Statistik

für Betriebswirtschaft und internationales Management

Sommersemester 2015

Prof. Dr. Stefan Etschberger
Hochschule Augsburg

Auswertungsmethoden für eindimensionales Datenmaterial

- Merkmal X wird an n Merkmalsträgern beobachtet ➠
  Urliste $(x_1, \ldots, x_n)$
  Im Beispiel: $x_1 = 4,\ x_2 = 11, \ldots, x_{12} = 6$
- Urlisten sind oft unübersichtlich, z.B.:

```
##  [1] 4 5 4 1 5 4 3 4 5 6 6 5 5 4 7 4 6 5 6 4 5 4 7 5 5 6 7 3
## [29] 7 6 6 7 4 5 4 7 7 5 5 5 5 6 6 4 5 2 5 4 7 5
```

- Dann zweckmäßig: Häufigkeitsverteilungen

Ausprägung (sortiert)	a_j	1	2	3	4	5	6	7	$\sum$
absolute Häufigkeit	$h(a_j) = h_j$	1	1	2	12	17	9	8	50
kumulierte abs. H.	$H(a_j) = \sum_{i=1}^{j} h(a_i)$	1	2	4	16	33	42	50	—
relative Häufigkeit	$f(a_j) = h(a_j)/n$	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{2}{50}$	$\frac{12}{50}$	$\frac{17}{50}$	$\frac{9}{50}$	$\frac{8}{50}$	1
kumulierte rel. H.	$F(a_j) = \sum_{i=1}^{j} f(a_i)$	$\frac{1}{50}$	$\frac{2}{50}$	$\frac{4}{50}$	$\frac{16}{50}$	$\frac{33}{50}$	$\frac{42}{50}$	1	—

Beispiel: Alter von Studierenden in Statistik Vorlesung

	$h(a_j)$	$H(a_j) = \sum_{i=1}^{j} h(a_j)$	$f(a_j) = \frac{h(a_j)}{n}$	$F(a_j) = \sum_{i=1}^{j} f(a_i)$
18	24	24	0.0637	0.0637
19	58	82	0.1538	0.2175
20	71	153	0.1883	0.4058
21	53	206	0.1406	0.5464
22	43	249	0.1141	0.6605
23	40	289	0.1061	0.7666
24	25	314	0.0663	0.8329
25	11	325	0.0292	0.8621
26	11	336	0.0292	0.8912
27	8	344	0.0212	0.9125
28	12	356	0.0318	0.9443
29	7	363	0.0186	0.9629
30	2	365	0.0053	0.9682
31	2	367	0.0053	0.9735
32	5	372	0.0133	0.9867
33	2	374	0.0053	0.9920
34	1	375	0.0027	0.9947
35	1	376	0.0027	0.9973
36	1	377	0.0027	1.0000

Empirische Verteilungsfunktion

- für metrische Merkmale
- Anteil der Ausprägungen, die höchstens so hoch sind wie x.
- Exakt:

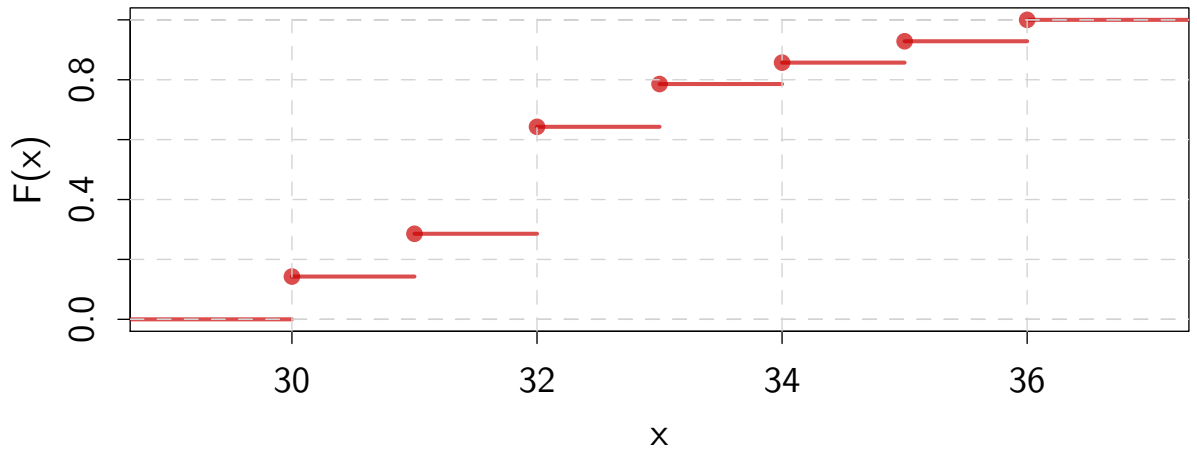
$$F(x) = \sum_{a_i \leqslant x} f(a_i)$$

Beispiel

```
Studenten.ueber.29 = sort(MyData$Alter[MyData$Alter > 29])
Studenten.ueber.29

##  [1] 30 30 31 31 32 32 32 32 32 33 33 34 35 36

# empirical cumulative distribution function (ecdf)
Studenten.F = ecdf(Studenten.ueber.29)
plot(Studenten.F, col=rgb(0.8,0,0,.7), lwd=3, main="", xlab="x", ylab="F(x)")
grid(lty=2) # Gitternetz
```



- für metrische Merkmale; Voraussetzung: sortierte Urliste
- Umkehrung der Verteilungsfunktion
- Anteil p gegeben, gesucht: $F^{-1}(p)$, falls vorhanden.
- Definition p-Quantil:

$$\tilde{x}_p = \begin{cases} \frac{1}{2}(x_{n\cdot p} + x_{n\cdot p+1}), & \text{wenn } n \cdot p \in \mathbb{N}_0 \\ x_{\lceil n\cdot p \rceil}, & \text{sonst} \end{cases}$$

aufrunden

$n \cdot p = 14 \cdot 0{,}2 = 2{,}8$, also keine ganze Zahl
$\Rightarrow \lceil n \cdot p \rceil = \lceil 2{,}8 \rceil = 3$

Beispiel

z.B. $\tilde{x}_{0,2} = x_3 = 31$ (x_3 = 3. Wert)

d.h. mind. 20% sind höchstens 31
(mind. 80% sind mind. 31)

```
Studenten.ueber.29
##  [1] 30 30 31 31 32 32 32 32 32 33 33 34 35 36
n = length(Studenten.ueber.29)
p = c(1,2,3.9,4,10,13.9)/n
 quantile(Studenten.ueber.29, prob=p, type=2)
## 7.142857% 14.28571% 27.85714% 28.57143% 71.42857% 99.28571%
##      30.0      30.5      31.0      31.5      33.0      36.0
```

Beispiel: $p = \frac{3}{7}$, $n = 14$
$n \cdot p = 6 \in \mathbb{N}_0$
$\Rightarrow \tilde{x}_{3/7} = \frac{1}{2}(x_6 + x_7)$
$= \frac{1}{2}(32 + 32) = 32$

Beispiel: Urliste $x = (1,1,1,2,3)$
gesucht $\tilde{x}_{0,2}$, $\tilde{x}_{0,25}$, $\tilde{x}_{0,5}$, $\tilde{x}_{0,75}$

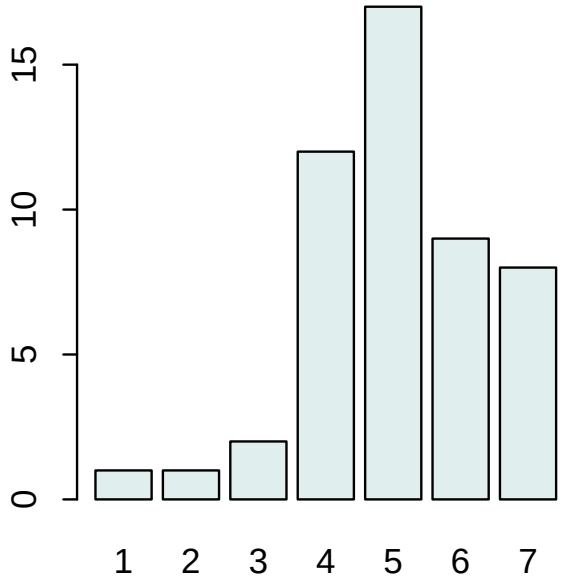
$\tilde{x}_{0,2} = \frac{1}{2}(x_1 + x_2) = 1$
$\tilde{x}_{0,25} = x_{\lceil 1{,}25 \rceil} = x_2 = 1$
$\tilde{x}_{0,5} = x_{\lceil 2{,}5 \rceil} = x_3 = 1$
$\tilde{x}_{0,75} = x_{\lceil 3{,}75 \rceil} = x_4 = 2$

❶ Balkendiagramm

```
table(x)

## x
##  1  2  3  4  5  6  7
##  1  1  2 12 17  9  8

barplot(table(x), col="azure2")
```



(Höhe proportional zu Häufigkeit)

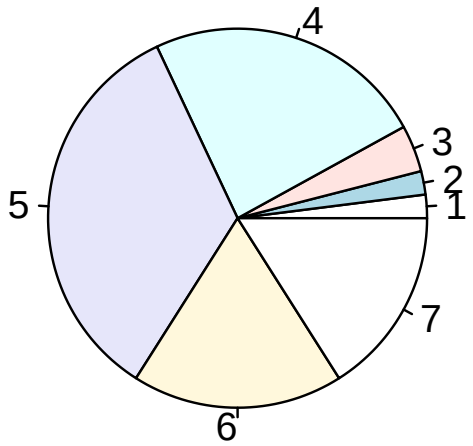
❷ Kreissektorendiagramm

Winkel: $w_j = 360° \cdot f(a_j)$

z.B. $w_1 = 360° \cdot \frac{1}{50} = 7{,}2°$

$w_7 = 360° \cdot \frac{8}{50} = 57{,}6°$

```
pie(table(x))
```

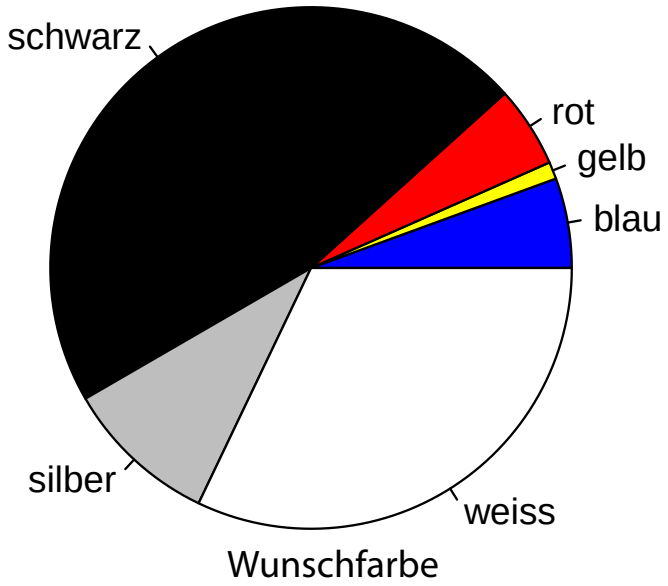


(Fläche proportional zu Häufigkeit)

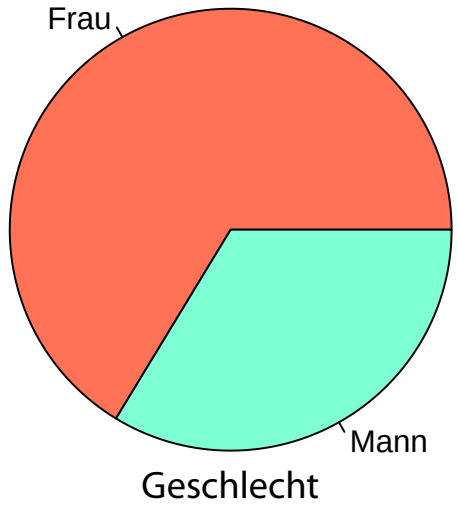
Umfrage: Kuchen

Kreissektorendiagramm

```
pie(table(MyData$Farbe),
    col=c("blue", "yellow", "red",
          "black", "grey", "white"))
```

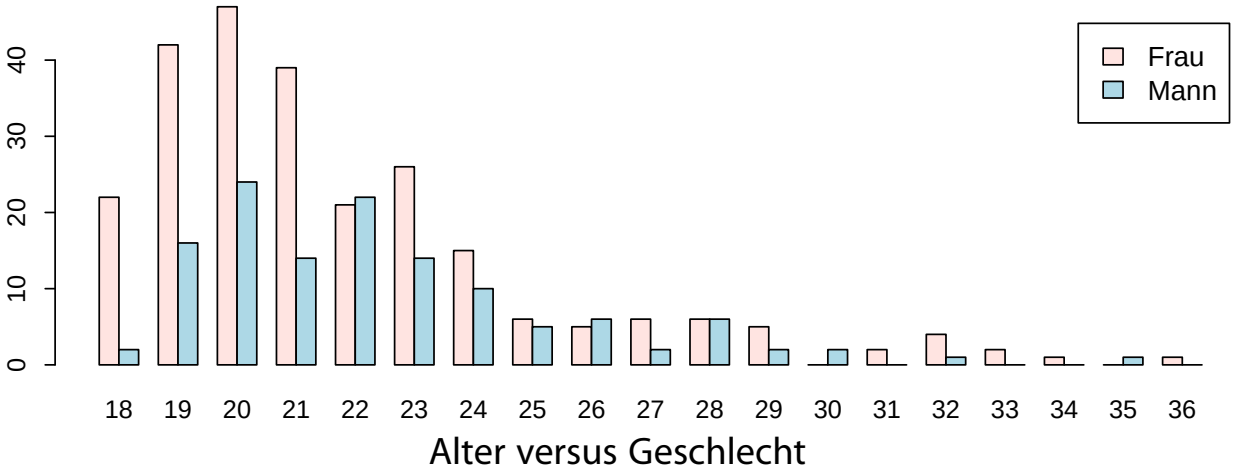



```
pie(table(MyData$Geschlecht),
    col=c("coral1", "aquamarine"))
```

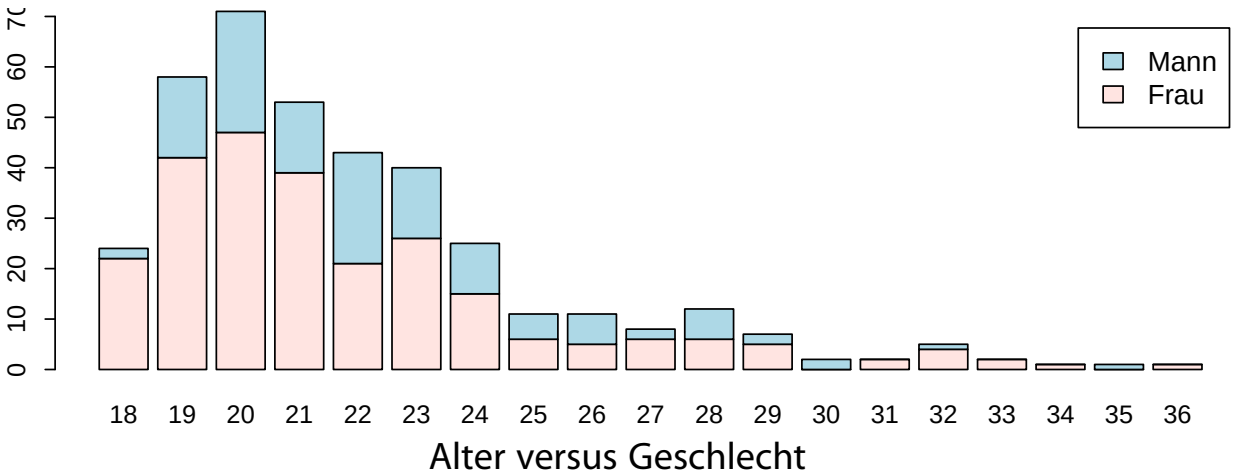


Balkendiagramm, Klassen getrennt oder gestapelt


```
barplot(xtabs(~ Geschlecht + Alter),
        legend=TRUE, beside=TRUE, col=c("mistyrose", "lightblue"))
```




```
barplot(xtabs(~ Geschlecht + Alter),
        legend=TRUE, beside=FALSE, col=c("mistyrose", "lightblue"))
```



❸ Histogramm

- für klassierte Daten
- Fläche proportional zu Häufigkeit:

$$\text{Höhe}_j \cdot \text{Breite}_j = c \cdot h(a_j)$$

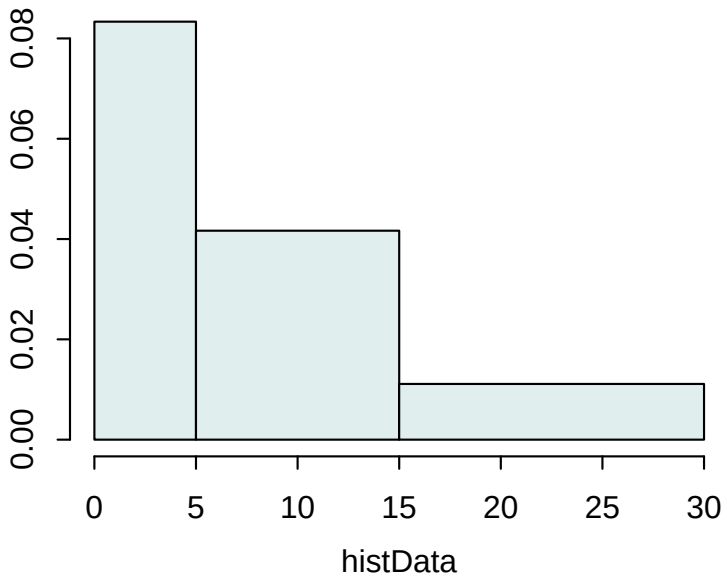
oft: $c = \frac{1}{n}$ oder $c = 1$

$$\Rightarrow \quad \text{Höhe}_j = c \cdot \frac{h(a_j)}{\text{Breite}_j}$$

- Im Beispiel mit $c = \frac{1}{12}$:

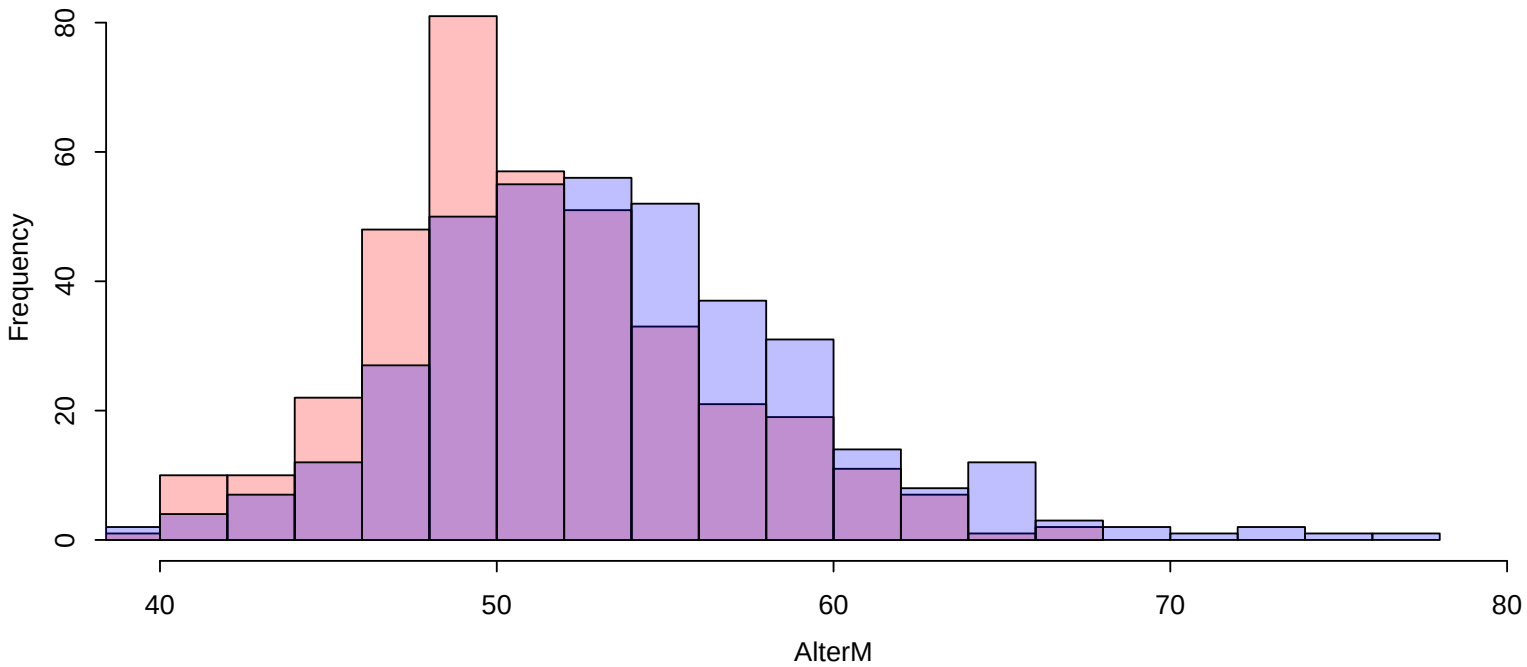
Klasse	$[0;5)$	$[5;15)$	$[15;30]$
$h(a_j)$	5	5	2
Breite_j	5	10	15
Höhe_j	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{90}$

```
histData <- c(0,1,2,3,4,
              5,6,7,10,14,
              15,30)
truehist(histData,
         breaks=c(0, 4.999, 14.999, 30),
         col="azure2", ylab='')
```



Histogramm

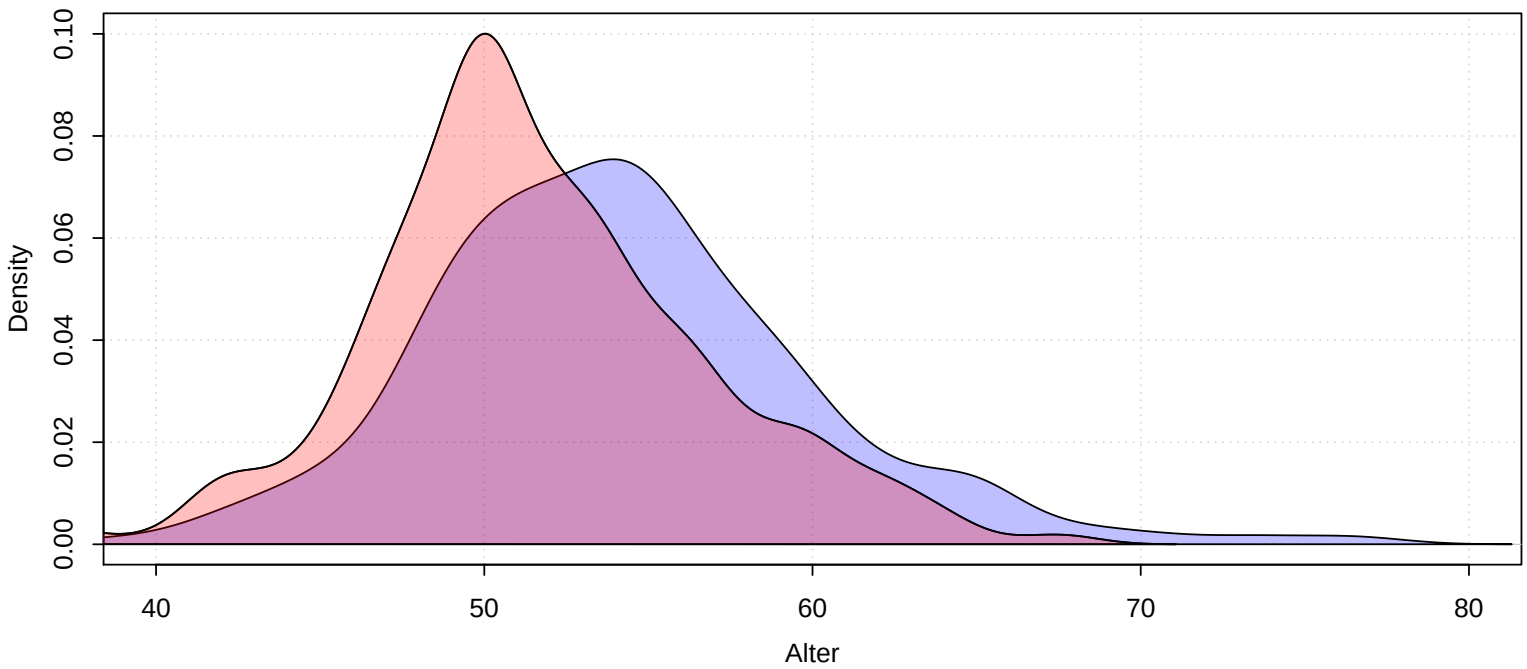
```
plot(hist(AlterM, plot=F, breaks=20),
          col=rgb(1,0,0,1/4), # make red transparent
          main="",
          xlim=c(40,80)) # draw from 40 to 80
plot(hist(AlterV, plot=F, breaks=20),
          col=rgb(0,0,1,1/4),
          add=TRUE)
```



Histogramm: Alter der Väter (blau) und Mütter (rosa)

Dichteplot

```
densMutter = density(AlterM)
densVater = density(AlterV)
plot(densMutter, main="", xlab="Alter",
xlim=c(40,80), # draw from 40 to 80
panel.first=grid()) # draw a grid
polygon(densVater, density=-1, col=rgb(0,0,1,1/4))
polygon(densMutter, density=-1, col=rgb(1,0,0,1/4))
```



Dichteplot: Alter der Väter (blau) und Mütter (rosa)

"Sollen wir das arithmetische Mittel als durchschnittliche Körpergröße nehmen und den Gegner erschrecken, oder wollen wir ihn einlullen und nehmen den Median?"

Modus x_{Mod}: häufigster Wert

Beispiel:

a_j	1	2	4
$h(a_j)$	4	3	1

$\Rightarrow x_{Mod} = 1$

Sinnvoll bei allen Skalenniveaus.

Median x_{Med}: $= \tilde{x}_{0,5}$ ‚mittlerer Wert', d.h.

1. Urliste aufsteigend sortieren: $x_1 \leqq x_2 \leqq \cdots \leqq x_n$
2. Dann

$$x_{Med} \begin{cases} = x_{\frac{n+1}{2}}, & \text{falls } n \text{ ungerade} \\ \in [x_{\frac{n}{2}}; x_{\frac{n}{2}+1}], & \text{falls } n \text{ gerade (meist } x_{Med} = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)) \end{cases}$$

Im Beispiel oben:
$1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 4 \Rightarrow x_{Med} \in [1;2]$, z.B. $x_{Med} = 1{,}5$

Sinnvoll ab ordinalem Skalenniveau.

- Arithmetisches Mittel $\bar{x}$: Durchschnitt, d.h.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{j=1}^{k} a_j \cdot h(a_j)$$

Im Beispiel:

$$\bar{x} = \tfrac{1}{8} \cdot (\underbrace{1+1+1+1}_{1\cdot 4} + \underbrace{2+2+2}_{2\cdot 3} + \underbrace{4}_{4\cdot 1}) = 1{,}75$$

Sinnvoll nur bei kardinalem Skalenniveau.
Bei klassierten Daten:

$$\bar{x}^* = \tfrac{1}{n} \sum \text{Klassenmitte} \cdot \text{Klassenhäufigkeit}$$

Im Beispiel:

$$\bar{x}^* = \tfrac{1}{12} \cdot (2{,}5 \cdot 5 + 10 \cdot 5 + 22{,}5 \cdot 2) = 8{,}96 \neq 7{,}5 = \bar{x}$$

Lageparameter

Ausgaben für Schuhe

```
median(AusgSchuhe)

## [1] 200

mean(AusgSchuhe)

## [1] 271.1353
```

Alter

```
median(Alter)

## [1] 21

mean(Alter)

## [1] 22.02122
```

~~Lieblingsfarbe~~

```
summary(Geschlecht)

## Frau Mann
##  250  127
```

Alter der Mutter

```
median(AlterM)

## [1] 51

mean(AlterM)

## [1] 51.61538
```

Streuungsparameter

- Voraussetzung: kardinale Werte $x_1, \ldots, x_n$
- **Beispiel:**

a)	x_i	1950	2000	2050	} je $\bar{x} = 2000$
b)	x_i	0	0	6000	}

- Spannweite: $SP = \max_i x_i - \min_i x_i$

  Im Beispiel:

a) $SP = 2050 - 1950 = 100$
b) $SP = 6000 - 0 = 6000$

- Mittlere quadratische Abweichung:

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

$\frac{1}{n}\sum(x_i - \bar{x}) = \frac{1}{n}\sum x_i - \frac{1}{n}\cdot\sum\bar{x} = \bar{x} - \frac{1}{n}\cdot n\bar{x} = 0$

$\frac{1}{n}\sum|x_i - \bar{x}|$

$\frac{1}{n}\sum(x_i - \bar{x})^2$

$\frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}\sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$
$= \frac{1}{n}\left[\sum(x_i^2) - 2\bar{x}\sum(x_i) + \bar{x}^2\sum_{i=1}^{n} 1\right]$
$= \frac{1}{n}\left[\sum(x_i^2) - 2\bar{x}\cdot n\bar{x} + \bar{x}^2\cdot n\right] = \frac{1}{n}\sum(x_i^2) - \bar{x}^2$

Streuungsparameter (2)

- **Mittlere quadratische Abweichung** im Beispiel:

a) $s^2 = \frac{1}{3} \cdot (50^2 + 0^2 + 50^2)$

$= \frac{1}{3} \cdot (1950^2 + 2000^2 + 2050^2) - 2000^2 = 1666{,}67$ $€^2$ (kann interpretiert werden)

b) $s^2 = \frac{1}{3} \cdot (2000^2 + 2000^2 + 4000^2)$

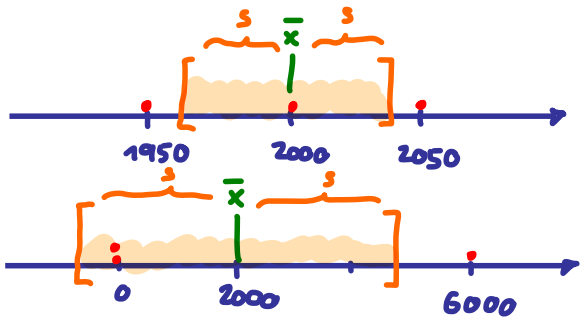
$= \frac{1}{3} \cdot (0^2 + 0^2 + 6000^2) - 2000^2 = 8000000$

- **Standardabweichung:** $s = \sqrt{s^2}$

Im Beispiel:

a) $s = \sqrt{1666{,}67} = 40{,}82$

b) $s = \sqrt{8000000} = 2828{,}43$



- **Variationskoeffizient:** $V = \frac{s}{\bar{x}}$ (maßstabsunabhängig)

Im Beispiel:

a) $V = \frac{40{,}82}{2000} = 0{,}02\ (\widehat{=} 2\,\%)$

b) $V = \frac{2828{,}43}{2000} = 1{,}41\ (\widehat{=} 141\,\%)$

Umsatz

Filiale No	Standard. abw.	arithm. Mittel	Variations-koeffizient $\frac{s}{\bar{x}}$
1	5000000	50 Mio	0,1
2	500000	1 Mio	0,5
3	⋮		
⋮			
100			

Lage und Streuung: Überblick


```
LageStreuung = function(x) {
  x=na.omit(x)  # ignoriere fehlende Werte
  n = length(x) # Anzahl nicht fehlender Werte
  popV = var(x)*(n-1)/n # var() ist nicht mittl. qu. Abweichung
  return(list(mean=mean(x),
              median=median(x),
              Variance=popV,
              StdDev=sqrt(popV),
              VarCoeff=sqrt(popV)/mean(x)))
}
mat1 = sapply(MyData[c("Alter","AlterV","AlterM", # sapply: pro Spalte anwenden
                       "Geschwister", "AnzSchuhe", "AusgSchuhe")],
              LageStreuung)
```

	Alter	AlterV	AlterM	Geschwister	AnzSchuhe	AusgSchuhe
mean	22.02	54.12	51.62	1.49	22.25	271.14
median	21.00	54.00	51.00	1.00	20.00	200.00
Variance	11.07	33.90	25.87	1.27	423.57	42806.14
StdDev	3.33	5.82	5.09	1.13	20.58	206.90
VarCoeff	0.15	0.11	0.10	0.75	0.93	0.76

Univariate Statistik mit dem Taschenrechner

Beispiel: Alter von 10 Studenten

21 29 19 19 21 22 23 33 24 23

gesucht: $\bar{x}$, s, V, x_{med}

Mode → STAT → 1-Var
→ SD

} Statistikmodus für 1 Merkmal

(Tabelle)
<Zahl> [Dt] <Zahl> [Dt] ....

} Dateneingabe

[AC] Shift [STAT] [Var]

} Ergebnisse

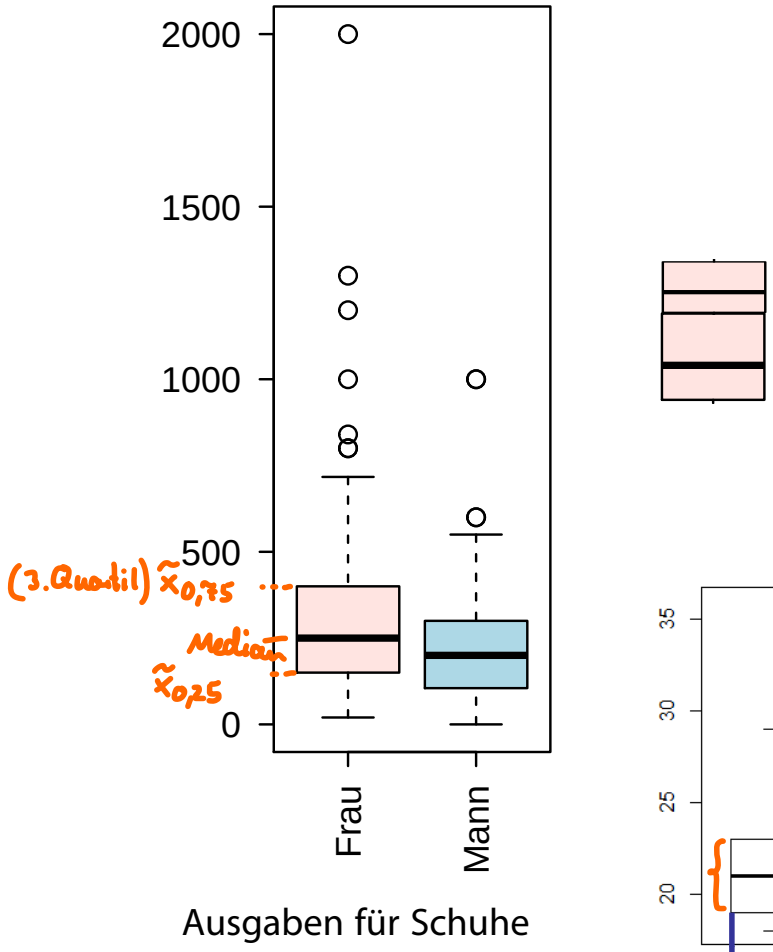
$\bar{x}$ = 23,4

x_{Bn}/b = 4,2

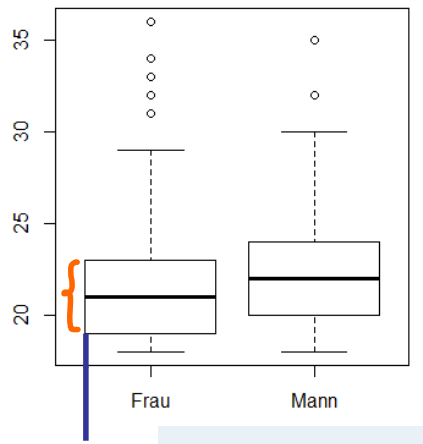
Lage und Streuung als Grafik: Boxplot

- Graphische Darstellung von Lage und Streuung
- Box: Oberer/Unterer Rand: 3. bzw. 1. Quartil ($\tilde{x}_{0,75}$ bzw. $\tilde{x}_{0,25}$),
- Linie in Mitte: Median
- Whiskers: Länge: Max./Min Wert, aber beschränkt durch das 1,5-fache des Quartilsabstands (falls größter/kleinster Wert größeren/kleineren Abstand von Box: Länge Whiskers durch größten/kleinsten Wert innerhalb dieser Schranken)
- Ausreißer: Alle Objekte außerhalb der Whisker-Grenzen

```
boxplot(AusgSchuhe ~ Geschlecht,
        col=c("mistyrose", "lightblue"),
        data=MyData, main="", las=2)
```



Ausgaben für Schuhe

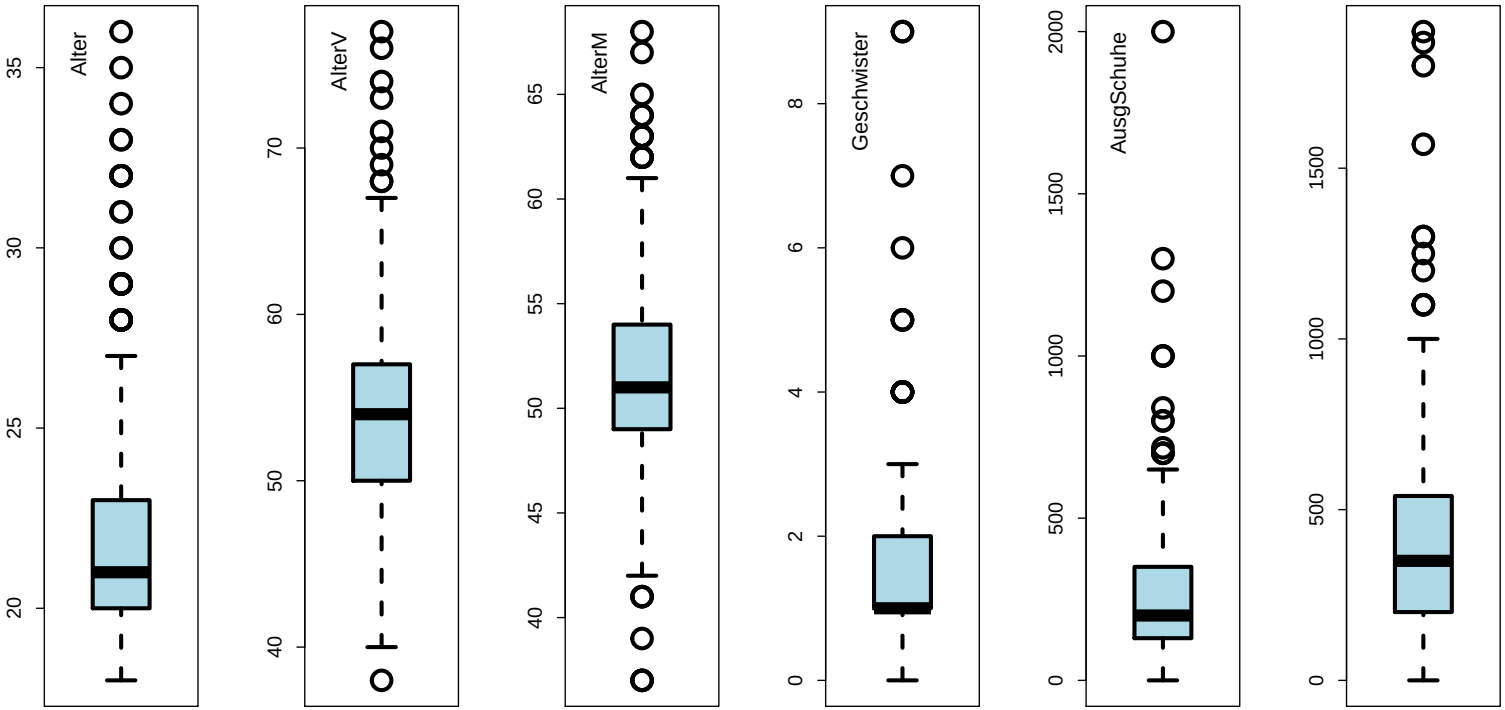



```
summary(MyData)

##     Jahrgang        Alter          Groesse    Geschlecht     AlterV
##  Min.   :2014   Min.   :18.00   Min.   :151   Frau:250   Min.   :38.00
##  1st Qu.:2014   1st Qu.:20.00   1st Qu.:165   Mann:127   1st Qu.:50.00
##  Median :2014   Median :21.00   Median :170              Median :54.00
##  Mean   :2014   Mean   :22.02   Mean   :172              Mean   :54.12
##  3rd Qu.:2015   3rd Qu.:23.00   3rd Qu.:178              3rd Qu.:57.00
##  Max.   :2015   Max.   :36.00   Max.   :197              Max.   :77.00
##
##      AlterM         GroesseV       GroesseM    Geschwister        Farbe
##  Min.   :37.00   Min.   :160.0   Min.   : 76   Min.   :0.000   blau   : 21
##  1st Qu.:49.00   1st Qu.:175.0   1st Qu.:162   1st Qu.:1.000   gelb   :  4
##  Median :51.00   Median :180.0   Median :165   Median :1.000   rot    : 19
##  Mean   :51.62   Mean   :179.4   Mean   :166   Mean   :1.493   schwarz:176
##  3rd Qu.:54.00   3rd Qu.:183.0   3rd Qu.:170   3rd Qu.:2.000   silber : 36
##  Max.   :68.00   Max.   :202.0   Max.   :192   Max.   :9.000   weiss  :121
##
##     AusgKomm      AnzSchuhe       AusgSchuhe          Essgewohnheiten Raucher
##  Min.   :   0   Min.   :  2.00   Min.   :   0.0              :205       :208
##  1st Qu.: 200   1st Qu.: 10.00   1st Qu.: 130.0   carnivor   :157    ja  : 21
##  Median : 350   Median : 20.00   Median : 200.0   fruktarisch :  1    nein:148
##  Mean   : 409   Mean   : 22.25   Mean   : 271.1    pescetarisch:  6
##  3rd Qu.: 540   3rd Qu.: 30.00   3rd Qu.: 350.0    vegan       :  1
##  Max.   :1900   Max.   :275.00   Max.   :2000.0    vegetarisch :  7
##  NA's   :1
##    NoteMathe        MatheZufr
##  Min.   :1.000   nicht    :  0
##  1st Qu.:2.300   geht so  : 81
##  Median :3.300   zufrieden: 68
##  Mean   :3.276   sehr     :  0
##  3rd Qu.:4.000   NA's     :228
##  Max.   :5.000
##  NA's   :73
```

Boxplots

```
for(attribute in c("Alter", "AlterV", "AlterM", "Geschwister",
                   "AusgSchuhe", "AusgKomm")) {
  data=MyData[, attribute]
  boxplot(data, # all rows, column of attribute
          col="lightblue",     # fill color
          lwd=3,               # line width
          cex=2,                # character size
          oma=c(1,1,2,1)
          )
  text(0.7,max(data), attribute, srt=90, adj=1)
}
```



- Gegeben: kardinale Werte $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n$
- **Achtung!** Die Werte müssen aufsteigend sortiert werden!
- Lorenzkurve:

> Wieviel Prozent der Merkmalssumme entfällt auf die x Prozent kleinsten Merkmalsträger?

- **Beispiel:** Die 90 % ärmsten besitzen 20 % des Gesamtvermögens.
- Streckenzug: $(0,0),\ (u_1,v_1),\ \ldots,\ (u_n,v_n) = (1,1)$ mit

$$v_k = \text{Anteil der } k \text{ kleinsten MM-Träger an der MM-Summe} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{n} x_i}$$

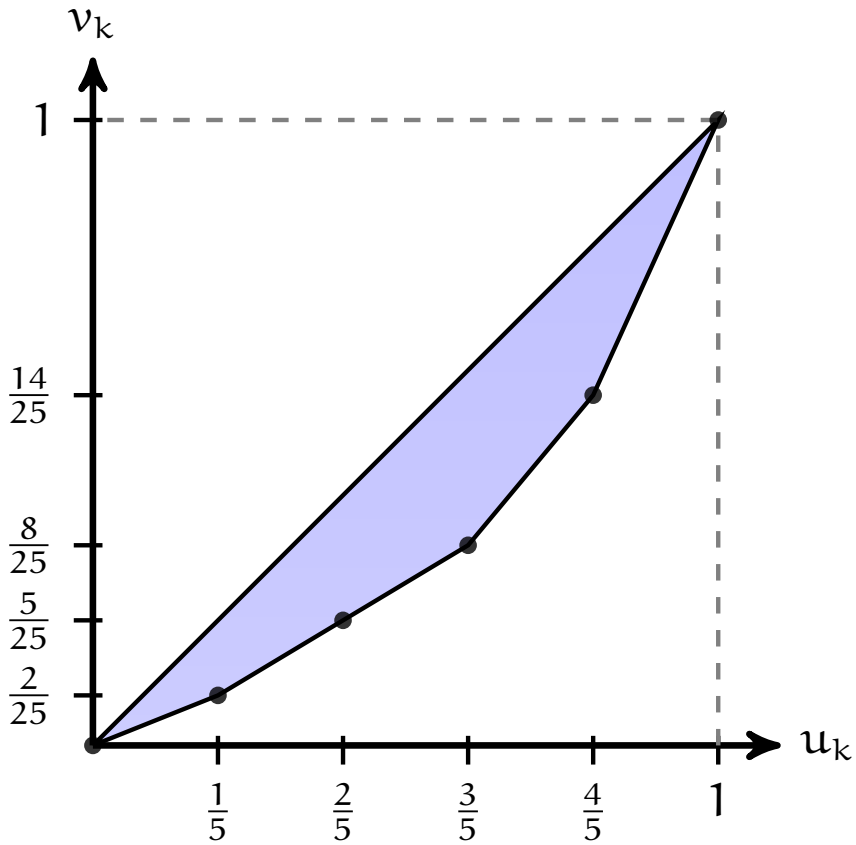
$$u_k = \text{Anteil der } k \text{ kleinsten an der Gesamtzahl der MM-Träger} = \frac{k}{n}$$

Lorenzkurve: Beispiel

Markt mit fünf Unternehmen; Umsätze: 6, 3, 11, 2, 3 (Mio. €)

$\Rightarrow n = 5,\ \sum_{k=1}^{5} x_k = 25$

k	1	2	3	4	5
x_k	2	3	3	6	11
p_k	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{11}{25}$
v_k	$\frac{2}{25}$	$\frac{5}{25}$	$\frac{8}{25}$	$\frac{14}{25}$	1
u_k	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1

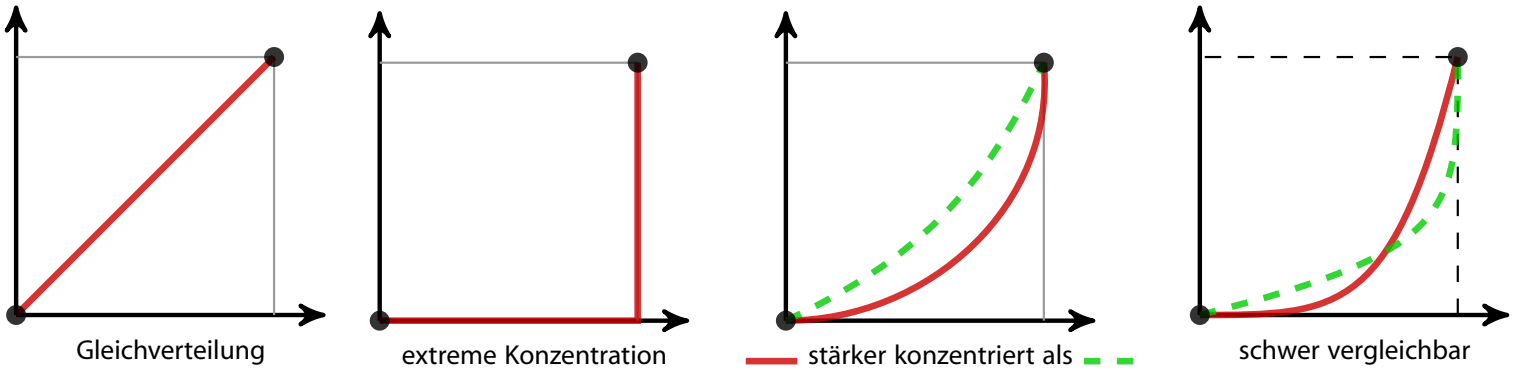


Knickstellen:

- Bei i-tem Merkmalsträger $\Longleftrightarrow$ $x_{i+1} > x_i$
- Empirische Verteilungsfunktion liefert Knickstellen:

a_j	2	3	6	11
$h(a_j)$	1	2	1	1
$f(a_j)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
$F(a_j)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1

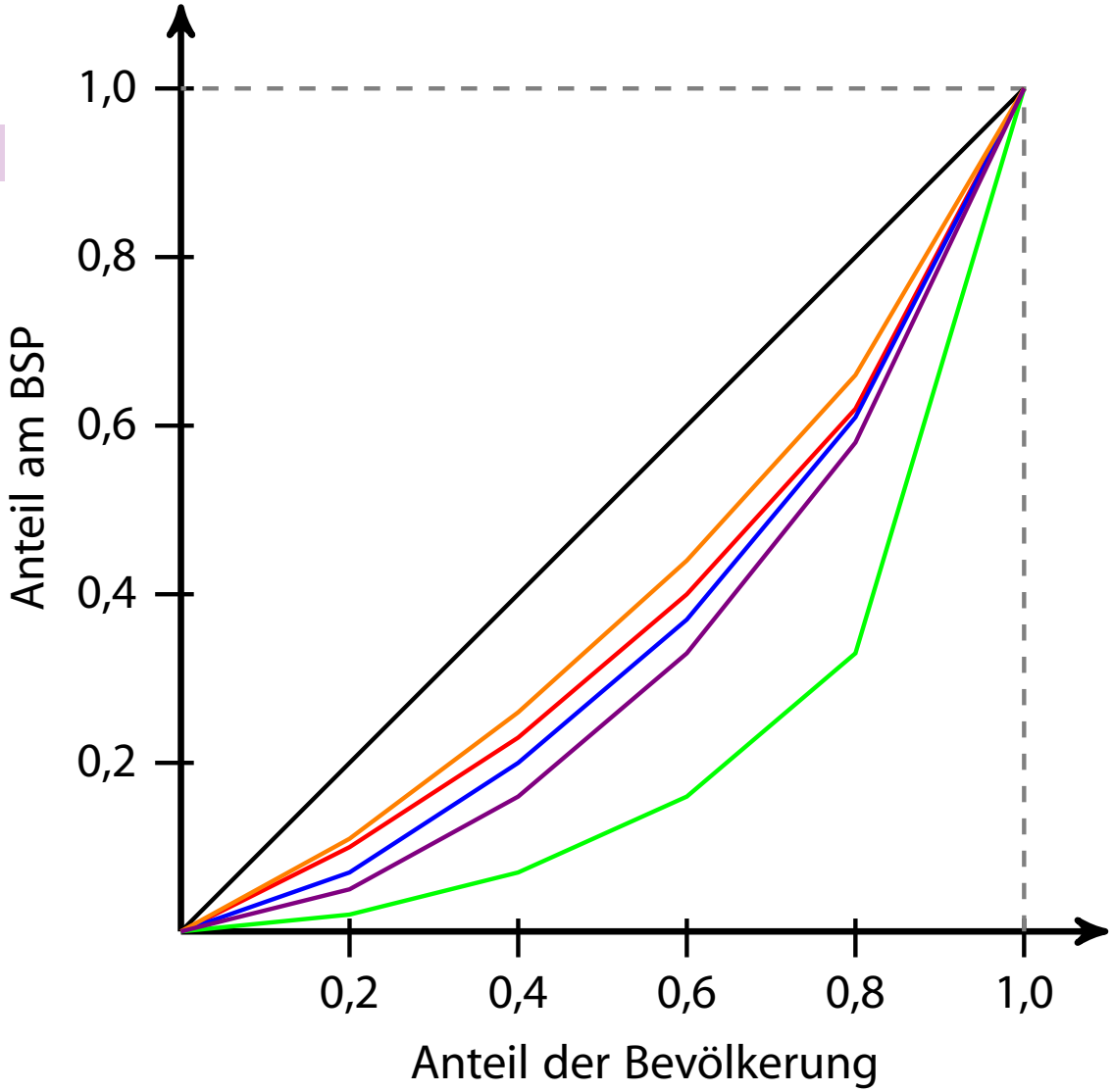
Vergleich von Lorenzkurven:



Lorenzkurve: Beispiel Bevölkerungsanteil gegen BSP

Bangladesch
Brasilien
Deutschland
Ungarn
USA

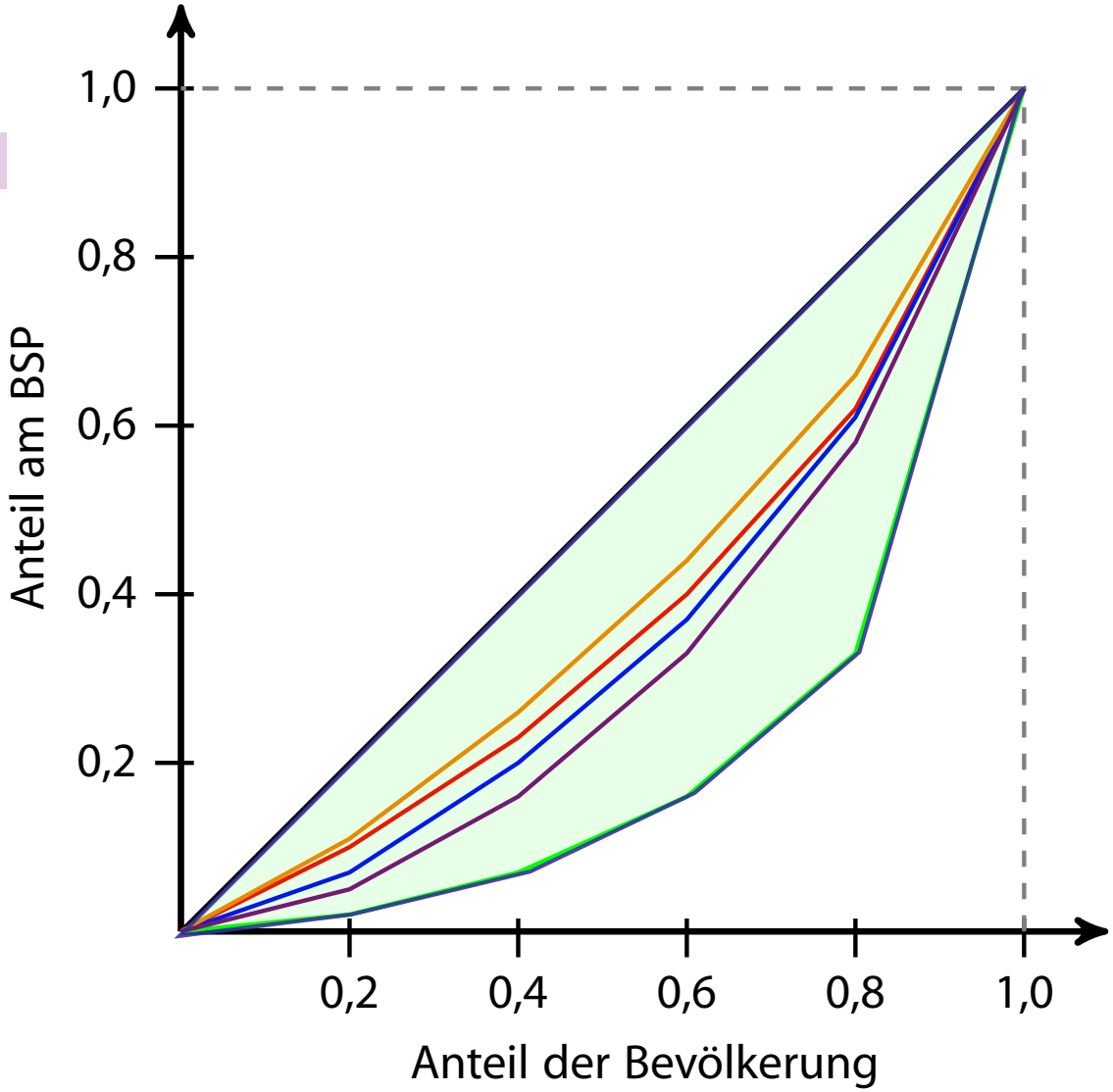
(Stand 2000)



Lorenzkurve: Beispiel Bevölkerungsanteil gegen BSP

Bangladesch
Brasilien
Deutschland
Ungarn
USA

(Stand 2000)



Gini-Koeffizient

- Numerisches Maß der Konzentration: Gini-Koeffizient G

$$G = \frac{\text{Fläche zwischen } 45°\text{-Linie und L}}{\text{Fläche unter } 45°\text{-Linie}} = \frac{\text{◢}}{\text{◢}} \quad = \text{◢} \cdot 2$$

- Aus den Daten:

$$G = \frac{2\sum_{i=1}^{n} i\,x_i - (n+1)\sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i} = \frac{2\sum_{i=1}^{n} i\,p_i - (n+1)}{n} \quad \text{wobei} \quad p_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

- Problem: $G_{max} = \frac{n-1}{n}$

- Normierter Gini-Koeffizient:

$$G_* = \frac{n}{n-1} \cdot G \in [0;1]$$

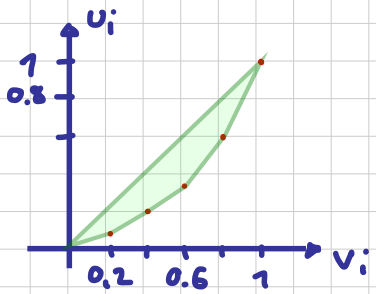
Beispiel (Gini-Koeffizient)

Urliste: 100 500 200 150 300

gesucht: Lorenzkurve, Gini-Koeffizient

$\sum x_i = 1250$

sortiert:	100	150	200	300	500
p_i	$\frac{10}{125}$	$\frac{15}{125}$	$\frac{20}{125}$	$\frac{30}{125}$	$\frac{50}{125}$
u_i	$\frac{10}{125}$	$\frac{25}{125}$	$\frac{45}{125}$	$\frac{75}{125}$	1
v_i	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$



$$G = \frac{2\sum_{i=1}^{n} i\,p_i - (n+1)}{n}$$

$$= \frac{2}{5}\left[1\cdot\frac{10}{125} + 2\cdot\frac{15}{125} + 3\cdot\frac{20}{125} + 4\cdot\frac{30}{125} + 5\cdot\frac{50}{125}\right] - \frac{6}{5}$$

$$= \frac{2}{125}\left[1\cdot 2 + 2\cdot 3 + 3\cdot 4 + 4\cdot 6 + 5\cdot 10\right] - \frac{6}{5}$$

$$= \frac{188}{125} - \frac{6}{5} = \frac{38}{125} \approx 0{,}304$$