalverdeling ($Z = \frac{X - \mu_X}{\sigma_X} \sim N(0,1)$)

- Ha(lternatieve) = test op de afwijking van de Ho

- Tweezijdig testen : zowel > als <
- Eenzijdig testen: of > of <

Toeststatistiek

- p of de overschrijdingskans bij B.I. (meeste 95%) aan een significantieniveau $\alpha = .05$ (meestal 5%)

  $p < .05$ (voor kleine steekproeven)
  $p < .01$ (voor grote steekproeven)
  $p < .001$ (voor zeer grote steekproeven)

Evaluatie

- Ho klopt en wordt weerhouden als $p > \alpha$
- Ho wordt verworpen, en Ha weerhouden als $p < \alpha$

- Twee mogelijke fouten:
  - Type I fout: een juiste Ho verwerpen
  - Type II fout: een foute Ho weerhouden (aanvaarden)

Deel 8 bis

Vergelijken van twee of meer groepen

Waneer verschillen tussen groepen belangrijk zijn

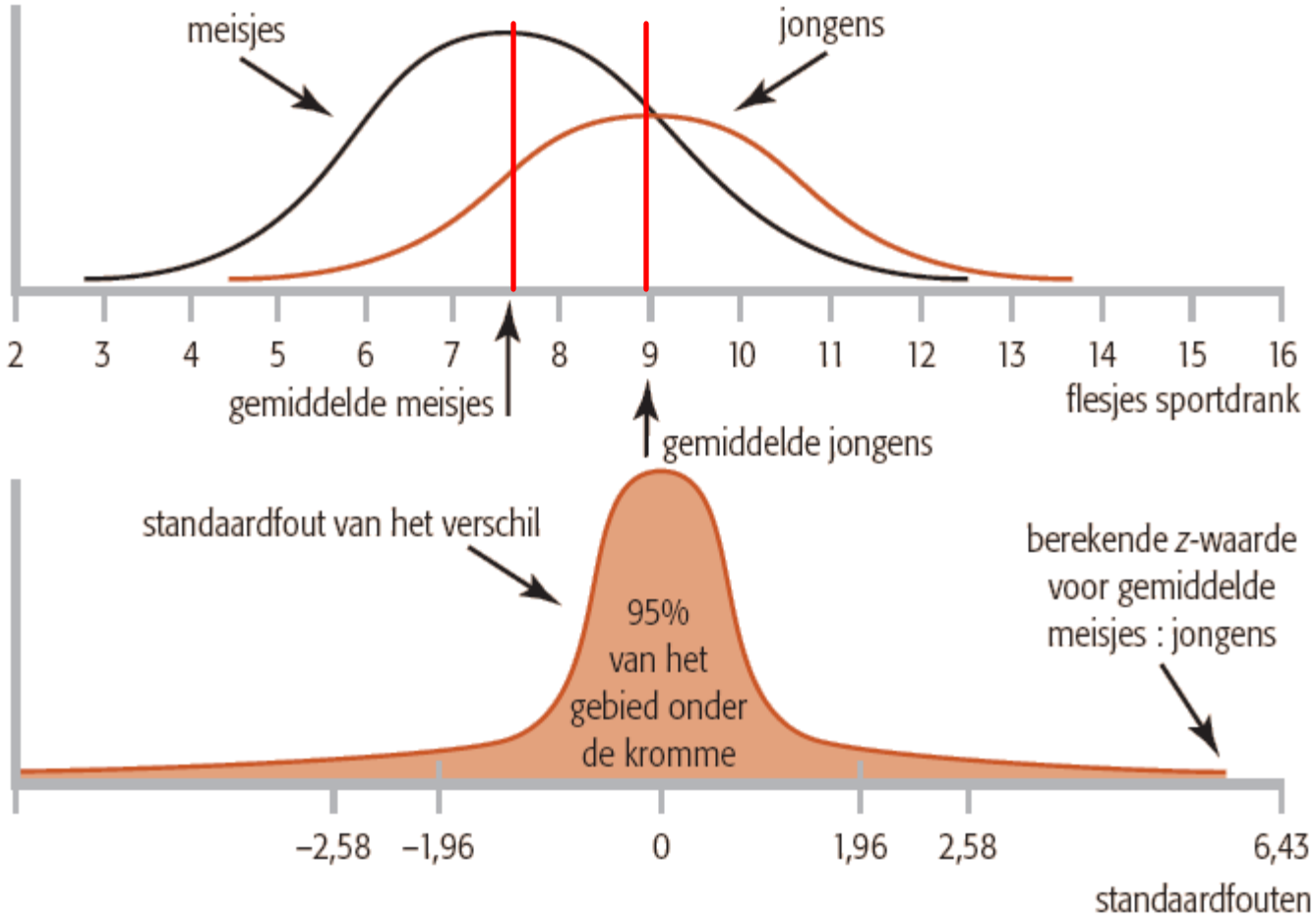
- Er moeten **aantoonbare verschillen** tussen de groepen bestaan.
  - De verschillen moeten (statistisch) **significant** of betekenisvol zijn.
  - De verschillen moeten zo **groot** zijn dat de onderzoeker er afzonderlijke doelgroepen of segmenten van kan maken.
    - °Deelgroepenonderzoek
    - Aparte behandeling van verschillen
  - De verschillen moeten **stabiel** zijn (=/= tijdelijk)

Wanneer is een verschil tussen groepen significant?

- De naam van die kritieke steekproefgrootheid verandert afhankelijk van de analyse en de onderliggende veronderstellingen, maar **meestal wordt de steekproefgrootheid aangeduid met een letter: z, t, F of iets vergelijkbaars.**

Toetsen op significante verschillen tussen twee groepen: percentages of gemiddelden?

- De formules bij de vergelijking van percentages of gemiddelden verschillen.
  - Bij een categorische schaal(nominaal en ordinaal) moeten er **percentages** worden vergeleken.
  - Bij een metrische schaal (interval en ratio) worden **gemiddelden** vergeleken.


meisjes
jongens
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
gemiddelde meisjes
gemiddelde jongens
flesjes sportdrank
standaardfout van het verschil
95%
van het
gebied onder
de kromme
berekende z-waarde
voor gemiddelde
meisjes : jongens
−2,58
−1,96
0
1,96
2,58
6,43
standaardfouten

(1) Verschillen tussen percentages bij twee groepen

- **Nulhypothese (Ho)**: de hypothese dat het verschil tussen de populatieparameters gelijk aan nul is.
- **Alternatieve hypothese**: er bestaat een reëel verschil tussen beide.
- Toetsen of er een echt verschil bestaat tussen de percentages van twee groepen.
  - Je moet eerst een 'vergelijking' tussen de twee percentages maken.
  - Dan wordt het verschil in **een aantal standaardfouten** verwijderd van de nulwaarde waarvan de hypothese uitgaat.
  - Als je eenmaal het aantal standaardfouten weet, levert **kennis van het gebied onder de normale verdeling** een beoordeling op van de kans dat de nulhypothese wordt ondersteund.

formule voor de toets op een significant verschil tussen twee percentages

$$z = \frac{p_1 - p_2}{s_{p_1 - p_2}}$$

waarbij

p_1 = percentage in steekproef 1
p_2 = percentage in steekproef 2
$s_{p_1 - p_2}$ = standaardfout van het verschil tussen twee percentages

Standaardfout vaak 5%

Een voorbeeld

- Uit de opiniepeiling van verleden jaar bij 300 parlementsleden bleek dat 40 procent aan de universiteit had gestudeerd, terwijl uit de peiling van dit jaar onder 100 parlementsleden blijkt dat 65 procent aan de universiteit heeft gestudeerd. (en de resp. rest dus niet)

- **Is het verschil tussen beide steekproeven significant?**

- Als je de formule toepast: P1 = 65(%), P2 = 40(%),
  - sP1-P2 = 5,54
  - z = 4,51
- Omdat de **z-waarde** groter is dan 1,96 is het verschil significant

SPSS gebruiken

- Net als de meeste andere statistische analyseprogramma's toetst SPSS niet of het verschil tussen de percentages van twee groepen statistisch significant is.
- Je kunt SPSS **wel gebruiken om het steekproefpercentage te bepalen van de variabele die je interesseert en de bijbehorende steekproefomvang.**

- Herhaal deze beschrijvende analyse voor de andere steekproef en je hebt alle waarden die je nodig hebt (p1, p2, n1, n2) om de berekeningen handmatig of met een spreadsheetprogramma uit te voeren.

(2) Verschillen tussen gemiddelden bij twee groepen

- De procedure om te toetsen of het verschil tussen twee gemiddelden significant is van twee verschillende groepen (steekproeven) is dezelfde procedure die wordt gebruikt bij het toetsen van twee percentages.
- Het is echter niet zo moeilijk te raden dat de vergelijkingen er anders uitzien omdat het nu om een metrische schaal gaat.
- Let op: Alleen te gebruiken *bij grote(re) steekproeven* (>30)

formule voor de toets op een significant verschil tussen twee gemiddelden

$$z = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

T-toets

- De **t-toets** is de *parametrische* statistische inferentietoets die wordt gebruikt **bij kleine steekproeven (n ≤ 30)**.
  - Reden: de t-toets vertrouwt niet op de normale verdeling maar op de **t-verdeling** of zogenoemde *Student-verdeling*. (genoemd naar William Sealy Gosset, die *Student* als pseudoniem droeg bij een publicatie over kwaliteitscontrole bij bierbrouwer Guinness)
  - De test wordt gebruikt om :
    - Na te gaan of het populatiegemiddelde (van een normaal verdeelde grootheid, bv. leeftijd of punten voor een opleidingsonderdeel) afwijkt van een bepaalde waarde
    - Na te gaan of **twee gemiddelden** van twee groepen in de populatie van elkaar verschillen (via vergelijking van de gemiddelden).

- Echter, de (soort) t-toets is afhankelijk van of de twee groepen "*onafhankelijk"* (impaired) zijn of "*gepaard*" (paired).
  - Als de twee groepen verschillen, bijvoorbeeld Belgen en Fransen, dan gebruik je **de onafhankelijke t-test**.
  - Als de twee groepen uit dezelfde steekproef , bijvoorbeeld mannen en vrouwen, komen dan gebruik je **de gepaarde t-test**.
- De 'Sig.'- of p-waarde is de vlag.

Opvragen in SPSS

- Via **Analyze** > **Compare Means**
  - Klik **One-sample t-test**
  - De **Independent Samples t-test**
  - Of klik **Paired Samples t-test**

- De vraag is "*Verschillen de onderzochte gemiddelden significant van elkaar?*"
  - **Nulhypothese van gelijkheid**: dwz. dat de beide gemiddelde aan elkaar gelijk zijn

- Oefening: vergelijken van gemiddelde van VAR *Vormingsniveau* obv. VAR *Geslacht* via *Independent Samples t-test*

*Output2 [Document2] - PASW Statistics Viewer

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

Output
- Log
- T-Test
  - Title
  - Notes
  - Active Dataset
  - Paired Samples S
  - Paired Samples C
  - Paired Samples T
- Log
- T-Test
  - Title
  - Notes
  - Active Dataset
  - Paired Samples S
  - Paired Samples C
  - Paired Samples T
- Log
- T-Test
  - Title
  - Notes
  - Active Dataset
  - Paired Samples S
  - Paired Samples C
  - Paired Samples T
- Log
- T-Test
  - Title
  - Notes
  - Active Dataset
  - Group Statistics
  - Independent Sam

		N	Correlation	Sig.
Pair 1	Wat is uw geslacht? & Deelnemer	160	-,002	,976

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Wat is uw geslacht? - Deelnemer	-79,200	46,336	3,663	-86,435	-71,965	-21,621	159	,000

```
T-TEST GROUPS=Geslacht(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=Vormingsniveau
  /CRITERIA=CI(.95).
```

T-Test

[DataSet1] K:\Opleidingsonderdelen\3de bachelo... ...7_tris (voorbeeldvragenlijst).sav

Group Statistics

	Wat is uw geslacht?	N			
Som van studiejaar en PC-ervaring	Mannelijk	112			
	Vrouwelijk	48	7,3333	1,26042	,18193

Independent-Samples T Test

Test Variable(s): Som van studiejaar ...

Grouping Variable: Geslacht(1 2)

Define Groups...

Options...

Deelnemer [Case]
Hoe oud bent u? ...
2010 min de opg...
Aan welke faculte...
In welk studiejaar...
Aan welke univer...
Op een schaal va...
Vertrouwen (laag,...
Hoeveel jaar com...
Studiejaar geagg...

OK Paste Reset Cancel Help

Independent-Samples T Test: Opti...

Confidence Interval Percentage: 95 %

Missing Values
- Exclude cases analysis by analysis
- Exclude cases listwise

Continue Cancel Help

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Som van studiejaar en PC-ervaring	Equal variances assumed	9,304	,003	6,388	158	,000	2,38095	,37273	1,64478	3,11712
	Equal variances not assumed			8,099	152,656	,000	2,38095	,29397	1,80019	2,96172

PASW Statistics Processor is ready

H: 3.1, W: 16.4 cr

Inbox in Tom.Scham... | 8.bis Vergelijken van ... | Dataset deel 7_tris (v... | *Output2 [Documen...

NL 14:37

*Output2 [Document2] - PASW Statistics Viewer

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

- Output
  - Log
  - T-Test
    - Title
    - Notes
    - Active Dataset
    - Paired Samples S
    - Paired Samples C
    - Paired Samples T
  - Log
  - T-Test
    - Title
    - Notes
    - Active Dataset
    - Paired Samples S
    - Paired Samples C
    - Paired Samples T
  - Log
  - T-Test
    - Title
    - Notes
    - Active Dataset
    - Paired Samples S
    - Paired Samples C
    - Paired Samples T
  - Log
  - T-Test
    - Title
    - Notes
    - Active Dataset
    - Group Statistics
    - Independent Sam

		N	Correlation	Sig.
Pair 1	Wat is uw geslacht? & Deelnemer	160	-,002	,976

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Wat is uw geslacht? - Deelnemer	-79,200	46,336	3,663	-86,435	-71,965	-21,621	159	,000

```
T-TEST GROUPS=Geslacht(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=Vormingsniveau
  /CRITERIA=CI(.95).
```

T-Test

[DataSet1] K:\Opleidingsonderdelen\3de bachelo… …eel 7_tris (voorbeeldvragenlijst).sav

Group Statistics

	Wat is uw geslacht?	N			
Som van studiejaar en PC-ervaring	Mannelijk	112			
	Vrouwelijk	48	7,3333	1,26042	,18193

Independent-Samples T Test

Test Variable(s): Som van studiejaar …

Grouping Variable: Geslacht(1 2)

Deelnemer [Case]
Hoe oud bent u? …
2010 min de opg…
Aan welke faculte…
In welk studiejaar…
Aan welke univer…
Op een schaal va…
Vertrouwen (laag,…
Hoeveel jaar com…
Studiejaar geagg…

Options… Define Groups… OK Paste Reset Cancel Help

Define Groups

Use specified values
Group 1: 1
Group 2: 2
Cut point:
Continue Cancel Help

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Som van studiejaar en PC-ervaring	Equal variances assumed	9,304	,003	6,388	158	,000	2,38095	,37273	1,64478	3,11712
	Equal variances not assumed			8,099	152,656	,000	2,38095	,29397	1,80019	2,96172

PASW Statistics Processor is ready H: 3.1, W: 16.4 cm

Inbox in Tom.Scham… 8.bis Vergelijken van … Dataset deel 7_tris (v… *Output2 [Documen… NL 14:37

Analyse output

- Presentatie gemiddelde

  VAR1 en VAR2

Group Statistics

	Wat is uw geslacht?	N	Mean	Std. Deviation	Std. Error Mean
Som van studiejaar en PC-ervaring	Mannelijk	112	9,7143	2,44370	,23091
	Vrouwelijk	48	7,3333	1,26042	,18193

- T-Test coëfficiënt

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Som van studiejaar en PC-ervaring	Equal variances assumed (H0)	9,304	,003	6,388	158	,000	2,38095	,37273	1,64478	3,11712
	Equal variances not assumed (Ha)			8,099	152,656	,000	2,38095	,29397	1,80019	2,96172

Conclusie: aan het ingestelde 95% B.I., wordt de **Ho van gelijkheid** van de varianties voor beide subgroepen (cf. *Levene's test*) en voor de gemiddelden *(t-test for equality of means*) van beide subgroepen **verworpen**.

Dus: op basis van het gemiddelde, verschillen de subgroepen mannen en vrouwen, en zijn de verschillen significant (niet toevallig).

Variantie-analyse

- **Variantieanalyse (ANOVA)**: moet bij meervoudige vergelijkingen worden gebruikt

  - Significant: als tussen **gemiddelden van minstens 3 (sub)groepen** een relevant verschil bestaat
  - De 'Sig.'- of p-waarde is de vlag.
  - Als de vlag wappert, is het gerechtvaardigd dat de onderzoeker vervolgens naar elk paar gemiddelden kijkt om te bepalen welk ervan significant van elkaar verschilt
    - Met **post hoc-toets van Duncan** kun je vaststellen waar het paar (de paren) statistisch significante verschillen tussen de gemiddelden zit(ten).

- Soorten:
  - **One Way ANOVA (een-weg-variantieanalyse) ***
  - Two way ANOVA

- Voorbeeld: We vragen ons af of er tussen **drie verschillende groepen** wat de lichaamslengte van de personen uit die groepen betreft, systematische verschillen zijn of dat eventuele verschillen zuiver op toeval berusten. We vergelijken Friezen, Hollanders en Limburgers. Is de lichaamslengte in deze groepen gemiddeld genomen dezelfde, of zijn er systematische verschillen?
  - Duidelijk is dat binnen elke groep verschillen in lengte zijn. Niet alle Hollanders zijn even lang en ook niet alle Friezen. De vraag is of er ook tussen de groepen verschillen zijn. Of de verschillende groepen een bron van variatie zijn.
  - Daartoe worden steekproeven genomen en **de totale "variantie",** die een maat is voor de variatie, uiteengelegd, geanalyseerd, in twee componenten, de variantie binnen de groepen en de variantie tussen de groepen.
  - Het bovenstaande is een voorbeeld van een eenweg-variantie-analyse.
    - Er is sprake van één factor (de groep waartoe iemand behoort: Fries, Hollander of Limburger)

- Samengevat, wat ANOVA onderzoekt:
  - SSBetween: Variantie tussen elke groep (mannen vs vrouwen)
  - SSWithin: Variantie in elke groep
  - SSTotal: Totale variantie van een variabele

F stijgt naarmate SSB(x,y)/SSW(x,y) stijgt !!

$$F = \frac{S_X^2}{S_Y^2}$$

= steekproefvariantie VARx
= steekproefvariantie VARy

Opvragen in SPSS

- Via **Analyze** > **Compare Means** > **One-Way ANOVA**
  - klik *Options* > *Descriptives (Statistics)*
  - klik *Post-hoc Multiple comparisons (Equal variances assumed)* > *Bonferroni*

- **Nulhypothese van gelijkheid**: dwz. dat de beide gemiddelde aan elkaar gelijk zijn

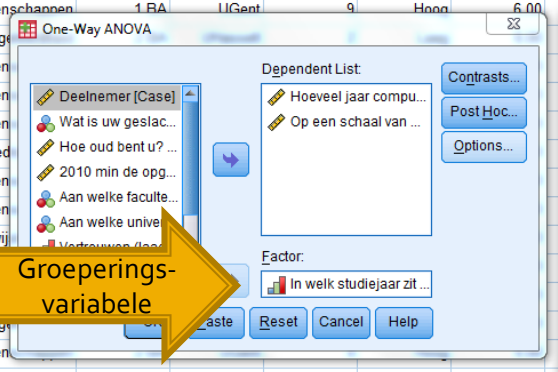
- Oefening: vergelijken van gemiddelde van VAR ***PC-ervaring*** en ***Vertrouwen*** obv. VAR ***Studiejaar***

Dataset deel 7_tris (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor

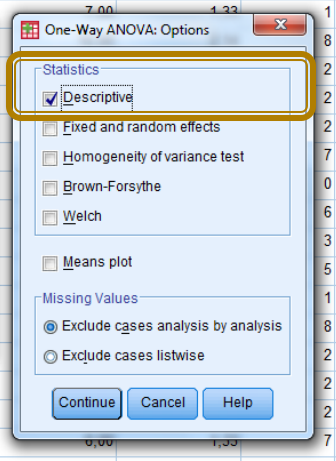
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Visible: 16 of 16 Variables

	Case	Geslacht	Leeftijd	Geboortejaar	Faculteit	Studiejaar	Instelling	Vertrouwen	VertrouwenLMH	PC_ervaring	Vormingsniveau	Studiejaar_mean	Wantrouwen	Vertrouwen_voldoende	Bachelorstudent	Vorn
1	1	Mannelijk	23,00	1987	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00	14,00	2,14	2	Voldoende vertrouwen	Master	
2	2	Mannelijk	22,00	1988	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00	10,00	2,14	2	Voldoende vertrouwen	Bachelor	
3	3	Vrouwelijk	20,00	1990	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00	6,00	1,33	7	Onvoldoende vertrouwen	Bachelor	
4	4	Mannelijk	20,00	1990	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00	7,00	2,14	0	Voldoende vertrouwen	Bachelor	
5	5	Vrouwelijk	21,00	1989	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00	9,00	1,33	6	Onvoldoende vertrouwen	Bachelor	
6	6	Mannelijk	20,00	1990	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00	6,00	2,14	3	Voldoende vertrouwen	Bachelor	
7	7	Mannelijk	22,00	1988	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00	10,00	2,14	5	Onvoldoende vertrouwen	Bachelor	
8	8	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00	7,00	1,33	1	Voldoende vertrouwen	Bachelor	
9	9	Mannelijk	21,00	1989	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00	10,00	2,14	8	Onvoldoende vertrouwen	Bachelor	
10	10	Mannelijk	22,00	1988	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00	11,00	2,14	2	Voldoende vertrouwen	Bachelor	
11	11	Mannelijk	23,00	1987	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00	14,00	2,14	2	Voldoende vertrouwen	Master	
12	12	Mannelijk	22,00	1988	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00	10,00	2,14	2	Voldoende vertrouwen	Bachelor	
13	13	Vrouwelijk	20,00	1990	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00	6,00	1,33	7	Onvoldoende vertrouwen	Bachelor	
14	14	Mannelijk	20,00	1990	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00	7,00	2,14	0	Voldoende vertrouwen	Bachelor	
15	15	Vrouwelijk	21,00	1989	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00	9,00	1,33	6	Onvoldoende vertrouwen	Bachelor	
16	16	Mannelijk	20,00	1990	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00	6,00	2,14	3	Voldoende vertrouwen	Bachelor	
17	17	Mannelijk	22,00	1988	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00	10,00	2,14	5	Onvoldoende vertrouwen	Bachelor	
18	18	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00	7,00	1,33	1	Voldoende vertrouwen	Bachelor	
19	19	Mannelijk	21,00	1989	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00	10,00	2,14	8	Onvoldoende vertrouwen	Bachelor	
20	20	Mannelijk	22,00	1988	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00	11,00	2,14	2	Voldoende vertrouwen	Bachelor	
21	21	Mannelijk	23,00	1987	Ingenieurswetenschappen	MA	UGent	8	Hoog	10,00	14,00	2,14	2	Voldoende vertrouwen	Master	
22	22	Mannelijk	22,00	1988	Wetenschappen	2 BA	VUBrussel	8	Hoog	8,00	10,00	2,14	2	Voldoende vertrouwen	Bachelor	
23	23	Vrouwelijk	20,00	1990	Economie en Bedrijfskunde	1 BA	UAntwerpen	3	Laag	5,00	6,00	1,33	7	Onvoldoende vertrouwen	Bachelor	
24	24	Mannelijk	20,00	1990	Politieke en Sociale Wetenschappen	1 BA	UGent	10	Hoog	6,00	7,00	2,14	0	Voldoende vertrouwen	Bachelor	
25	25	Vrouwelijk	21,00	1989	Wetenschappen	2 BA	UAntwerpen	4	Midden	7,00	9,00	1,33	6	Onvoldoende vertrouwen	Bachelor	
26	26	Mannelijk	20,00	1990	Letteren en wijsbegeerte	1 BA	KUBrussel	7	Hoog	5,00	6,00	2,14	3	Voldoende vertrouwen	Bachelor	
27	27	Mannelijk	22,00	1988	Diergeneeskunde	3 BA	KULeuven	5	Midden	7,00	10,00	2,14	5	Onvoldoende vertrouwen	Bachelor	
28	28	Vrouwelijk	20,00	1990	Geneeskunde en gezondheidswetenschappen	1 BA	UGent	9	Hoog	6,00	7,00	1,33	1	Voldoende vertrouwen	Bachelor	
29	29	Mannelijk	21,00	1989	Rechtsgeleerdheid	2 BA	UHasselt	2	Laag	8,00	10,00	2,14	8	Onvoldoende vertrouwen	Bachelor	
30	30	Mannelijk	22,00	1988	Farmaceutische Wetenschappen	2 BA	UGent	8	Hoog	9,00	11,00	2,14	2	Voldoende vertrouwen	Bachelor	



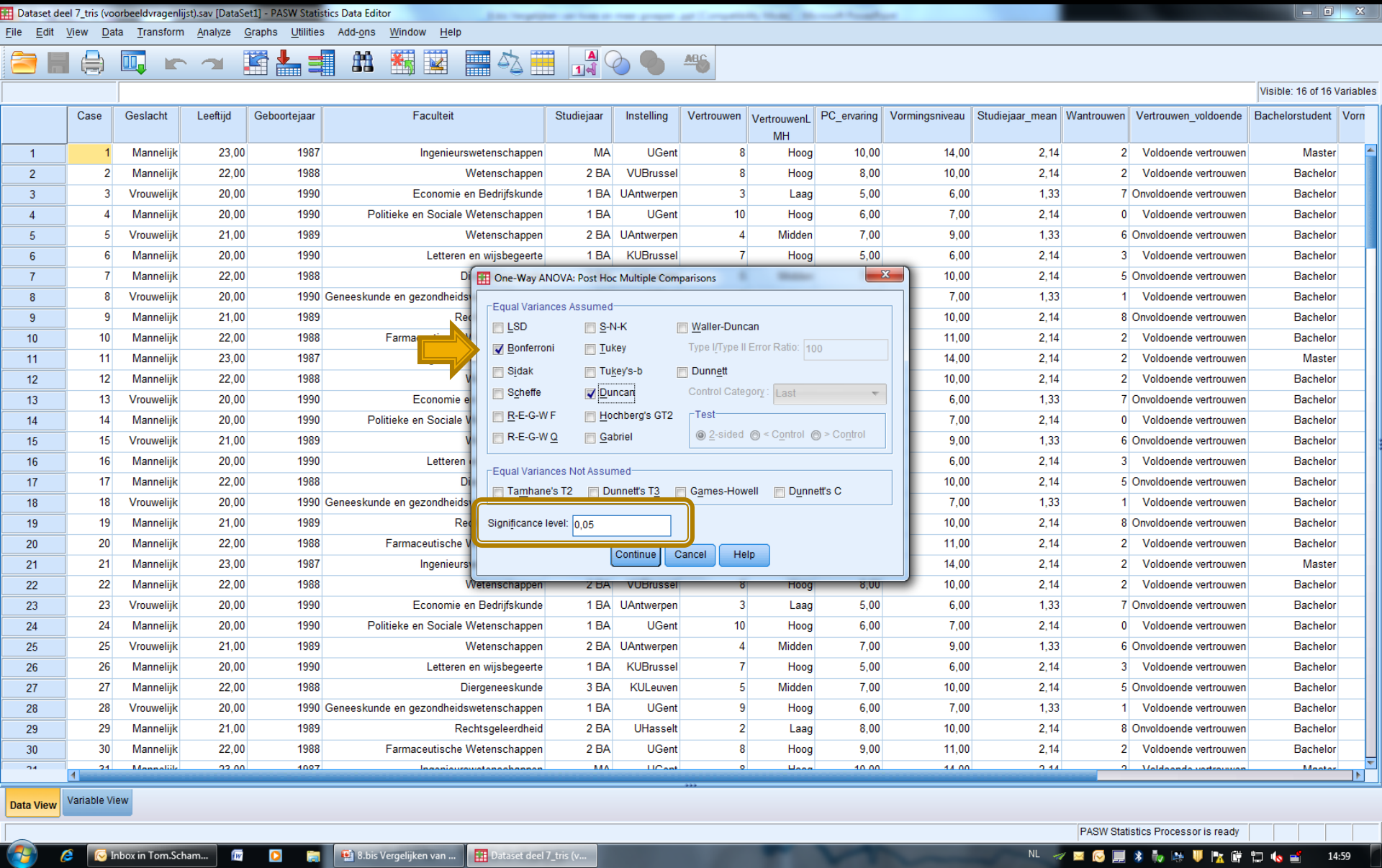
Groeperings-variabele



Data View | Variable View

PASW Statistics Processor is ready

Variantie-analyse - ... | Inbox in Tom.Scham... | 8.bis Vergelijken van ... | *Output1 [Documen... | Dataset deel 7_tris (v... | NL | 15:55


Dataset deel 7_tris (voorbeeldvragenlijst).sav [DataSet1] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Visible: 16 of 16 Variables
Case Geslacht Leeftijd Geboortejaar Faculteit Studiejaar Instelling Vertrouwen VertrouwenLMH PC_ervaring Vormingsniveau Studiejaar_mean Wantrouwen Vertrouwen_voldoende Bachelorstudent
1 1 Mannelijk 23,00 1987 Ingenieurswetenschappen MA UGent 8 Hoog 10,00 14,00 2,14 2 Voldoende vertrouwen Master
2 2 Mannelijk 22,00 1988 Wetenschappen 2 BA VUBrussel 8 Hoog 8,00 10,00 2,14 2 Voldoende vertrouwen Bachelor
3 3 Vrouwelijk 20,00 1990 Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 5,00 6,00 1,33 7 Onvoldoende vertrouwen Bachelor
4 4 Mannelijk 20,00 1990 Politieke en Sociale Wetenschappen 1 BA UGent 10 Hoog 6,00 7,00 2,14 0 Voldoende vertrouwen Bachelor
5 5 Vrouwelijk 21,00 1989 Wetenschappen 2 BA UAntwerpen 4 Midden 7,00 9,00 1,33 6 Onvoldoende vertrouwen Bachelor
6 6 Mannelijk 20,00 1990 Letteren en wijsbegeerte 1 BA KUBrussel 7 Hoog 5,00 6,00 2,14 3 Voldoende vertrouwen Bachelor
7 7 Mannelijk 22,00 1988 10,00 2,14 5 Onvoldoende vertrouwen Bachelor
8 8 Vrouwelijk 20,00 1990 Geneeskunde en gezondheids 7,00 1,33 1 Voldoende vertrouwen Bachelor
9 9 Mannelijk 21,00 1989 10,00 2,14 8 Onvoldoende vertrouwen Bachelor
10 10 Mannelijk 22,00 1988 11,00 2,14 2 Voldoende vertrouwen Bachelor
11 11 Mannelijk 23,00 1987 14,00 2,14 2 Voldoende vertrouwen Master
12 12 Mannelijk 22,00 1988 10,00 2,14 2 Voldoende vertrouwen Bachelor
13 13 Vrouwelijk 20,00 1990 6,00 1,33 7 Onvoldoende vertrouwen Bachelor
14 14 Mannelijk 20,00 1990 7,00 2,14 0 Voldoende vertrouwen Bachelor
15 15 Vrouwelijk 21,00 1989 9,00 1,33 6 Onvoldoende vertrouwen Bachelor
16 16 Mannelijk 20,00 1990 6,00 2,14 3 Voldoende vertrouwen Bachelor
17 17 Mannelijk 22,00 1988 10,00 2,14 5 Onvoldoende vertrouwen Bachelor
18 18 Vrouwelijk 20,00 1990 Geneeskunde en gezondheids 7,00 1,33 1 Voldoende vertrouwen Bachelor
19 19 Mannelijk 21,00 1989 10,00 2,14 8 Onvoldoende vertrouwen Bachelor
20 20 Mannelijk 22,00 1988 Farmaceutische 11,00 2,14 2 Voldoende vertrouwen Bachelor
21 21 Mannelijk 23,00 1987 14,00 2,14 2 Voldoende vertrouwen Master
22 22 Mannelijk 22,00 1988 10,00 2,14 2 Voldoende vertrouwen Bachelor
23 23 Vrouwelijk 20,00 1990 Economie en Bedrijfskunde 1 BA UAntwerpen 3 Laag 5,00 6,00 1,33 7 Onvoldoende vertrouwen Bachelor
24 24 Mannelijk 20,00 1990 Politieke en Sociale Wetenschappen 1 BA UGent 10 Hoog 6,00 7,00 2,14 0 Voldoende vertrouwen Bachelor
25 25 Vrouwelijk 21,00 1989 Wetenschappen 2 BA UAntwerpen 4 Midden 7,00 9,00 1,33 6 Onvoldoende vertrouwen Bachelor
26 26 Mannelijk 20,00 1990 Letteren en wijsbegeerte 1 BA KUBrussel 7 Hoog 5,00 6,00 2,14 3 Voldoende vertrouwen Bachelor
27 27 Mannelijk 22,00 1988 Diergeneeskunde 3 BA KULeuven 5 Midden 7,00 10,00 2,14 5 Onvoldoende vertrouwen Bachelor
28 28 Vrouwelijk 20,00 1990 Geneeskunde en gezondheidswetenschappen 1 BA UGent 9 Hoog 6,00 7,00 1,33 1 Voldoende vertrouwen Bachelor
29 29 Mannelijk 21,00 1989 Rechtsgeleerdheid 2 BA UHasselt 2 Laag 8,00 10,00 2,14 8 Onvoldoende vertrouwen Bachelor
30 30 Mannelijk 22,00 1988 Farmaceutische Wetenschappen 2 BA UGent 8 Hoog 9,00 11,00 2,14 2 Voldoende vertrouwen Bachelor
One-Way ANOVA: Post Hoc Multiple Comparisons
Equal Variances Assumed
LSD S-N-K Waller-Duncan
Bonferroni Tukey Type I/Type II Error Ratio: 100
Sidak Tukey's-b Dunnett
Scheffe Duncan Control Category: Last
R-E-G-W F Hochberg's GT2 Test
R-E-G-W Q Gabriel 2-sided < Control > Control
Equal Variances Not Assumed
Tamhane's T2 Dunnett's T3 Games-Howell Dunnett's C
Significance level: 0,05
Continue Cancel Help
Data View Variable View
PASW Statistics Processor is ready
Inbox in Tom.Scham...
8.bis Vergelijken van ...
Dataset deel 7_tris (v...
NL
14:59

Analyse van de output

- Descriptives

Analyse van de variantie

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Hoeveel jaar computerervaring heb je?	1 BA	64	5,5000	,50395	,06299	5,3741	5,6259	5,00	6,00
	2 BA	64	8,0000	,71270	,08909	7,8220	8,1780	7,00	9,00
	3 BA	16	7,0000	,00000	,00000	7,0000	7,0000	7,00	7,00
	MA	16	10,0000	,00000	,00000	10,0000	10,0000	10,00	10,00
	Total	160	7,1000	1,58293	,12514	6,8528	7,3472	5,00	10,00
Op een schaal van 1 tot 10 met 10 het meest, hoeveel vertrouwen heb je erin dat je SPSS zal leren?	1 BA	64	7,25	2,702	,338	6,58	7,92	3	10
	2 BA	64	5,50	2,619	,327	4,85	6,15	2	8
	3 BA	16	5,00	,000	,000	5,00	5,00	5	5
	MA	16	8,00	,000	,000	8,00	8,00	8	8
	Total	160	6,40	2,585	,204	6,00	6,80	2	10

ANOVA (F-toets)

- Aan het ingestelde 95% B.I., wordt de **Ho van gelijkheid** van de varianties van beide subgroepen **verworpen**.
- Dus: op basis van de gemiddelden verschillen de subgroepen mannen en vrouwen significant van elkaar (niet toevallig).

Tussenvariantie en binnenvariantie

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Hoeveel jaar computerervaring heb je?	Between Groups	350,400	3	116,800	379,600	,000
	Within Groups	48,000	156	,308		
	Total	398,400	159			
Op een schaal van 1 tot 10 met 10 het meest, hoeveel vertrouwen heb je erin dat je SPSS zal leren?	Between Groups	170,400	3	56,800	9,934	,000
	Within Groups	892,000	156	5,718		
	Total	1062,400	159			

Er zijn verschillen, maar welke ?

Zie **Bonferroni test**

Bonferroni test

Multiple Comparisons

Dependent Variable		(I) In welk studiejaar zit u?	(J) In welk studiejaar zit u?	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Hoeveel jaar computerervaring heb je?	Bonferroni	1 BA	2 BA	-2,50000*	,09806	,000	-2,7620	-2,2380
			3 BA	-1,50000*	,15504	,000	-1,9143	-1,0857
			MA	-4,50000*	,15504	,000	-4,9143	-4,0857
		2 BA	1 BA	2,50000*	,09806	,000	2,2380	2,7620
			3 BA	1,00000*	,15504	,000	,5857	1,4143
			MA	-2,00000*	,15504	,000	-2,4143	-1,5857
		3 BA	1 BA	1,50000*	,15504	,000	1,0857	1,9143
			2 BA	-1,00000*	,15504	,000	-1,4143	-,5857
			MA	-3,00000*	,19612	,000	-3,5241	-2,4759
		MA	1 BA	4,50000*	,15504	,000	4,0857	4,9143
			2 BA	2,00000*	,15504	,000	1,5857	2,4143
			3 BA	3,00000*	,19612	,000	2,4759	3,5241
Op een schaal van 1 tot 10 met 10 het meest, hoeveel vertrouwen heb je erin dat je SPSS zal leren?	Bonferroni	1 BA	2 BA	1,750*	,423	,000	,62	2,88
			3 BA	2,250*	,668	,006	,46	4,04
			MA	-,750	,668	1,000		
		2 BA	1 BA	-1,750*	,423	,000		
			3 BA	-,500	,668	1,000	-1,29	2,29
			MA	-2,500*	,668	,002	-4,29	-,71
		3 BA	1 BA	-2,250*	,668	,006	-4,04	-,46
			2 BA	-,500	,668	1,000	-2,29	1,29
			MA	-3,000*	,845	,003	-5,26	-,74
		MA	1 BA	,750	,668	1,000	-1,04	2,54
			2 BA	2,500*	,668	,002	,71	4,29
			3 BA	3,000*	,845	,003	,74	5,26

*. The mean difference is significant at the 0.05 level.

Tests op basis van paarsgewijze vergelijkingen tussen categorieën om te achterhalen of er afwijkingen in de verschillen zijn

Bonferroni test enkel gebruiken bij een beperkt aantal categorieën: 4 of 5.

Doorstreept = niet significante verschillen !

Deel 10

Verklarende statistiek – een inleiding

- Verklaren
- Voorspellen
- Projecties
- Modellen

- **Voorspelling**: een uitspraak over wat er naar verwachting in de toekomst zal gebeuren op basis van ervaringen uit het verleden of voorafgaande observatie.
- **Twee methoden** van voorspellen
  - Extrapolatie: bij extrapolatie gebruik je een ervaring uit het verleden als middel om de toekomst te voorspellen.
  - Voorspellend model: in een voorspellend model zijn de omstandigheden opgenomen die naar verwachting een rol spelen en de factor of variabelen beïnvloeden die je wilt voorspellen.

Hoe toetsen ?

- Ongeacht de voorspellingsmethode wil je altijd beoordelen hoe goed je voorspelling is, ofwel hoe goed je methode of model is in het doen van die voorspellingen.

- Als je vergelijkt hoe ver de voorspelde waarden afliggen van de feitelijke of waargenomen waarden, doe je een analyse van residuen.

Enkelvoudige regressieanalyse

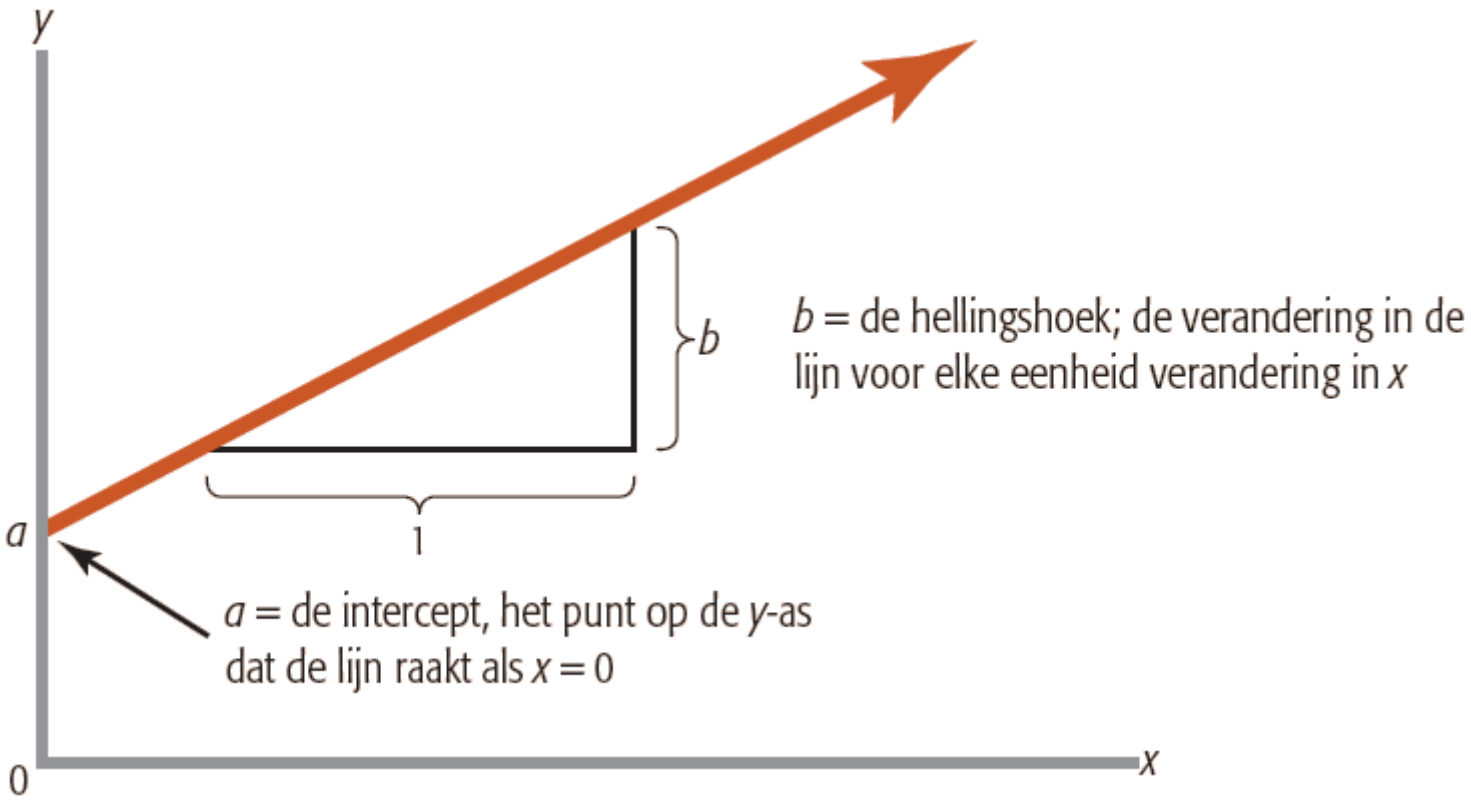
- Een voorspellende analysetechniek waarbij één variabele wordt gebruikt om het niveau van een andere variabele te voorspellen met de formule voor de rechte lijn.

formule voor een rechtlijnig verband $y = a + bx$

waarbij
y = de te voorspellen variabele
x = de variabele die wordt gebruikt om y te voorspellen
a = de **intercept**, ofwel het punt waar de lijn de y-as doorsnijdt $(x = 0)$
b = de **hellingshoek**, ofwel de verandering in y als x één eenheid verandert

Zie ook Deel 7. Eenvoudige correlaties

- Aan **regressie** ligt een rechtlijnig verband ten grondslag en het is een krachtig voorspellend model.


y
b
b = de hellingshoek; de verandering in de lijn voor elke eenheid verandering in x
1
a
a = de intercept, het punt op de y-as dat de lijn raakt als x = 0
0
x

Basisprocedures bij ERA

- **Onafhankelijke variabele**: de variabele die wordt gebruikt om de afhankelijke variabele te voorspellen en deze heet x in de regressieformule.

- **Afhankelijke variabele**: de variabele die wordt voorspeld en deze heet gewoonlijk y in de regressievergelijking voor een rechte lijn.

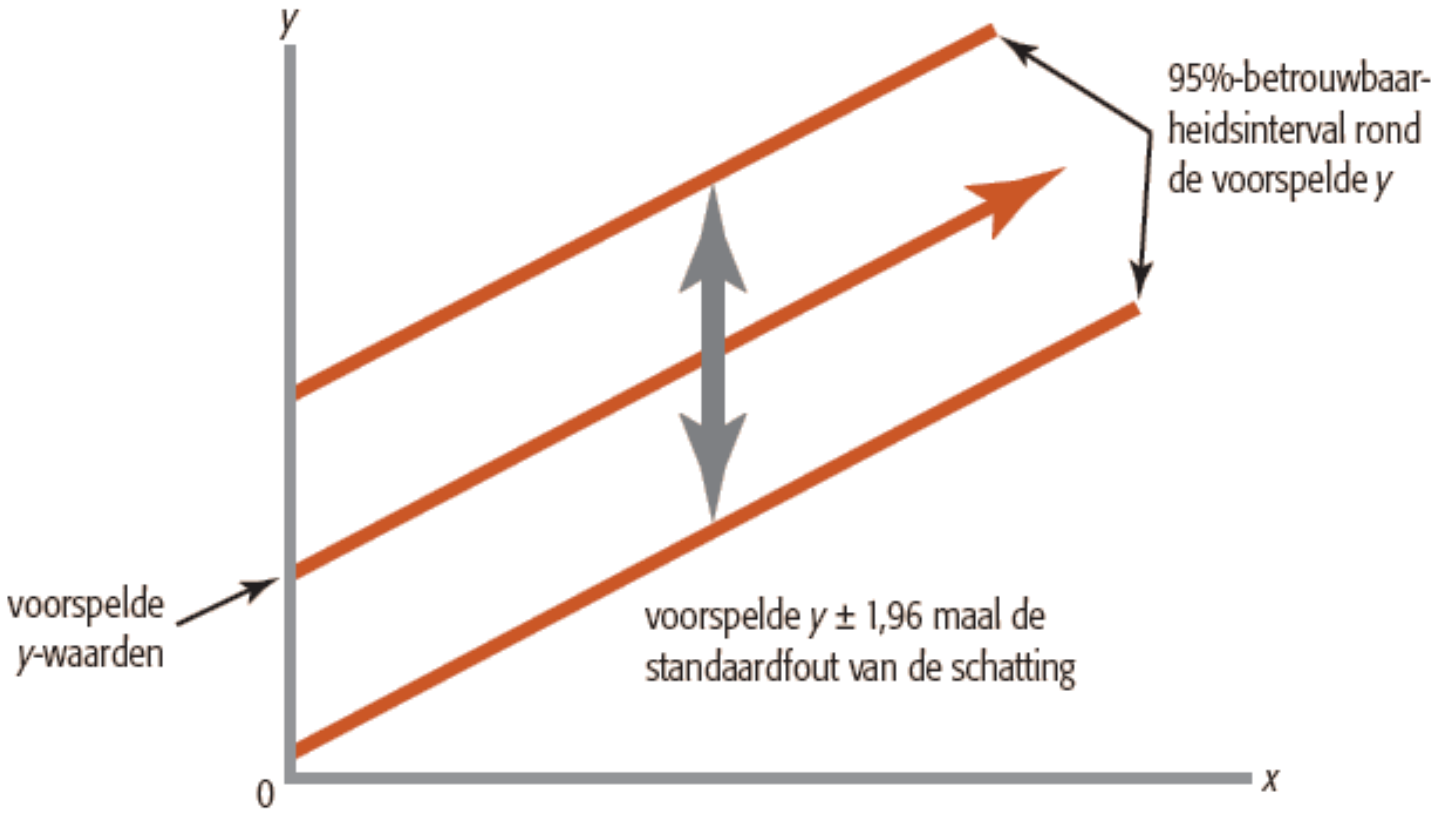
- **Het kleinstekwadratencriterium**: een manier om te garanderen dat de rechte lijn die door de punten in het *spreidingsdiagram* loopt, zo is gepositioneerd dat de verticale afstanden van de verschillende punten tot de lijn zo klein mogelijk zijn.

Basisprocedures bij ERA - 2

- Het is bij regressieanalyse niet voldoende om alleen de waarden voor a en b te berekenen, omdat de twee waarden nog moeten worden getoetst op statistische significantie.

- Het feit dat de lijn (slechts) de beste benadering is van de punten betekent dat we rekening moeten houden met *een bepaalde hoeveelheid fouten* als we de lijn voor onze voorspelling gebruiken.

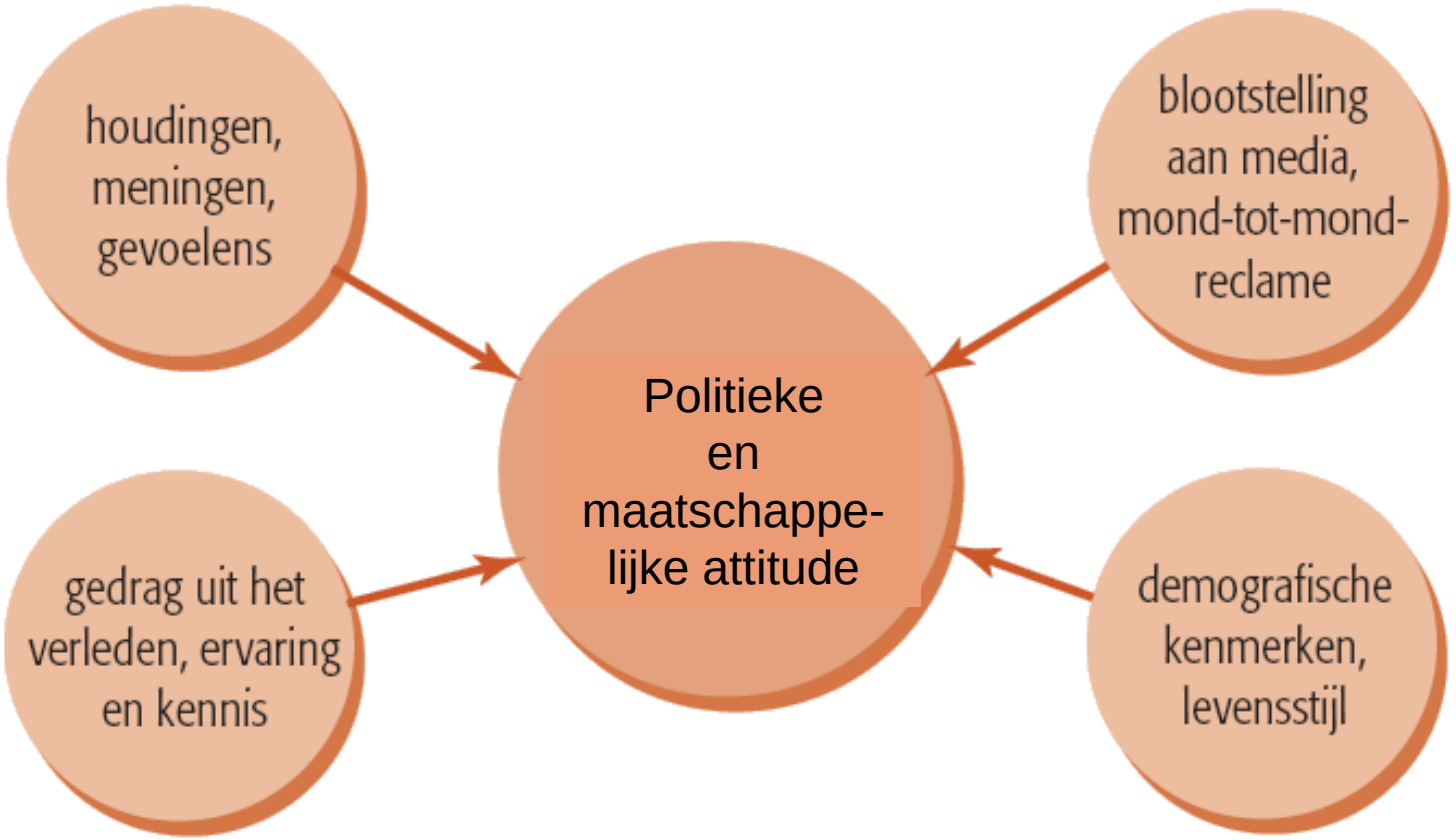
- **Standaardfout van de schatting**: analoog aan de standaardfout van het gemiddelde die je hebt gebruikt om een populatiegemiddelde op basis van een steekproef te schatten.


y
95%-betrouwbaar-
heidsinterval rond
de voorspelde y
voorspelde
y-waarden
voorspelde y ± 1,96 maal de
standaardfout van de schatting
0
x

Meervoudige regressieanalyse

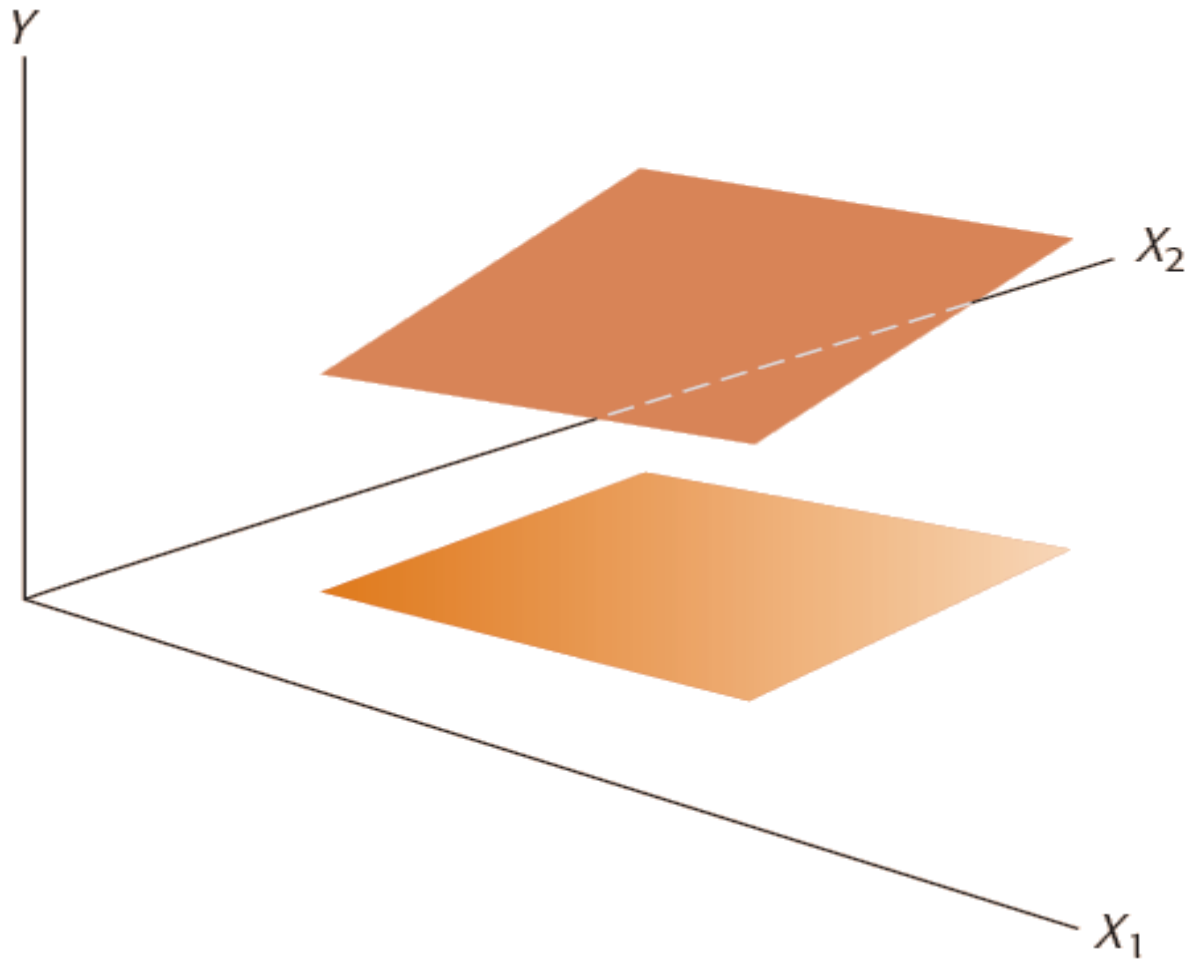
- Meervoudige regressieanalyse is een uitbreiding van enkelvoudige regressieanalyse in zoverre dat er meer dan één onafhankelijke variabele wordt gebruikt in de regressievergelijking.

- **Een onderliggend conceptueel model**: in een algemeen conceptueel model zijn onafhankelijke en afhankelijke variabelen opgenomen waarbij wordt aangegeven hoe ze met elkaar in verband staan.


houdingen, meningen, gevoelens
blootstelling aan media, mond-tot-mond-reclame
Politieke en maatschappe-lijke attitude
gedrag uit het verleden, ervaring en kennis
demografische kenmerken, levensstijl

Basisveronderstellingen bij MRA

- De regressielijn verandert in **een regressie-vlak**.
- De R-waarde, ook wel *de determinatiecoëfficiënt* genoemd, is een handige maat voor de sterkte van het hele lineaire verband.
- **De veronderstelling van onafhankelijkheid**: de onafhankelijke variabelen statistisch onafhankelijk van elkaar moeten zijn (geen correlatie!).
  - **Variance inflation factor** (VIF): zolang de VIF < 10 -> *multicollineariteit* geen probleem is.


Y
X2
X1

meervoudige regressievergelijking $y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots . b_m x_m$

waarbij

y = de afhankelijke of te voorspellen variabele

x_1 = de onafhankelijke variabele i

a = de intercept

b_i = de hellingshoek van de onafhankelijke variabele i

m = het aantal onafhankelijke variabelen in de vergelijking

Speciale toepassingen op MRA

- Een '**dummy**' als onafhankelijke variabele gebruiken
- **Gestandaardiseerde bèta's**: gebruiken om het belang van de onafhankelijke variabelen te vergelijken
- Meervoudige regressieanalyse als screenings-instrument gebruiken

Stapsgewijze MRA

- Wanneer? (***stepwise regression***)
  - Bij MRA als screeningsinstrument
  - Bij een groot aantal onafhankelijke variabelen in het conceptuele model
  - Doel = het aantal onafhankelijke variabelen terug te brengen tot een beheersbaar aantal.
- Wat en hoe?
  - De (ene) onafhankelijke variabele die statistisch significant is en de meeste variantie in de afhankelijke variabele verklaart wordt vastgesteld in de meervoudige regressievergelijking ingevoerd.

Stapsgewijze MRA - 2

- Vervolgens wordt de statistisch significante onafhankelijke variabele die het meest bijdraagt aan de verklaring van de overblijvende onverklaarde variantie in de afhankelijke variabele vastgesteld en in de vergelijking ingevoerd.
- Alle niet-significante onafhankelijken worden geëlimineerd.

Waarschuwingen ivm. MRA

- Regressieanalyse is niets meer dan een statistisch instrument waarbij een lineair verband wordt 'voor'ondersteld tussen twee variabelen.

- Je moet regressieanalyse niet toepassen om dingen te voorspellen die buiten de grenzen van de data liggen die je voor het regressiemodel hebt gebruikt.

Wat hebben we geleerd ?

- **Regressie-analyse uitvoeren**: Analyze, Regression, Lineair, aanduiden van afhankelijke en onafhankelijke variabele, OK
- **Regressie-analyse uitvoeren en determinatiecoëfficiënt opvragen**: Analyze, Regression, Lineair, aanduiden van afhankelijke en onafhankelijke variabele, statistics, aanduiden van 'R square changed', Continue, OK
- **Regressie-analyse uitvoeren en ongestandaardiseerde richtingscoëfficiënt opvragen**: Analyze, Regression, Lineair, aanduiden van afhankelijke en onafhankelijke variabele, statistics, aanduiden van 'Model fit', Continue, OK (in de output verschijnt dan de parameter B die aangeeft hoe stijl de relatie is, hoeveel eenheden we verwachten dat de afhankelijke variabele stijgt als de onafhankelijke variabele met 1 eenheid toeneemt. Uit het teken van B (+ of -) kunnen we ook de richting van de relatie afleiden.)

Wat geleerd ? - 2

- **Regressie-analyse uitvoeren en de constante in de regressievergelijking opvragen**: Analyze, Regression, Lineair, aanduiden van afhankelijke en onafhankelijke variabele, statistics, aanduiden van 'Model fit', Continue, OK (in de output verschijnt dan de waarde van de constante (=A) in de kolom B naast '(Constant)'; de constante in de regressievergelijking geeft aan welke waarde we op de afhankelijke variabele verwachten als de onafhankelijke variabele gelijk is aan 0.)

- **Regressie-analyse uitvoeren en scatterplot met regressielijn opvragen**: Analyze, Regression, Lineair, aanduiden van afhankelijke en onafhankelijke variabele, Plot, aanduiden van optie 'normal probability plot', Continue, OK