



**UNIVERSITEIT  
GENT**

# UNIVARIATE STATISTIEK

Statistiek in de criminologie, hoofdstuk 3 / prof. dr. Christophe Vandeviver

# INLEIDING

# INLEIDING

- Univariate statistiek.
  - 1 variabele.
- Frequentietabellen.
  - Absoluut ( $f_i$ ), relatief ( $p_i$ ), cumulatief ( $\sum n$ ).
- Grafieken.
  - Staafdiagrammen, histogrammen.
  - Box plot.
- Centrummaten.
  - Modus ( $M_o$ ), mediaan ( $M_d$ ), kwartieren ( $Q$ ), rekenkundig gemiddelde ( $\bar{x}$ ).
- Spreidingsmaten (variabiliteit, verschillen).
  - Variantie ( $s^2$ ), standaarddeviatie (standaardafwijking) ( $s$ ).

# KEUZE VAN STATISTISCHE PARAMETERS

- Welke statistische parameters je kan gebruiken is afhankelijk van het meetniveau van de variabele.
- Vuistregel:
  - Statistische parameters die op een lager meetniveau kunnen worden gebruikt, kunnen ook op een hoger meetniveau worden gebruikt.
  - Maar in principe nooit omgekeerd.

# ABSOLUTE & RELATIEVE FREQUENTIE

- Absolute frequentie.
  - Aantal elementen met een bepaalde waarde ( $f_i$ ).
  - Bv. Het aantal inbraken per straat, hoe vaak je het afgelopen jaar slachtoffer werd van een seksueel delict.
- Relatieve frequentie (proportie).
  - Aantal elementen met een bepaalde waarde ( $f_i$ ) gerelateerd aan het totaal aantal elementen ( $n$ ):
$$p_i = \frac{f_i}{n}; f_i = p_i n$$
  - Percentage: proportie vermenigvuldigd met honderd.

# EEN VOORBEELD

**Tabel 1. Seksueel slachtofferschap onder studenten aan UGent (n = 1200)**

	Absolute frequentie	Proportie	Percentage (%)	Cumulatief percentage (%)
<b>Geen seksueel slachtoffer</b>	666	$666/1200 \rightarrow 0,55$	$0,55*100 \rightarrow 55,5$	55,5
<b>Wel seksueel slachtoffer</b>	534	$534/1200 \rightarrow 0,44$	$0,44*100 \rightarrow 44,5$	100
<b>Totaal</b>	1200	1	100	

**Tabel 2. Studiejaar slachtoffers van seksueel geweld onder studenten aan UGent (n = 534)**

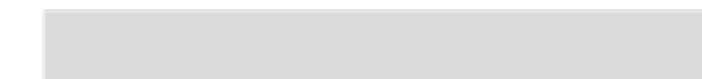
	Absolute frequentie	Proportie	Percentage (%)	Cumulatief percentage (%)
<b>1<sup>e</sup> bachelor</b>	387	$387/534 \rightarrow 0,73$	$0,73*100 \rightarrow 72,47$	72,47
<b>2<sup>e</sup> bachelor</b>	49	0,09	9,18	81,65
<b>3<sup>e</sup> bachelor</b>	39	0,07	7,30	88,95
<b>1<sup>e</sup> master</b>	24	0,05	4,49	93,45
<b>2<sup>e</sup> master</b>	13	0,02	2,43	95,88
<b>Schakeljaar</b>	22	0,04	4,12	100
<b>Totaal</b>	534	1,00	100	

## Most dangerous cities

Total murders in 2014

**WRONG**

Chicago



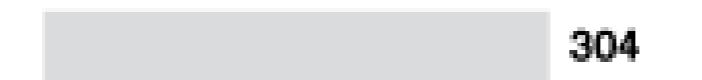
407

New York



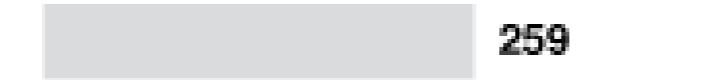
328

Detroit



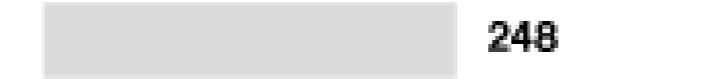
304

Los Angeles



259

Philadelphia



248

## Most dangerous cities

Murder rate in major US cities in 2014,  
per 100,000 people

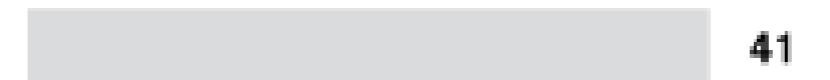
**RIGHT**

Detroit



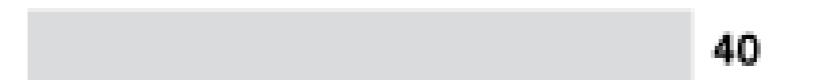
45

New Orleans



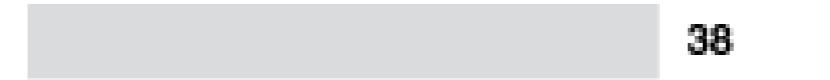
41

Newark



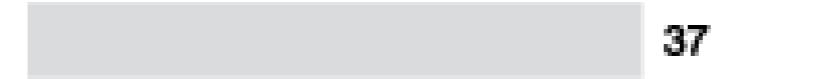
40

St. Louis



38

Baltimore



37

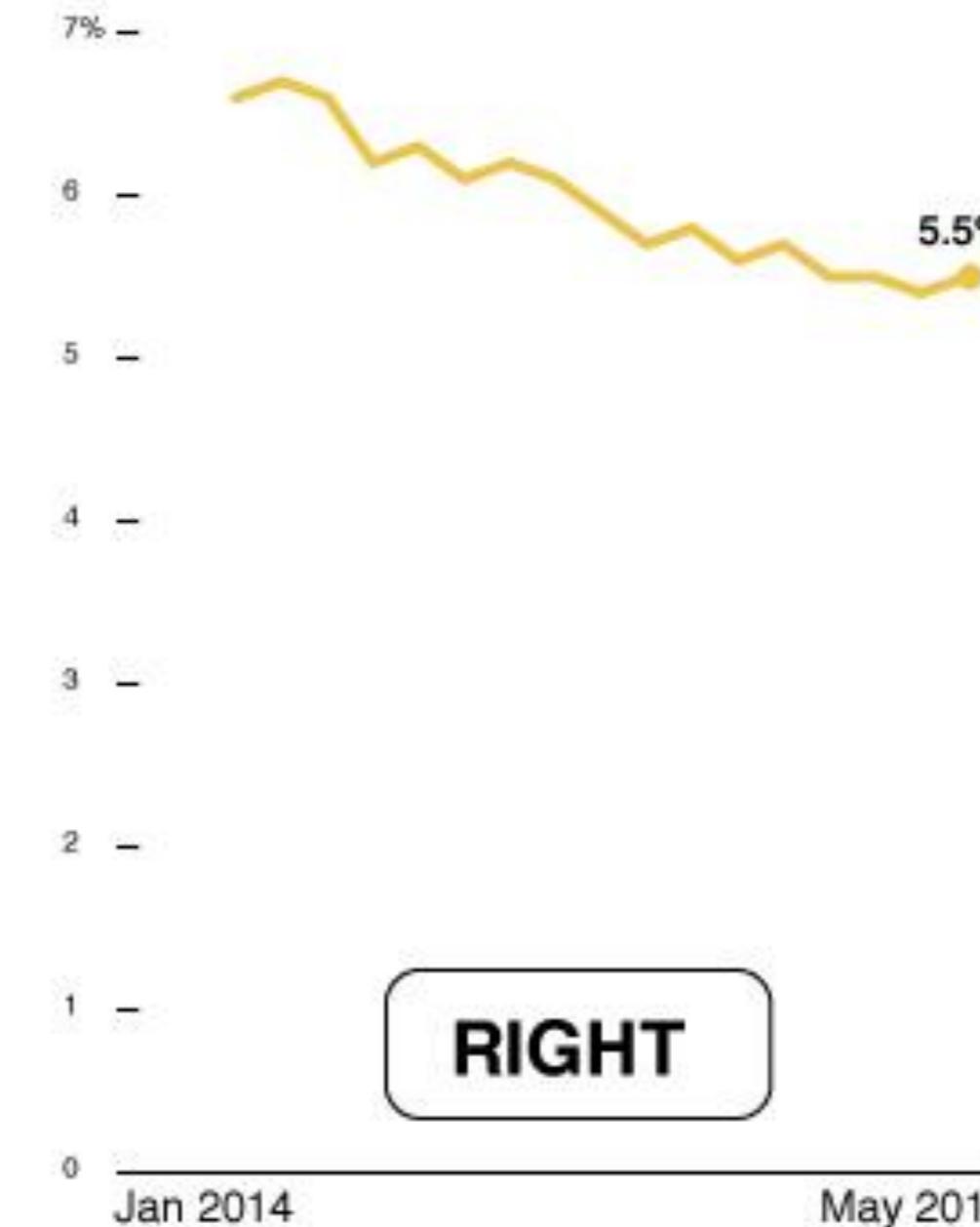
# GRAFIEKEN

# OVER DE SCHAAL EN IJKING VAN X- EN Y-AS

US GDP



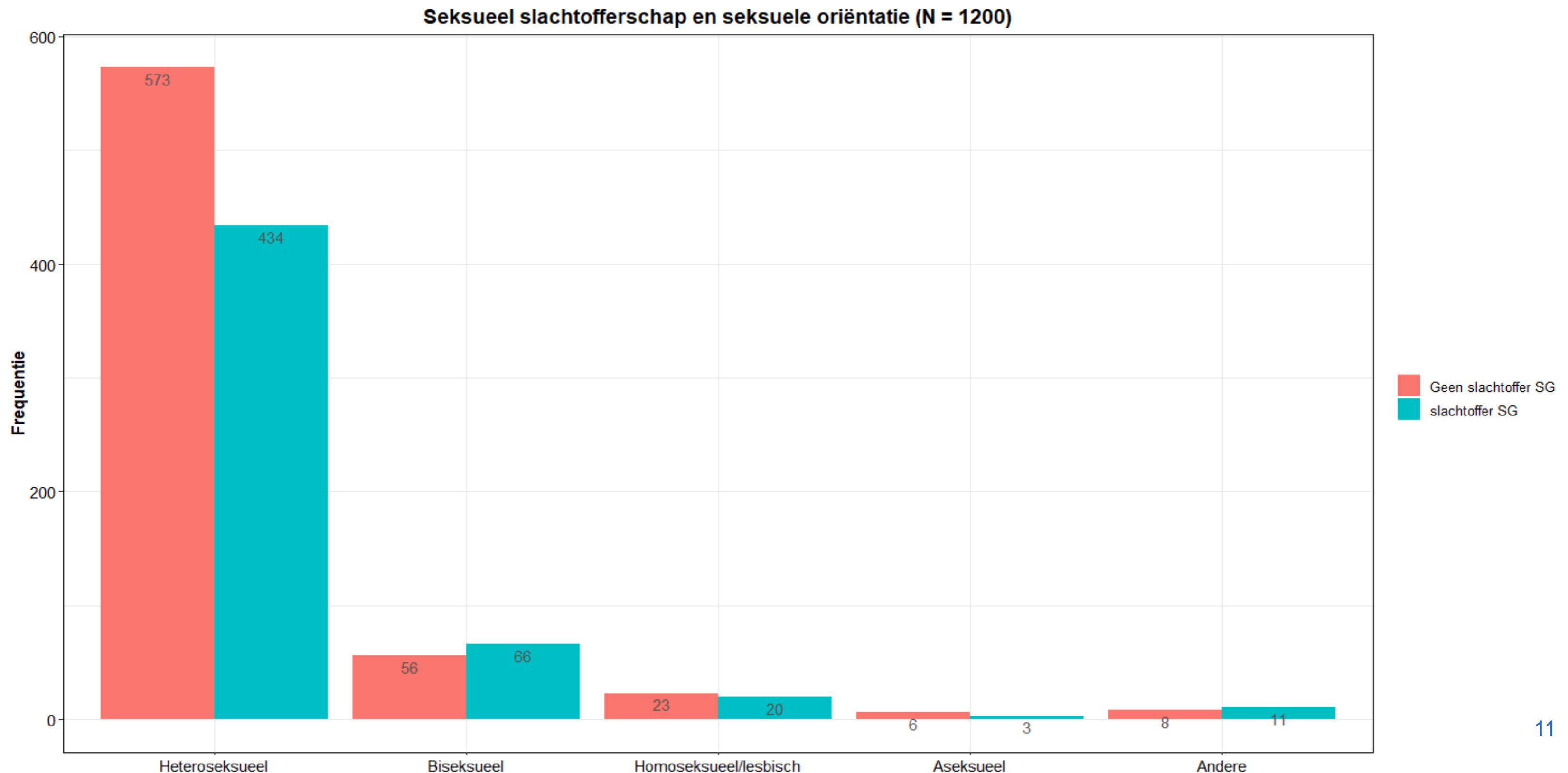
US GDP



Source: [National Geographic](#)

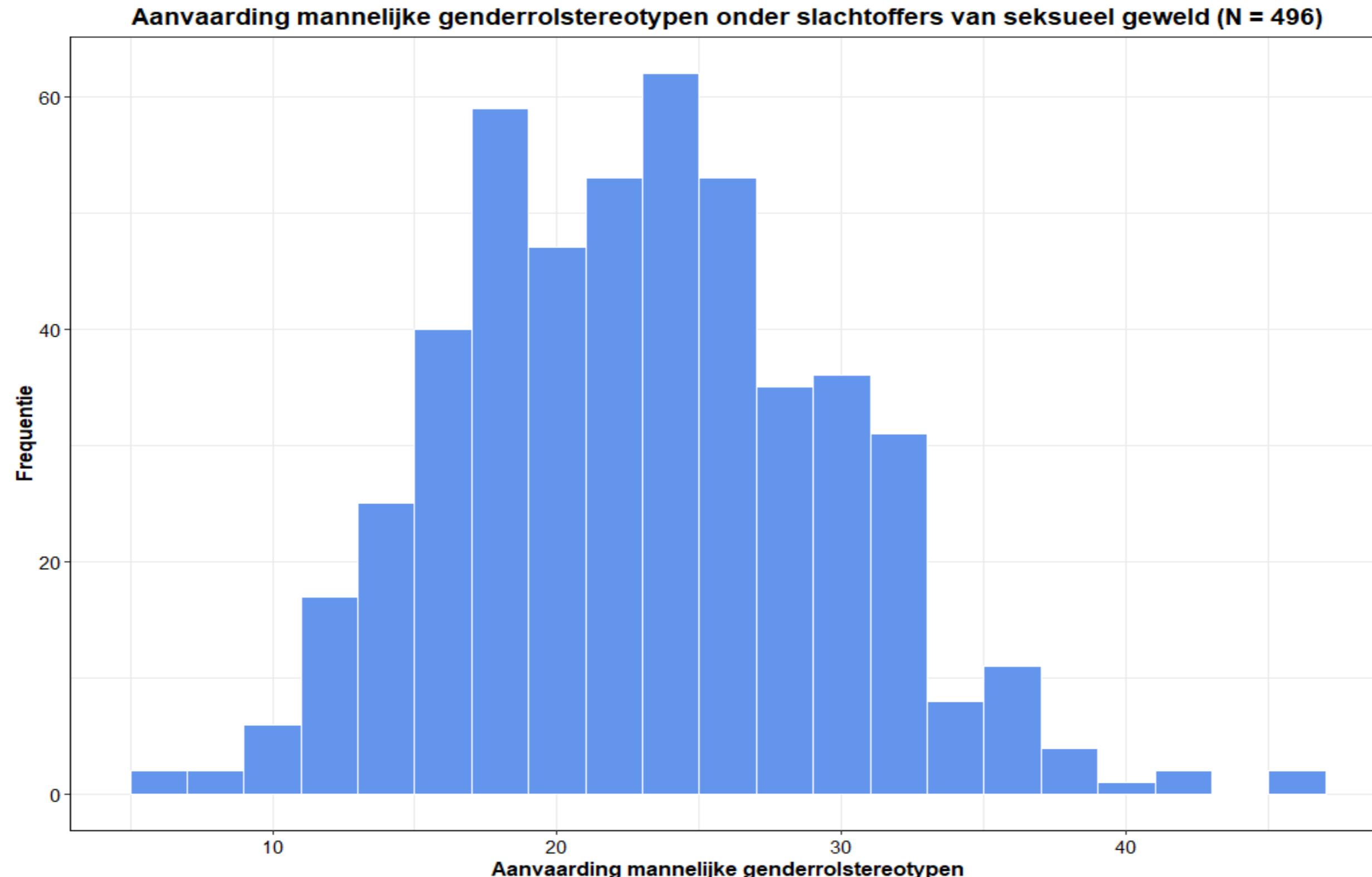
# STAAFDIAGRAM (BAR CHART)

## – Nominale en ordinale variabelen



# HISTOGRAM

## – Metrische variabelen



# PARAMETERS VAN CENTRALITEIT

# MODUS (MODE)

- Nominale, ordinale & metrische variabelen.
- Modale waarde  $M_o$ .
  - Waarde met vaakst voorkomende frequentie in een set waarden.
  - Zgn. ‘zwaartepunt’ van verdeling.

$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \rightarrow \begin{array}{cc} x & f_i \\ 0 & 3 \\ 1 & 4 \\ 2 & 2 \end{array} \rightarrow M_o = 1$$

# MEDIAAN (MEDIAN)

- Ordinale & metrische variabelen.
- Mediaan  $M_d$ .
  - Middelste waarde van een oplopend geordende set warden.
  - Zgn. ‘middenpunt’ van verdeling.
  - 50% observaties  $< M_d$ .
  - 50% observaties  $> M_d$ .
- Relatief ongevoelig voor extreme waarden (i.t.t. rekenkundig gemiddelde).
- Hoe berekenen?
  1. Rangschik alle observaties (cases)  $n$  oplopend (van laag naar hoog).
  2. Bepaal de waarde van  $\frac{n+1}{2}$ -de observatie.
    1. Indien  $n$  oneven getal:  $M_d$  direct afleesbaar.
    2. Indien  $n$  even getal:  $M_d$  gemiddelde van twee middelste observaties.

# MEDIAAN (MEDIAN)

$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$



$(0, 0, 0, 1, 1, 1, 1, 2, 2)$



$$M_d = 1$$

# KWARTIELEN (QUARTILES)

- Ordinale & metrische variabelen.
- Oplopend geordende reeks observaties opgedeeld in vier groepen met zo gelijk mogelijk aantal elementen per groep.
  - $Q_1 = 25\% n$  kleiner en  $75\% n$  groter.
  - $Q_2 = 50\% n$  kleiner en  $50\% n$  groter =  $M_d$ .
  - $Q_3 = 75\% n$  kleiner en  $25\% n$  groter.
- Vijfgetallensamenvatting:
  - $Q_0 = \min$
  - $Q_1, M_d, Q_3$
  - $Q_4 = \max$

# REKENKUNDIG GEMIDDELDE (MEAN)

- Metrische variabelen.
- Rekenkundig gemiddelde  $\bar{x}$ .
  - ‘Gemiddelde’ waarde van een set waarden.
  - Zgn. ‘evenwichtspunt’ van verdeling.
- Som van alle voorkomende waarden  $x$  gedeeld door het totaal aantal observaties  $n$ :

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

# REKENKUNDIG GEMIDDELDE (MEAN)

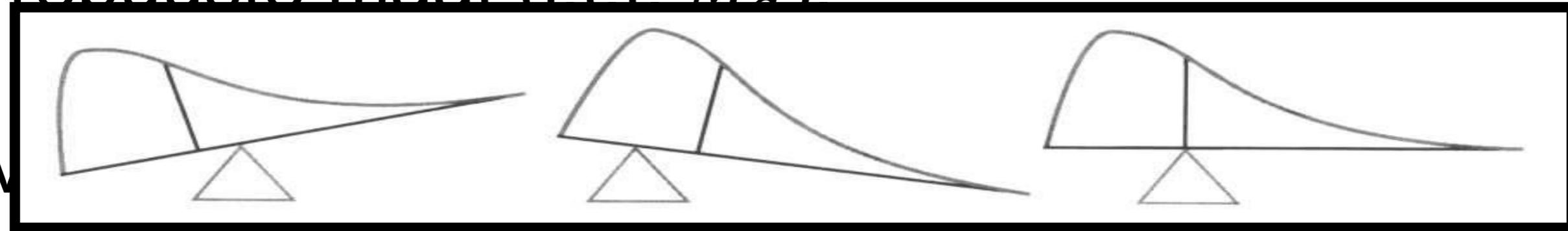
$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

↓

$$\bar{x} = \frac{0 + 0 + 0 + 1 + 1 + 1 + 1 + 2 + 2}{9} = \frac{8}{9} = 0.89$$

# KENMERKEN VAN REKENKUNDIG GEMIDDELDE

- $\bar{x}$  is het evenwichtspunt van de verdeling.
- $\bar{x}$  is gevoelig voor extreme waarden en bijgevolg  
**geen** robuuste maat (i t t  $M_d$ )
- Som v



$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

# VOORBEELD

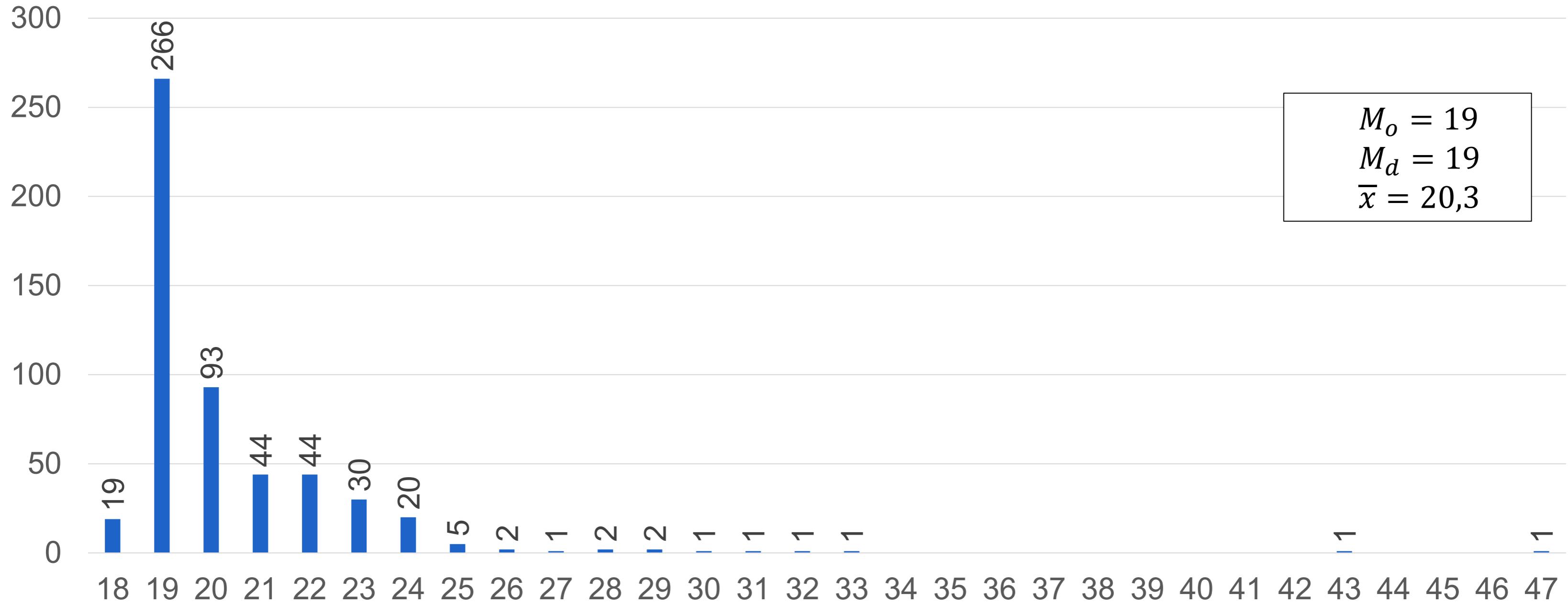
- Frequentieverdeling leeftijd slachtoffers seksueel geweld.
- $M_o = 19$
- $M_d = 19$
- $\bar{x} = 20,3$

Tabel 3. Leeftijd slachtoffers seksueel geweld bij studenten aan UGent (n = 534)

Leeftijd	Frequentie
18	19
19	266
20	93
21	44
22	44
23	30
24	20
25	5
26	2
27	1
28	2
29	2
30	1
31	1
32	1
33	1
43	1
47	1

# VOORBEELD

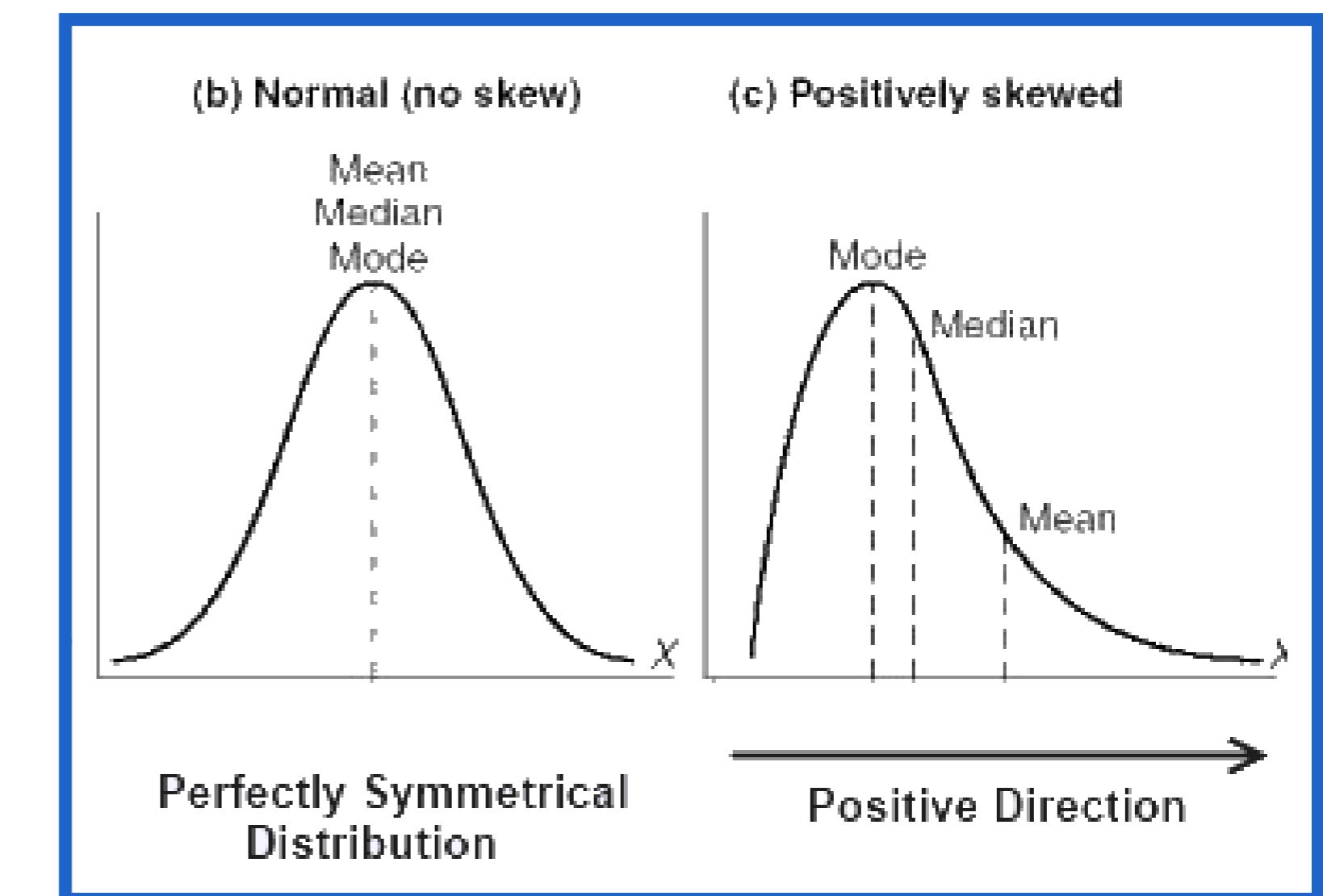
Leeftijd slachtoffers seksueel geweld bij studenten aan UGent (n = 534)



$$\begin{aligned}M_o &= 19 \\M_d &= 19 \\\bar{x} &= 20,3\end{aligned}$$

# KIEZEN VAN GEPASTE CENTRUMMAAT?

- Vuistregel:
  - Informatiegehalte maximaliseren en centrummaten combineren.
  - Onderzoeks vraag is bepalend.
- Elke centrummaat drukt iets anders uit én vorm van verdeling beïnvloedt centrummaten:



# PARAMETERS VAN SPREIDING

# VARIATIEBREEDE

- Ordinale variabelen.
- Variatiebreedte  $V$ .
  - Verschil tussen grootste en kleinste waargenomen waarde.
  - Gebruikt zeer beperkte hoeveelheid beschikbare info.

$$V = \max_i(x_i) - \min_i(x_i)$$

## INTERKWATIELAFSTAND (INTER QUARTILE RANGE)

- Ordinale variabelen.
- Interkwartielafstand *IKA*.
  - Verschil tussen het derde en eerste kwartiel.
  - Gebruikt 50% van de beschikbare info.

$$IKA = Q_3 - Q_1$$

# (STEEKPROEF)VARIANTIE (VARIANCE)

- Metrische variabelen.
- $\text{Var } s^2$ .
  - Mate waarin steekproefwaarden afwijken van het rekenkundig gemiddelde  $\bar{x}$ .
  - Uitdrukking van de mate waarin observaties onderling verschillen.

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

## (STEEKPROEF)STANDAARDAFWIJKING (STANDARD DEVIATION)

- Metrische variabelen.
- Standaardafwijking  $s$ .
  - Vierkantswortel van de variantie.

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

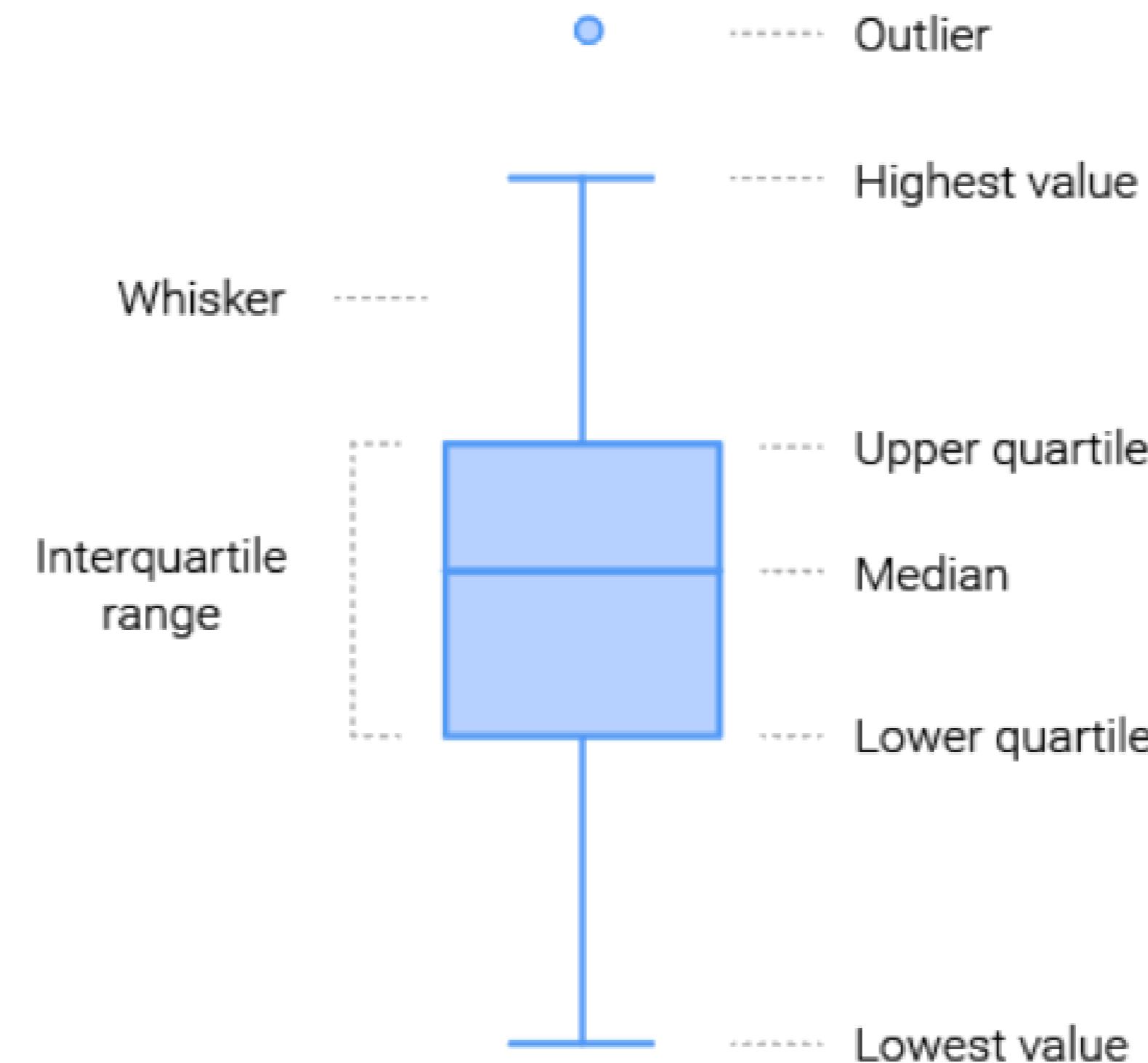
- Twee zeer aantrekkelijke eigenschappen:
  - Uitgedrukt in dezelfde eenheid als de waarden van de variabele, en
  - Uitvergrote afwijkingen beperkt.

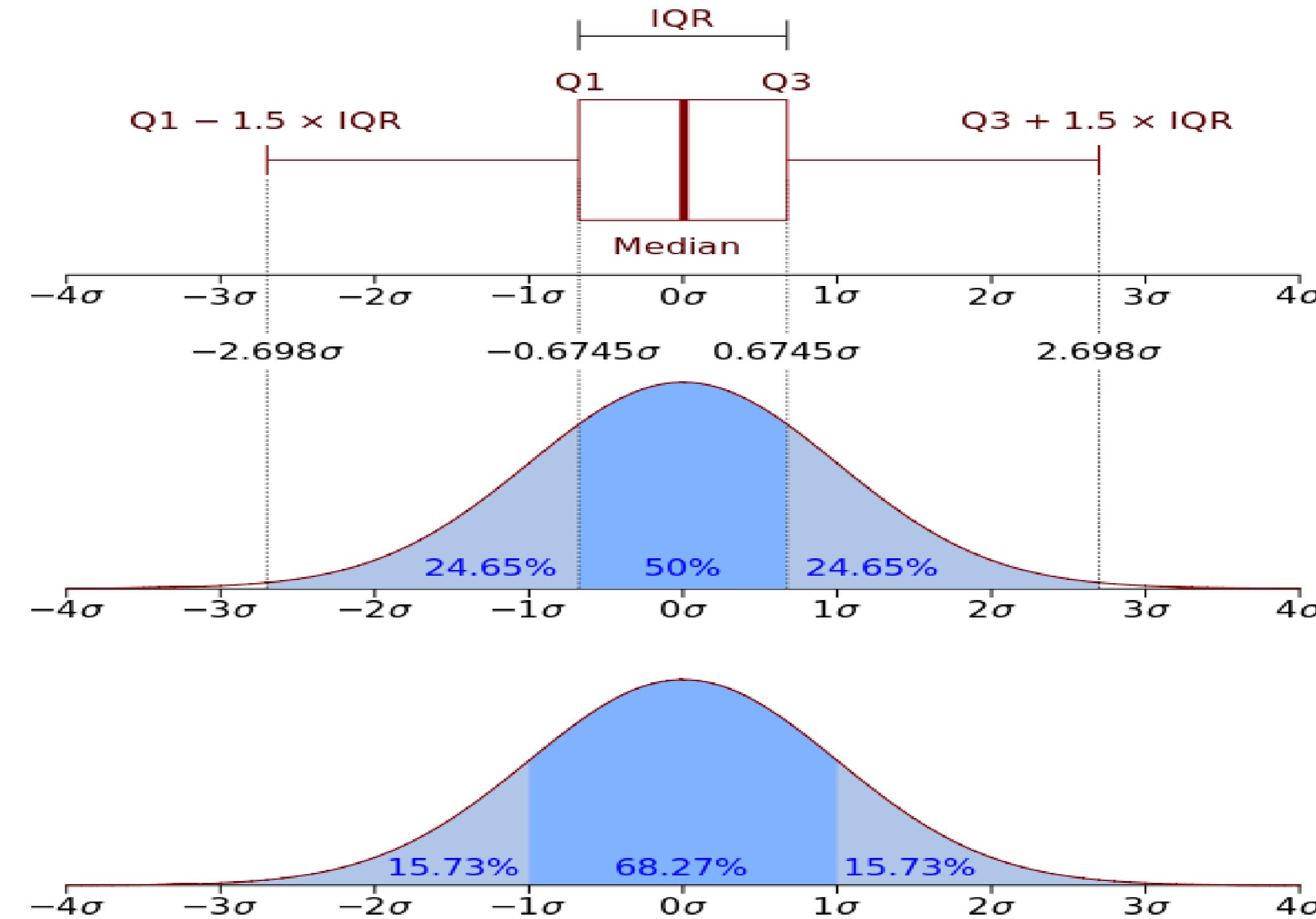
# BOX PLOT

# BOX PLOT

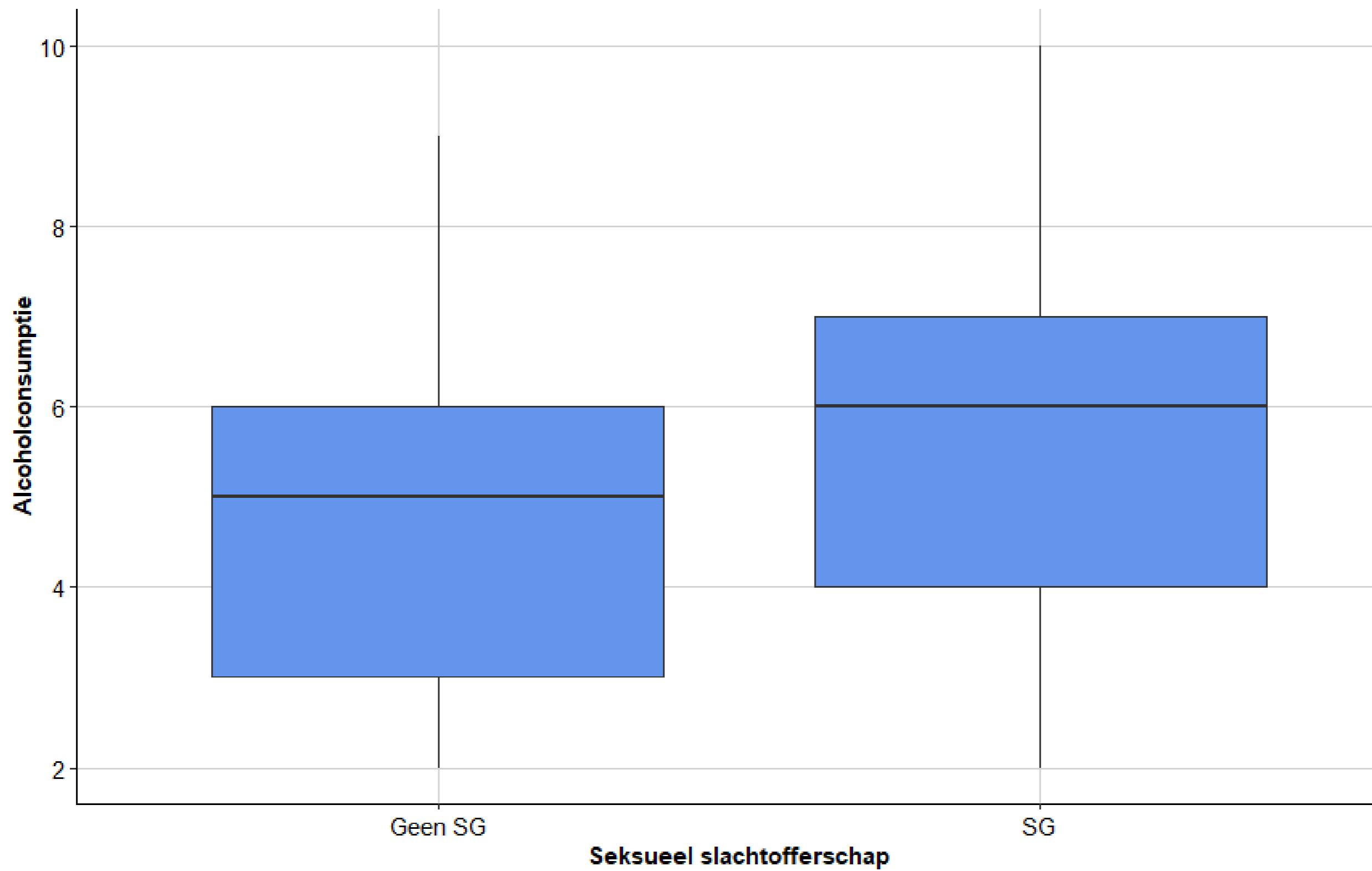
- Ordinale en metrische variabelen.
- Visuele evaluatie van informatie m.b.t. centraliteit en spreiding van frequentieverdeling variabele in één grafiek.
- Zgn. vijfgetallensamenvatting:
  - Mediaan:  $Q_2 = M_d$ .
  - Kwartielen:  $Q_1$  en  $Q_3$ .
  - Minimale en maximale niet-uitschietende waarden:  
$$Q_1 - 1.5 * (Q_3 - Q_1) = \min \text{ en } Q_3 + 1.5 * (Q_3 - Q_1) = \max.$$

# BOX PLOT

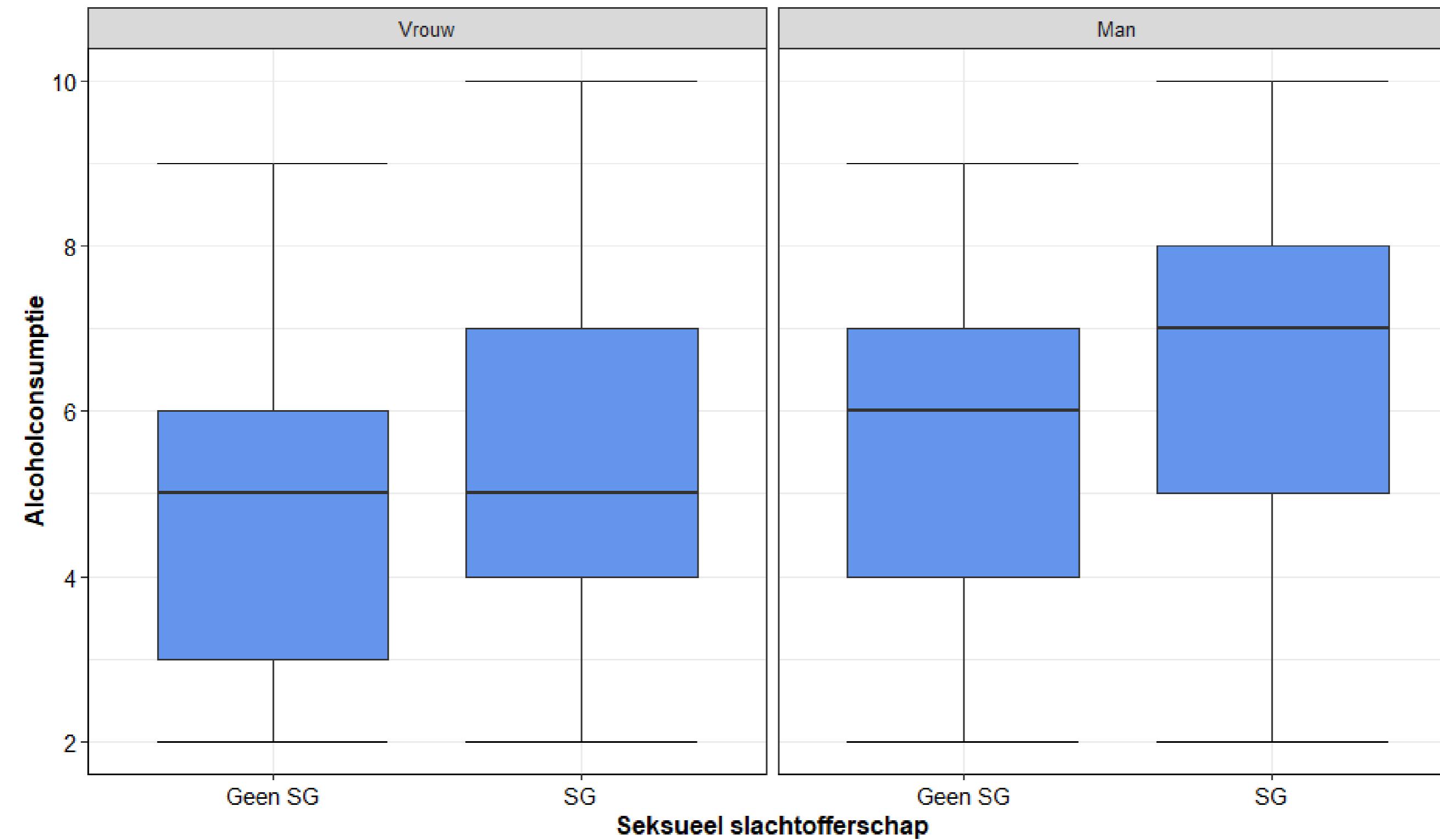




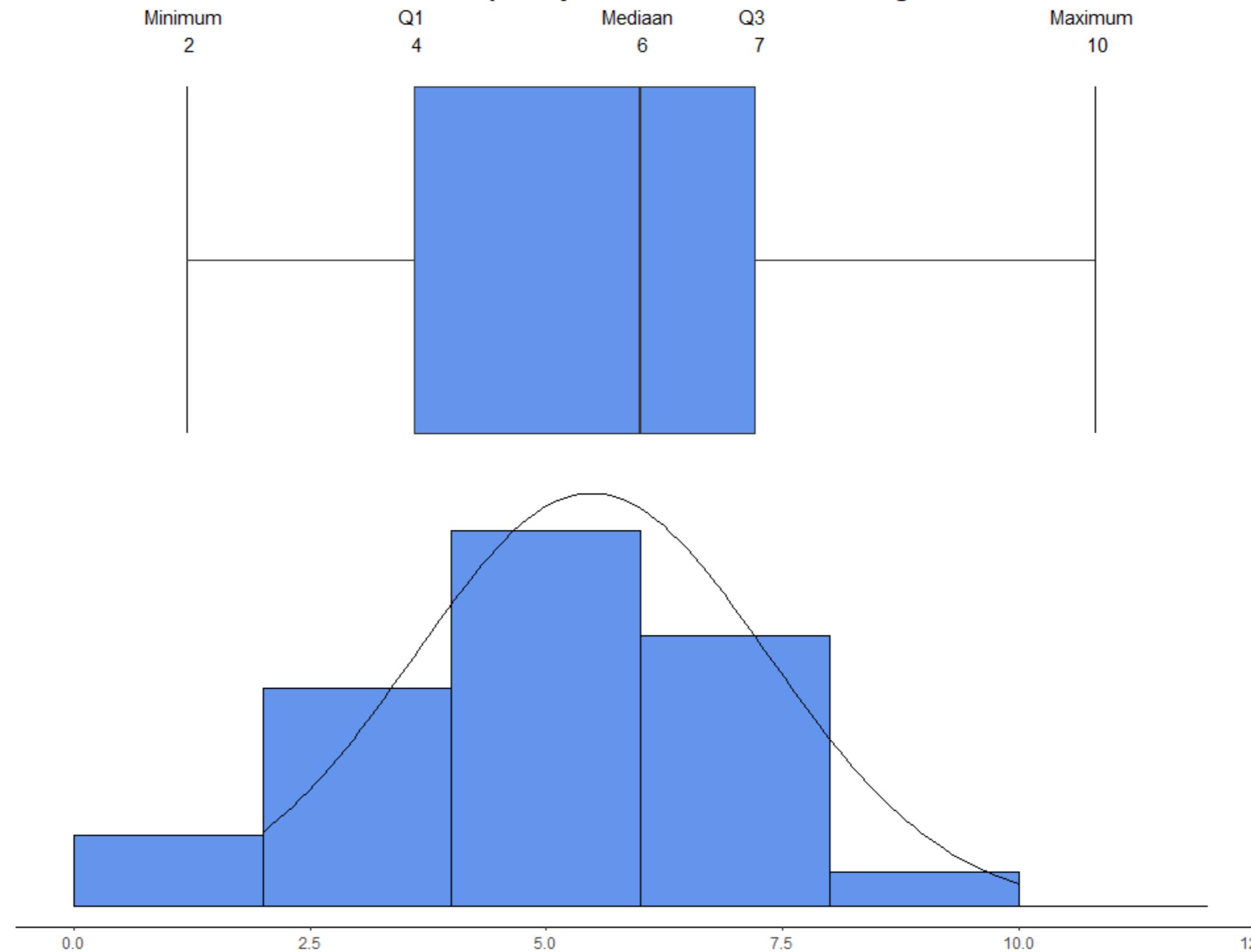
## Alcoholconsumptie en seksueel slachtofferschap



## Alcoholconsumptie en seksueel slachtofferschap



### Alcoholconsumptie bij slachtoffers van seksueel geweld



# Prof. Dr. Christophe Vandeviver

Onderzoeksprofessor Criminologie UGent

VAKGROEP CRIMINOLOGIE, STRAFRECHT &  
SOCIAAL RECHT

E [christophe.vandeviver@ugent.be](mailto:christophe.vandeviver@ugent.be)

T +32 9 264 67 38

[www.ugent.be](http://www.ugent.be)

 Universiteit Gent

 @ugent

 Ghent University