

## APPENDIX

### BASISHANDLEIDING BIJ SPSS

---

#### 1. DOELSTELLINGEN

Deze appendix is een eerste inleiding tot het statistische softwareprogramma SPSS en is bedoeld voor eerstejaarsstudenten Criminologische Wetenschappen. De nadruk ligt op het stap voor stap aanleren van de basisprincipes voor het efficiënt en verantwoord uitvoeren van eenvoudige statistische analyses. Deze basishandleiding SPSS geeft geïnteresseerde studenten de mogelijkheid op zelfstandige basis reeds kennis te maken met de meest recente versie van SPSS Statistics 25, de opbouw, werkwijze en mogelijkheden. Na het doorlopen van de instructies zijn studenten instaat om een eigen gegevensbestand op te maken en enkele eenvoudige analyses uit te voeren. In dit hoofdstuk worden ook voorbeelden gegeven hoe SPSS-output dient geïnterpreteerd en gerapporteerd. Echter, een grondiger inleiding volgt later tijdens de opleiding Criminologische Wetenschappen in het opleidingsonderdeel ‘Kwantitatieve Toegepaste data analyse en rapportage’.

#### 2. WAT IS SPSS ?

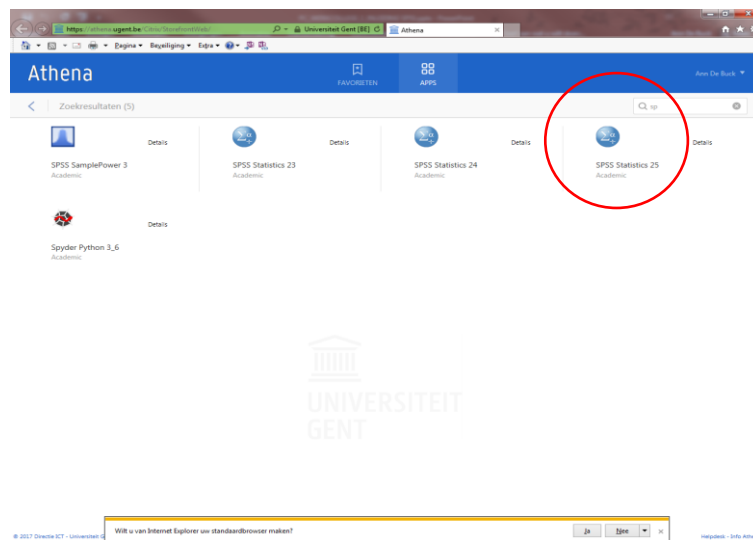
SPSS staat voor *Statistical Package for the Social Sciences* en is een software programma om statistische analyses uit te voeren dat op maat werd geschreven voor gebruik in de sociale wetenschappen. Omwille van het succes ervan in andere wetenschappelijke disciplines en in zowel commerciële als niet-commerciële organisaties wordt nu gewoon van SPSS gesproken. Het is een zeer krachtig programma dat met een hoge snelheid uitgebreide datasets kan analyseren. Op een paar seconden tijd kan SPSS het werk verrichten waarvoor een onderzoeker een tiental jaren terug nog meerdere dagen nodig had.

Van het programma SPSS bestaan verschillende versies bedoeld voor verschillende types van computers en gericht naar verschillende gebruikersgroepen. Wij maken gebruik van SPSS for Windows. De meest recente versie SPSS Statistics 27 is voor UGent studenten beschikbaar via Athena.



### 3. SPSS OPENEN- OPSLAAN VAN GEGEVENS

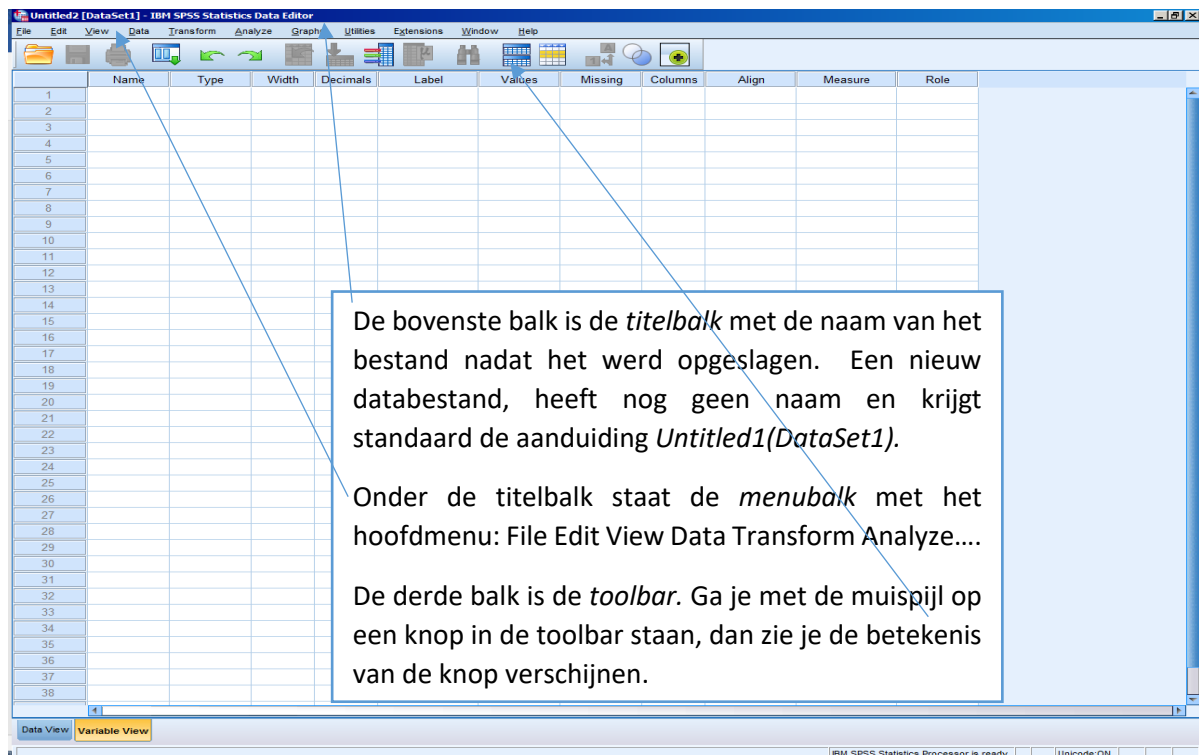
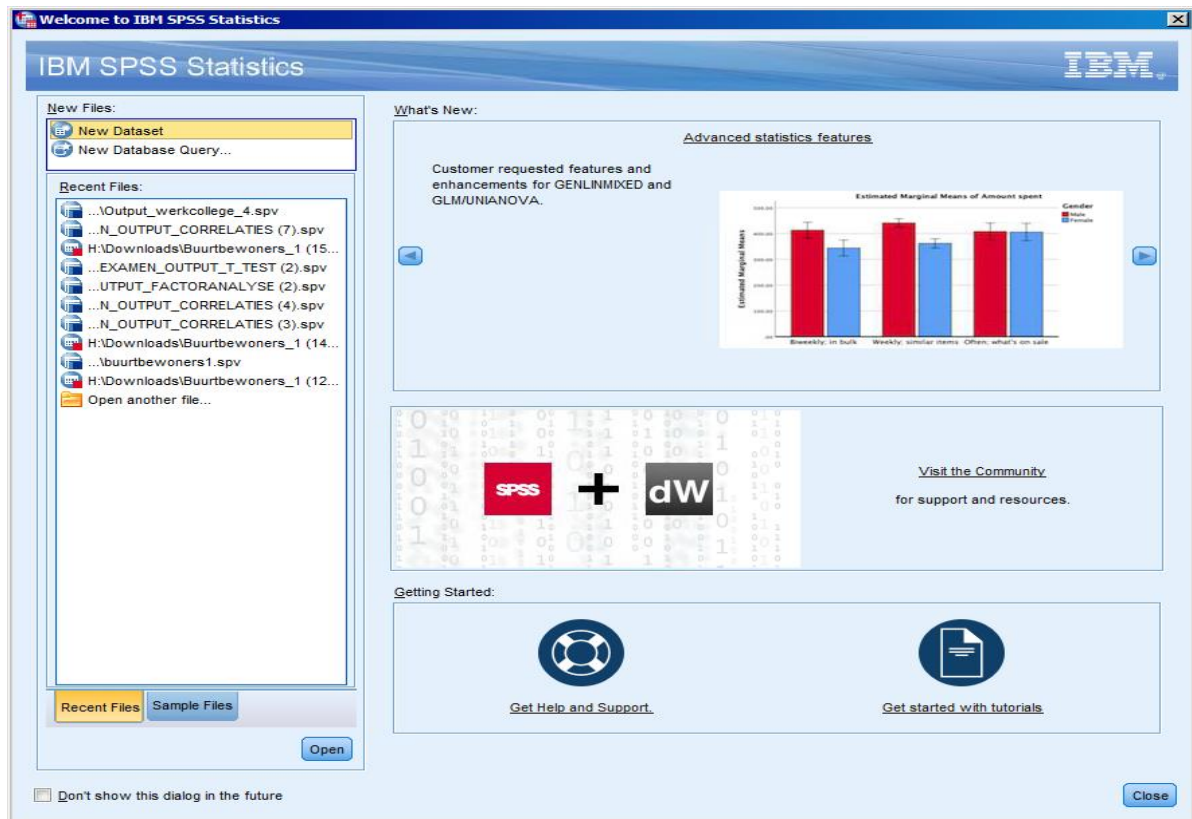
SPSS openen via Athena icoontje SPSS Statistics 27 en het programma start op.



Pop-up verschijnt met advies om data op te slaan als een .zsav bestand. Ok aanvinken.



Standaard toont SPSS de laatst gebruikte bestanden. Een ander bestand opvragen kan via *Open another file*. Een programma openen zonder databestand actief te maken, kan door *New dataset* aan te vinken.



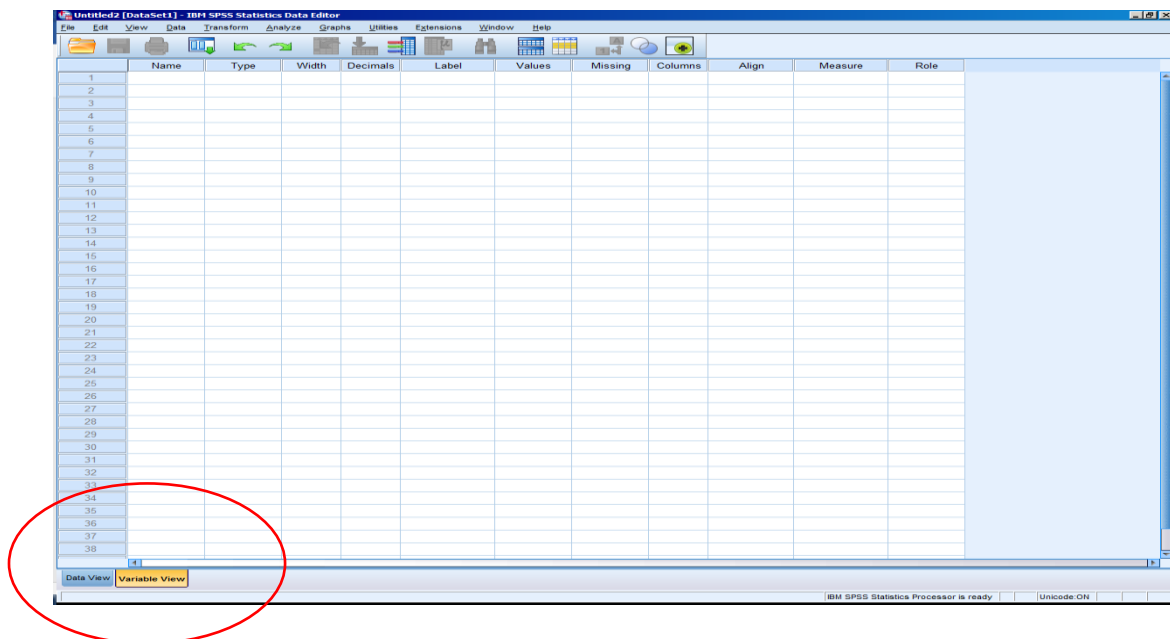
SPSS heeft verschillende *vensters* die gekoppeld zijn aan een apart type bestand dat je afzonderlijk moet openen en afzonderlijk moet opslaan.

Vensters openen: *File/Open/Data*    *File/Open/Syntax*    *File/Open/Output*

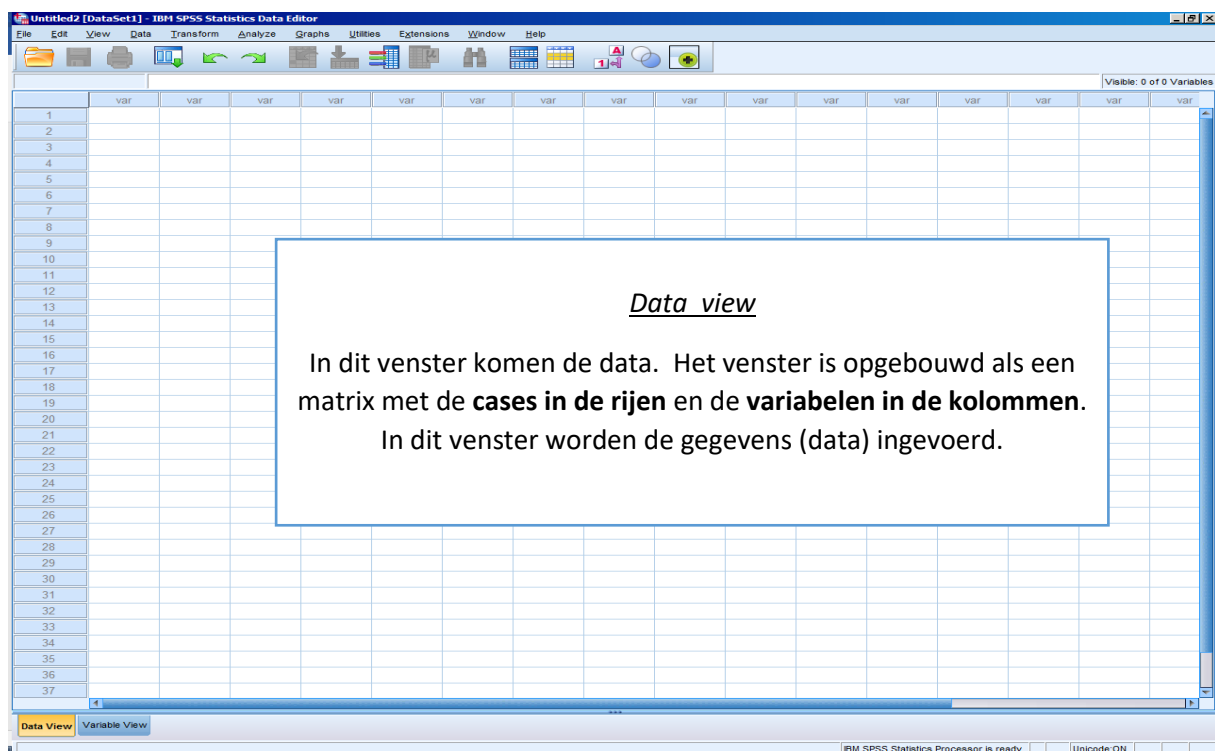
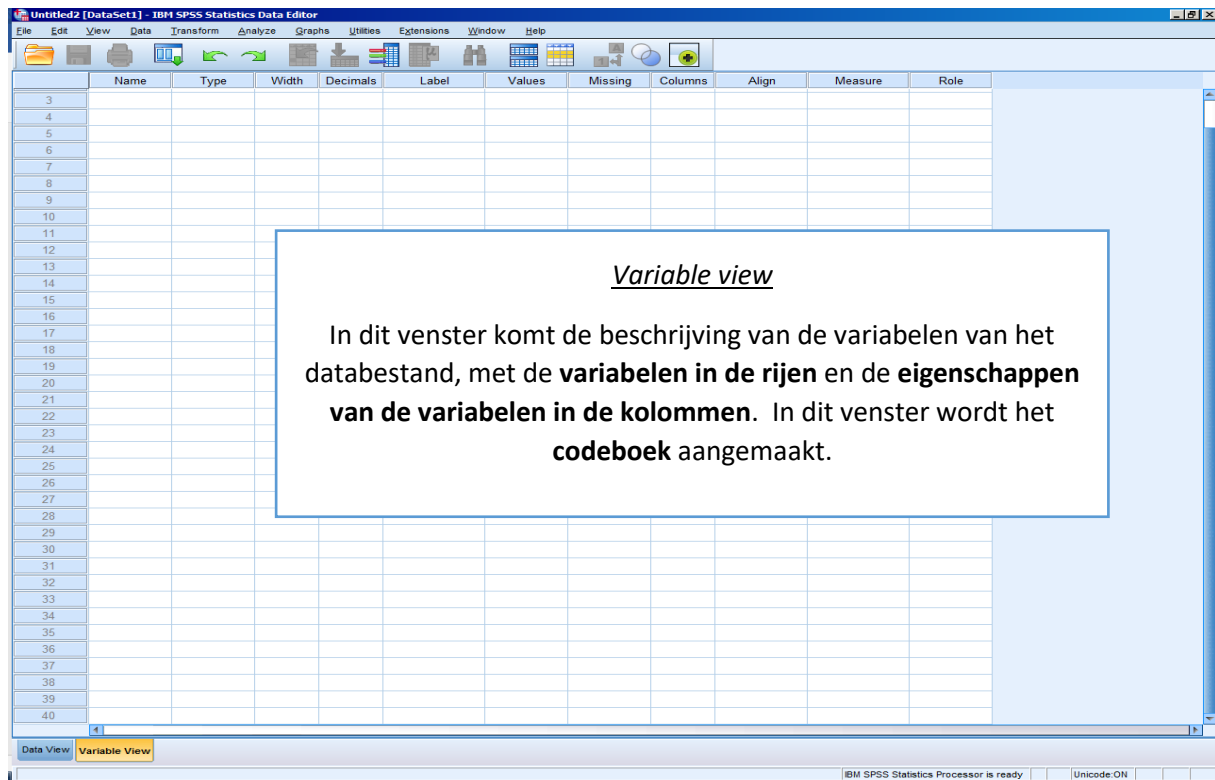
VENSTER	FUNCTIE INHOUD	BESTAND	EXTENSIE
DATA-editor	Hierin staan de data en variabelen	DATA	.zsav
SYNTAX-venster	Hierin sla je de syntax op van SPSS-opdrachten zodat je een eerdere opdracht nogmaals kunt laten uitvoeren. De syntax is de formule waarmee een statistische toets werd uitgevoerd	SYNTAX	.sps
OUTPUT-venster	Hierin komen de resultaten van de analyses en kun je alleen openen nadat je de analyses hebt gedaan	OUTPUT	.spo

### Data-editor

De inhoud van de data-editor is de datamatrix of dataset. De dataset in SPSS bestaat uit **waarden** (*values*) van **statistische eenheden** of **onderzoekseenheden** (*cases*) (vb. individuen, maar ook aggregaten zoals scholen, buurten, landen, gevangenis, ...) op **variabelen** (*variables*). Data-editor is onderverdeeld in twee deelvenster: *dataview* en *variable view*.



Verzamelde gegevens worden opgenomen in een gegevensbestand dat SPSS kan analyseren. Het SPSS-gegevensbestand is een matrix waarbij de waarnemingen in de rijen staan en de variabelen in de kolommen. Het maken van een gegevensbestand bestaat uit twee stappen: ten eerste het omzetten van de metingen in variabelen en ten tweede het invoeren van de gegevens in het gegevensbestand.



*Toelichting bij Variable View*

## SPSS-VENSTER 1: VARIABLE VIEW

Toelichting bij kolommen in <i>variable view</i>	
Name	Naam van de variabele. Kies een <b>kernwoord</b> dat goed bij de variabele past of een vraagnummer. Bijv. <i>IDnummer of casenummer</i> als volgnummer voor de respondent of <i>V1_bandouders</i> (vraag 1 band met de ouders).
Type	Type variabele: - <i>numeriek</i> : variabele bestaat uit cijfers - <i>alfanumeriek (Eng: strings)</i> : variabele bestaat uit letters of andere tekens. Standaard staat het veld op numeriek.
Width	Het aantal karakters dat een ingevoerde waarde mag hebben. Bijv. 91,850 bestaat uit 6 karakters. Standaard staat het veld op Width 8: voor numerieke variabelen is dat meestal voldoende.
decimals	Het aantal decimalen van de ingevoerde waarde. Zijn er geen decimalen nodig, zet deze waarde dan op 0. Dit is overzichtelijker.  Bijv. variabele 'gender' Standaard staat het veld op 2 decimalen.
Label	Uitgebreidere omschrijving voor de variabele. Hier kan je bijvoorbeeld de vraag uit de survey noteren. Handig om variabelen tijdens analyses beter te herkennen, vooral als meerdere onderzoekers op 1 dataset werken.

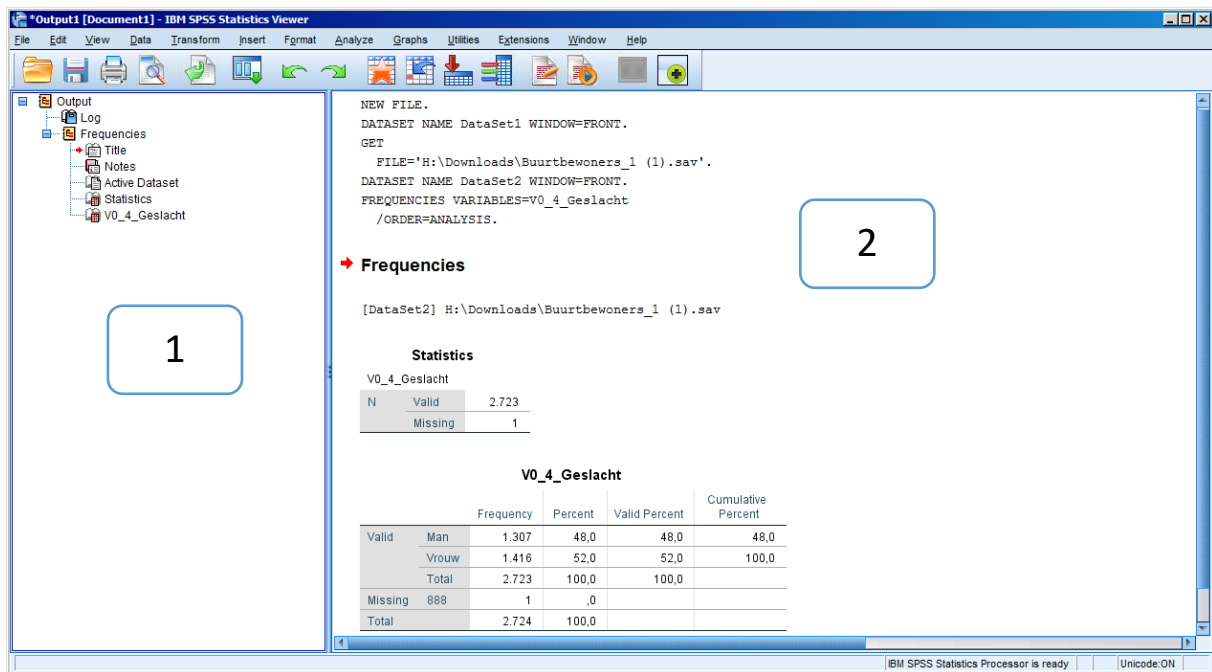
## SPSS-VENSTER 1: VARIABLE VIEW

Toelichting bij kolommen in <i>variable view</i>	
Values	Codes die aan waarden van een variabele worden toegekend. Bijv. bij de variabele 'gender' staat code '1' voor jongens en code '2' voor meisjes. Standaard staat Values op None.
Missing	waarden die in het bestand ontbreken bijv. een respondent heeft een vraag niet ingevuld. Ofwel vul je hier niets in (no missing values) ofwel geef je aan de ontbrekende waarde een onmogelijke waarde zoals -999. Dit is ' <i>discrete missing value</i> '. Kies een waarde waarvan je onmiddellijk ziet dat het om een ontbrekende waarde gaat.
Columns	Kolombreedte van de variabele op het tabblad Data View Heeft geen invloed op de statistische analyse.
Align	Uitlijning van het tabblad Variable view Heeft geen invloed op de statistische analyse
Measure	De meetschaal van de variabele: nominaal, ordinaal of scale.

## Output venster

In dit venster verschijnen de uitkomsten van de analyses in

1. Het **navigatievenster** : overzicht van de output. Klik op de uitkomst die je wilt zien.
2. Het **uitvoervenster**: hier lees je de uitkomsten van de analyses (grafieken en tabellen).



Zolang dit venster niet wordt afgesloten, worden de resultaten van opdrachten onder elkaar geplaatst in het openstaande outputvenster.

In het rechtergedeelte vinden we de output zelf terug. Resultaten van opeenvolgende analyses worden onder elkaar in dit venster weergegeven. In het linkergedeelte staat een navigatiestructuur. Die laat ons toe om snel van het ene naar een ander stuk van de output te springen. Stukken output kunnen ook tijdelijk verborgen worden door op het minteken te klikken bij de betreffende vermelding in de navigatiestructuur.

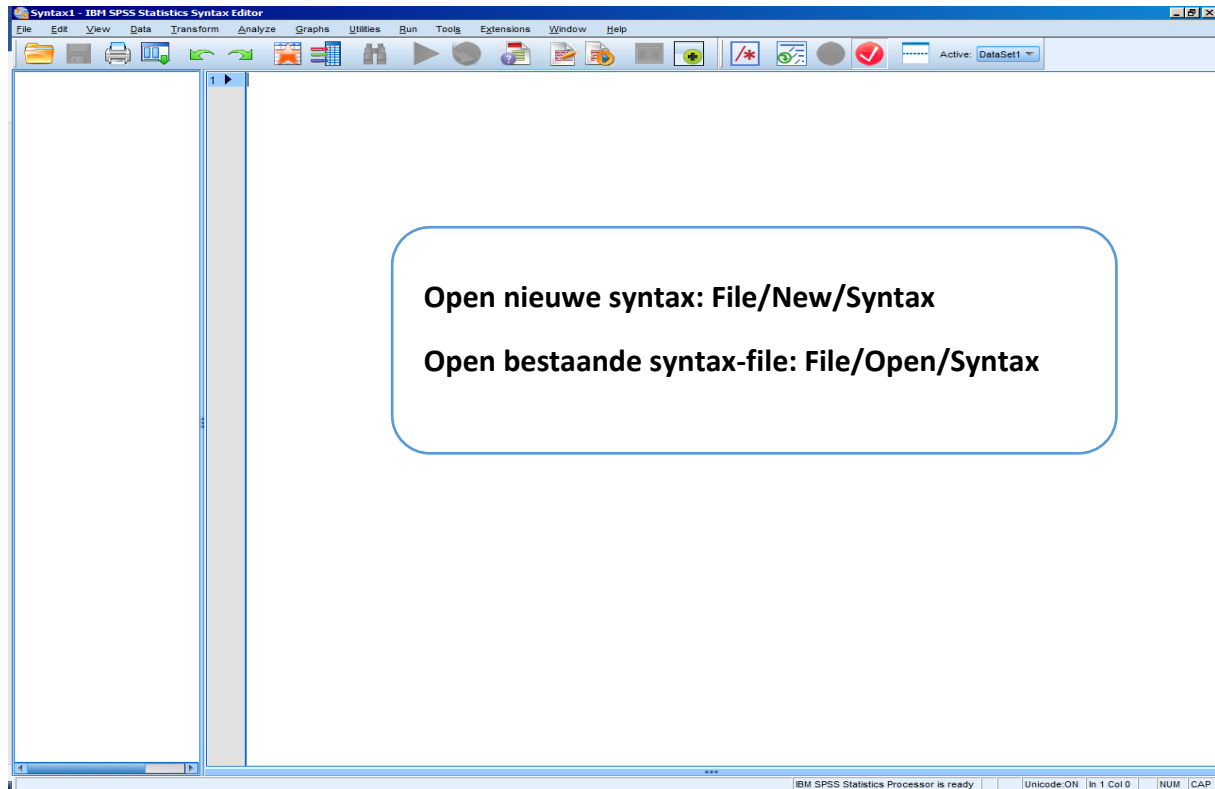
Boven in het output-venster vinden we de menubalk. We vinden er dezelfde elementen als in het data-editor-venster, met uitzondering van de items 'data' en 'transform'. Het bevat daarentegen wel twee nieuwe items:

- **Insert:** Hiermee kunnen bepaalde elementen (zoals een titel) worden toegevoegd aan de output.
- **Format:** Hiermee kan bij het printen de positie worden gewijzigd van de elementen in de output.

De lay-out van elementen in de output kan aangepast worden door op het betreffende element te dubbelklikken.

### *Syntax venster*

Om instructies te geven door commando's in te tikken (in plaats van keuzes te maken uit de menu's) staat een syntax-venster ter beschikking.



### *Waarom syntax gebruiken ?*

- **Herhaling van analyses**: wil je een analyse herhalen met andere variabelen, dan is het eenvoudiger om de syntax aan te passen. Anders moet je alle dialoogvensters en submenu's terug instellen
- **Bewaren**: vaak worden veel analyses na elkaar uitgevoerd en weet je niet meer welke instructies zijn gegeven. In een afzonderlijk opgeslagen syntaxbestand staan alle commando's netjes onder elkaar zodat raadplegen en reconstructie steeds mogelijk is.

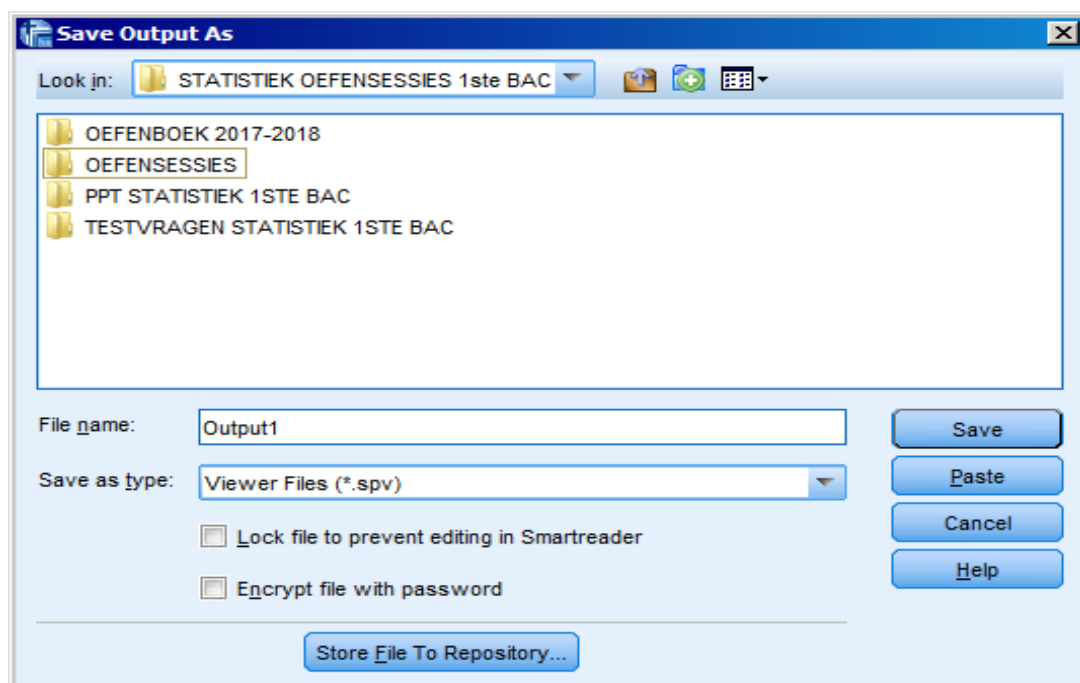
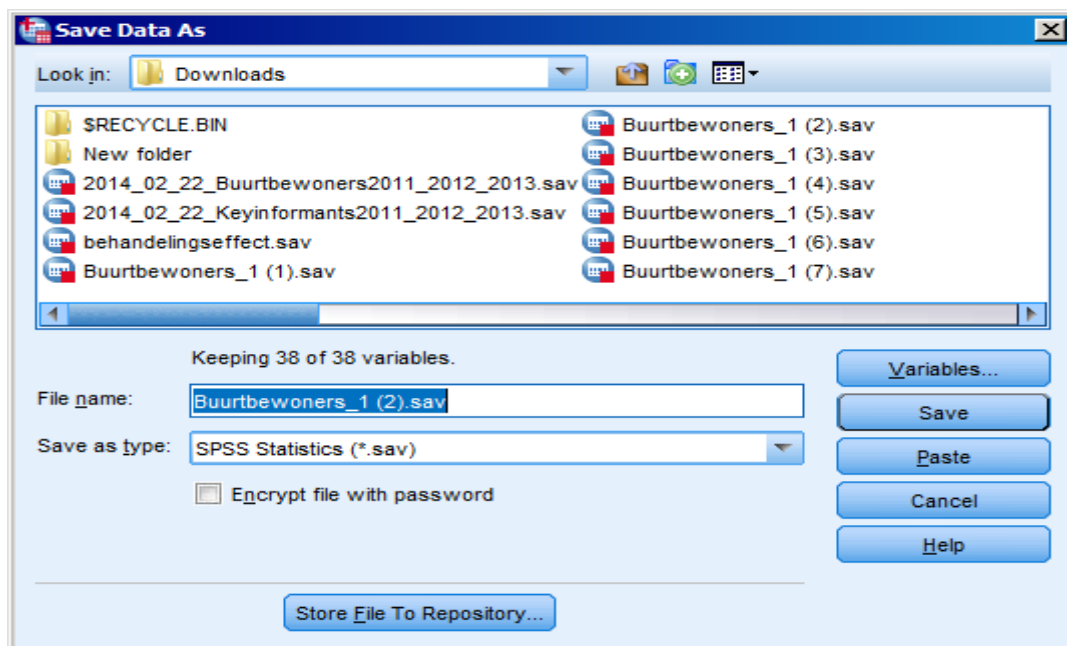
In het onderste gedeelte kunnen de commando's ingetikt worden. Die commando's kunnen uitgevoerd worden door in het menu *Run* te kiezen voor: All (voert alle commando's uit), Selection (voert de geselecteerde commando's uit), Current (voert commando uit waarop de cursor staat), of To end (voert commando uit waarop de cursor staat en alle volgende commando's).



## Opslaan van gegevens

Wil je je werk bewaren dan moet je de ingevoerde gegevens plus de output opslaan.

- **Opslaan gegevens (data) in data-editor** : ga naar *File/save as* en kies het bestand waar je wilt opslaan; **Opgelet : sla op als .zsav**. Klik op save. De tabbladen Data View en Variable View zijn beide opgeslagen in 1 bestand.
- **Opslaan van de output**: herhaal bovenstaande procedure in het outputvenster. Outputfile wordt altijd opgeslagen met de extensie .spv. Klik op save.



Vooraleer een SPSS-sessie te beëindigen moet steeds de bedenking worden gemaakt of alles wat van belang kan zijn voor later gebruik, op schijf werd bewaard. Als SPSS wordt beëindigd zonder dat dit gebeurde, kan belangrijke informatie verloren gaan waardoor reeds verrichte analyses moeten worden overgedaan. Dit kan zowel gaan om databestanden waarin wijzigingen werden aangebracht, als om de output die op het scherm verscheen. Als de output van belang is, moet die worden bewaard. **SPSS vraagt voor ieder geopend venster met nog niet bewaarde wijzigingen of de inhoud ervan op schijf moet worden bewaard.**

#### 4. INVOEREN VAN GEGEVENS

Aan de hand van onderstaand voorbeeld kun je zelf een mini-databestand opbouwen.

### ZELF AAN DE SLAG

**FICTIEF VOORBEELD: SURVEY**

Respondentnummer \_\_\_\_\_

Gender: M    V                      Leeftijd: \_\_\_\_\_ (in jaren)    Geboortjaar: \_\_\_\_\_

Faculteit:    1. Recht en Criminologie  
              2. Letteren en Wijsbegeerte

Op een schaal van 1 tot 4, hoe zeker ben je dat je voor het opleidingsonderdeel Toegepaste data-analyse zal slagen ?

1    absoluut niet zeker  
2    niet zeker  
3    vrij zeker  
4    absoluut zeker

Hoeveel jaar ervaring met SPSS heb je al ? (omcirkel slechts 1 antwoord)

0    geen ervaring  
1    1 jaar ervaring  
2    2 jaar of meer ervaring

18

#### **STAP 1: creatie van het codeboek/aanmaken van variabelen**

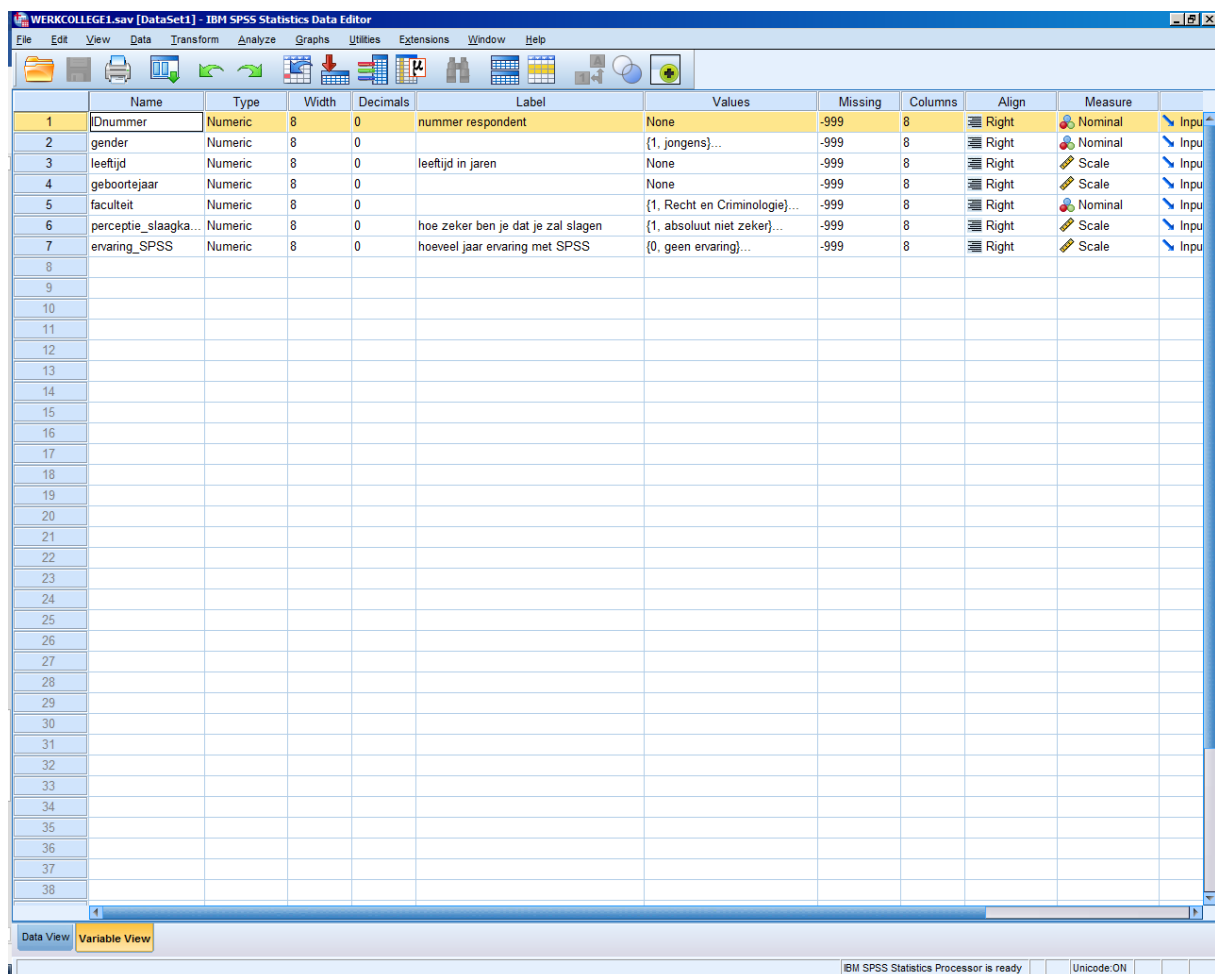
Start SPSS op en ga naar data-editor, tabblad Variable View.

Alle vragen uit het fictieve survey-voorbeeld peilen naar een kenmerk van de respondent en worden beschouwd als een variabele. In bovenstaand voorbeeld is er sprake van 7 variabelen.

Bedenk hoe je de variabelen in SPSS wilt benoemen, blijf zo dicht mogelijk bij de oorspronkelijke vraag.

Op de eerste rij in het tablad Variable View komt het respondentnummer of IDnummer. Typ dit in onder Name en druk op enter. In de overige kolommen zal SPSS automatisch een standaardwaarde invullen. Ga verder naar rij 2 en voer gender of geslacht in.

Rij 3: leeftijd, rij 4: geboortejaar, rij 5: faculteit, rij 6: perceptie\_slaagkans en rij 7: ervaring\_SPSS.



Verder invullen van de kolommen:

**Type** staat standaard door SPSS ingevuld als *Numeric*

**Width** staat standaard op '8'. Je kan dit veranderen door in de cel te klikken

**Decimals** staat standaard op '2'. Tip: zet dit op 0 om alles overzichtelijk te houden

**Label**

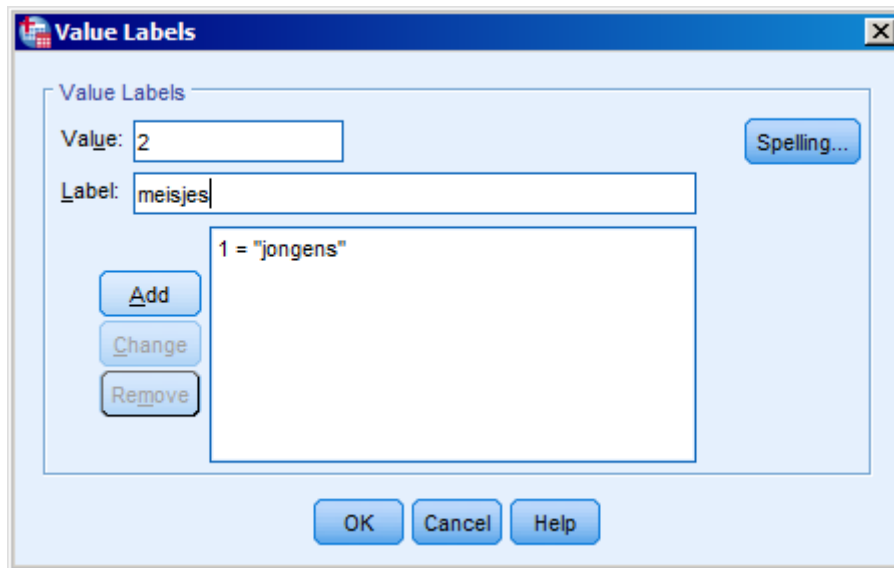
hoeft niet ingevuld om analyses uit te voeren maar geeft duidelijkheid. Hier kan je bijvoorbeeld de vraag uit de vragenlijst invoeren zodat je achteraf variabelen gemakkelijker kunt herkennen/opzoeken.

**Values**

Hoewel het in SPSS ook mogelijk is om te werken met niet-numerieke waarden van variabelen (in de voorbeeldtabel: de naam en het geslacht van de respondent) is het in de praktijk vaak makkelijker om niet-numerieke waarden door numerieke codes te vervangen (bvb. voor GESLACHT: '1' voor vrouwen en '0' voor mannen). Dit vereenvoudigt het intikwerk bij het ingeven van een datamatrix en vermindert de kans op fouten daarbij. **Let op:** Dit houdt echter in dat de numerieke score op zo een variabele eigenlijk niet meer is dan een code, die niet mag geïnterpreteerd worden als de echte waarde. De code dient immers alleen om de verschillende categorieën van elkaar te onderscheiden. Omdat de gebruikte codes in zo een geval niet erg betekenisvol zijn, bestaat de mogelijkheid om aan elke score een 'label' (beschrijving) toe te kennen die de echte waarde weergeeft en die in de output van de analyses wordt vermeld.

Pas vanaf het intervalniveau zijn de toegekende scores ook de echte waarden. Dat alle scores numeriek zijn, houdt ook in dat SPSS bereid is om alle analysetechnieken met om het even welke veranderlijken uit te voeren. Zo zouden we aan SPSS zonder problemen kunnen vragen om het gemiddelde geslacht te berekenen, hoewel dit onzin is. Het is dan aan de gebruiker om het meetniveau van de variabelen waarmee hij werkt te kennen en aan de hand daarvan de geschikte procedures te selecteren.

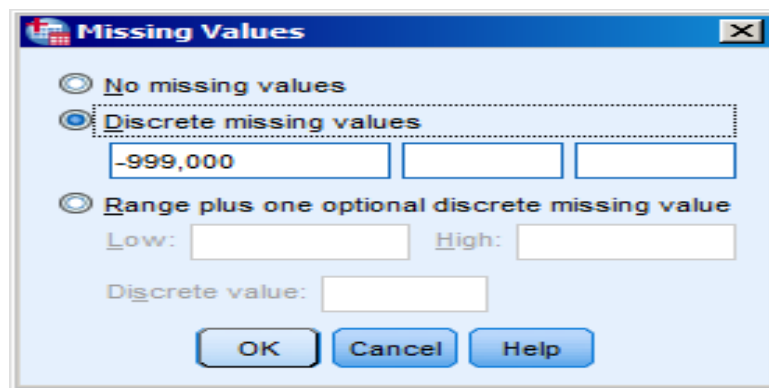
Klik in de kolom Values op de cel van de variabele *gender* en vervolgens op het blauwe vierkantje. Hiermee bekom je het venster **Value Labels**. Type in het vak achter Value: 1 en achter Label: jongens. Klik op Add en herhaal voor meisjes (value: 2 en label: meisjes).



Geef ook de variabelen *faculteit*, *perceptie\_slaagkans* en *ervaring\_SPSS* een code. Gebruik bij het coderen van de categorieën van de variabelen dezelfde numerieke waarden als in de vragenlijsten.



## Missing



Respondenten beantwoorden niet steeds alle vragen. Heeft een respondent nagelaten een vraag te beantwoorden, dan is het vaak handig om een waarde te kiezen voor het missende antwoord. Kies, om verwarring te voorkomen, een waarde waarvan je meteen weet dat het om een ontbrekende waarde gaat.

Er zijn twee soorten missing values:

- **User missing values:** is een respondent vergeten om een vraag te beantwoorden, dan kan hier een fictief getal ingevuld worden dat buiten de range van de variabele ligt (vb. 999)
- **System missing values:** waarden zelf door SPSS gecreëerd wanneer een case of een cel niet is ingevuld en weergegeven door een puntje

### Columns

staat standaard op 8 en is over het algemeen voldoende

### Align

staat standaard op *left* bij stringvariabelen en op *right* bij numerieke

### Measure

geef aan om welk type variabelen het gaat: nominaal, ordinaal of scale

## **STAP 2: het invoeren van data in Data View**

Als het codeboek in Variabele View is aangemaakt, kunnen de data werkelijk worden ingevoerd in het tabblad Data View. (klik linksonder in het tabblad Variabele View).

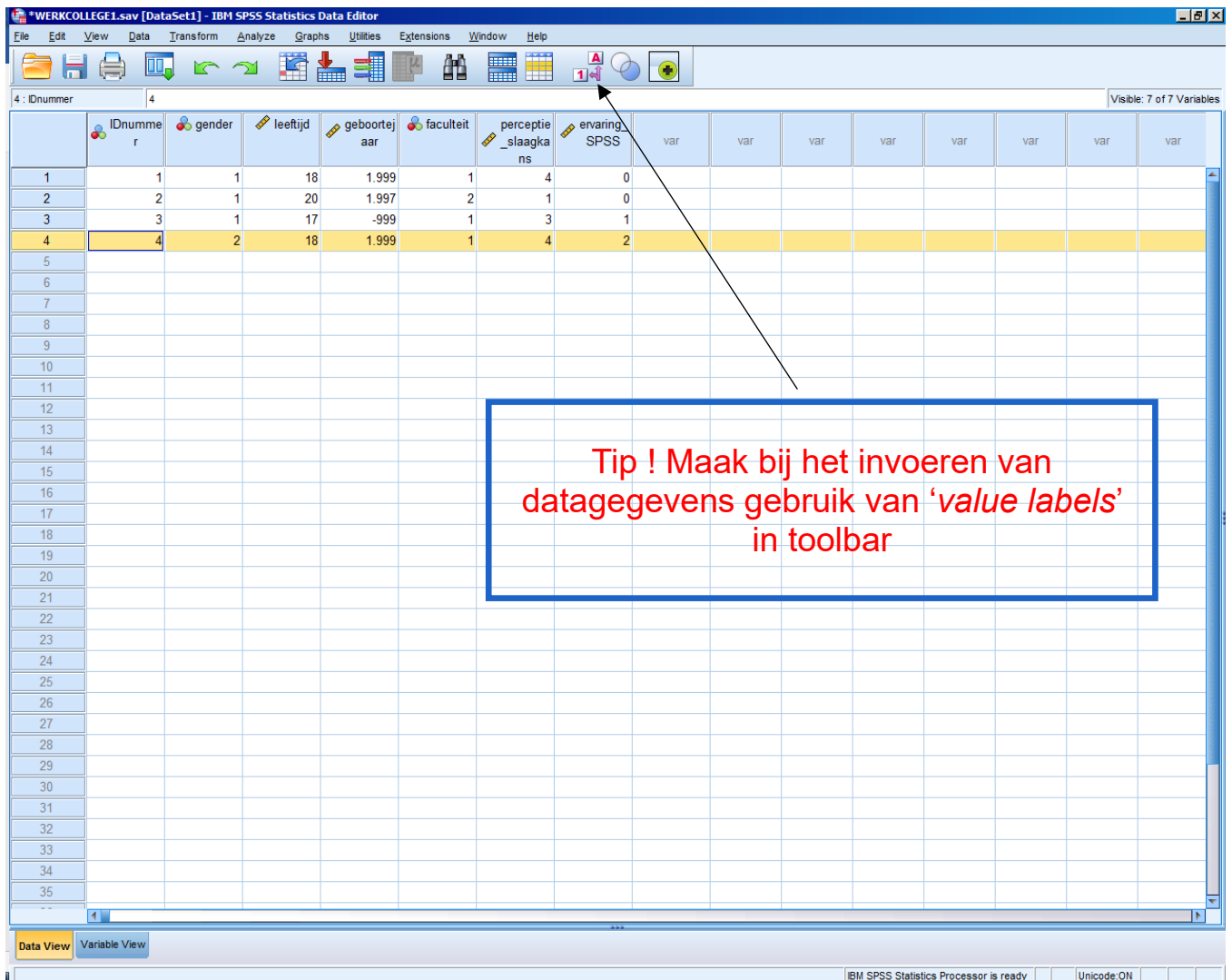
In het tabblad Data View staan de aangemaakte variabelen in de kolommen en de rijen komen de respondenten met daarachter de antwoorden op de respectievelijke variabelen.

Hieronder vind je fictieve data voor 4 respondenten.

## ZELF AAN DE SLAG: INVOEREN DATA

Casenummer	Gender	Leeftijd	Geboortejr	Faculteit	Slaagkans 1 tot 4	Ervaring SPSS
1	Jongen	18	1999	Recht en Crim	Absoluut zeker	Geen
2	Jongen	20	1997	Letteren & Wijsbegeerte	Absoluut niet zeker	Geen
3	Jongen	17	Geen antwoord	Recht en Crim	Vrij zeker	1jaar
4	meisje	18	1999	Recht en Crim	Absoluut zeker	2jaar

- Ga naar tabblad *data view* en vul datagegevens uit tabel hierboven in.



Tip ! Maak bij het invoeren van datagegevens gebruik van 'value labels' in toolbar

## 5. HET BEWERKEN VAN WAARNEMINGEN

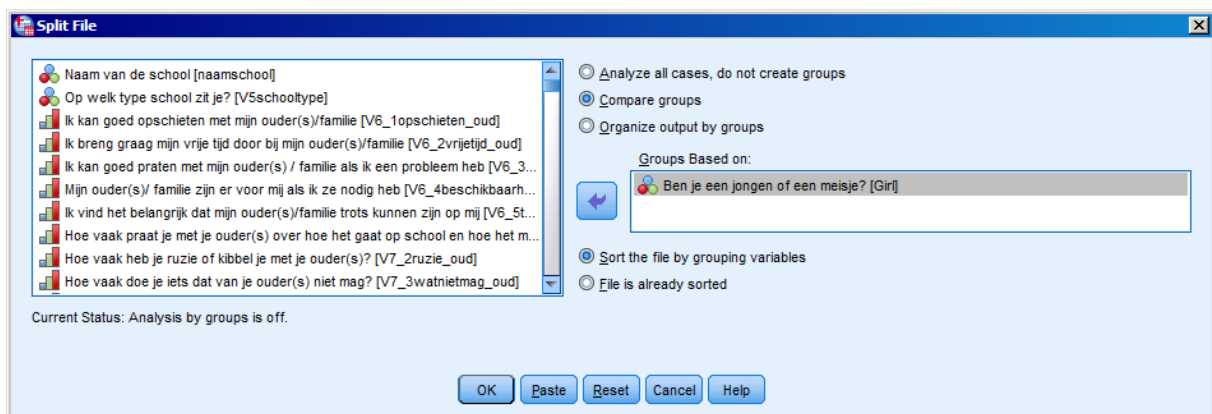
Met bewerkingsprocedures kunnen in SPSS:

- groepen opgesplitst worden voor identieke analyses per groep : **Split File**
- waarnemingen geselecteerd worden : **Select Cases**
- hercoderen van variabelen : **Recode into Same/Different Variables**

### *Split File*

Met de opdracht Split File kan een groepsindeling opgegeven worden in SPSS waarna elke volgende analyse automatisch voor elke groep afzonderlijk wordt uitgevoerd.

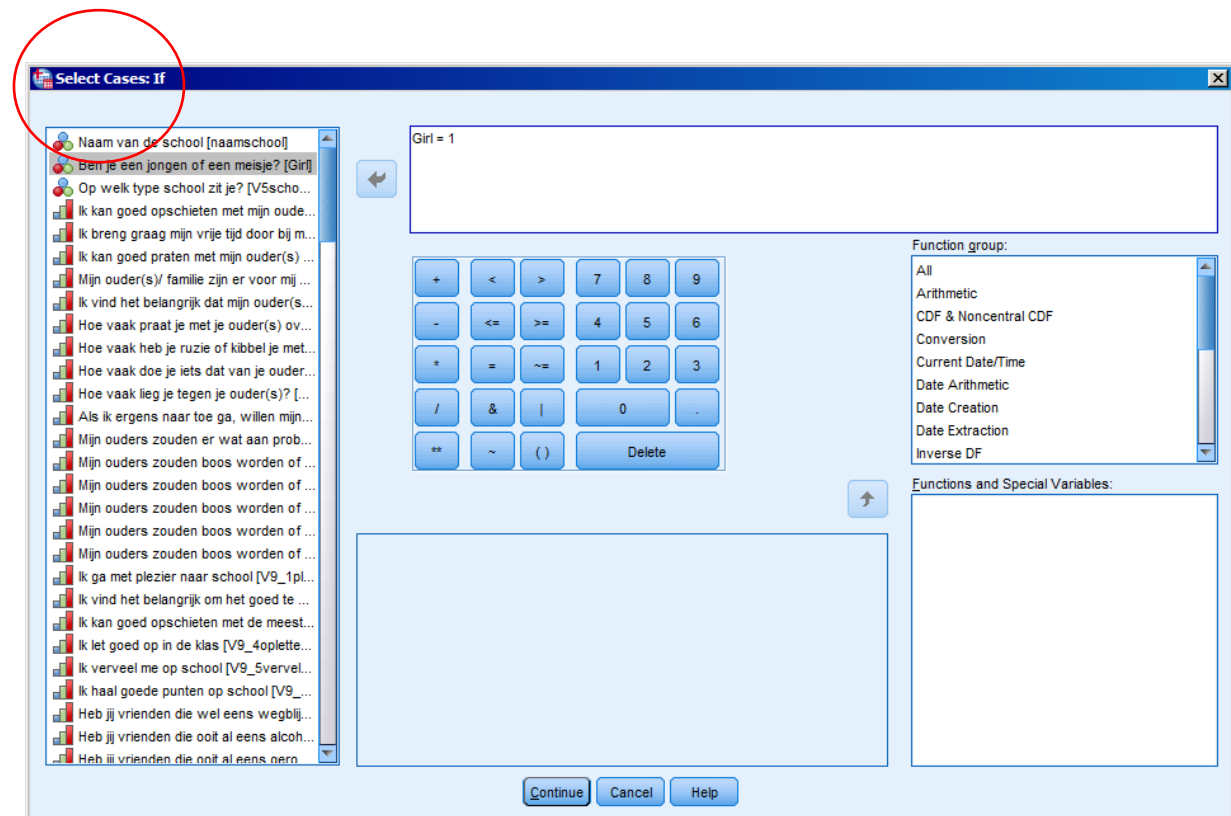
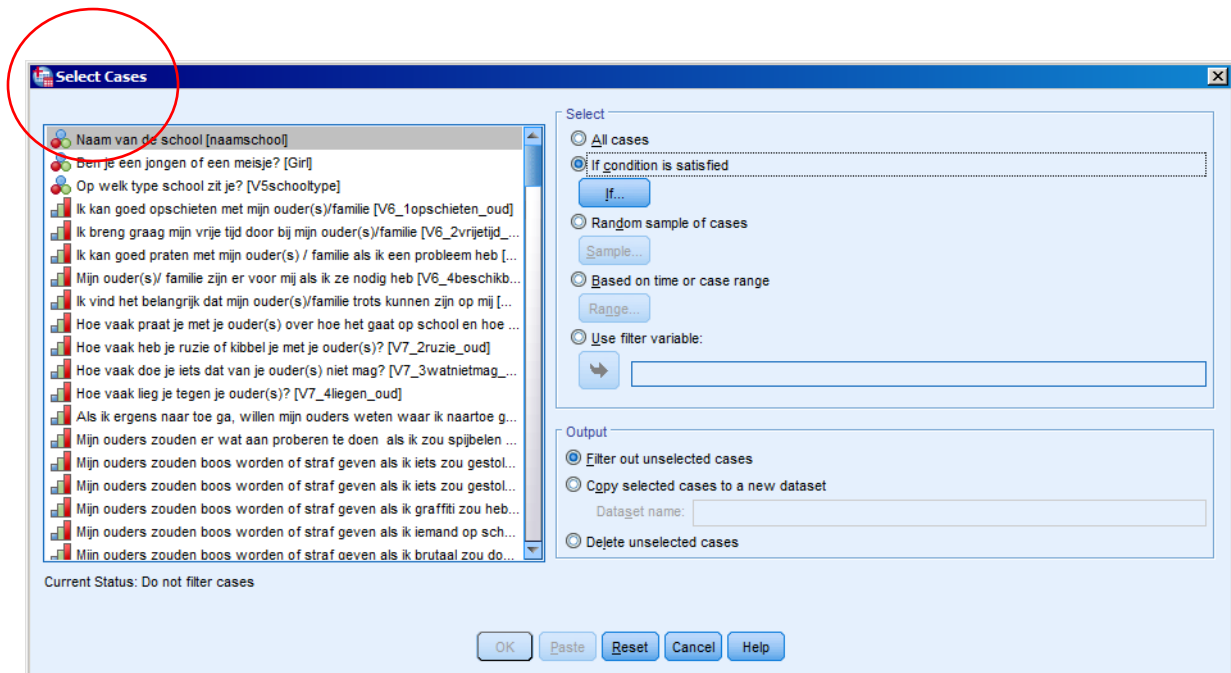
Stel dat we de beschrijvende statistieken voor jongens en meisjes apart willen presenteren, dan geven we via de opdracht *Data / Split File* een groepsdeling op (*compare groups of via organize output by groups*). Alle analyses die we daarna uitvoeren, wordt voor elke groep afzonderlijk uitgevoerd.



### *Select Cases*

De opdracht Select cases selecteert een bepaalde groep waarnemingen waarna de volgende analyses alleen voor deze groep waarnemingen worden uitgevoerd. Waarnemingen kunnen geselecteerd worden op basis van een voorwaarde, toeval of case-nummer. Bijvoorbeeld: via de opdracht *Data / Select Cases / If condition is satisfied / variabele 'geslacht' = 1* kunnen alle meisjes in de dataset worden geselecteerd (op voorwaarde dat meisjes als '1' werden gecodeerd). Met Select Cases kan ook een aselechte steekproef worden getrokken. Willen we enkel de 50 eerste waarnemingen betrekken in onze analyses, dan kan dit ook via Select Cases.





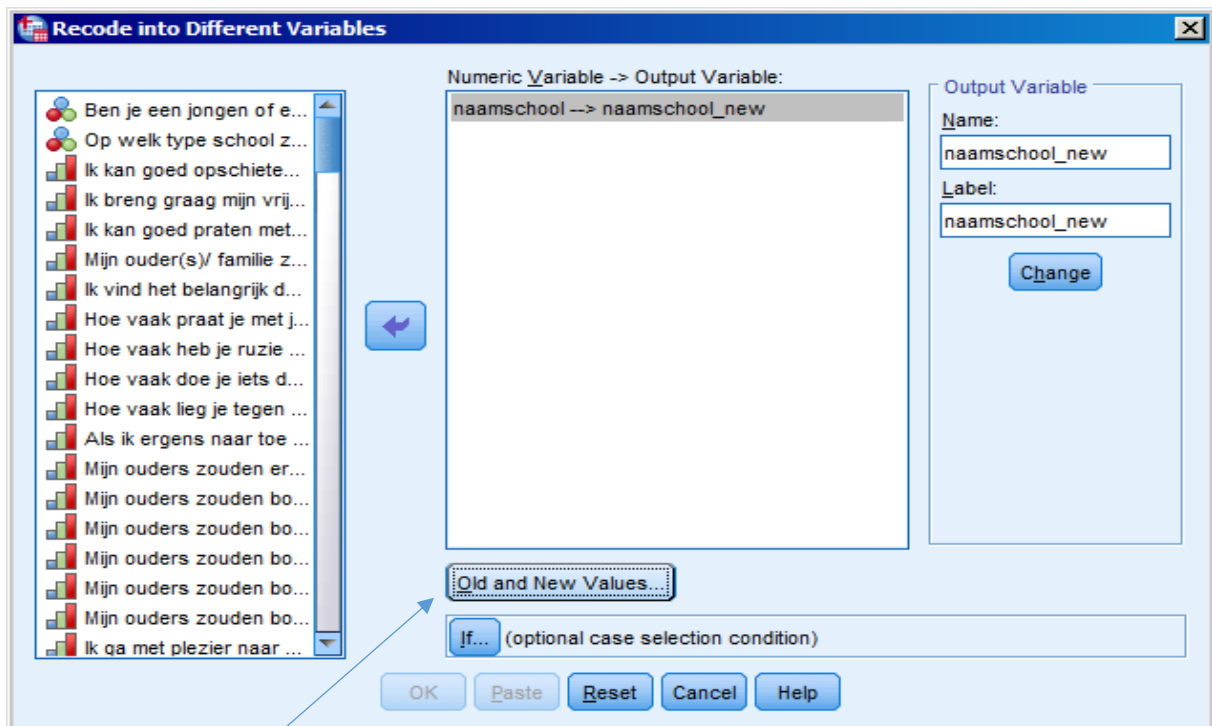
**Opgelet ! zolang de opdracht in SPSS niet werd 'gereset', zullen alle volgende analyses uitgevoerd worden voor groepen apart of geselecteerde waarnemingen.**

### ***Recode into Same/Different Variables***

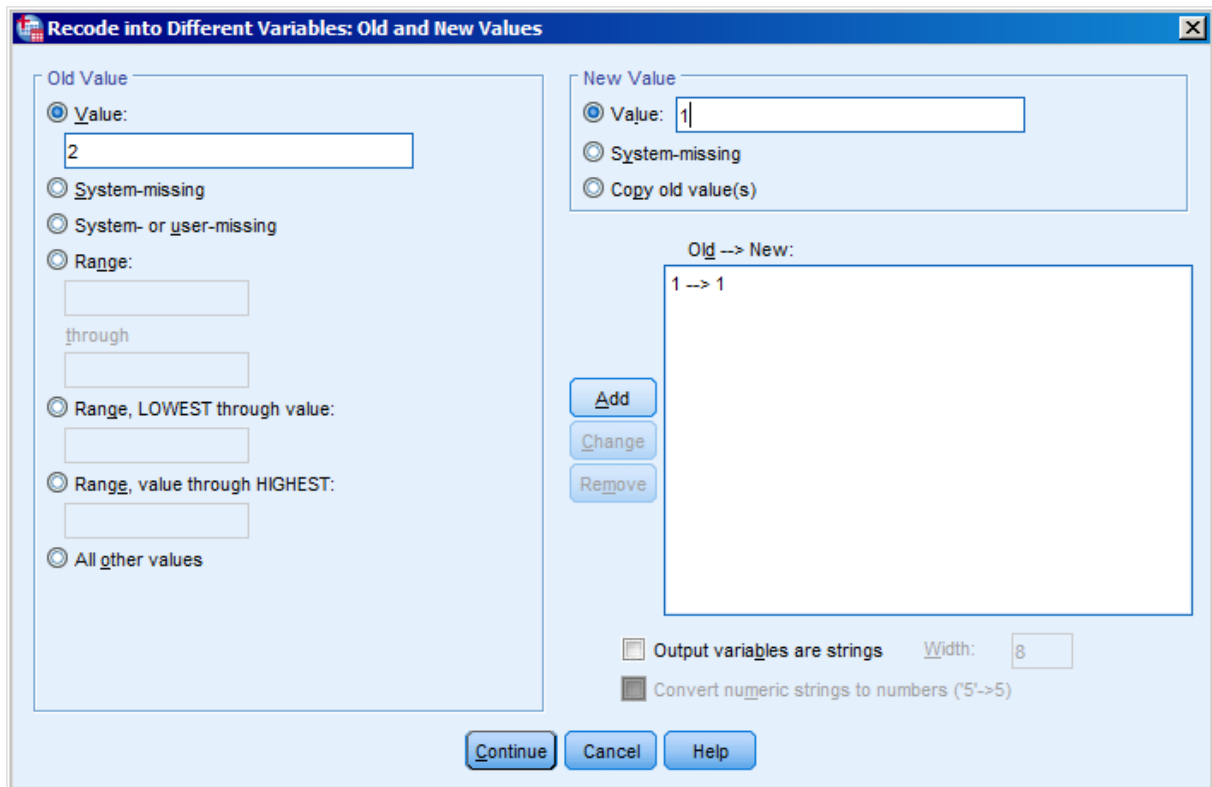
Via de opdracht *Recode into Same/Different Variables* (onder menuknop *Transform*) kunnen antwoordcategorieën van variabelen hercodeerd worden. Bijvoorbeeld: we willen van een variabele met 4 antwoordcategorieën (1 tot 4) een dichotome variabele maken waarbij we de waarden 1 en 2 gaan hercoderen naar waarde 1 en waarden 3 en 4 gaan hercoderen naar waarde 2.

*Recode into Same Variable* verandert een oorspronkelijke variabele naar een nieuwe variabele, met *Recode into Different Variable* behoud je de oorspronkelijke variabele en wordt een nieuwe variabele gecreëerd. We raden aan om bij voorkeur deze laatste optie te gebruiken. Op die manier vermijdt je dat oorspronkelijke variabelen verloren gaan.

Selecteer de variabele die je wilt hercoderen, geef onder *output variable* een nieuwe naam en label, klik op *change*.



Onder *Old and New Values* geven we aan de oude (originele) waarden een nieuwe waarde.



Door op *Add* te klikken wordt de hercodering als uit te voeren actie toegevoegd. Klik op *Continue* en *OK*. De bestaande variabele wordt nu vervangen door een nieuwe met nieuwe codes. Vergeet niet om onder *Values* de categorieën van de nieuwe variabele aan te geven.

## 6. UITVOEREN VAN STATISTISCHE ANALYSES

In deze kennismaking met SPSS gaan we kort in op een aantal basis analysetechnieken. In hetgeen volgt komen volgende analyses aan bod:

1. Univariate beschrijvende statistieken
2. Samenhang tussen twee variabelen
3. Verklaren van een variabele op basis van één of meerdere onafhankelijke variabelen

### 6.1. Univariate beschrijvende statistieken

Het opvragen van frequentietabellen, parameters van centraliteit, spreiding en vorm gebeurt in SPSS via *Analyze Descriptive Statistics Frequencies*. Frequenties kunnen getoond worden in een tabel of een grafiek (*Charts*).

Hierna tonen we hoe je de univariate beschrijvende statistieken opvraagt voor nominale, ordinale en metrische variabelen. Eerst wordt een woordje uitleg gegeven bij het ingeven van instructies in dialoogvensters in SPSS.

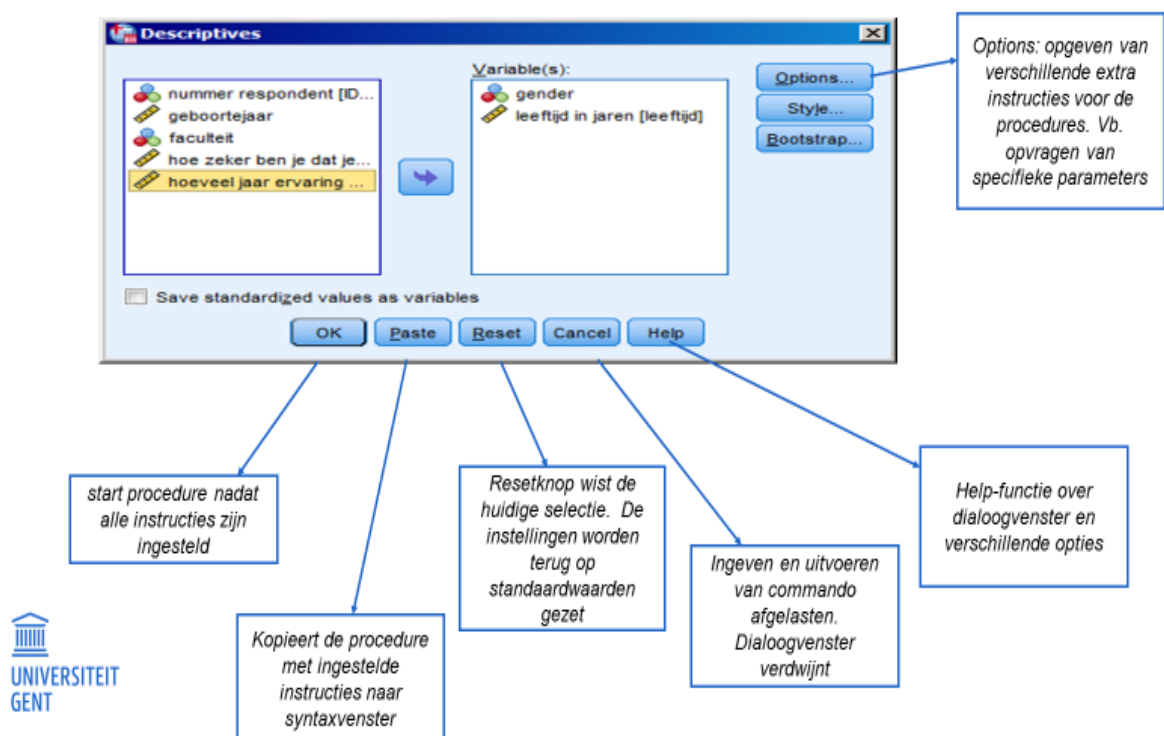
#### *Ingeven van instructies in dialoogvensters*

Twee werkwijzen kunnen gevolgd worden om SPSS aan te zetten om instructies te geven: kiezen van opties in de menustructuur of intikken van commando's in een syntax-venster. Het uitvoeren van commando's via syntax betekent dat we een bepaalde commandotaal gebruiken om SPSS berekeningen te laten doen. We behandelen in deze inleidende syllabus enkel het werken met de menustructuur.

In alle SPSS-dialoogvensters vinden we volgende commandoknoppen:

## DIALOOGVENSTERS IN SPSS: VOORBEELD

### • Selecteer *gender* en *leeftijd*

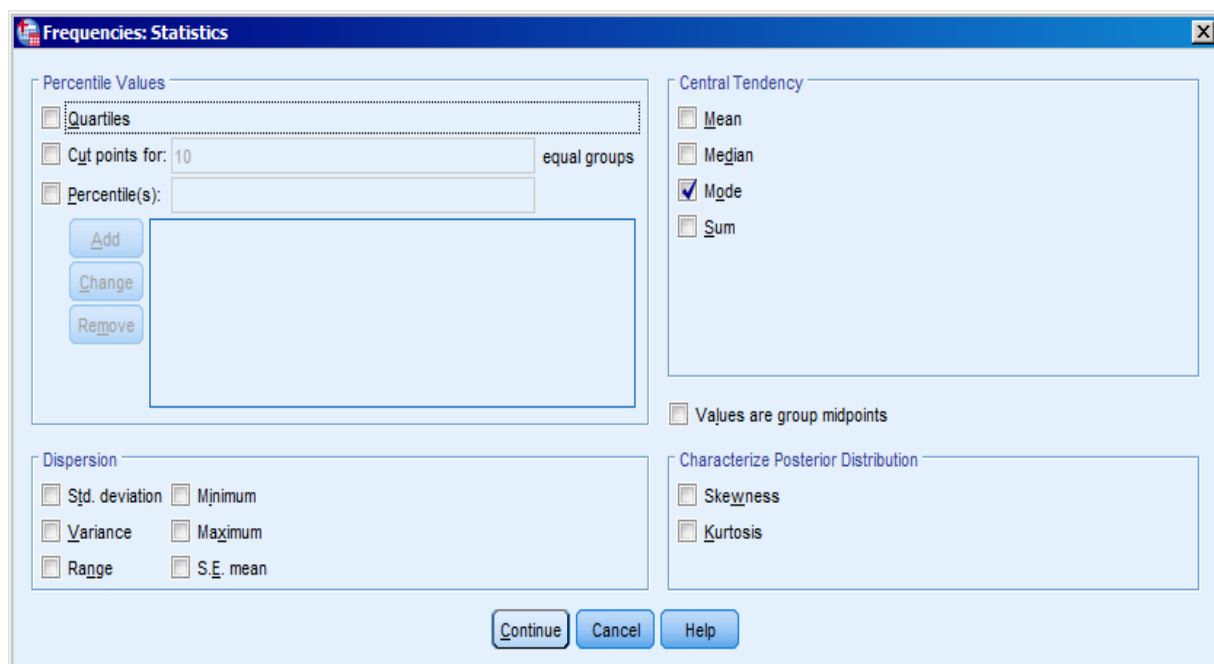
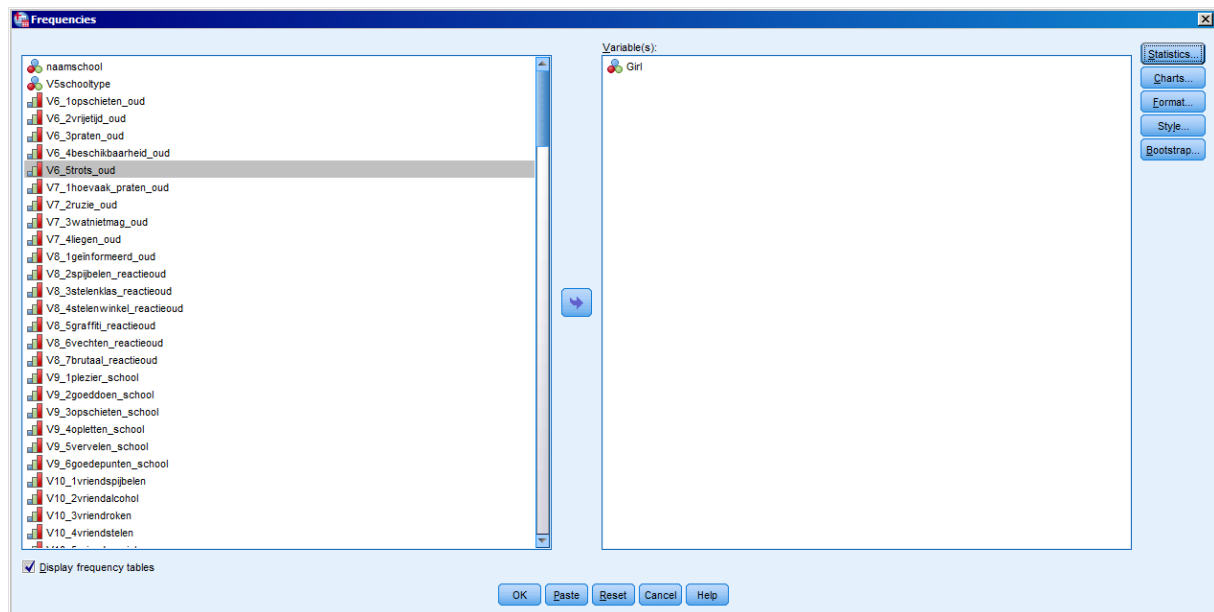


36

**Let op!** De selecties die gemaakt werden in een dialoogvenster blijven gedurende een hele SPSS-werksessie onveranderd, tenzij die door de gebruiker zelf worden gewijzigd. Dit noemt men '**persistentie van de instellingen**' en heeft tot gevolg dat een gebruiker steeds goed moet nakijken welke opties aan en uit staan, en of dit inderdaad overeenkomt met de wensen van de analyse. De 'Reset'-toets zorgt ervoor dat de instellingen van een vorig gebruik van een dialoogvenster worden gewist.

### *Parameters van centraliteit en spreiding voor nominale variabelen*

Ga naar *Analyze* *Descriptive Statistics* *Frequencies*. Kies onder *Statistics* *Mode* *Continue* *OK*



**Ben je een jongen of een meisje?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	jongen	386	47,4	47,9	47,9
	meisje	420	51,5	52,1	100,0
	Total	806	98,9	100,0	
Missing	System	9	1,1		
Total		815	100,0		

*Valid* categorieën van de variabele met de geldige waarden

*Missing* item-nonrespons of ontbrekende waarden

*Frequency* absolute frequenties

*Percent* absolute frequenties gepercenteerd

*Valid percent* geldige percentages

*Cumulative percent* cumulatieve, geldige percentages

In een zelfrapportage survey werd aan 815 leerlingen uit de basisschool gevraagd aan te duiden of zij een jongen of een meisje waren. Het meetniveau van deze variabele is **nominaal**. De antwoordcategorieën zijn ‘jongen’ en ‘meisje’. In totaal hebben 806 leerlingen de vraag beantwoord, 9 leerlingen deden dat niet. Er zijn 9 missings (1.1%). De analyse is dus gebaseerd op 98.9% (806 respondenten) van de respondenten.

**Statistics**

Ben je een jongen of een meisje?

N	Valid	806
	Missing	9
Mode		1

Bekijken we de parameter van centraliteit: de modus. De modale waarde bedraagt 1. Code 1 staat voor meisjes (jongens werden gecodeerd met code ‘0’). Dit betekent dat de meeste respondenten meisjes zijn. En wat met de spreiding? We kunnen kijken naar het percentage dat niet tot de modus behoort, of we kunnen de variatieratio, de index van diversiteit of de spreidingsmaat ‘d’ zelf berekenen.

### Parameters van centraliteit en spreiding voor ordinale variabelen

Ga naar *Analyze* *Descriptive Statistics* *Frequencies*. Kies onder *Statistics* *Quartiles* *Mode* *Median* *Range* *Minimum* *Maximum* *Continue* *OK*

**Frequencies: Statistics**

**Percentile Values**

☒ **Quartiles**

☐ Cut points for: 10 equal groups

☐ Percentile(s):

**Central Tendency**

☐ Mean

☒ Median

☒ Mode

☐ Sum

☐ Values are group midpoints

**Dispersion**

☐ Std. deviation

☒ Minimum

☐ Variance

☒ Maximum

☒ Range

☐ S.E. mean

**Characterize Posterior Distribution**

☐ Skewness

☐ Kurtosis

### Hoe vaak lieg je tegen je ouder(s)?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(bijna) nooit	515	63,2	63,3	63,3
	paar keer per maand	219	26,9	26,9	90,3
	paar keer per week	68	8,3	8,4	98,6
	(bijna) elke dag	11	1,3	1,4	100,0
	Total	813	99,8	100,0	
Missing	System	2	,2		
Total		815	100,0		



We vroegen aan 815 leerlingen uit de basisschool ‘hoe vaak lieg je tegen je ouders?’ De antwoordcategorieën zijn ‘(bijna) nooit’ (code 0), ‘een paar keer per maand’ (code 1), ‘een paar keer per week’ (code 2), ‘(bijna) elke dag’ (code 3). We kunnen in deze situatie de respondenten ordenen op een niet-nauwkeurig gedefinieerde schaal. De afstanden tussen de categorieën zijn immer niet geïjkt.

We interpretern deze tabel als volgt. Een overgrote meerderheid van de leerlingen (63.3 % of 515 van de bevraagde leerlingen) zegt (bijna) nooit te liegen tegen de ouders. Niettemin zegt 26.9% (of 219 respondenten) dat toch een paar keer per maand te doen. Samen maken deze 2 categorieën 90.3% van de geldige respons uit. Dit betekent dat er slechts heel weinig leerlingen aangeven een paar keer per week of (bijna) elke dag te liegen tegen de ouders. Wat zegt deze tabel nu inhoudelijk voor de criminoloog-onderzoeker? Uiteindelijk gaat het in onderzoek steeds om de inhoudelijke informatie die voortkomt uit de tabel. Als criminoloog zien we verschillen en grote tendensen onder het fenomeen ‘liegen tegen de ouders’. De grote tendens is dat de meeste jongeren (bijna) nooit liegen tegen de ouders (tenminste dat rapporteren de leerlingen in de steekproef), niettemin zijn er **verschillen (variatie of spreiding)** waar te nemen. Het zijn precies die verschillen die ons later zullen interesseren. Criminologen willen hiervoor immers een verklaring bieden en deze toetsen. Op dit ogenblik krijgen we enkel zicht op de aard van deze spreiding.

### Statistics

Hoe vaak lieg je tegen je ouder(s)?

N	Valid	813
	Missing	2
Median		,00
Mode		0
Range		3
Minimum		0
Maximum		3
Percentiles	25	,00
	50	,00
	75	1,00

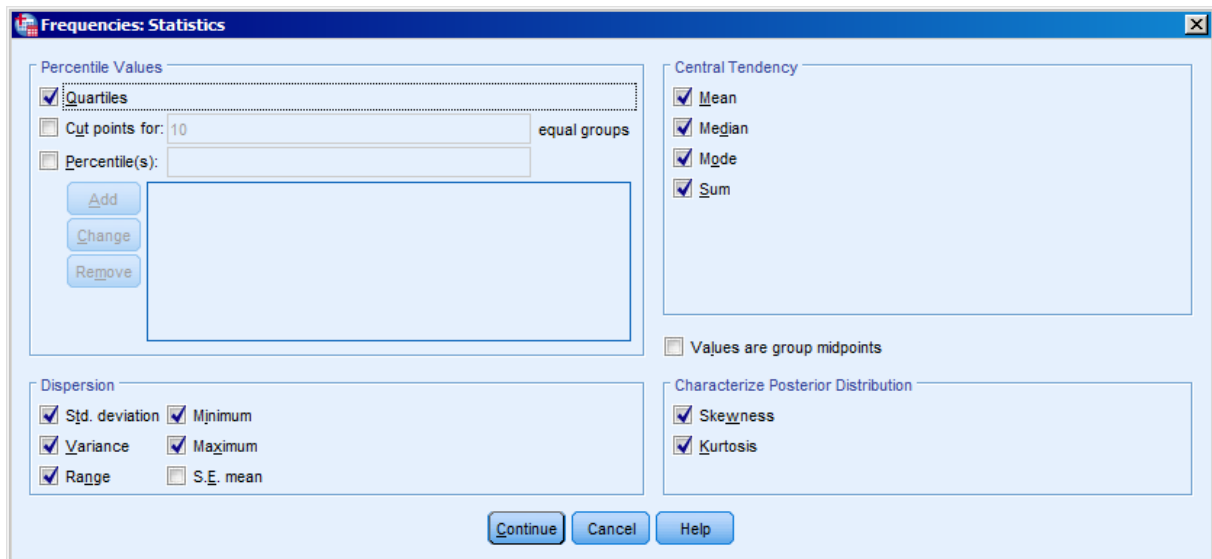
We kunnen dankzij de beschrijvende statistieken nog meer inhoudelijke informatie distilleren uit de ruwe data dan de informatie die we uit de frequentietabel kunnen opmaken. Uit de tabel met de beschrijvende statistieken kunnen we volgende parameters opvragen: de mediaan en de



modus, de variatiebreedte (range), de minimumwaarde, de maximumwaarde en de spreiding over de antwoordcategorieën op basis van de percentielen.

### *Parameters van centraliteit, spreiding en vorm voor metrische variabelen*

Ga naar *Analyze Descriptive Statistics Frequencies*. We vragen alle univariate beschrijvende statistieken op van centraliteit (centrale tendenzen), spreading (dispersion) en vorm (skewness en kurtosis).



We vroegen 4758 Belgische adolescenten naar hun bier/alcoholconsumptie van de afgelopen maand (*Denk terug aan de voorbije 30 dagen. Tijdens hoeveel verschillende gelegenheden heb je volgende zaken (bier of alcopops, wijn, sterke drank) gedronken ?*). Hieronder presenteren we de frequentietabel en de univariate beschrijvende statistieken. De variabele 'bier/alcohol consumptie afgelopen maand' is van het metrische niveau.



**Bier/alcoholconsumptie afgelopen 30 dagen**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3.125	65,7	69,9	69,9
	1	472	9,9	10,6	80,5
	2	291	6,1	6,5	87,0
	3	188	4,0	4,2	91,2
	4	89	1,9	2,0	93,2
	5	110	2,3	2,5	95,6
	6	35	,7	,8	96,4
	7	24	,5	,5	97,0
	8	17	,4	,4	97,3
	9	10	,2	,2	97,6
	10	60	1,3	1,3	98,9
	12	7	,1	,2	99,1
	13	2	,0	,0	99,1
	14	1	,0	,0	99,1
	15	11	,2	,2	99,4
	16	2	,0	,0	99,4
	20	15	,3	,3	99,8
	25	2	,0	,0	99,8
	30	3	,1	,1	99,9
	40	1	,0	,0	99,9
	48	1	,0	,0	99,9
	50	3	,1	,1	100,0
	52	1	,0	,0	100,0
	Total	4.470	93,9	100,0	
Missing	>= 96	2	,0		
	ambiguous answer	27	,6		
	no answer	259	5,4		
	Total	288	6,1		
Total		4.758	100,0		

4758 adolescenten werden bevraagd naar hun bier/alcoholconsumptie de afgelopen 30 dagen. Van deze 4758 hebben 288 jongeren (6.1%) de vraag niet of niet goed beantwoord. Dit betekent dat de rapportage van de resultaten gebaseerd zal zijn op de geldige antwoorden van 4470 adolescenten.

De overgrote meerderheid of 69.9% (3125) van de jongeren geeft geen enkele bier/alcoholconsumptie aan de afgelopen 30 dagen. 10.6% (472 respondenten) geeft aan tijdens 1 gelegenheid bier of alcohol te hebben gedronken in de afgelopen 30 dagen. 1.3% of 60 respondenten geven aan de afgelopen dertig dagen tijdens 5 gelegenheden bier of alcohol te hebben gedronken.

### Statistics

Bier/alcoholconsumptie afgelopen maand		
N	Valid	4.470
	Missing	288
Mean		1,07
Median		,00
Mode		0
Std. Deviation		3,040
Variance		9,244
Skewness		7,711
Std. Error of Skewness		,037
Kurtosis		93,982
Std. Error of Kurtosis		,073
Range		52
Minimum		0
Maximum		52
Percentiles	25	,00
	50	,00
	75	1,00

*Wat kunnen we nu aflezen uit de tabel met de beschrijvende statistieken ?*

**N Valid** (geldig): Aantal observaties (respondenten) waarvoor we een antwoord op de vraag hebben. 4470 respondenten hebben een ‘geldig’ antwoord gegeven op deze vraag.

**Missing:** aantal observaties waarvoor we geen antwoord op de vraag hebben. 288 respondenten hebben geen antwoord gegeven op deze vraag. Dit betekent dat  $288/(4470 + 288)$  of 0.06 % van het totaal aantal bevraagde jongeren deze vraag niet beantwoord heeft. Noot: **alle berekeningen die de PC maakt, gebeuren op het aantal geldige observaties**

**Percentiles:** we krijgen hier de kwartielen te zien. De eerste 25 % van de observaties heeft een score 0 (geen bier/alcoholconsumptie de afgelopen 30 dagen), de middelste van de

gerangschikte observaties heeft een score van 0 en de respondenten die het derde kwartiel uitmaken, hebben een score van 1 (1 bier/alcoholconsumptie de afgelopen 30 dagen).

### *Parameters van centraliteit*

Welke zijn nu de grote tendensen inzake bier/alcoholconsumptie gedurende de voorbije 30 dagen? Aangezien de variabele van het metrisch niveau is, kunnen we preciezere informatie verkrijgen dan het geval is voor variabelen gemeten op een nominaal of ordinaal meetniveau. De meest precieze informatie die we kunnen verkrijgen is het rekenkundig gemiddelde of mean. (We herhalen: het rekenkundig gemiddelde is de som van alle waarden gedeeld door het aantal respondenten)

Het **rekenkundig** gemiddelde bedraagt 1.07. Respondenten rapporteren een gemiddelde bier/alcoholconsumptie de afgelopen 30 dagen van 1.07.

De **mediaan** bedraagt .00. Als we de observaties zouden rangschikken op basis van het kenmerk ‘bier/alcoholconsumptie gedurende de afgelopen 30 dagen’ dan is de waarde die overeenkomt met de observatie die de groep netjes in twee gelijke delen verdeelt, .00. De mediaan ligt iets lager dan het gemiddelde.

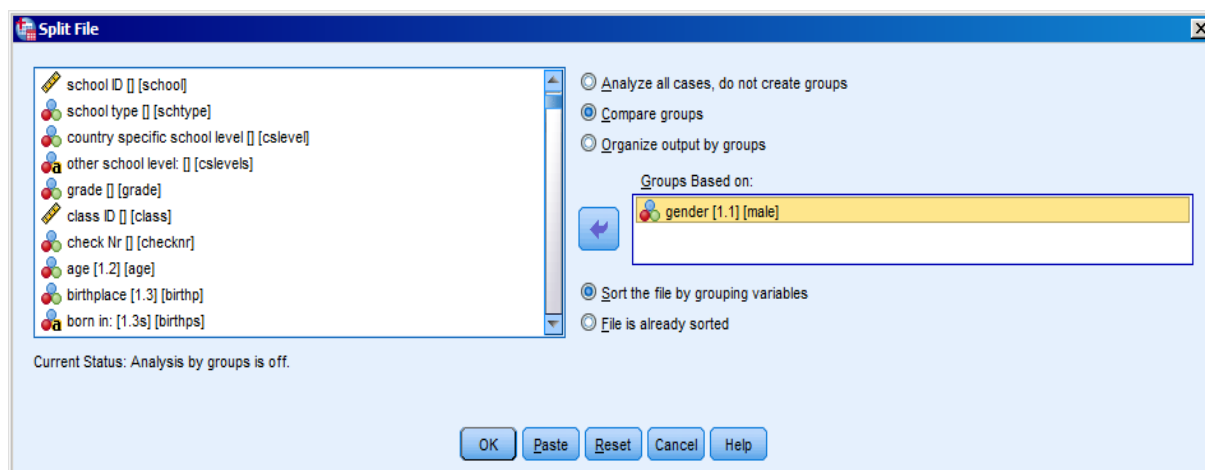
De **modus** bedraagt 0. De meest voorkomende waarde is 0. De meest typische jongere in deze steekproef heeft een bier/alcoholconsumptie gedurende de afgelopen 30 dagen, van 0.

### *Parameters van spreiding*

Er zijn duidelijk grote verschillen inzake bier/alcoholconsumptie gedurende de afgelopen 30 dagen onder de respondenten in deze steekproef. De minimumwaarde bedraagt 0 en de maximumwaarde bedraagt 52. Daarom is de **variatiebreedte** 52. Dit wil zeggen dat bier/alcoholconsumptie gedurende de afgelopen 30 dagen varieert van 0 tot 52. Van een grote variabiliteit gesproken. Aangezien deze variabele van het metrische niveau is, kunnen we de **variantie** en **standaardafwijking** berekenen. Hieruit leiden we af hoe sterk onze respondenten van het gemiddelde verschillen. Is de variantie nul, dan is er geen spreiding en dan verschillen de respondenten niet van het gemiddelde. In dat heeft iedereen dezelfde waarde. Uiteraard kunnen we ook hier de kwartielen opvragen als we dat willen en op basis daarvan het verschil tussen de waarde die overeenkomt met het hoogste kwartiel en het laagste kwartiel bestuderen. De variantie bedraagt 9.244 en de standaardafwijking bedraagt 3.04. De respondenten verschillen dus van het rekenkundig gemiddelde.

Om de betekenis van de begrippen standaardafwijking en variantie ten volle te begrijpen, is het nog steeds zinvol om deze zelf te berekenen. Dit wordt in dit oefenboek behandeld in hoofdstuk II : univariate statistiek.

Je vraagt je wellicht af wat deze gegevens op zich zeggen. Op zich lijken sommige parameters van centraliteit en spreiding niet erg veel te zeggen. Ze vatten informatie samen uit een hele reeks van observaties over een bepaald kenmerk. Toch zijn deze voor beleidsmakers en onderzoekers belangrijk. Ongelijke spreiding roept veel criminologische vragen op. Waarom zijn er verschillen tussen individuen waar te nemen? Verder is het ook mogelijk dat er verschillen zijn tussen groepen. We presenteren hierna de beschrijvende statistieken van bier/alcoholconsumptie gedurende de afgelopen 30 dagen voor jongens en meisjes apart. We doen dit aan de hand van de bewerking *Split File*, waarna we de beschrijvende statistieken opvragen via *Analyze Descriptive Statistics Descriptives*.



### Beschrijvende statistieken

geslacht		N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
female	Bier/alcoholconsumptie voorbij 30 dagen	2.245	40	0	40	,81	2,120	4,495
male	Bier/alcoholconsumptie voorbij 30 dagen	2.213	52	0	52	1,33	3,738	13,975

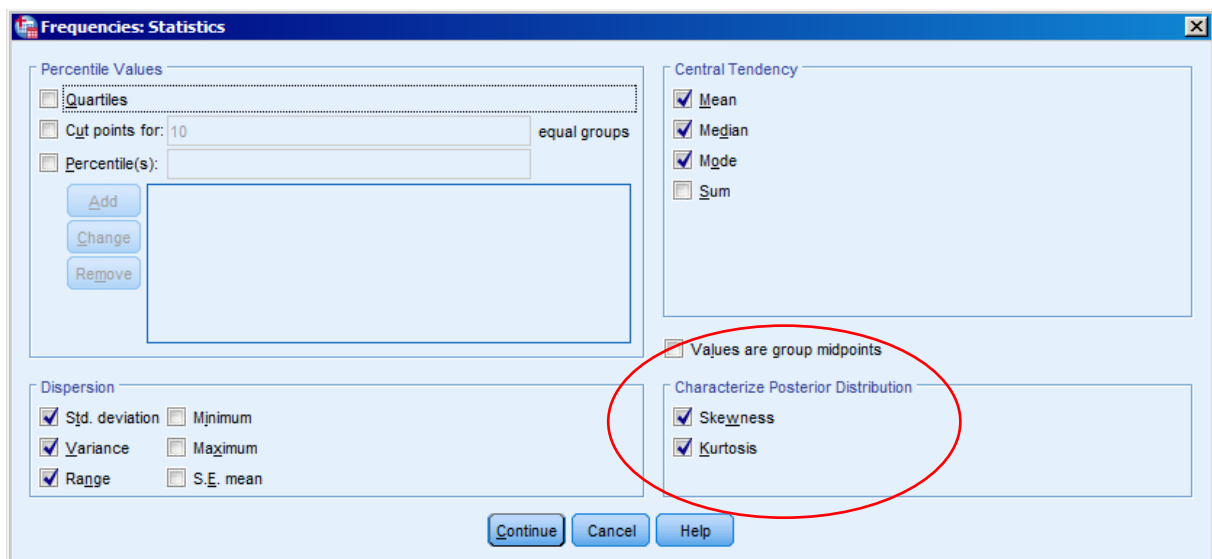
Hoe rapporteren we de resultaten ?

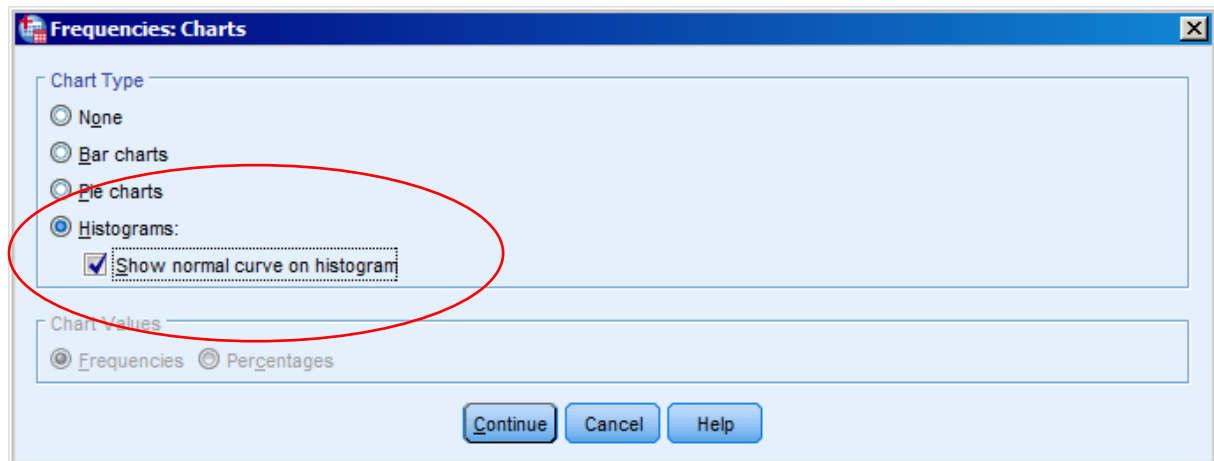
*‘Uit de beschrijvende statistieken kunnen we afleiden dat meisjes ( $M=.81$ ) gemiddeld minder bier/alcoholconsumptie gedurende de laatste 30dagen rapporteren dan jongens ( $M=1.33$ ). De standaardafwijking bij meisjes is lager dan bij jongens (respectievelijk:  $SD: 2.120$  en  $SD: 3.74$ ).*

*Bij jongens worden grotere verschillen waargenomen dan bij meisjes. De variatiebreedte bij meisjes bedraagt 40 en bij jongens 52. Beide groepen kenmerken zich door individuen die geen bier/alcoholconsumptie gedurende de laatste 30 dagen rapporteren, tot individuen die tot 40 verschillende gelegenheden met bier/alcoholconsumpties gedurende de afgelopen 30 dagen rapporteren bij meisjes en tot 52 bij de jongens.'*

### **Parameters van vorm: skewness en kurtosis**

De verdeling van kenmerken in de populatie kan verschillende vormen aannemen. De meest gekende verdeling is de symmetrische verdeling, ook wel normale verdeling genoemd. In het theoretische deel van deze syllabus wordt heel specifiek aandacht besteed aan de eigenschappen ervan. De verdeling van een kenmerk kan echter sterk afwijken van een normale verdeling. Er zijn verschillende manieren om inzicht te krijgen in de mate waarin een waargenomen verdeling van een variabele overeenkomt met een normale verdeling. We kunnen dit doen door parameters of grafieken op te vragen. Parameters die inzicht geven in de overeenkomst met een normale verdeling zijn scheefheid (skewness) en kurtosis.





We vragen een grafische voorstelling op van de variabele ‘bier/alcoholconsumptie gedurende afgelopen 30 dagen’ voor jongens en meisjes apart. We doen dit via *Analyze Descriptive Statistics Frequencies Charts*. Kies voor histogram (aangezien de variabele van het metrische niveau is) en vink aan *show normal curve on histogram*. Op die manier kunnen we visueel inspecteren in welke mate deze verdeling afwijkt van een normale verdeling. We vroegen ook de beschrijvende statistieken voor jongens en meisjes op met inbegrip van de parameters van vorm.

### Statistics

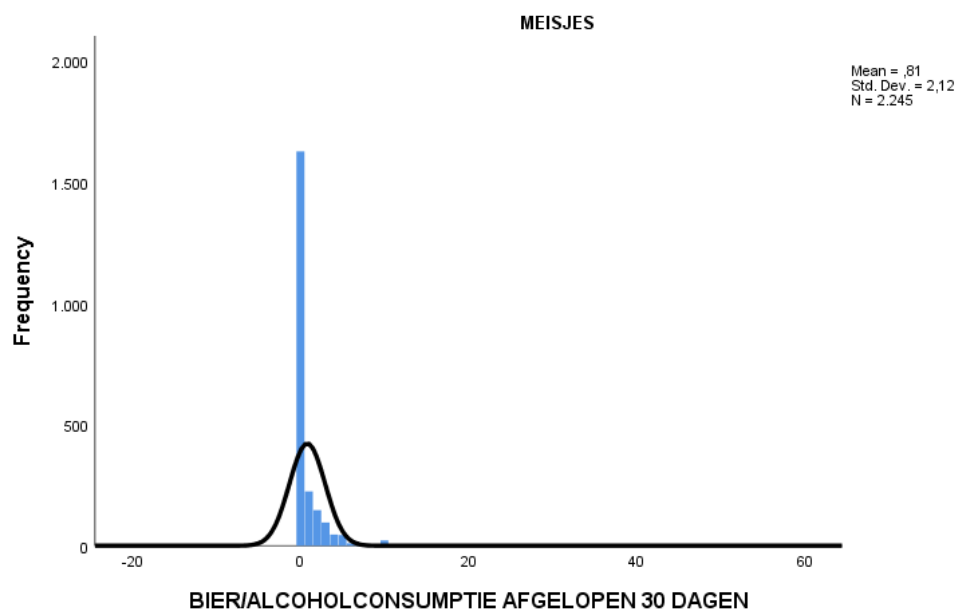
Bier/alcoholconsumptie afgelopen 30dagen

female	N	Valid	2.245
		Missing	125
	Mean		,81
	Median		,00
	Mode		0
	Std. Deviation		2,120
	Variance		4,495
	Skewness		6,266
	Std. Error of Skewness		,052
	Kurtosis		71,199
	Std. Error of Kurtosis		,103
	Range		40
male	N	Valid	2.213
		Missing	162
	Mean		1,33
	Median		,00
	Mode		0
	Std. Deviation		3,738
	Variance		13,975

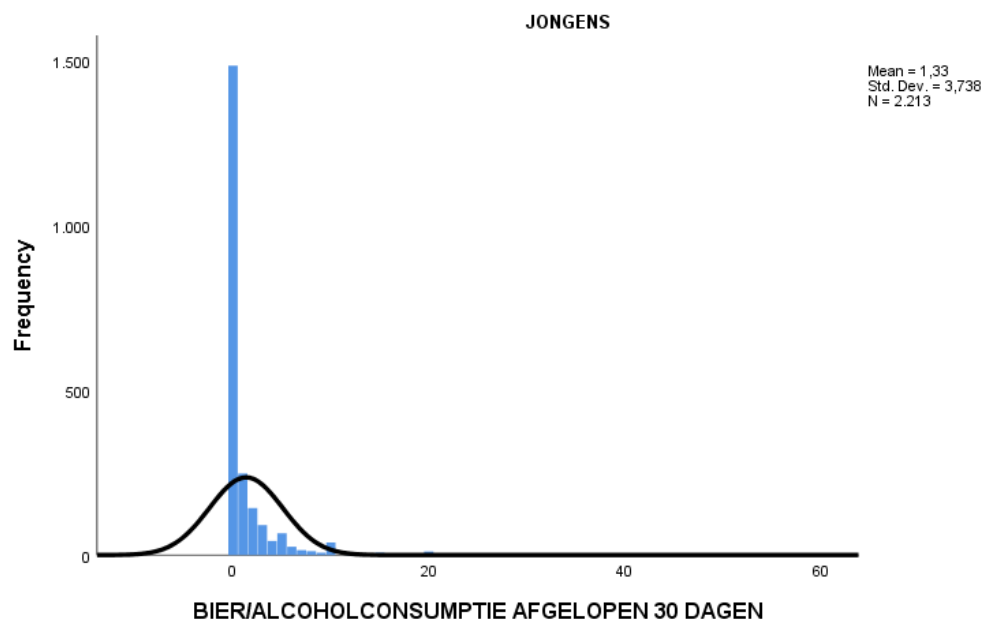
Skewness	7,082
Std. Error of Skewness	,052
Kurtosis	73,355
Std. Error of Kurtosis	,104
Range	52

Onderstaande histogrammen tonen grafisch hoe het kenmerk ‘bier/alcoholconsumptie gedurende afgelopen 30 dagen’ verdeeld is onder jongens en meisje apart. Visuele voorstellingen zijn heel belangrijk. Ze tonen ons in een oogopslag hoe sterk een kenmerk varieert. Onmiddellijk valt op dat de meeste respondenten geen enkele bier/alcoholconsumptie gedurende de afgelopen 30 dagen rapporteren en dit geldt zowel voor jongens als voor meisjes. Daarentegen rapporteert een kleine groep herhaald bier/alcoholconsumptie en ook dit geldt voor beide geslachten. Hoe komt dit ? Welke zijn nu de achtergrondkenmerken die dergelijke verschillen tussen individuen kunnen verklaren ? Of zijn de verschillen eerder toevallig ? Dit zijn vragen die sociaal wetenschappers interesseren.

Het fenomeen is statistisch niet normaal verdeeld. De verdeling neemt geen klokvorm aan. Hoe sterk wijkt deze verdeling dan af van de klokvorm ? De staart van de verdeling kan langer zijn langs de rechterzijde, maar kan ook langer zijn langs de linkerzijde. Dit bestuderen we aan de hand van de scheefheid “skewness”. De verdeling kan ook platter zijn dan de eentoppige normaalverdeling. Dit bestuderen we aan de hand van de “kurtosis”.







### Statistics

Bier/alcoholconsumptie afgelopen 30dagen

female	N	Valid	2.245
		Missing	125
	Skewness		6,266
	Std. Error of Skewness		,052
	Kurtosis		71,199
	Std. Error of Kurtosis		,103
male	N	Valid	2.213
		Missing	162
	Skewness		7,082
	Std. Error of Skewness		,052
	Kurtosis		73,355
	Std. Error of Kurtosis		,104

De skewness (scheefheid) geeft weer in welke mate een verdeling symmetrisch is. Bij de normale verdeling zijn modus, gemiddelde en mediaan gelijk aan elkaar en zijn de beide helften aan weerszijden van het gemiddelde elkaars spiegelbeeld en is de verdeling volledig symmetrisch. Als een variabele normaal verdeeld is, dan is de waarde van de skewness (scheefheid) en kurtosis gelijk aan nul. Positieve waarden voor de scheefheidsparameter en kurtosis parameter betekenen dat een bestudeerd kenmerk rechts schever en rechts platter is dan de situatie bij een normale verdeling, die overigens op het histogram werd aangebracht. We

krijgen hier de belangrijke informatie dat de verdeling van bier/alcoholconsumptie als kenmerk niet statistisch normaal verdeeld is, maar positief asymmetrisch is en dus een langere staart naar rechts heeft. Tevens is de verdeling rechts platter. De kurtosis (welling) is een maatstaf voor de relatieve platheid van de verdeling ten opzichte van de normale verdeling. De kurtosis is positiever als de top hoger ligt en negatiever als de top lager ligt. De normale verdeling heeft een kurtosis van nul. We laten de informatie over de “standard error of skewness” en “standard error of kurtosis” achterwege in dit hoofdstuk.

### ***Zelf uitrekenen van de variantie en standaardafwijking in SPSS***

Om de begrippen standaardafwijking en variantie ten volle te begrijpen, gaan we deze ontleden. De betekenis van deze abstracte begrippen vatten, kan nog altijd het best door deze eens zelf na te rekenen.

**De variantie is de som van de gekwadrateerde afwijkingen van elke observatie tegenover het gemiddelde, gedeeld door n-1.<sup>1</sup> De variantie is dus de variatie gedeeld door n-1. De standaardafwijking is de vierkantswortel uit de variantie.**

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

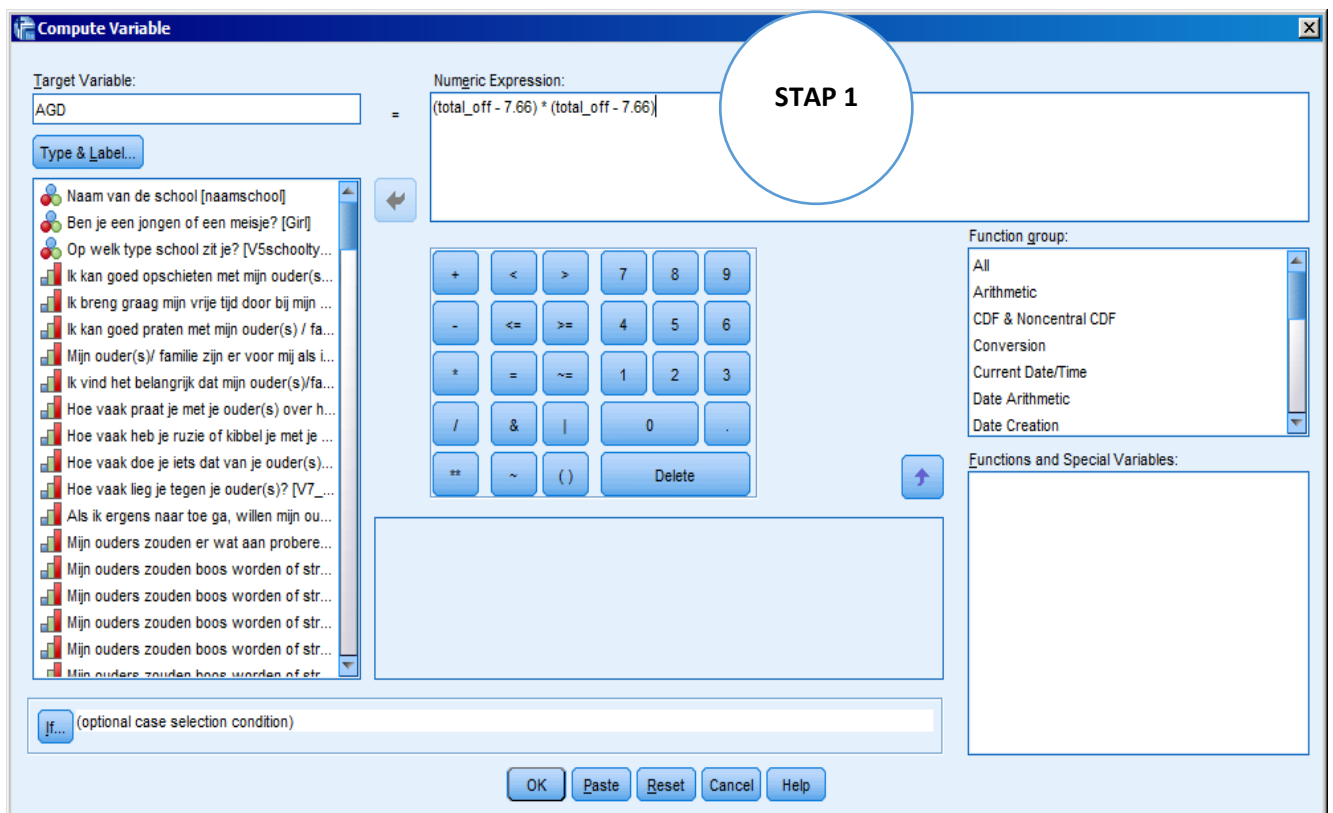
Teller: sum of squares (variatie)

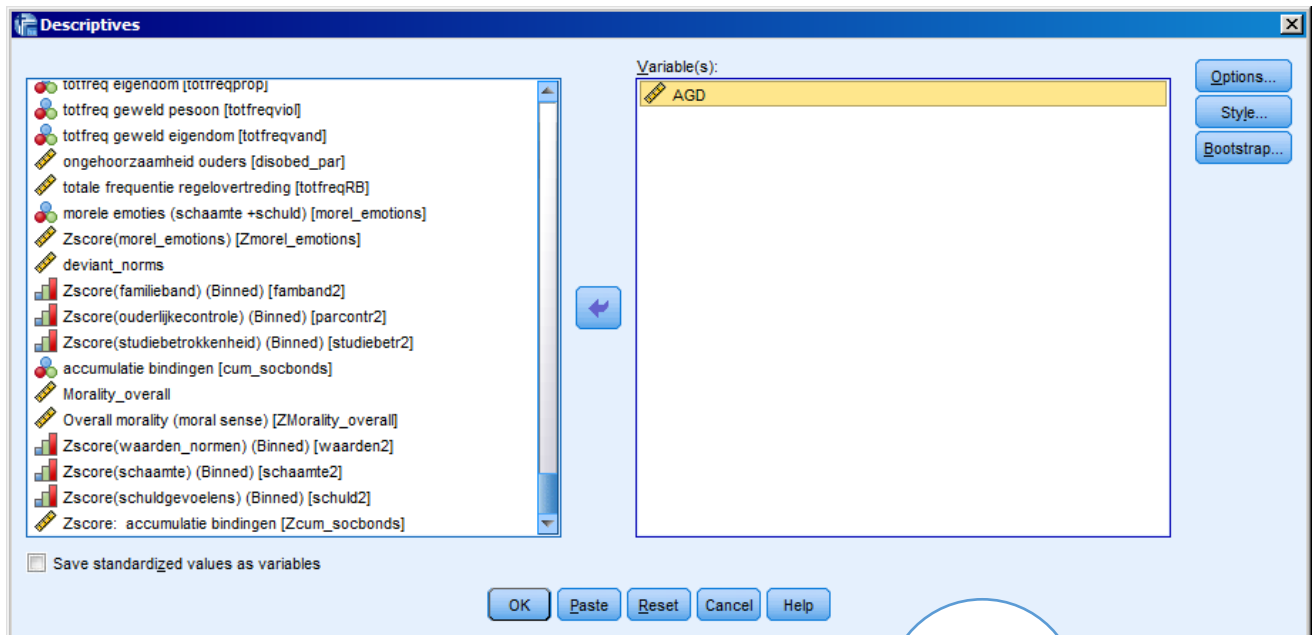
Noemer: n-1

<sup>1</sup> We moeten wel kwadrateren, want de som van de afwijkingen tegenover het gemiddelde is steeds nul. Door het kwadraat van elke afwijking te nemen, krijgen we zicht op hoe groot de afwijking is, de richting speelt immers geen rol (hoger of lager dan het gemiddelde). We nemen n-1 als noemer wanneer de resultaten verkregen worden via een steekproef. Bij populatiegegevens is dit steeds n. In criminologisch onderzoek werken we bijna altijd met steekproeven.

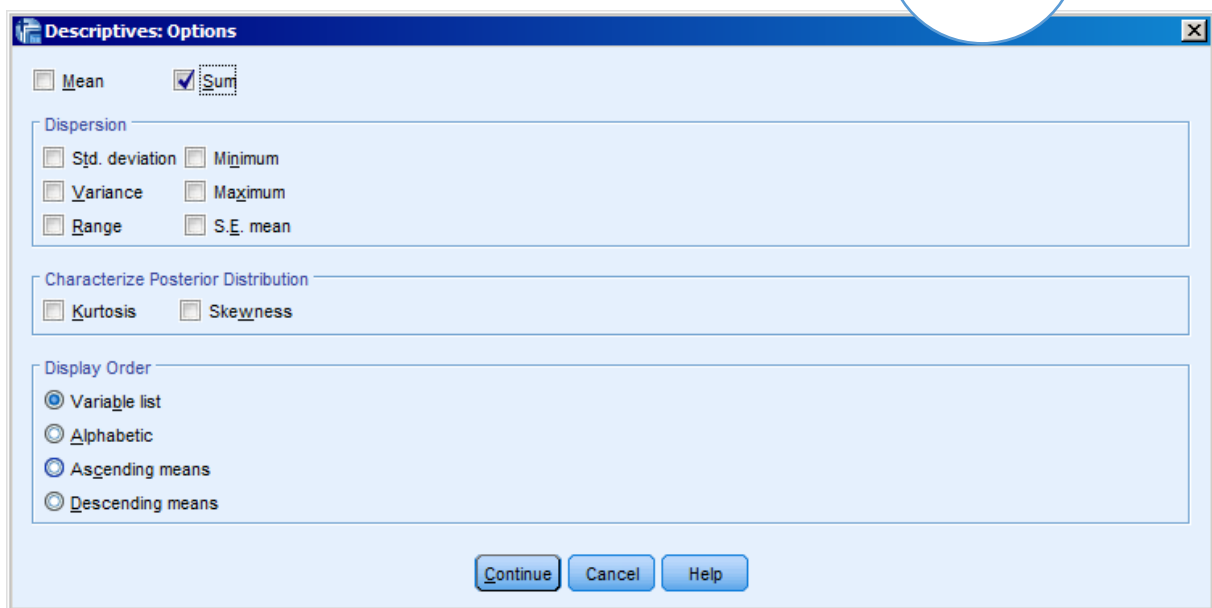
We vroegen aan 815 basisschoolkinderen hoeveel delicten zij al hadden gepleegd. 752 respondenten beantwoordden de vraag geldig. Het gemiddelde aantal gepleegde delicten bedroeg 7.66, de variantie 89.425 en de standaardafwijking 9.456. We kunnen zelf met SPSS narekenen of dit klopt, en dit in verschillende stappen:

- ➔ Het rekenkundig gemiddelde van de variabele 'daderschap' bedraagt: 7.66
- ➔ Laten we nu de afwijkingen van elke observatie tegenover het rekenkundig gemiddelde berekenen aan de hand van SPSS en deze vermenigvuldigen met zichzelf. We hoeven daarvoor een nieuwe variabele aan te maken die we AGD noemen (afwijking gemiddelde daderschap). We klikken op OK en we krijgen een nieuwe variabele (stap 1)
- ➔ Van deze variabele hebben we de som nodig volgens de definitie van de variantie: in SPSS doen we dit als volgt (stap 2)
- ➔ Het resultaat is de **sum of squares** of de variatie of de teller in de formule voor de berekening van de variantie (stap 3)





STAP 2



### Descriptive Statistics

	N	Sum
AGD	752	67.157,89
Valid N (listwise)	752	

STAP 3

- ➔ Delen we de sum of squares door n-1, dan bekomen we:  $67.157,89 / 751 = 89.425$
- ➔ De standaardafwijking is de vierkantswortel uit dit getal en dit bedraagt 9.456.
- ➔ Ter controle zie hieronder de beschrijvende statistieken van de variabele *daderschap*.

Statistics		
<i>daderschap</i>		
N	Valid	752
	Missing	63
Mean		7,66
Std. Deviation		9,456
Variance		89,425

## 6.2. Samenhang tussen twee variabelen

Om inzicht te krijgen in de samenhang tussen twee variabelen kunnen we in SPSS opvragen:

- **Kruistabel** voor de samenhang tussen nominale en ordinale variabelen (via opdracht *Analyze Descriptive Statistics Crosstabs*)
- **Spreidingsdiagram** of **scatterplot** waarbij de onafhankelijke variabele op de X-as wordt geplaatst en de afhankelijke op de Y-as. De waarnemingen worden gepresenteerd in de vorm van een puntenwolk. Spreidingsdiagrammen worden gemaakt via de opdracht *Graphs / Chart Builder* en vereist een minimaal ordinaal meetniveau. Spreidingsdiagrammen geven een goed inzicht in de mate waarin twee variabelen samenhangen en in de vorm van de samenhang (rechtlijnig of kromlijnig)

Samenhang tussen twee variabelen kan ook weergegeven worden in de vorm van een kengetal:

- **Chi-kwadraattoets:** om voor nominale en ordinale variabelen te bepalen of twee variabelen al dan niet onafhankelijk zijn van elkaar (*Analyze Descriptive Statistics Crosstabs Statistics*)
- **Phi en Cramers' V:** om voor nominale variabelen de mate van samenhang te bepalen (*Analyze Descriptive Statistics Crosstabs Statistics*)
- **Spearman's Rho en Kendall's Tau-b:** om voor ordinale variabelen de mate van samenhang te bepalen (*Analyze Descriptive Statistics Crosstabs Statistics*)
- **Pearson correlatiecoëfficiënt :** om voor metrische variabelen de mate van samenhang te bepalen (*Analyze Correlate Bivariate*)
- **Partiële correlatie:** als we willen controleren voor het effect van een derde variabele (*Analyze Correlate Partial*)

***Zelf narekenen van de correlatiecoëfficiënt van Pearson aan de hand van SPSS (fictief voorbeeld)***

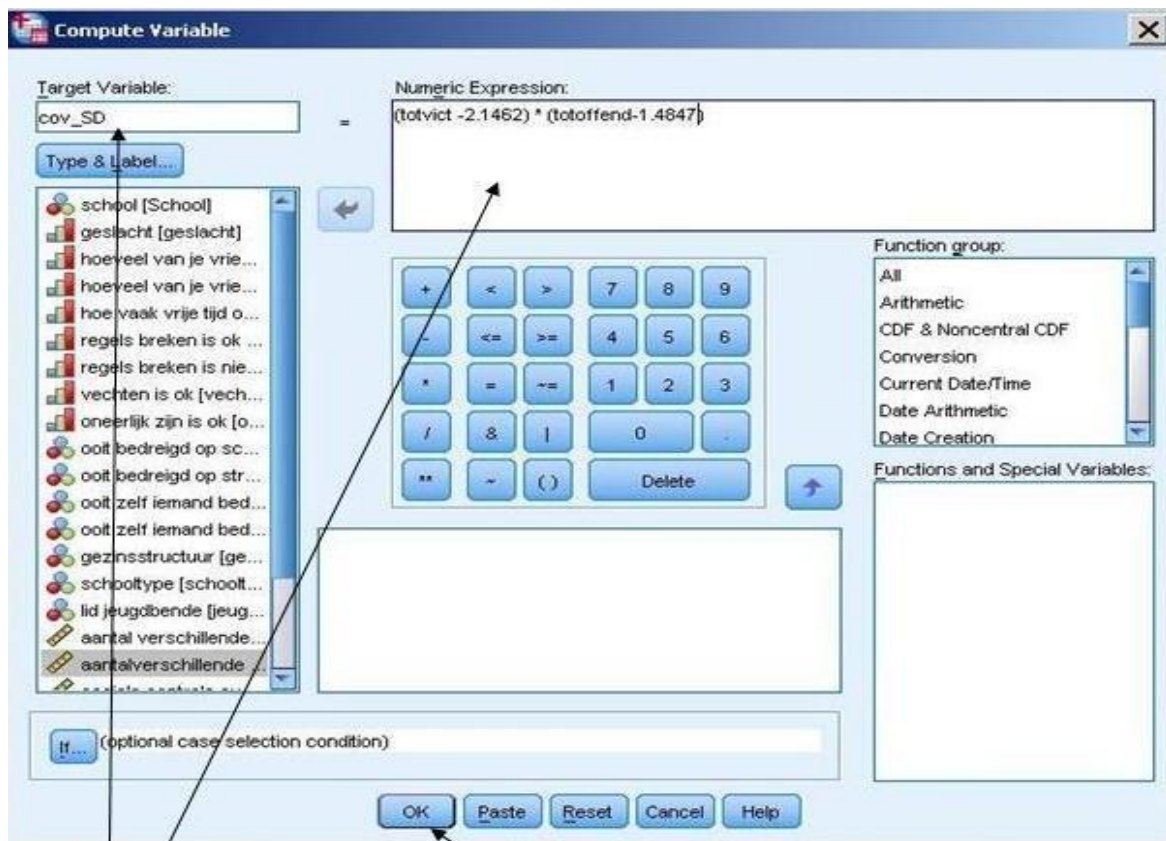
<b>Covariatie of kruisproductensom of sum of squares (SS<sub>xy</sub>)</b>	$SS_{xy} = \sum (X - \bar{X})(Y - \bar{Y})$
<b>Covariantie (S<sub>xy</sub>)</b>	$s_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1}$
<b>Correlatiecoëfficiënt van Pearson (R<sub>xy</sub>)</b>	$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$

In de teller van de formule van de correlatiecoëfficiënt staat de covariatie tussen X en Y. We beginnen dus met de covariatie te nemen tussen X en Y. Daartoe hebben we de gemiddelde waarden nodig van de beide variabelen. Gemiddelden, varianties en standaardafwijkingen verkrijgen we via de procedure ‘descriptives’.

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
aantal verschillende delicten slachtoffer	1491	,00	10,00	2,1462	2,05203	4,211
aantal verschillende delicten gepleegd	1504	,00	13,00	1,4847	1,92052	3,688
Valid N (listwise)	1453					

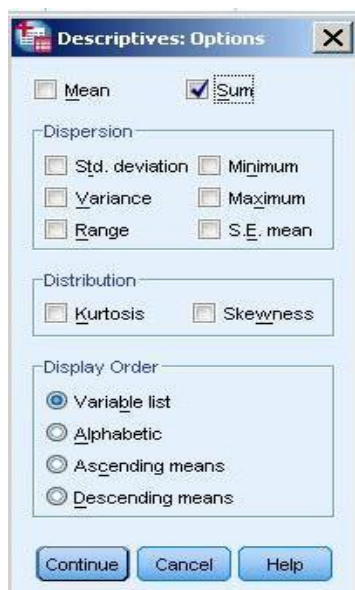
Om de covariatie te berekenen doen we het volgende: we vermenigvuldigen de afwijkingen van beide variabelen tov hun respectievelijke gemiddelden. We gebruiken de procedure “compute”.



Definieer de nieuwe variabele en zeg SPSS hoe te rekenen. Hier berekenen we het product van de individuele afwijkingen van X en Y tov hun gemiddelde

Klik op OK en SPSS maakt de nieuwe variabele aan

We moeten de som nemen van dit product. In SPSS doen we dat via “sum”



De som van de het product van de afwijkingen tov de respectievelijke gemiddelden is de volgende:

Descriptive Statistics		
	N	Sum
cov_SD	1453	2532,45
Valid N (listwise)	1453	

Deze som is de kruisproductensom of covariantie, en moet eerst gedeeld worden door (n-1) om tot de covariantie te komen.

De covariantie is dus  $2532/1452 = 1.743$

Deze waarden moeten we nog delen door het product van de standaardafwijkingen van de variabelen X en Y ofwel  $(2.05 * 1.92) = 3.93 \rightarrow$  daarom is de correlatie  $1.743/3.93 = 0.44$

Zoals je ziet, klopt onze eigen berekening met hetgeen SPSS doet! Met deze stapsgewijze naberekening hopen we dat je de filosofie achter een correlatie-analyse begrijpt. Hoewel we het meeste rekenwerk aan SPSS hebben overgelaten, hebben we je toch trachten te illustreren wat er eigenlijk gebeurt door een eenvoudige druk op de knop. Veel belangrijker is nu de vraag hoe we een correlatiecoëfficiënt inhoudelijk interpreteren vanuit een theoretische criminologische achtergrond. Vergeet niet dat je verwacht wordt als criminoloog een hypothese te kunnen formuleren over de aard van deze samenhang. De associatiemaat is symmetrisch. Maar wat denk je zelf over de relatie tussen daderschap en slachtofferschap? Beschouw je deze statistische relatie als uni-causaal? Zo ja, welke is oorzaak en gevolg? Of denk je eerder in termen van niet uni-causaal? Of gaat het hier om een schijneffect? Zo ja, wat zou dan de gemeenschappelijke oorzaak kunnen zijn? Wat denk je op basis van de criminologische theoretische bagage waarover je op dit moment beschikt?

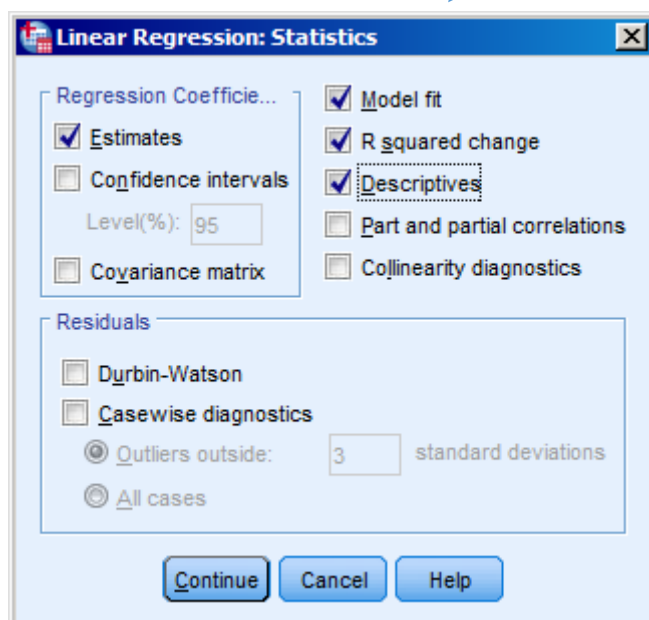
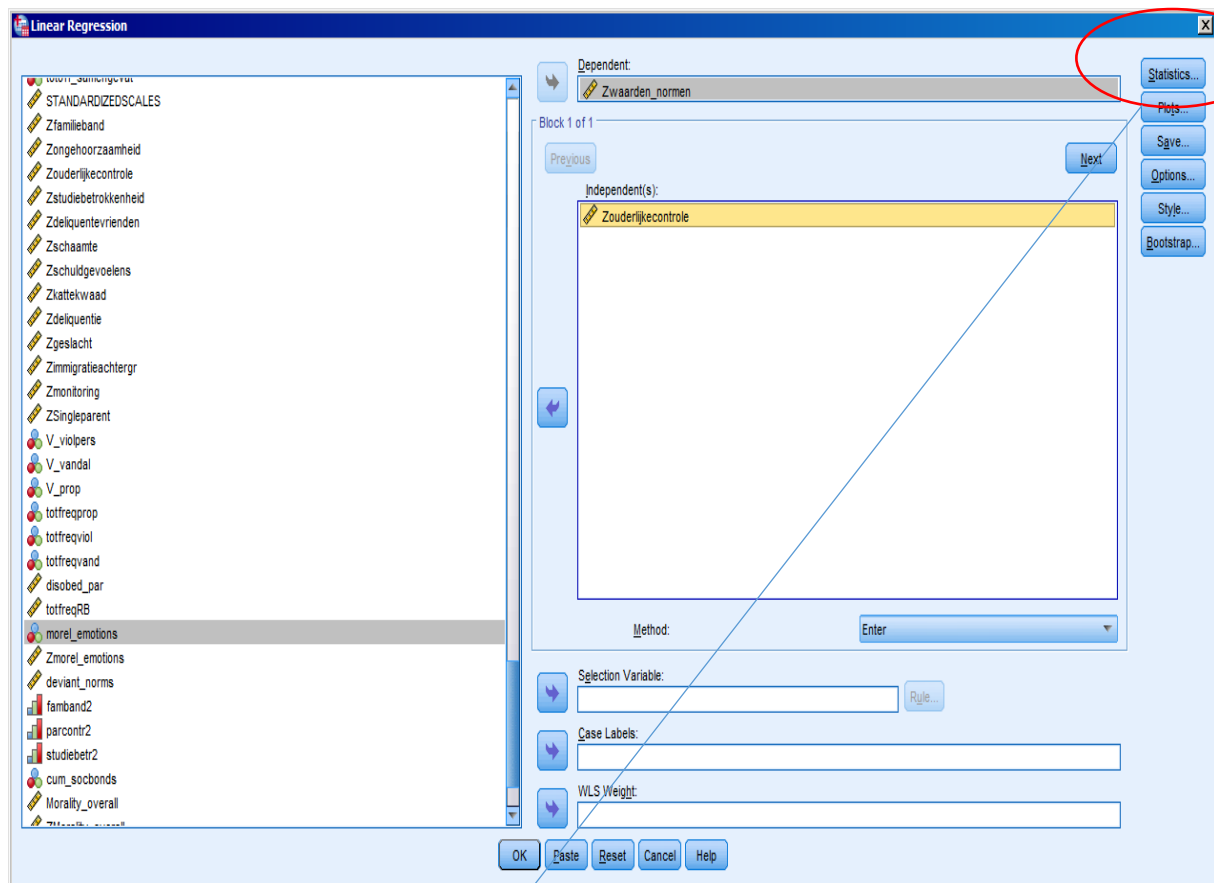


### 6.3. Verklaren van een variabele op basis van één of meerdere onafhankelijke variabelen

Om te toetsen of er een lineair verband bestaat tussen een afhankelijke variabele en één of meerdere onafhankelijke variabelen wordt een regressie analyse gebruikt (respectievelijk enkelvoudige lineaire regressie of meervoudige of multiple lineaire regressie). De regressie analyse levert een vergelijking op waarmee de afhankelijke variabele numeriek kan worden verklaard. Belangrijke assumptie is dat er sprake is van een lineair (rechtlijnig) verband tussen metrische variabelen (interval of ratio-niveau). Een lineaire regressie analyse voeren we in SPSS uit via de opdracht *Analyze Regression Linear*.

The screenshot displays the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the 'Regression' option is selected. The 'Linear...' option is highlighted. The variable list on the left includes 'geanticipeerde schuld' and various 'M\_V' variables. The bottom status bar shows 'Linear...' and 'IBM SPSS Statistics Processor is ready'.

Name	Label	Values	Missing	Columns	Align	Measure	Role
120	guilt	None	None	17	Right	Scale	Input
121	total_offending	None	None	17	Right	Scale	Input
122	kattekwaad	None	None	12	Right	Scale	Input
123	deliquentie	None	None	13	Right	Scale	Input
124	tot_off2	None	None	10	Right	Nominal	Input
125	counttot_off	None	None	14	Right	Nominal	Input
126	M_samenwonen	one	None	14	Right	Nominal	Input
127	M_V6	one	None	10	Right	Nominal	Input
128	M_V7	one	None	10	Right	Nominal	Input
129	M_V8	one	None	10	Right	Nominal	Input
130	M_V9	one	None	10	Right	Nominal	Input
131	M_V10	one	None	10	Right	Nominal	Input
132	M_V11	one	None	10	Right	Nominal	Input
133	M_V12	one	None	10	Right	Nominal	Input
134	M_V13	one	None	10	Right	Nominal	Input
135	M_V14	one	None	10	Right	Nominal	Input
136	totoff_samengevat	one	None	19	Right	Nominal	Input
137	STANDARDIZEDSCALES	one	None	8	Right	Unknown	Input
138	Zfamilieband	one	None	14	Right	Scale	Input
139	Zongehooraamheid	one	None	19	Right	Scale	Input
140	Zouderlijkecontrole	one	None	21	Right	Scale	Input
141	Zstudiebetrokkenheid	Zscore(studiebetrokkenheid)	None	22	Right	Scale	Input
142	Zdeliquentenvrienden	Zscore(deliquentenvrienden)	None	21	Right	Scale	Input
143	Zwaarden_normen	Zscore(waarden_normen)	None	17	Right	Scale	Input
144	Zschaamte	Zscore(schaamte)	None	13	Right	Scale	Input
145	Zschuldgevoelens	Zscore(schuldgevoelens)	None	18	Right	Scale	Input
146	Zkatttekwaad	Zscore(katttekwaad)	None	13	Right	Scale	Input
147	Zdeliquentie	Zscore(deliquentie)	None	14	Right	Scale	Input
148	Zgeslacht	geslacht	None	13	Right	Scale	Input
149	Zimmigratieachtergr	Zscore(immigratieachtergr)	None	21	Right	Scale	Input
150	Zmonitoring	Zscore(monitoring)	None	13	Right	Scale	Input
151	ZSingleparent	Zscore(Singleparent)	None	15	Right	Scale	Input
152	V_violpers	varscale violence persone	None	12	Right	Nominal	Input
153	V_vandal	varscale vandalisme	None	10	Right	Nominal	Input
154	V_prop	varscale eigendomsdelicten	None	10	Right	Nominal	Input
155	totfreqprop	totfreq eigendom	None	11	Right	Nominal	Input
156	totfreqviol	totfreq geweld pesoon	None	11	Right	Nominal	Input
157	totfreqvand	totfreq geweld eigendom	None	11	Right	Nominal	Input
158	risicobed_nac	onnehooorzaamheid ouders	None	12	Right	Scale	Input



**Model Summary**

		Adjusted R Square		R Square		Change Statistics			
Model	R	R Square		Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	,386 <sup>a</sup>	,149	,148	,92471461	,149	129,326	1	739	,000

a. Predictors: (Constant), Zouderlijkecontrole Zscore(ouderlijkecontrole)

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	110,587	1	110,587	129,326	,000 <sup>b</sup>
	Residual	631,917	739	,855		
	Total	742,503	740			

a. Dependent Variable: Zwaarden\_normen Zscore(waarden\_normen)

b. Predictors: (Constant), Zouderlijkecontrole Zscore(ouderlijkecontrole)

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,005	,034		-,140	,889
	Zouderlijkecontrole Zscore(ouderlijkecontrole)	,386	,034	,386	11,372	,000

a. Dependent Variable: Zwaarden\_normen Zscore(waarden\_normen)

*‘Uit de resultaten van de bivariate lineaire regressie-analyse blijkt dat ‘ouderlijke controle’ een matig sterk en positief effect heeft op ‘morele normen en waarden’ ( $\beta$ : ,386). Dit effect is significant ( $t=11.372$ ,  $p < 0,001$ ). 14.8% van de variantie in ‘morele normen en waarden’ kan verklaard worden door ‘ouderlijke controle’ ( $R^2$ : ,148). ‘*