Erasmus MC
University Medical Center Rotterdam

SPSS Introductiecursus

Sanne Hoeks

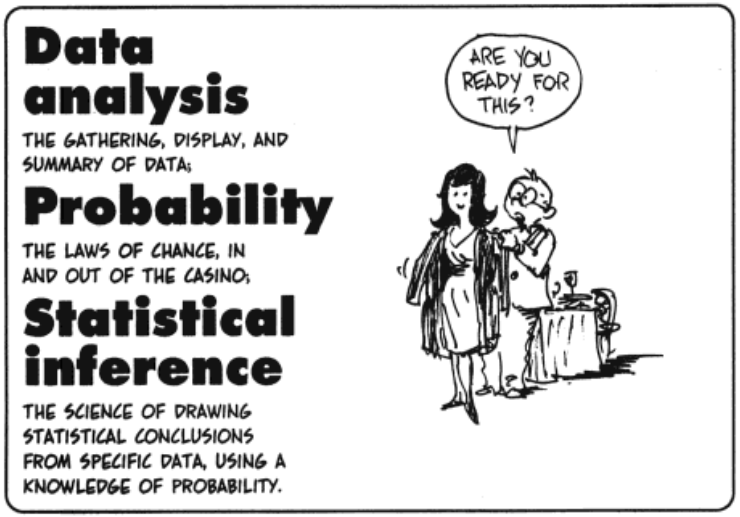
Mattie Lenzen

Statistiek, waarom?

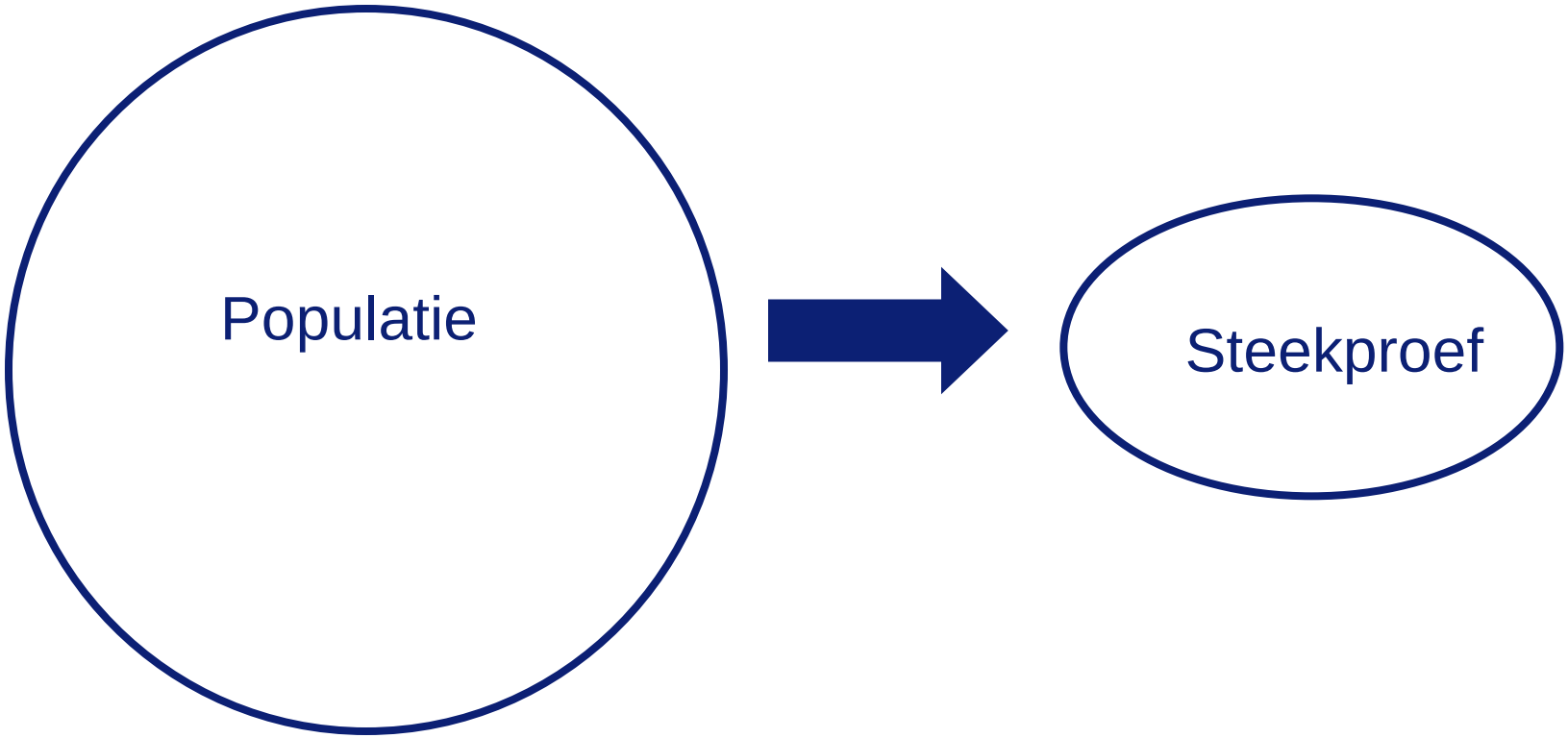
- Doel van het onderzoek om nieuwe feiten van de werkelijkheid vast te stellen door middel van systematisch onderzoek en empirische verzamelen van kwantitatieve informatie.

- Dit proces omvat meestal het samplen van patiënten en meten van bepaalde karakteristieken. Met zoveel metingen over vele patiënten, wordt het onmogelijk om elke patiënt afzonderlijk te onderzoeken.

- Statistiek is een verzameling van wetenschappelijke methoden die zich bezighoudt met getallen, een samenvatting van de gemeenschappelijke kenmerken van de steekproefpopulatie, en onderzoeken van onderliggende patronen.



Steekproef



+ Volledige informatie
- Niet haalbaar

+ Haalbaar
- Verlies van informatie

Populaties en steekproef

- Interesse gaat meestal uit naar de totale populatie (goed definiëren).
- Het is meestal niet haalbaar onderzoek te verrichten onder de totale populatie.
- Oplossing → een (representatieve) steekproef van de populatie.
- Conclusies gebaseerd op een steekproef gelden idealiter voor de totale populatie (externe validiteit)

Dataverzameling

- Meten/verzamelen van relevante kwantitatieve informatie van deelnemers aan het onderzoek.
- Voorbeeld gegevensverzameling:
  - vragenlijsten (KvL)
  - pijnscores
  - bloedonderzoek
  - lichamelijk onderzoek

Het soort gegevens dat je wilt verzamelen bepaald hoe om te gaan met deze gegevens, selectie van geschikte statistische toetsen en het interpreteren van de resultaten.

Data Analyse

Bij bewerken van onderzoeksgegevens zijn drie belangrijke stappen te onderscheiden:

1) Organiseren van gegevens

Data invoer, controle en transformatie

2) Beschrijvende Statistiek

Krijgen van inzicht in populatie en verkregen gegevens

3) Toetsende Statistiek

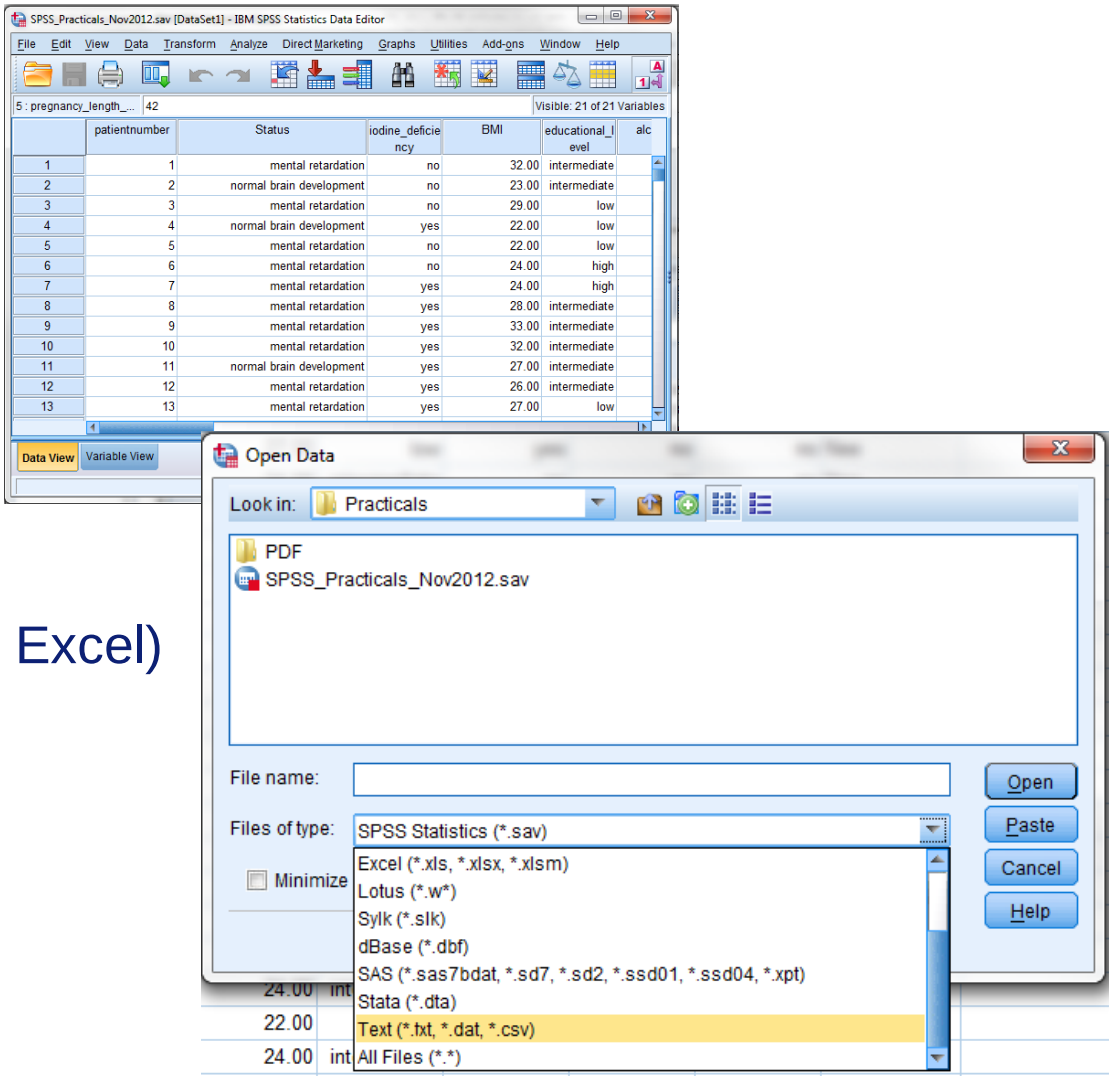
Testen van hypothese en trekken van conclusie(s) die verder gaat dan de onderzochte populatie (indien steekproef representatief)

SPSS windows

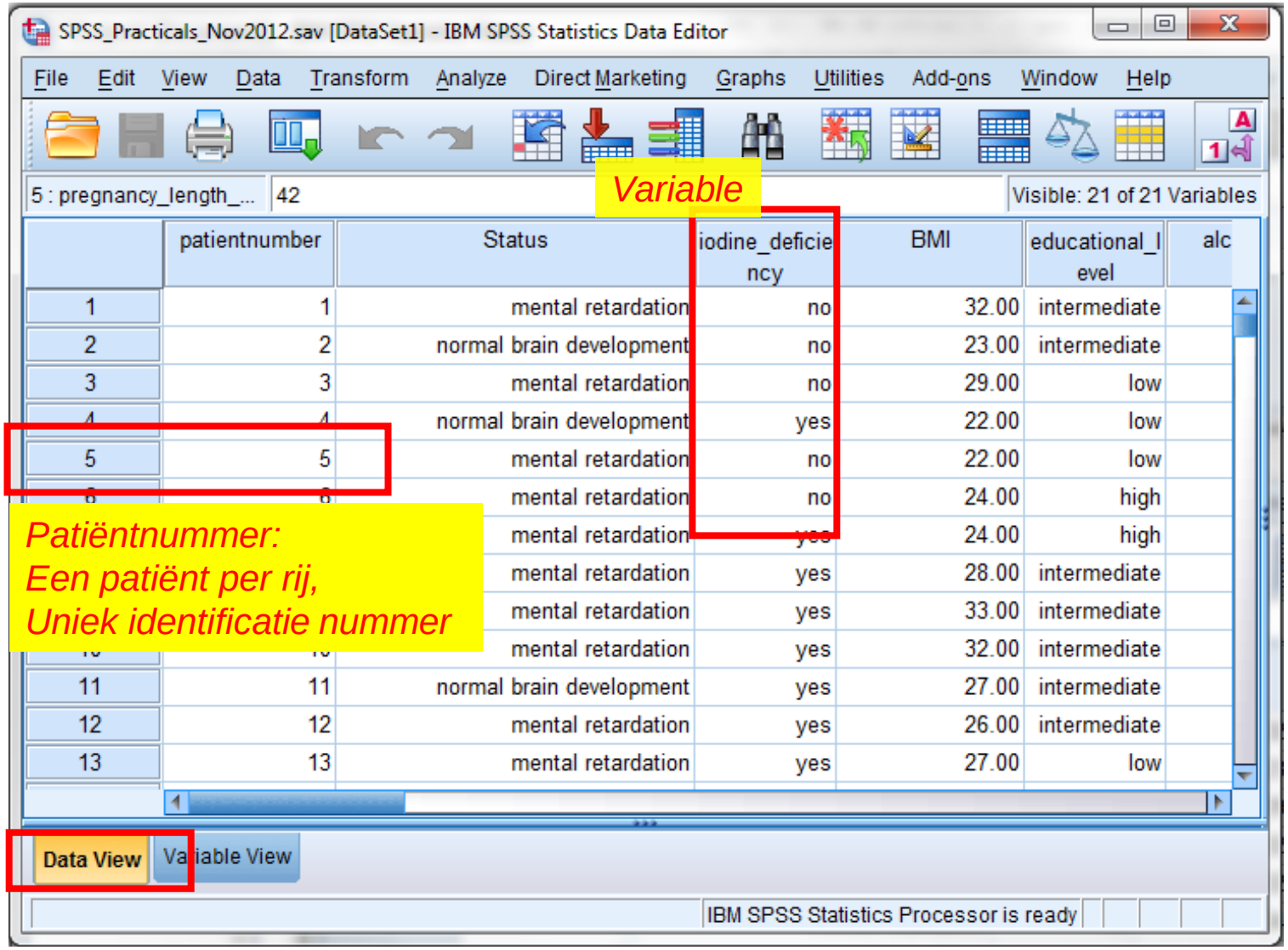
- **Data Editor *(.sav)***
  - Data matrix
  - Variable View
- **Syntax Editor *(.sps)***
  - Commando's
- **Viewer *(.spo)***
  - Resultaten van de analyses

Data Entry

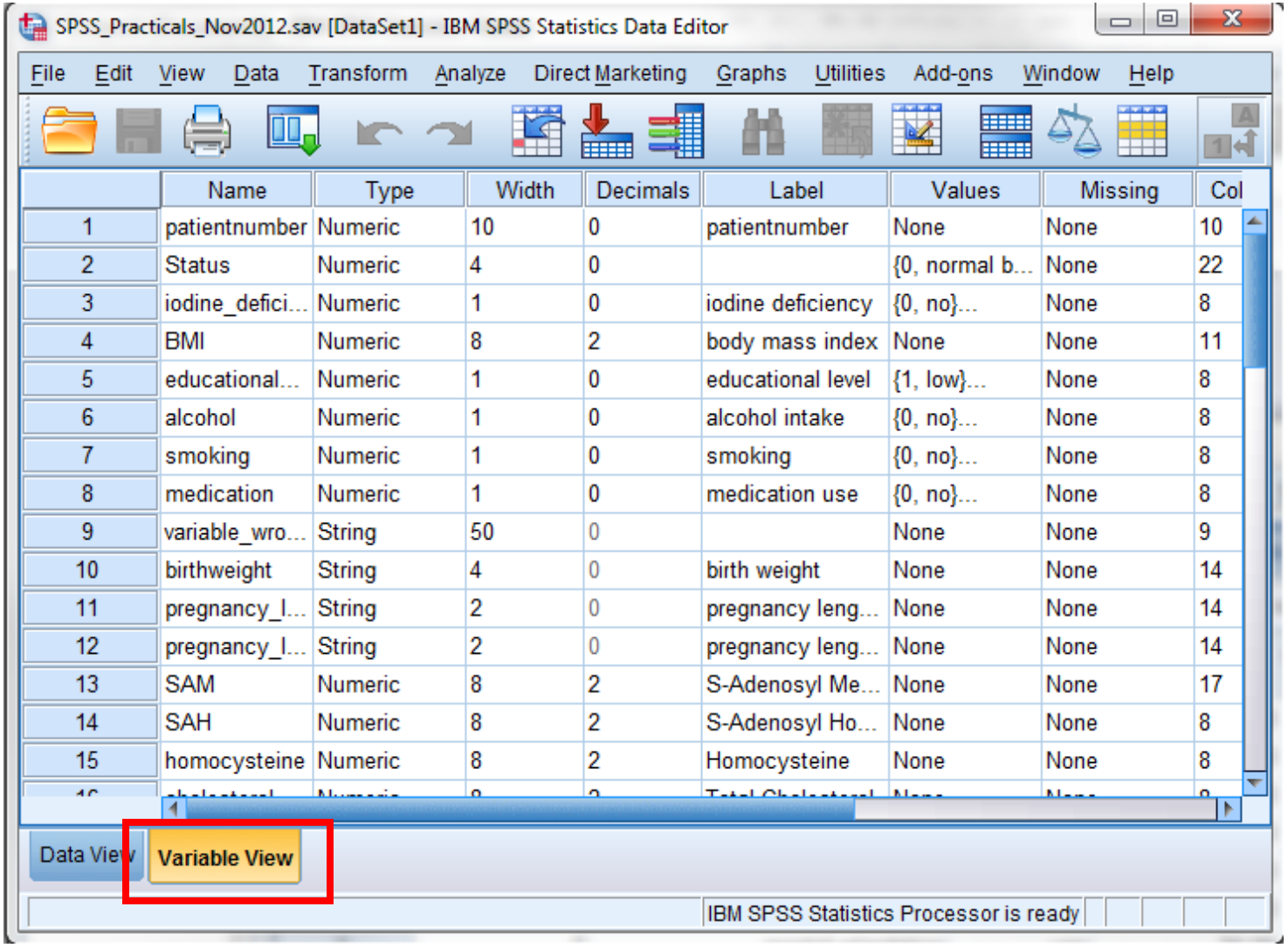
Inlezen data uit een ander bestand (bijv. Excel)



SPSS – Data View

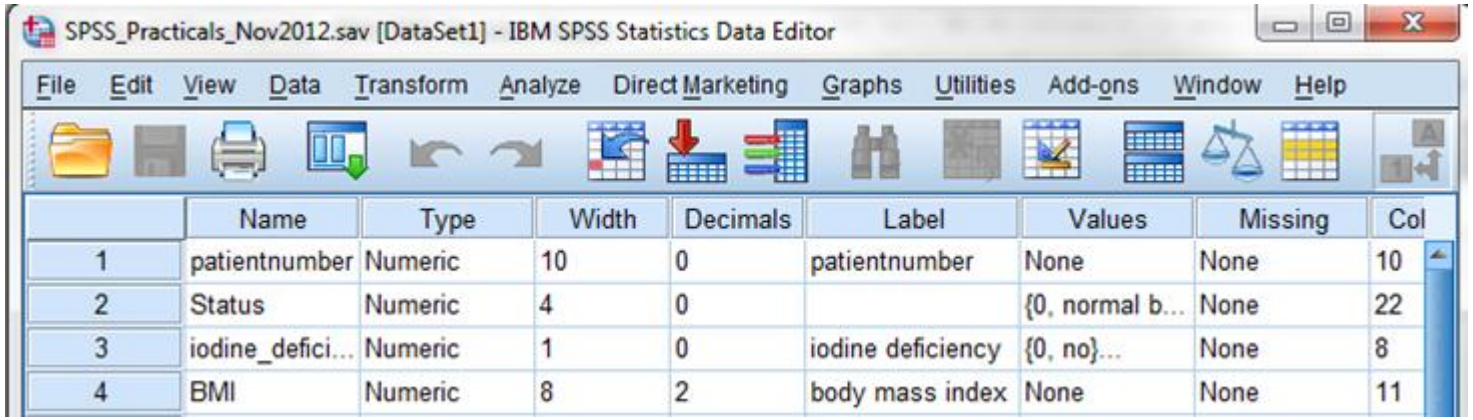


SPSS – Variable View

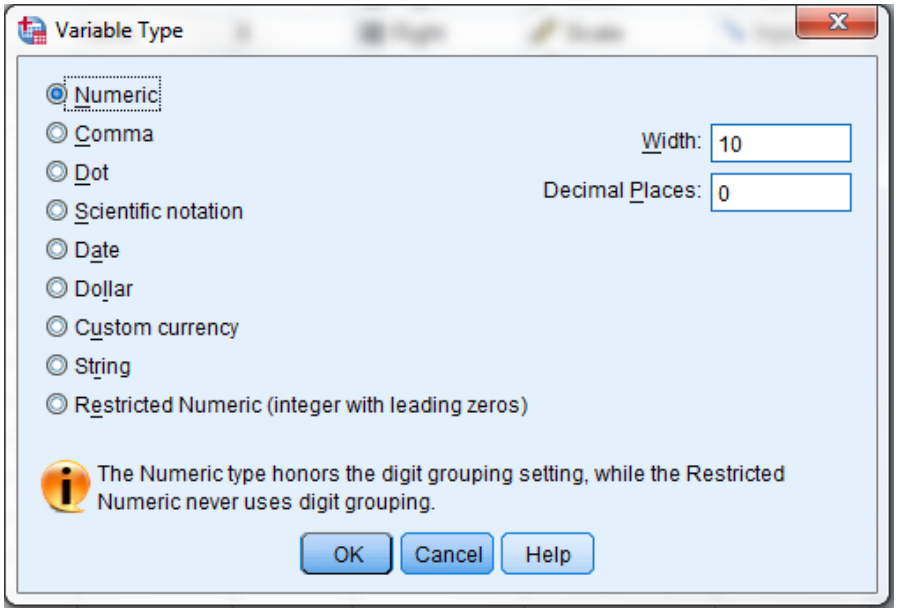


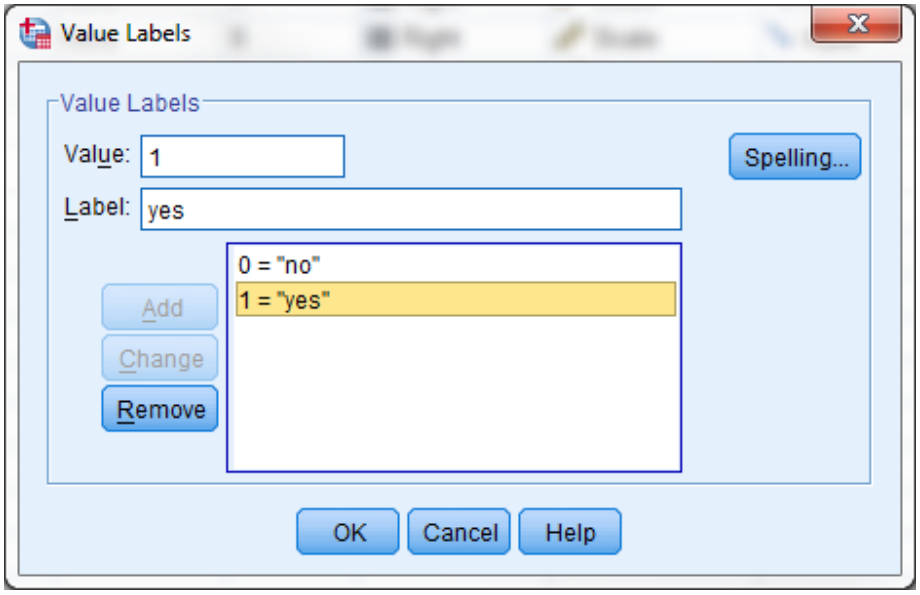
	Name	Type	Width	Decimals	Label	Values	Missing	Col
1	patientnumber	Numeric	10	0	patientnumber	None	None	10
2	Status	Numeric	4	0		{0, normal b...	None	22
3	iodine_defici...	Numeric	1	0	iodine deficiency	{0, no}...	None	8
4	BMI	Numeric	8	2	body mass index	None	None	11
5	educational...	Numeric	1	0	educational level	{1, low}...	None	8
6	alcohol	Numeric	1	0	alcohol intake	{0, no}...	None	8
7	smoking	Numeric	1	0	smoking	{0, no}...	None	8
8	medication	Numeric	1	0	medication use	{0, no}...	None	8
9	variable_wro...	String	50	0		None	None	9
10	birthweight	String	4	0	birth weight	None	None	14
11	pregnancy_l...	String	2	0	pregnancy leng...	None	None	14
12	pregnancy_l...	String	2	0	pregnancy leng...	None	None	14
13	SAM	Numeric	8	2	S-Adenosyl Me...	None	None	17
14	SAH	Numeric	8	2	S-Adenosyl Ho...	None	None	8
15	homocysteine	Numeric	8	2	Homocysteine	None	None	8

SPSS – variable properties



	Name	Type	Width	Decimals	Label	Values	Missing	Col
1	patientnumber	Numeric	10	0	patientnumber	None	None	10
2	Status	Numeric	4	0		{0, normal b...	None	22
3	iodine_defici...	Numeric	1	0	iodine deficiency	{0, no}...	None	8
4	BMI	Numeric	8	2	body mass index	None	None	11





Beschrijven van onderzoeksgegevens

- Uiteindelijk zal de dataverzameling resulteren in een enorme hoeveelheid gegevens van individuele patiënten en gegevens per patiënt.

- Beschrijvende statistiek wordt gebruikt om de verzamelde gegevens in een onderzoek te presenteren. Het betreft meestal een eenvoudige opsomming van kenmerken van de onderzochte populatie (steekproef) en verrichte metingen.

- Met behulp van simpele grafische analyses vormen zij de basis van vrijwel elke kwantitatieve analyse.

Distributie van de steekproef

- Een variabele heeft een bestaande populatie en een steekproefdistributie
- De distributie geeft informatie over het aantal keer dat een variabele voorkomt in de populatie / steekproef
- Bij een willekeurige steekproef, mag worden aangenomen dat dit representatief is voor de (onbekende) populatie

Beschrijvende en toetsende statistiek

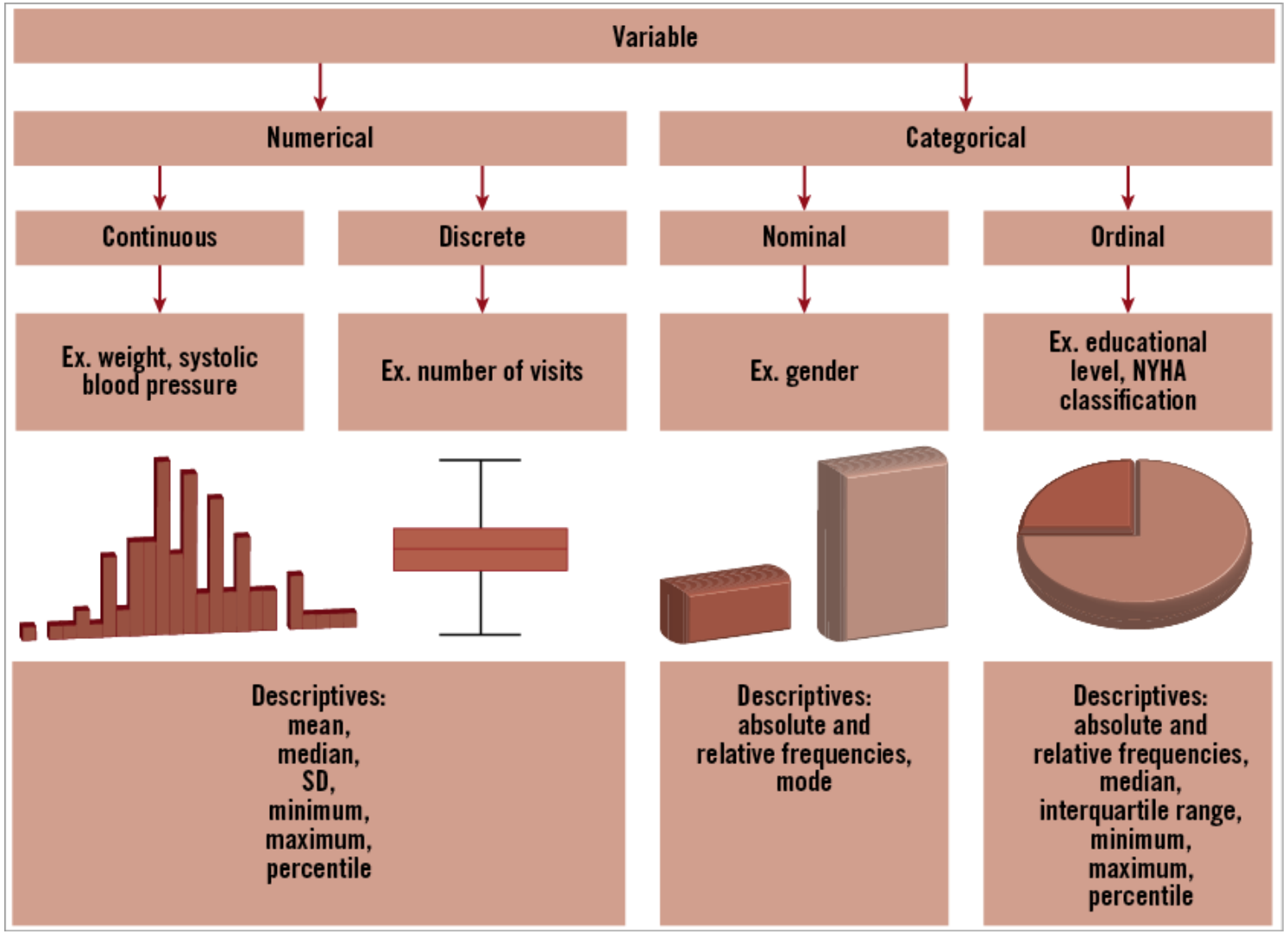
- **Beschrijvende statistiek**

Samenvatting van de gegevens van de steekproef (o.a. gemiddelde, standaard deviatie, tabellen, figuren)

- Toetsende statistiek

Statements over onbekende populatie parameters (o.a. betrouwbaarheidsinterval, statistische toetsen, p-waarde)

Variabelen/ metingen



Variabelen / metingen

- Categorische (kwalitatieve)
  De waarde van een variabele heeft geen numerieke waarde, bijv: *geslacht (0=man, 1=vrouw), opleidingsniveau (1=basis, 2=middelbaar, 3=hoger)*

- Numeriek (kwantitatief)
  leeftijd, bloeddruk, tellen

Categorische data

- Nominaal

  Geen natuurlijke ordening van categorieën

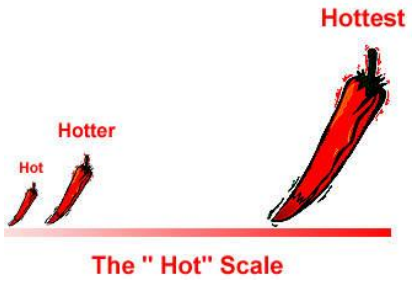
  geslacht, roken/niet-roken (dichotoom)
  bloedgroep, gehuwd/ alleenstaand/ gescheiden (polytoom)

- Ordinaal

  Natuurlijke ordening

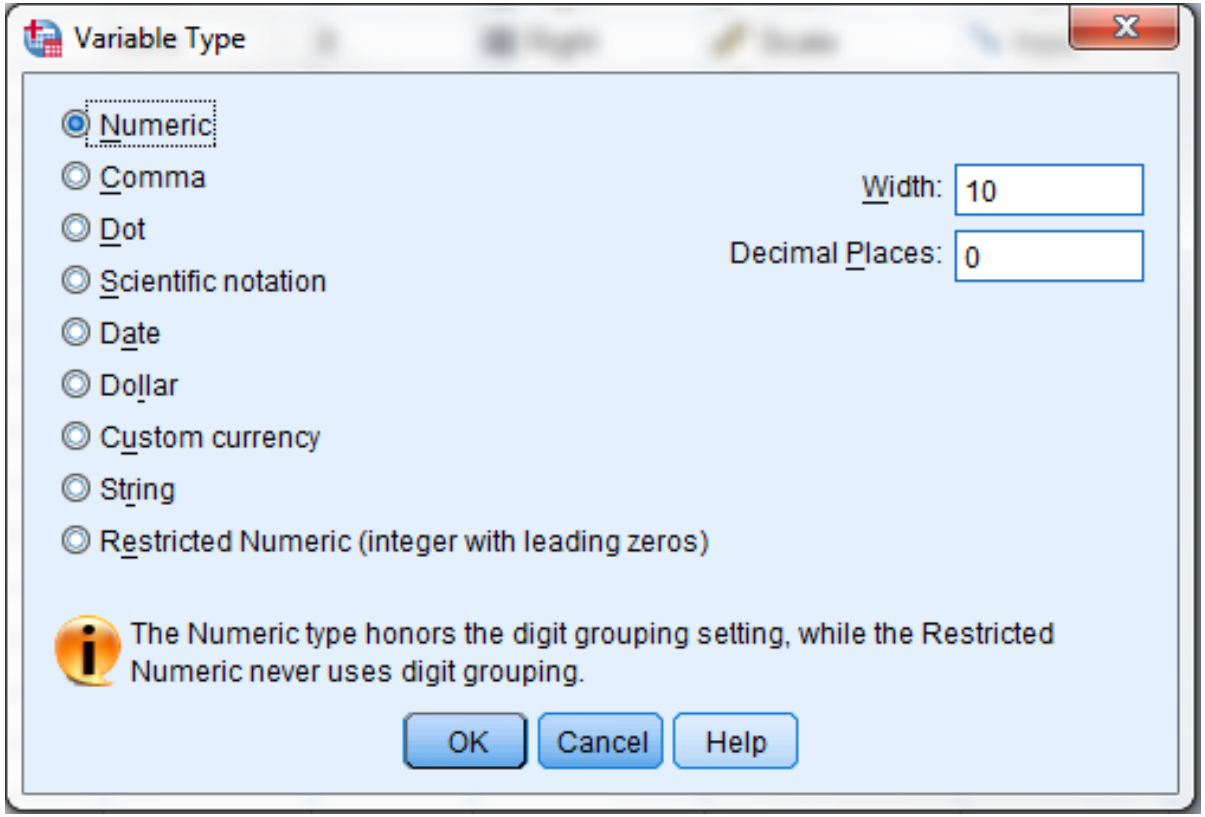
  opleidingsniveau: basis/middelbaar/ hoger sociale klasse: laag/midden/hoog

  NYHA klasse: I, II, III, IV



Hoe in te voeren SPSS?

- Categorisch / numeriek = Numeric
- Tekst = String



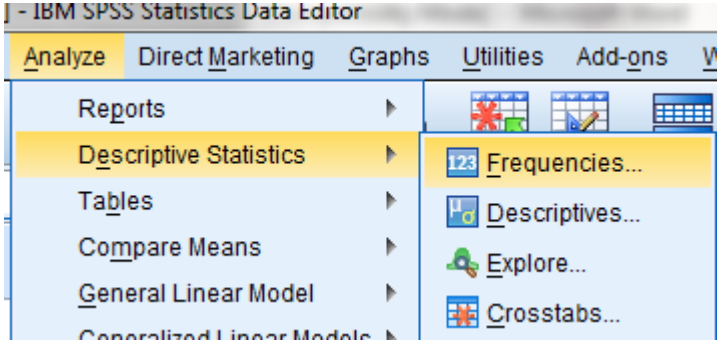
Categorische variabelen (presentatie)

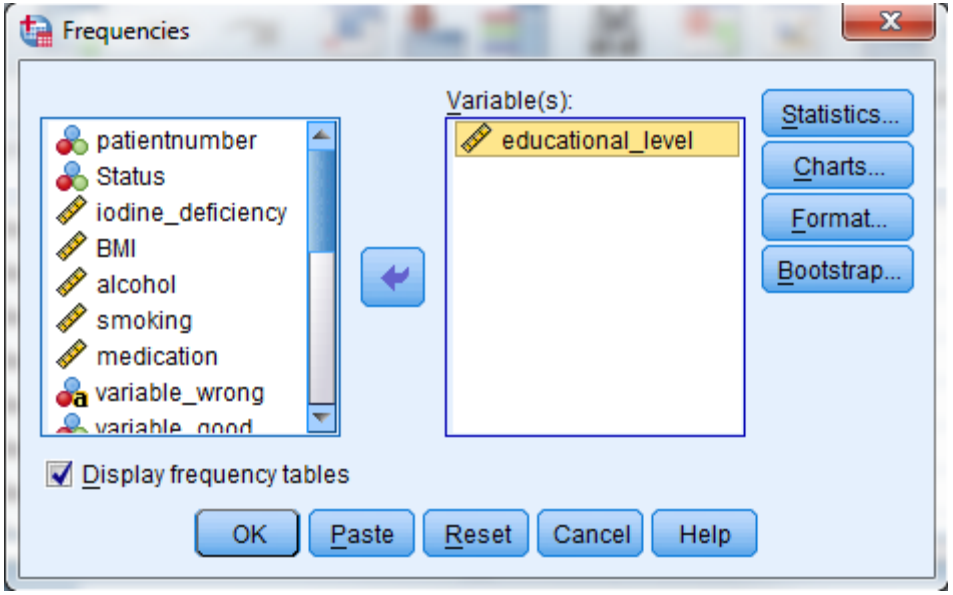
- Berekenen van het absolute, relatieve & cumulatieve relatieve frequenties

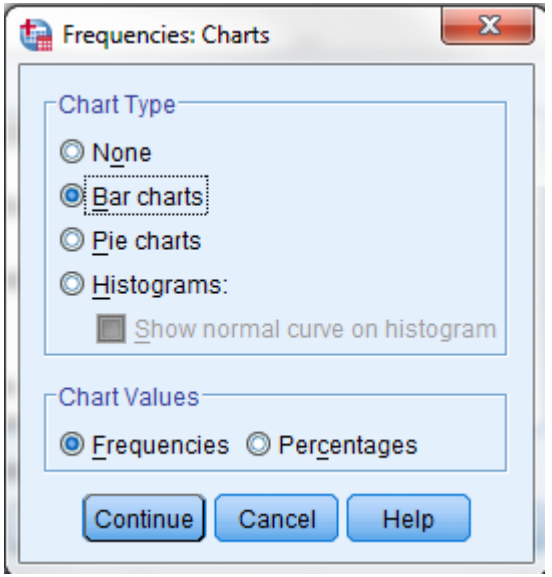
Opleidings-niveau	Absolute Frequentie	Relatieve Frequentie	Cumulatieve Relatieve Frequentie
Laag	39	39 / 181 = 21.6%	21.6%
Middelbaar	88	88 / 181 = 48.6%	21.6 + 48.6 = 70.2%
Hoger	54	54 / 181 = 29.8%	70.2 + 29.8 = 100%
Totaal	181	100%	100%

Categorische Variabele – SPSS

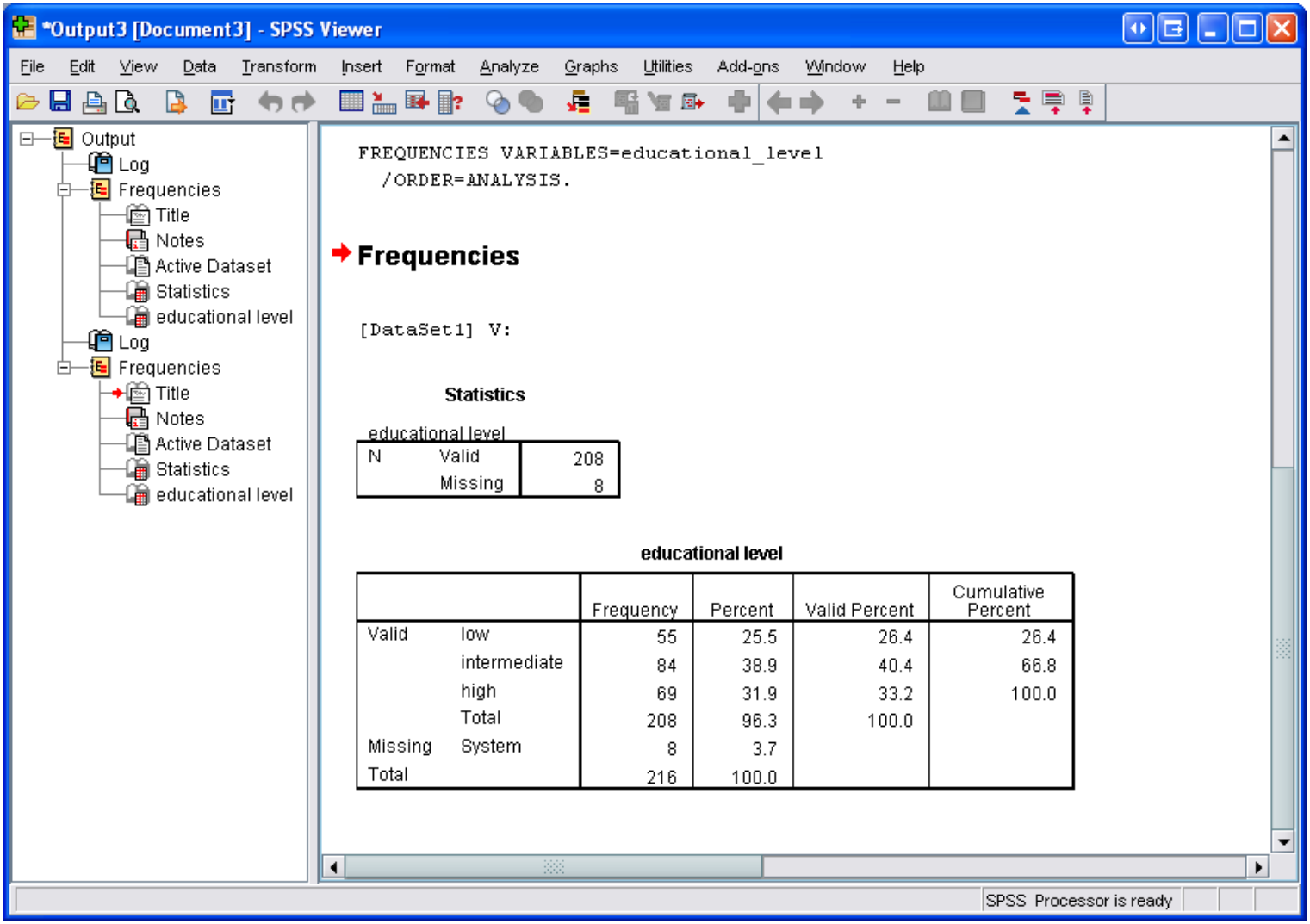
- Analyze – Descriptive Statistics – Frequencies







Categorische variabele – SPSS output



Variabelen / metingen

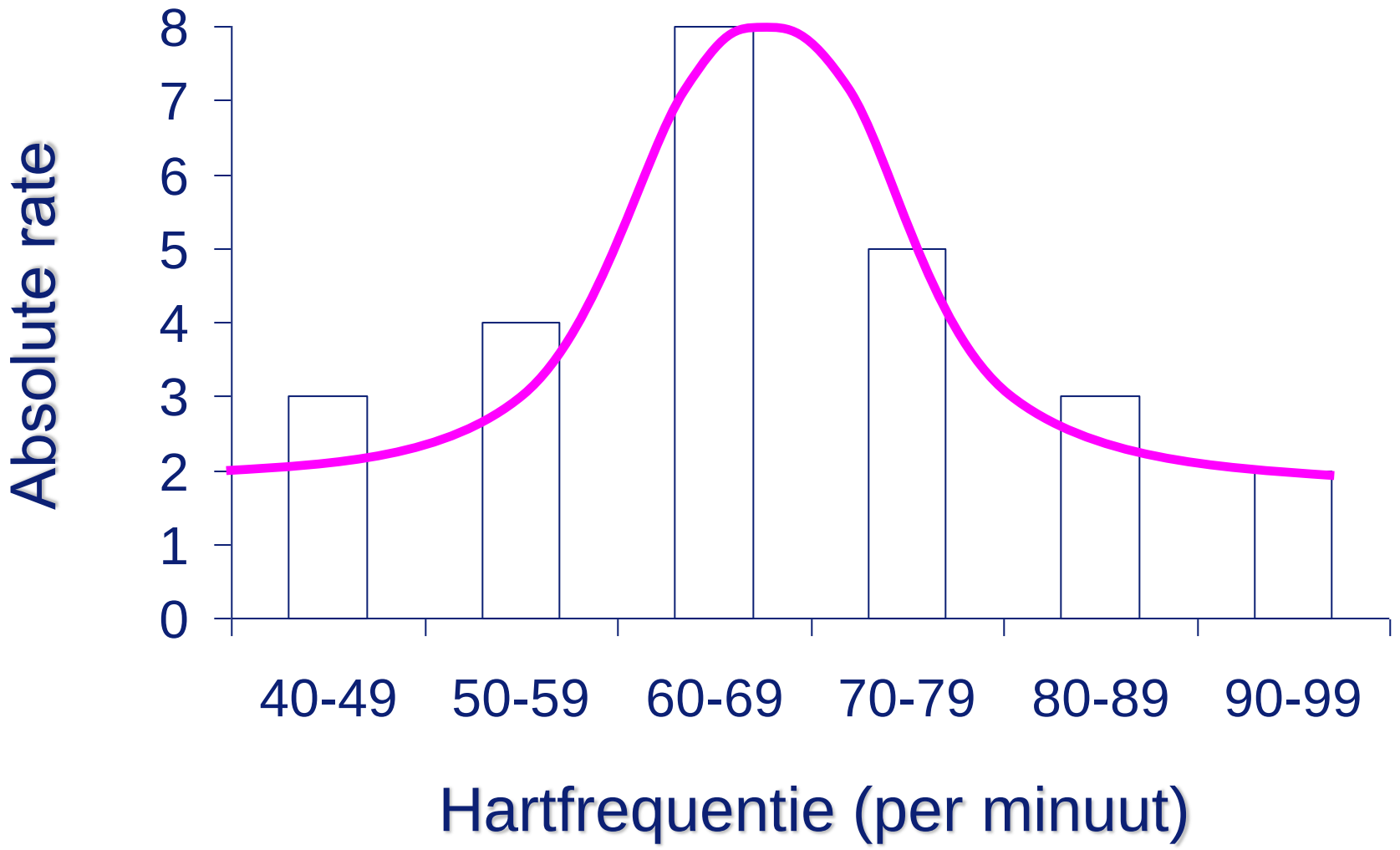
- Categorische (kwalitatieve)

  De waarde van een variabele heeft geen numerieke waarde, bijv:
  geslacht (0=man, 1=vrouw),
  opleidingsniveau (1=basis, 2=middelbaar, 3=hoger)

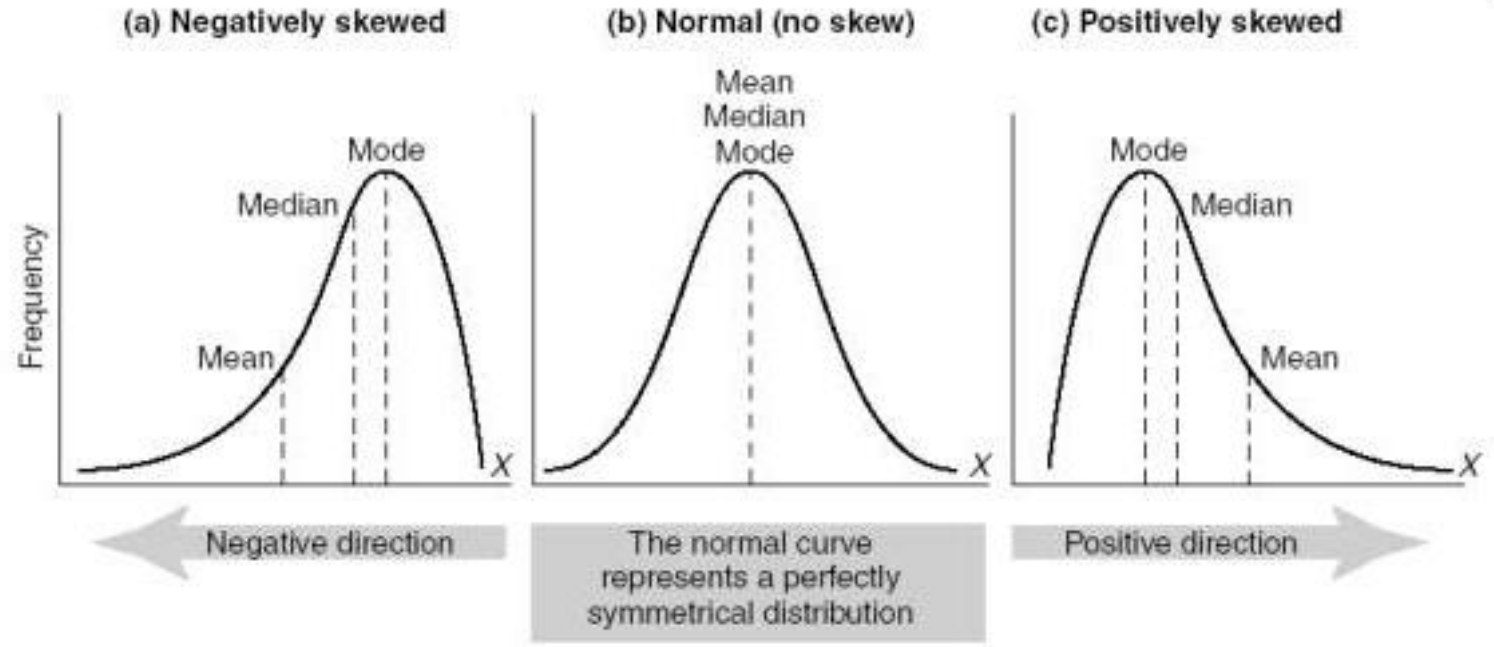
- Numeriek (kwantitatief)

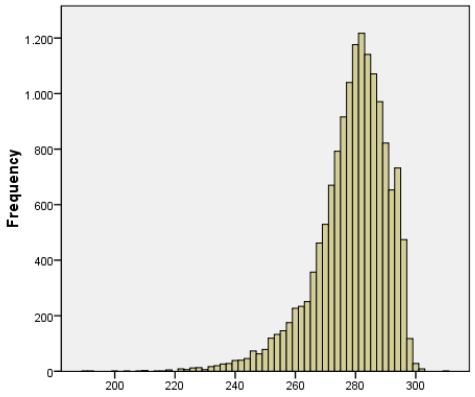
  Discrete variabele (beperkt aantal variabelen en daartussen niets)
  bijv: a*antal cursisten bij deze CNE*
  Continue variabele (kan elke waarde aannemen, incl. decimalen)
  bijv: *leeftijd, BMI, bloedverlies*

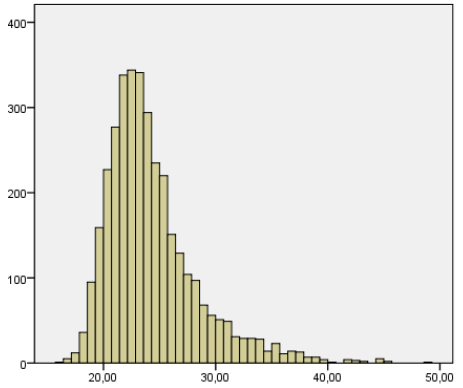
Voorbeeld: Hartfrequentie



Verdeling: normaal en scheef







Normale verdeling

Klokvormige verdeling, die voldoet aan:

- Gemiddelde in het midden.
- Hoe verder van het gemiddelde hoe lager de frequentie.
- Afstand tot gemiddelde is gelijk verdeeld over beide kanten (symmetrie).

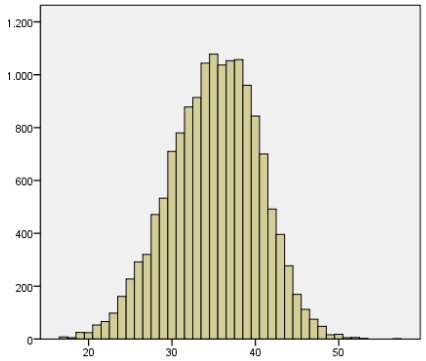
Normale verdeling is:

- Geschikt om distributie van een grote variatie aan variabelen te beschrijven (bijv: leeftijd, lengte, gewicht, RR, etc.)
- ‘Normale’ verdeling = ‘Gaussian’ – distribution
- Aanname voor uitvoeren van parametrische toetsen!

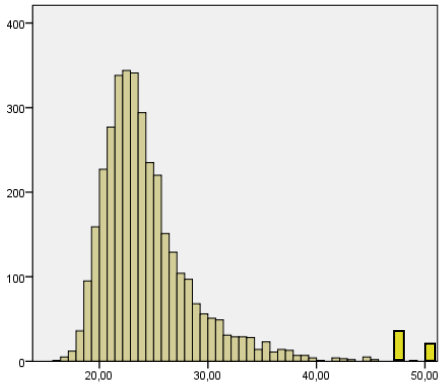
Hoe resultaten te presenteren?

Vuistregel:

Bij normaal verdeelde variabelen wordt het gemiddelde en de standaard deviatie gepresenteerd



Wanneer er sprake is van een scheve verdeling (skewed) en/of veel uitschieters, wordt de mediaan met percentielen (met name de interquartiel range) gepresenteerd



Normaliteitstoets

- Quick-and-dirty

  Steekproef gemiddelde = mediaan van steekproef

- Grafische weergave

  Histogram

- Formele statistische toets

  Kolmogorov-Smirnov

  Shapiro-Wilk

Voorbeeld

	inkomsten	log10
1	1000	3,00
2	1001	3,00
3	1002	3,00
4	999	3,00
5	998	3,00
6	1000	3,00
7	1001	3,00
8	1002	3,00
9	1035	3,01
10	997	3,00
11	1000	3,00
12	1001	3,00
13	1003	3,00
14	999	3,00
15	998	3,00
16	1002	3,00
17	965	2,98
18	1002	3,00
19	1001	3,00
20	1001000	6,00
mean	51000	3,15

Berekenen van het gemiddelde

- Populatie gemiddelde

$$\mu = \frac{\text{som van x aantallen in de populatie}}{\text{populatie grootte N}}$$

- Steekproef gemiddelde

$$\bar{x} = \frac{\text{som van x aantallen in de steekproef}}{\text{steekproef grootte n}}$$

- Aanname: variabele laat een normale distributie zien

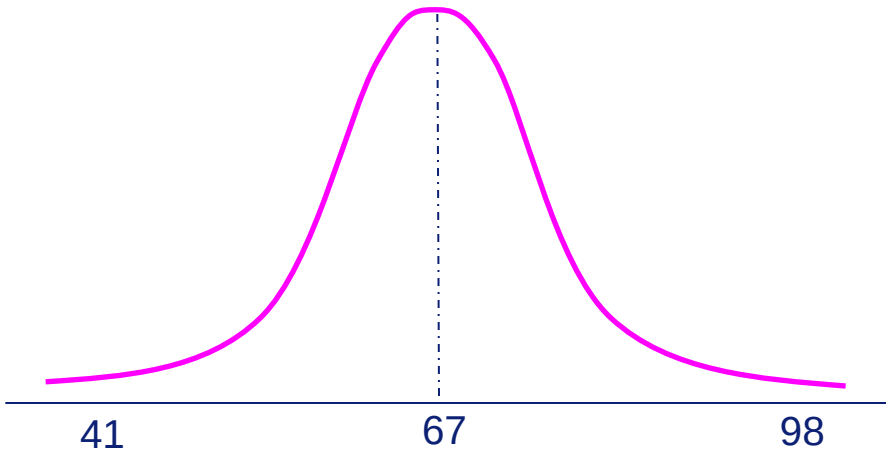
Gemiddelde berekenen

41 + 44 + 49 + 50 + 51
52 + 58 + 61 + 63 + 66
66 + 66 + 67 + 67 + 68
70 + 71 + 73 + 74 + 79
80 + 84 + 88 + 91 + 98 = 1676

$\overline{X} = \Sigma X_i / n$

1676 / 25 = 67,08

Mean = 67,08

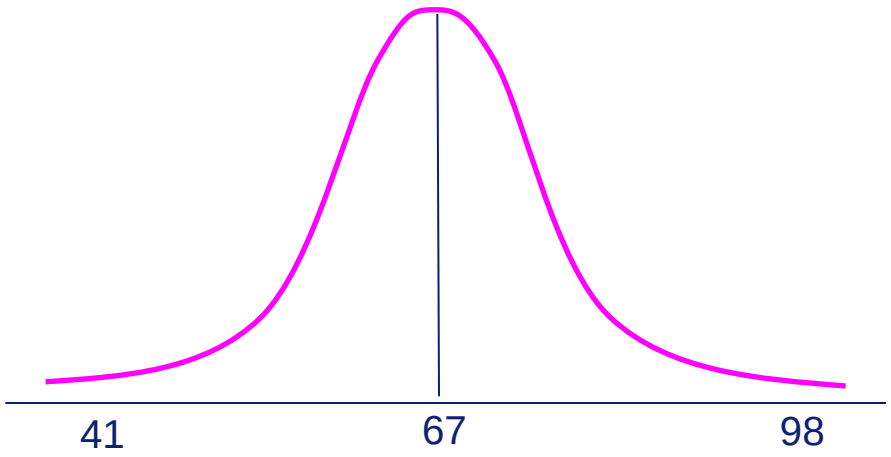


Gemiddelden – Mediaan

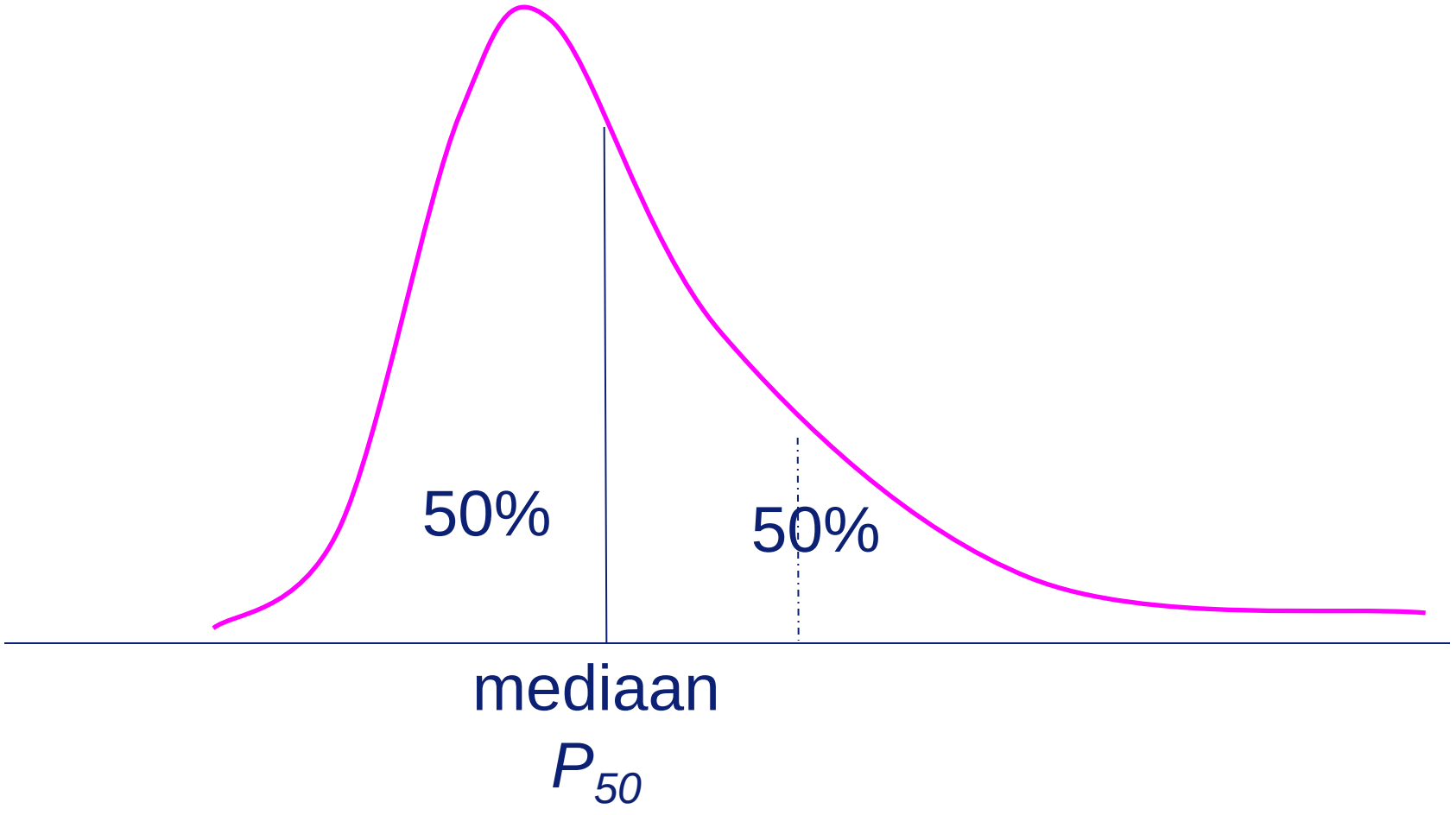
41	44	49	50	51
52	58	61	63	66
66	66	67	67	68
70	71	73	74	79
80	84	88	91	98

In dit voorbeeld:

gemiddelde = mediaan

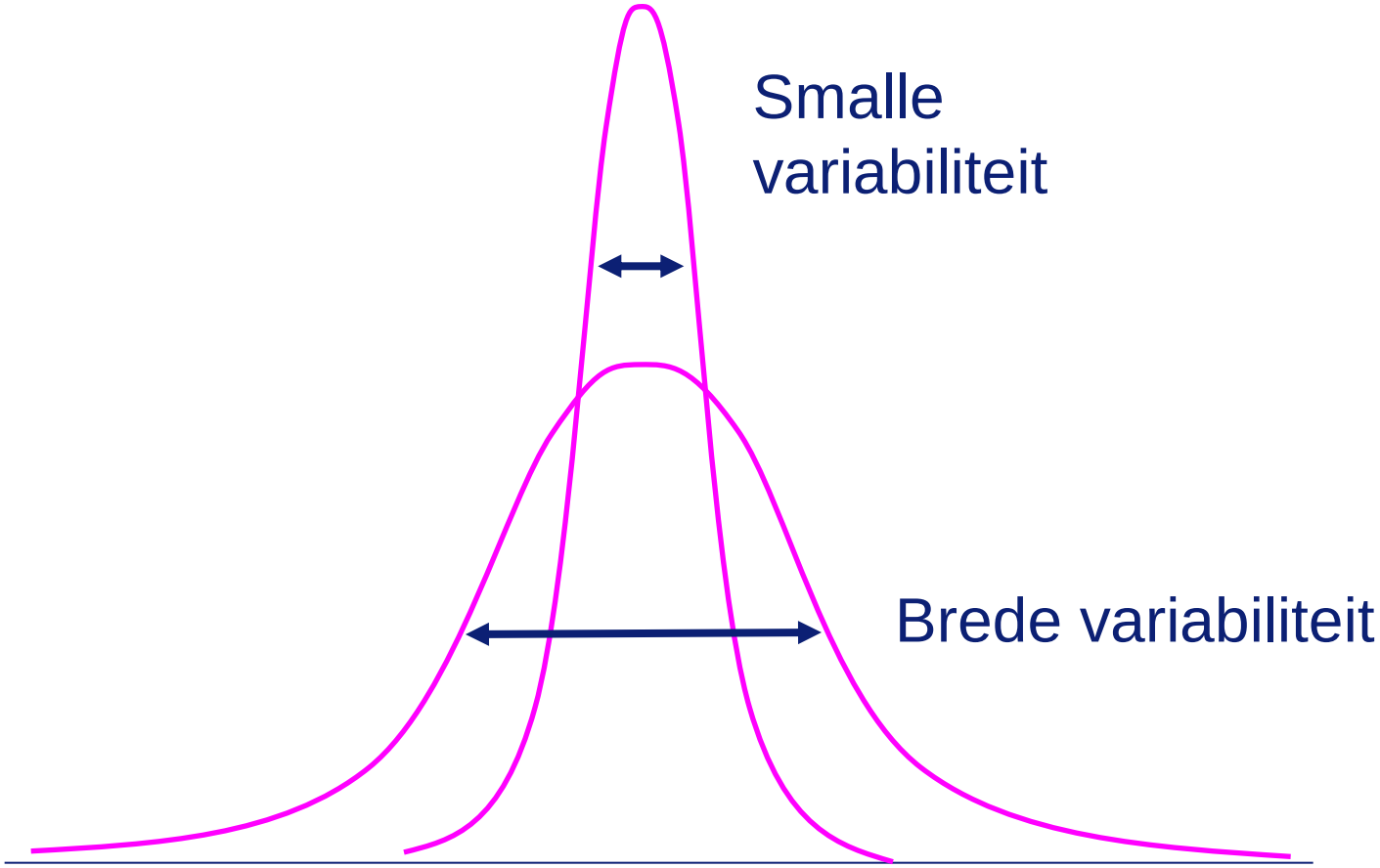


Gemiddelden – Mediaan



Indien scheve verdeling: mediaan is betrouwbaarder

Variabiliteit - Standaard Deviatie



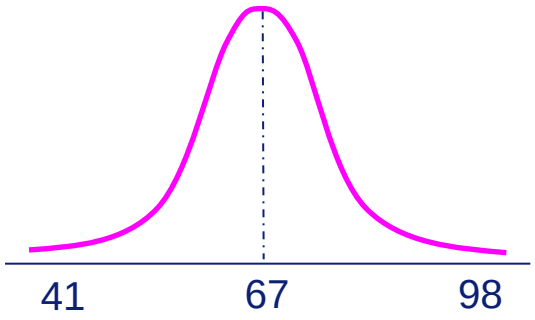
Variabiliteit - Range

41	44	49	50	51
52	58	61	63	66
66	66	67	67	68
70	71	73	74	79
80	84	88	91	98

Minimum = 41

Maximum = 98

Range = 98 – 41 = 57



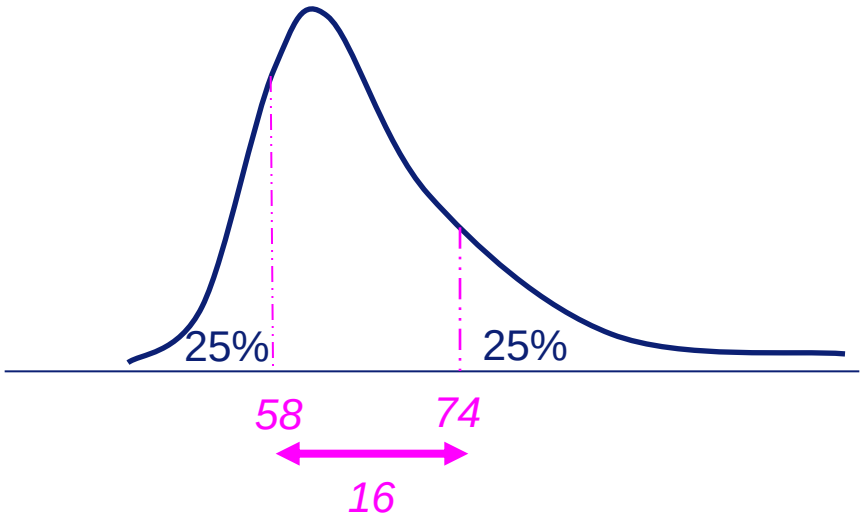
Variability – Inter-quartile range [IQR]

41	44	49	50	51
52	58	61	63	66
66	66	67	67	68
70	71	73	74	79
80	84	88	91	98

25th percentile (1st quartile) → 58

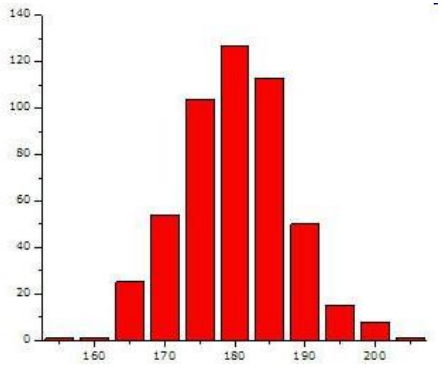
75th percentile (3rd quartile) → 74

Inter-quartile range = 74 - 58 = 16

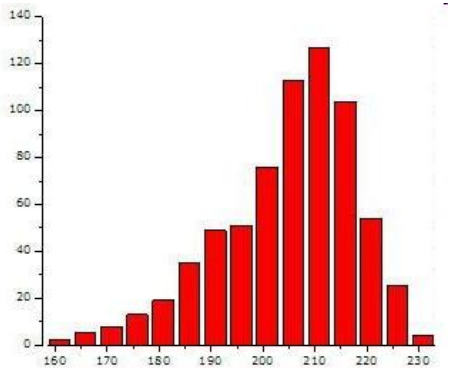


Samenvatting beschrijvende statistiek: wanneer gebruik je wat?

- Indien normale distributie:
  - gemiddelde
  - standaard deviatie

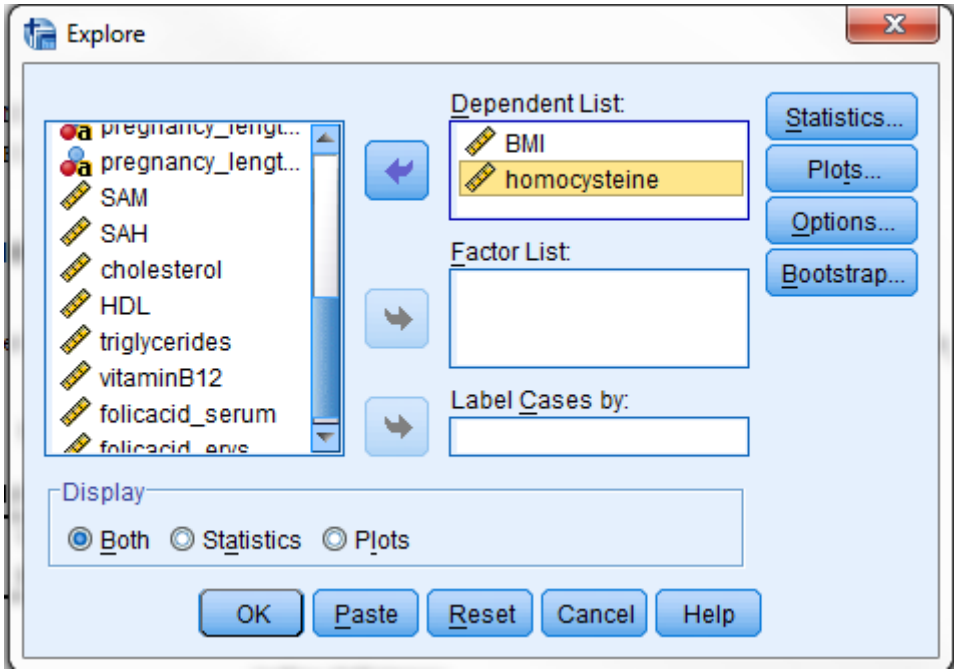


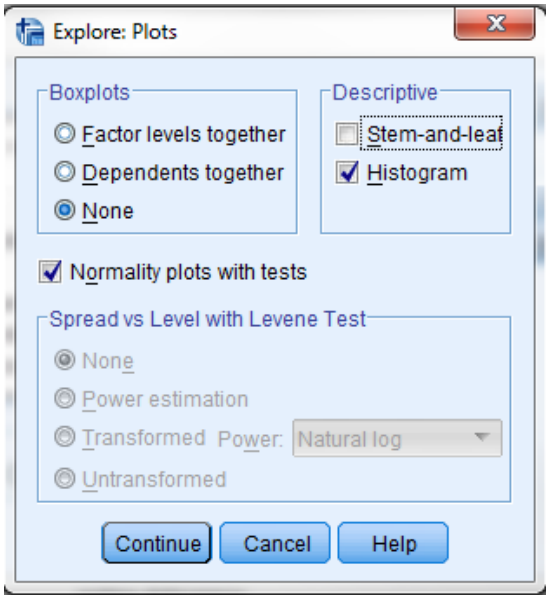
- Anders:
  - mediaan
  - range
  - percentielen (IQR)

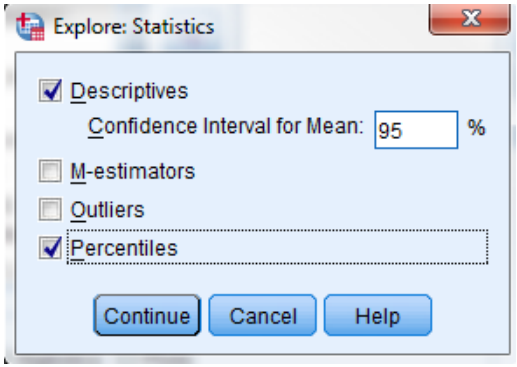


Beschrijvende statistiek - SPSS

- Maak (ook) gebruik van de EXPLORE functie







Beschrijvende statistiek - SPSS

Descriptives

			Statistic	Std. Error
Heart_Frequency	Mean		67,0800	2,92285
	95% Confidence Interval for Mean	Lower Bound	61,0475	
		Upper Bound	73,1125	
	5% Trimmed Mean		66,8556	
	Median		67,0000	
	Variance		213,577	
	Std. Deviation		14,61426	
	Minimum		41,00	
	Maximum		98,00	
	Range		57,00	
	Interquartile Range		21,50	
	Skewness		,176	,464
	Kurtosis		-,326	,902

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	Heart_Frequency	41,9000	47,0000	55,0000	67,0000	76,5000	89,2000	95,9000
Tukey's Hinges	Heart_Frequency			58,0000	67,0000	74,0000		