Wirtschaftsmathematik

Einführung in einige Teilbereiche der Wirtschaftsmathematik

Wintersemester 2016

HSA Wing Sessionlist WS 2016				
			52	
Datum	N.	Zeit	UE	Themen
Dienstag, 20. September 2016	1	18.00-21.15	4	Einführung, Zinsen, Renten
Dienstag, 27. September 2016	2	18.00-21.15	4	Tilgung, Festverz. Wertpapiere
Dienstag, 4. Oktober 2016	3	18.00-21.15	4	Lineare Optimierung: Einführung, Lösungsmethoden
Samstag, 8. Oktober 2016	4	08.00-11.45	4	Lineare Optimierung: Standardmaximumproblem, Simplex
Dienstag, 11. Oktober 2016	5	18.00-21.15	4	Gewöhnliche Differentialgleichungen
Samstag, 15. Oktober 2016	6	08.00-11.45	4	Analytische Lösung linearer DGLs
Dienstag, 18. Oktober 2016	7	18.00-21.15	4	Einführung, univ. Statistik, Konzentration
Samstag, 22. Oktober 2016	8	11.45-15.00	4	Korrelation, Regression, Preisindizes
Dienstag, 25. Oktober 2016	9	18.00-21.15	4	Kombinatorik, Wahrscheinlichkeiten; Binomial, Hypergeo, Poisson
Samstag, 29. Oktober 2016	10	08.00-11.15	4	Zufallsvariablen, Lage- und Streuung, Stetige ZV, Gleich-vtlg.
Dienstag, 15. November 2016	11	18.00-21.15	4	Normalvtlg., Schätzen und Eigenschaften von Punktschätzern
Samstag, 26. November 2016	12	08.00-11.15	4	Konfidenzintervalle, t-Test
Dienstag, 29. November 2016	13	18.00-21.15	4	Puffer, Wiederholung Besprechung Probeklausur
Samstag, 3. Dezember 2016		09.30-11.00		Klausur (regulärer Termin 90 Min., mit Aufsicht)

Prof. Dr. Stefan Etschberger
HSA

Lageparameter

"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

$\Rightarrow x_{\text{Mod}} = 1$

Sinnvoll bei allen Skalenniveaus.

Median x_{Med}: ‚mittlerer Wert', d.h.

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

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

Im Beispiel oben:

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

Sinnvoll ab ordinalem Skalenniveau.

Lageparameter (2)

- 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(na.exclude(AusgSchuhe))

## [1] 200

mean(na.exclude(AusgSchuhe))

## [1] 270.4529
```

Alter

```
median(Alter)

## [1] 21

mean(Alter)

## [1] 22.12537
```

~~Lieblingsfarbe~~

```
summary(Geschlecht)

## Frau Mann
##  389  281
```

Alter der Mutter

```
median(na.exclude(AlterM))

## [1] 51

mean(na.exclude(AlterM))

## [1] 51.63677
```

Streuungsparameter

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

$$\left.\begin{array}{ll|ccc} \text{a)} & x_i & 1950 & 2000 & 2050 \\ \hline \text{b)} & x_i & 0 & 0 & 6000 \end{array}\right\} \text{je } \bar{x} = 2000$$

- Spannweite: $\text{SP} = \max\limits_i x_i - \min\limits_i x_i$

  Im Beispiel:

$$\begin{array}{lll} \text{a)} & \text{SP} = 2050 - 1950 & = 100 \\ \text{b)} & \text{SP} = 6000 - 0 & = 6000 \end{array}$$

- Mittlere quadratische Abweichung:

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

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

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

$$= \frac{1}{n}\sum x_i^2 - 2\bar{x}\underbrace{\frac{1}{n}\sum x_i}_{\bar{x}} + \overbrace{\frac{1}{n}\sum^{n} \bar{x}^2}^{\frac{1}{n}\cdot n\cdot\bar{x}^2} = \frac{1}{n}\sum x_i^2 - 2\bar{x}^2 + \bar{x}^2$$

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

- **Mittlere quadratische Abweichung** im Beispiel:

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

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

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

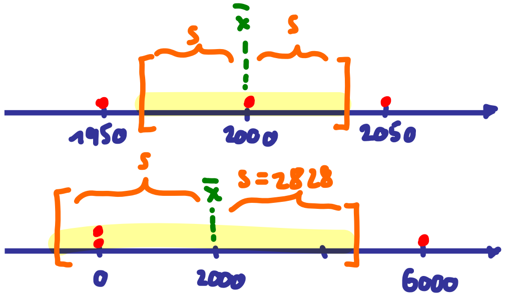
  $\quad = \frac{1}{3} \cdot (0^2 + 0^2 + 6000^2) - 2000^2 = 8000000$

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

Beispiel: Umsatzschwankungen

Filiale	s	$\bar{x}$	$s/\bar{x}$
1	500000	5 Mio	0.1
2	1000000	2000000	0.5
⋮	⋮	⋮	⋮
100			

Univariate Statistik mit TR

Mode → SD
→ STAT → 1-Var

Daten 1950
2000
2050

AC

Shift → STAT → Var → $\bar{x}$ 2000
→ ~~$s_{x n}$~~
$\sigma_{x n}$ 40.82

Lage und Streuung: Überblick

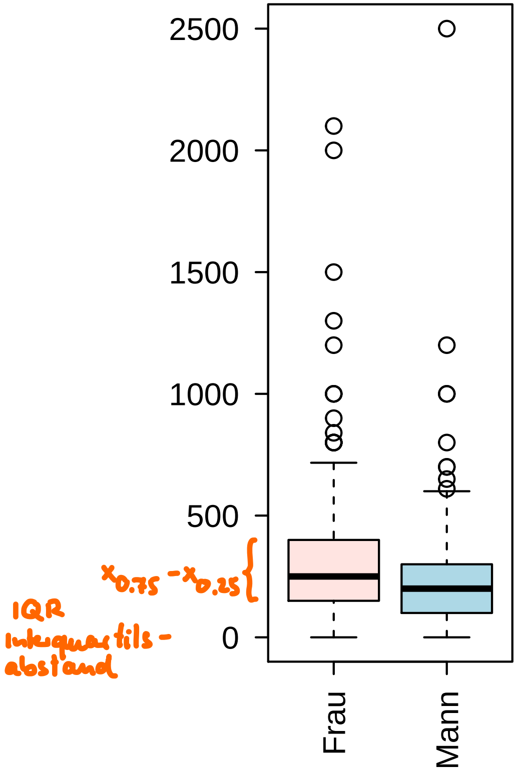
```
LageStreuung = function(x) {
  x=na.omit(x)  # ignoriere fehlende Werte
  n = length(x) # Anzahl nicht fehlender Werte
  popV = var(x)*(n-1)/n # var() ist nicht mittl. qu. Abweichung
  return(list(mean=mean(x),
              median=median(x),
              Variance=popV,
              StdDev=sqrt(popV),
              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	22.13	54.28	51.64	1.51	21.22	270.45
median	21.00	54.00	51.00	1.00	16.00	200.00
Variance	11.36	35.35	25.74	1.18	415.51	56333.39
StdDev	3.37	5.95	5.07	1.08	20.38	237.35
VarCoeff	0.15	0.11	0.10	0.72	0.96	0.88

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

Beispiel: Bargeld:

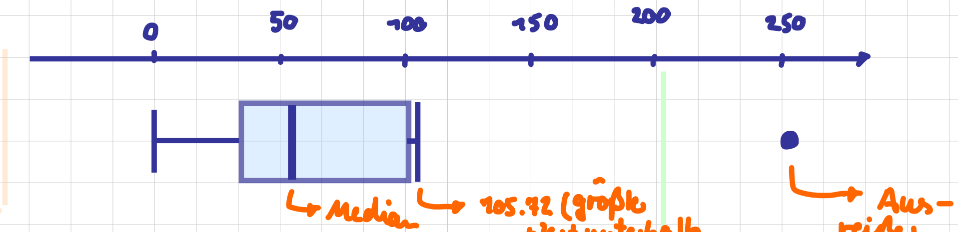
```
  0.00  29.47  36.64  51.25
 53.00  59.34  66.44 103.31
105.72 255.49
```

Aufgabe: Boxplot

$x_{0.25}$: $n \cdot p = 10 \cdot 0.25 = 2.5$ $\quad \tilde{x}_{0.25} = x_{\lceil 2.5 \rceil} = x_3 = 36.64$

$x_{0.5}$: $n \cdot p = 10 \cdot 0.5 = 5$ $\quad \tilde{x}_{0.5} = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(53 + 59.34) = 56.17$

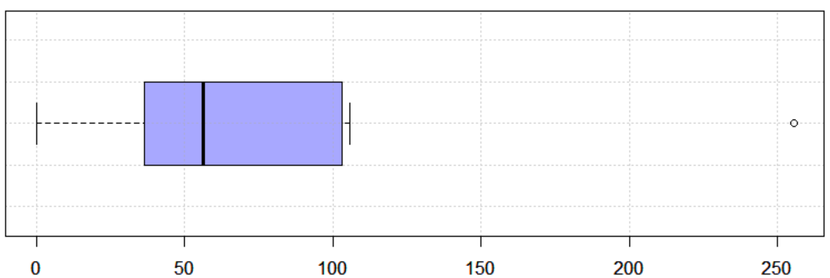
$x_{0.75}$: $\quad \tilde{x}_{0.75} = x_8 = 103.31$



$1.5 \cdot (x_{0.75} - x_{0.25}) = 100.005$

$x_{0.75} + 1.5 \cdot IQR = 103.31 + 100.005 = 203.315$

$x_{0.25} - 1.5 \cdot IQR = 36.64 - 100.005 = -63.36$

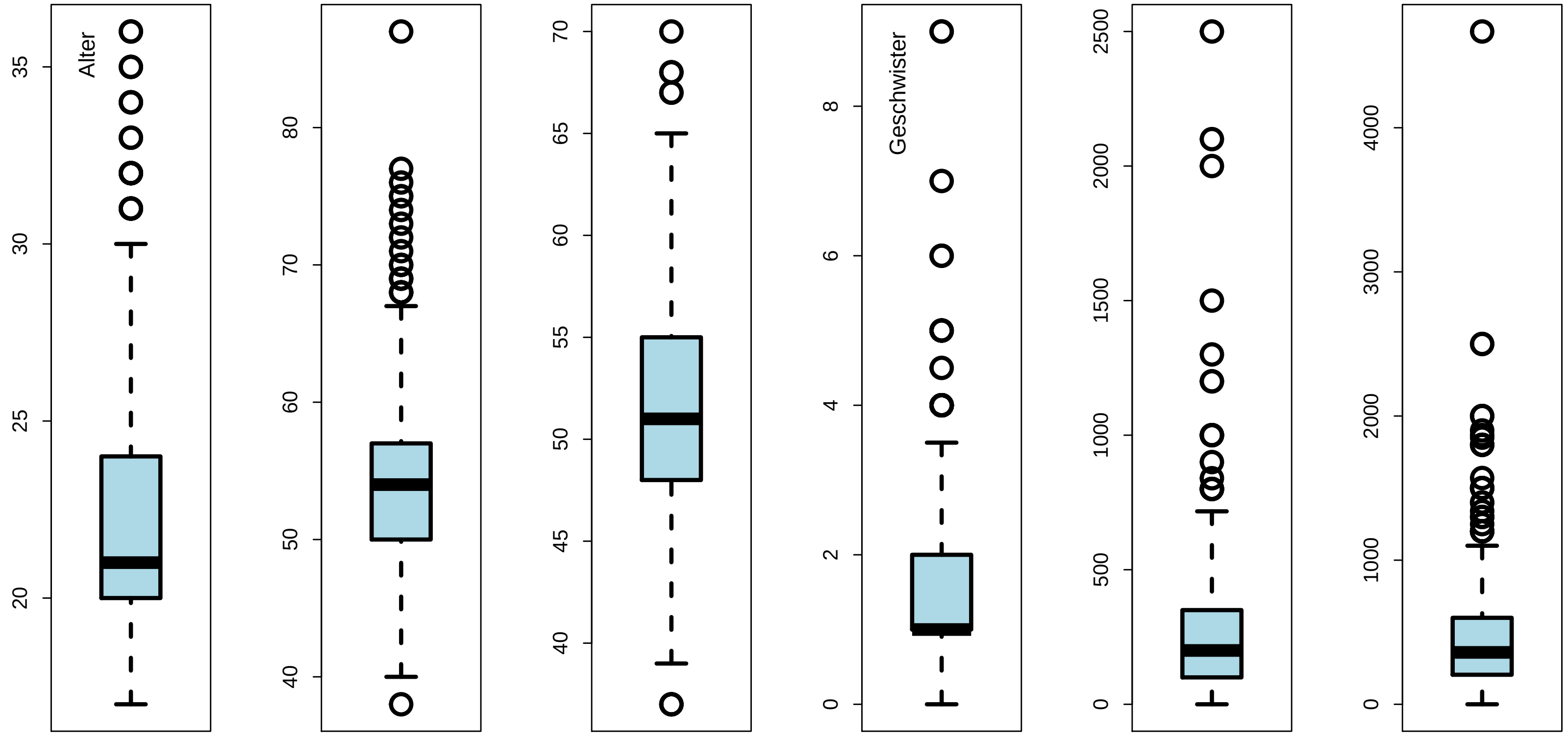


```r
summary(MyData)

##     Jahrgang        Alter          Groesse      Geschlecht     AlterV          AlterM     
##  Min.   :2014   Min.   :17.00   Min.   :150.0   Frau:389   Min.   :38.00   Min.   :37.00  
##  1st Qu.:2014   1st Qu.:20.00   1st Qu.:166.0   Mann:281   1st Qu.:50.00   1st Qu.:48.00  
##  Median :2015   Median :21.00   Median :172.0              Median :54.00   Median :51.00  
##  Mean   :2015   Mean   :22.13   Mean   :173.1              Mean   :54.28   Mean   :51.64  
##  3rd Qu.:2016   3rd Qu.:24.00   3rd Qu.:180.0              3rd Qu.:57.00   3rd Qu.:55.00  
##  Max.   :2016   Max.   :36.00   Max.   :198.0              Max.   :87.00   Max.   :70.00  
##                                                            NA's   :1       NA's   :1      
##     GroesseV        GroesseM      Geschwister        Farbe        AusgKomm        AnzSchuhe     
##  Min.   :160.0   Min.   : 76.0   Min.   :0.000   blau   : 31   Min.   :   0.0   Min.   :  2.00  
##  1st Qu.:175.0   1st Qu.:162.0   1st Qu.:1.000   gelb   :  5   1st Qu.: 207.5   1st Qu.:  8.00  
##  Median :180.0   Median :165.0   Median :1.000   rot    : 24   Median : 360.0   Median : 16.00  
##  Mean   :179.1   Mean   :166.2   Mean   :1.509   schwarz:333   Mean   : 458.1   Mean   : 21.22  
##  3rd Qu.:183.0   3rd Qu.:170.0   3rd Qu.:2.000   silber : 82   3rd Qu.: 600.0   3rd Qu.: 30.00  
##  Max.   :204.0   Max.   :192.0   Max.   :9.000   weiss  :195   Max.   :4668.0   Max.   :275.00  
##  NA's   :11      NA's   :8                                     NA's   :2                        
##    AusgSchuhe          Essgewohnheiten Raucher       NoteMathe               MatheZufr    Studiengang
##  Min.   :   0.0   carnivor     :420    ja  : 81   Min.   :1.000   unzufrieden    :185   BW   :107  
##  1st Qu.: 100.0   fruktarisch  :  1    nein:381   1st Qu.:2.650   geht so        :151   ET   :  1  
##  Median : 200.0   pescetarisch: 26     NA's:208   Median :3.300   zufrieden      :114   IM   : 74  
##  Mean   : 270.5   vegan        :  3               Mean   :3.233   sehr zufrieden: 74    Inf  : 48  
##  3rd Qu.: 350.0   vegetarisch : 15                3rd Qu.:4.000   NA's           :146   WI   : 59  
##  Max.   :2500.0   NA's         :205               Max.   :5.000                         NA's:381  
##  NA's   :1                                        NA's   :162                                     
```

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)
}
```



Konzentrationsmaße

- Gegeben: kardinale Werte $0 \leqq x_1 \leqq x_2 \leqq \dots \leqq x_n$
- **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: $(0,0),\ (u_1, v_1),\ \dots,\ (u_n, v_n) = (1,1)$ mit

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

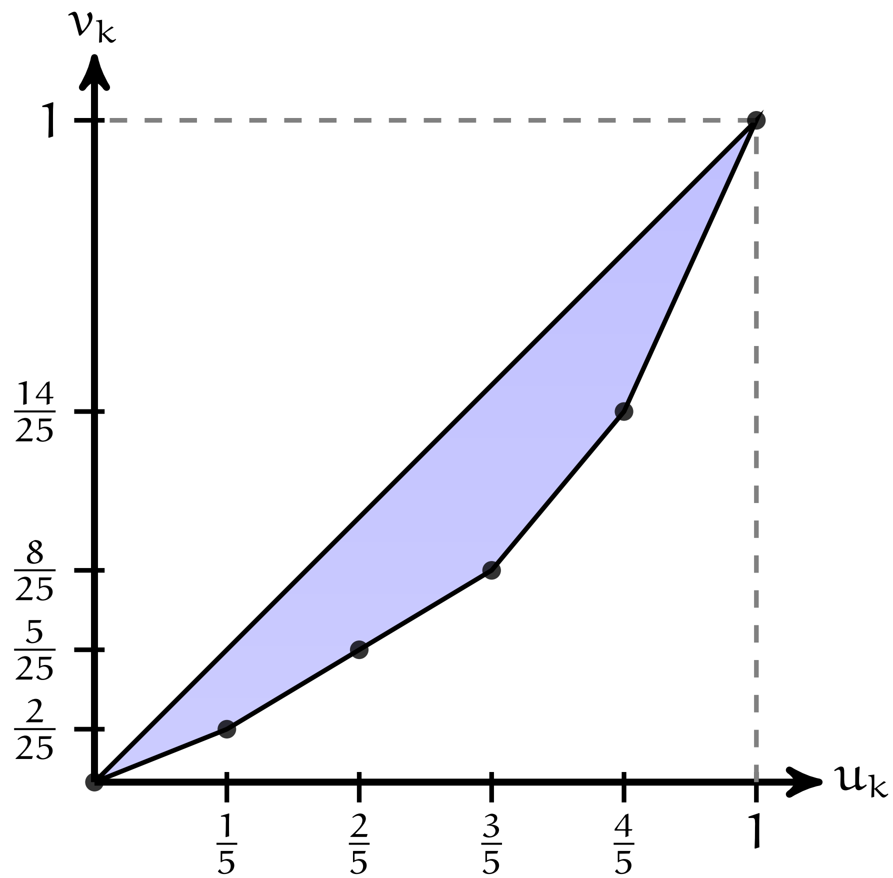
$$u_k = \text{Anteil der } k \text{ kleinsten an der Gesamtzahl der MM-Träger} = \frac{k}{n}$$

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

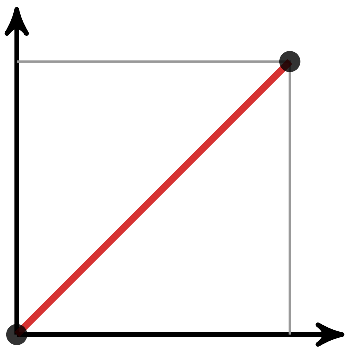


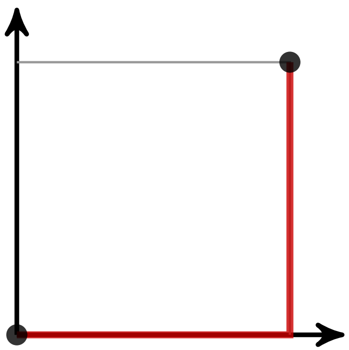
Knickstellen:

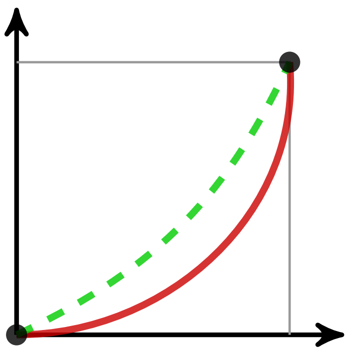
- Bei i-tem Merkmalsträger $\iff x_{i+1} > x_i$
- Empirische Verteilungsfunktion liefert Knickstellen:

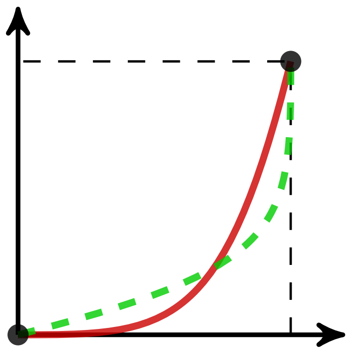
a_j	2	3	6	11
$h(a_j)$	1	2	1	1
$f(a_j)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
$F(a_j)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1

Vergleich von Lorenzkurven:


Gleichverteilung


extreme Konzentration

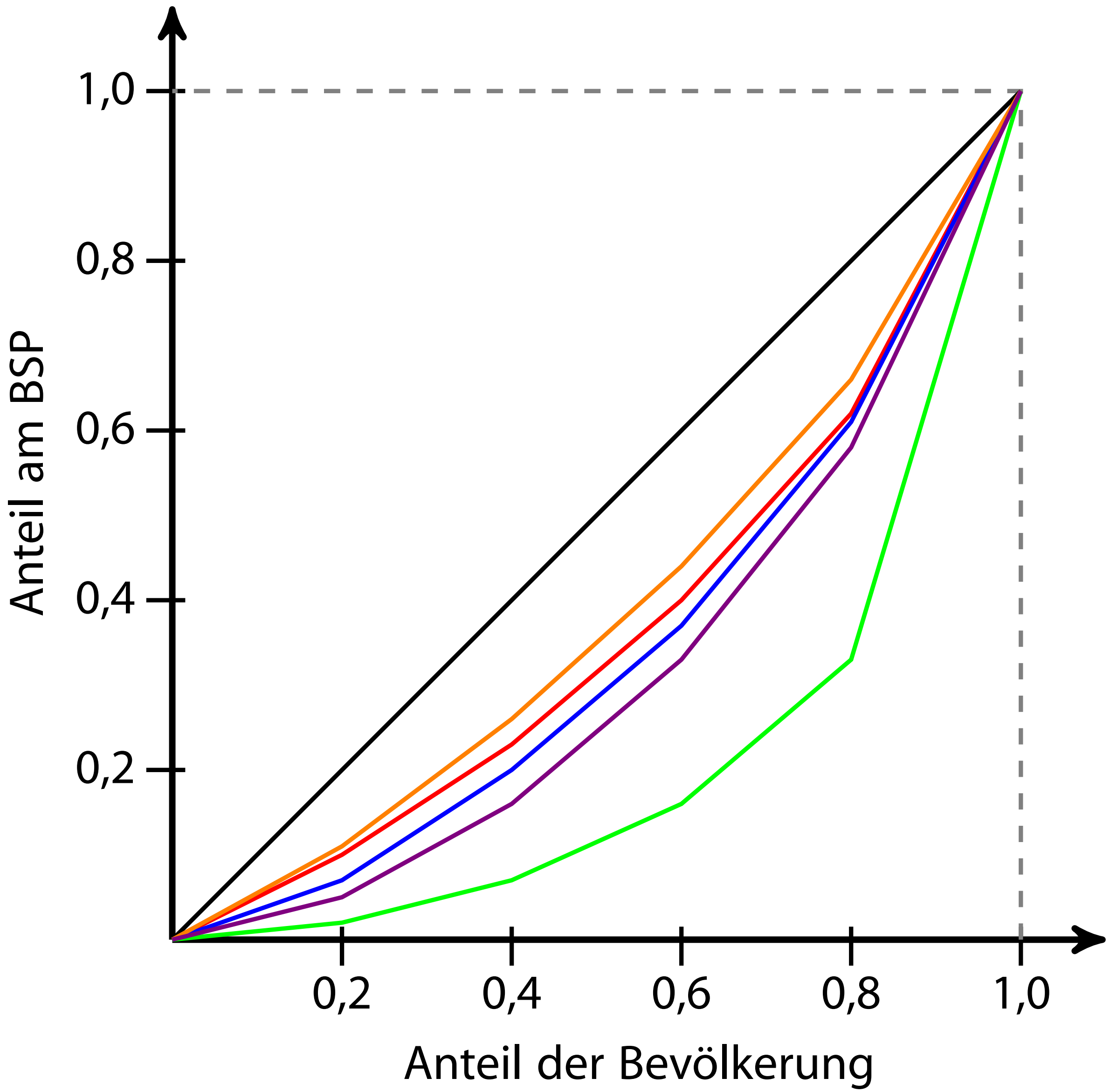

stärker konzentriert als


schwer vergleichbar

Lorenzkurve: Beispiel Bevölkerungsanteil gegen BSP

Bangladesch
Brasilien
Deutschland
Ungarn
USA

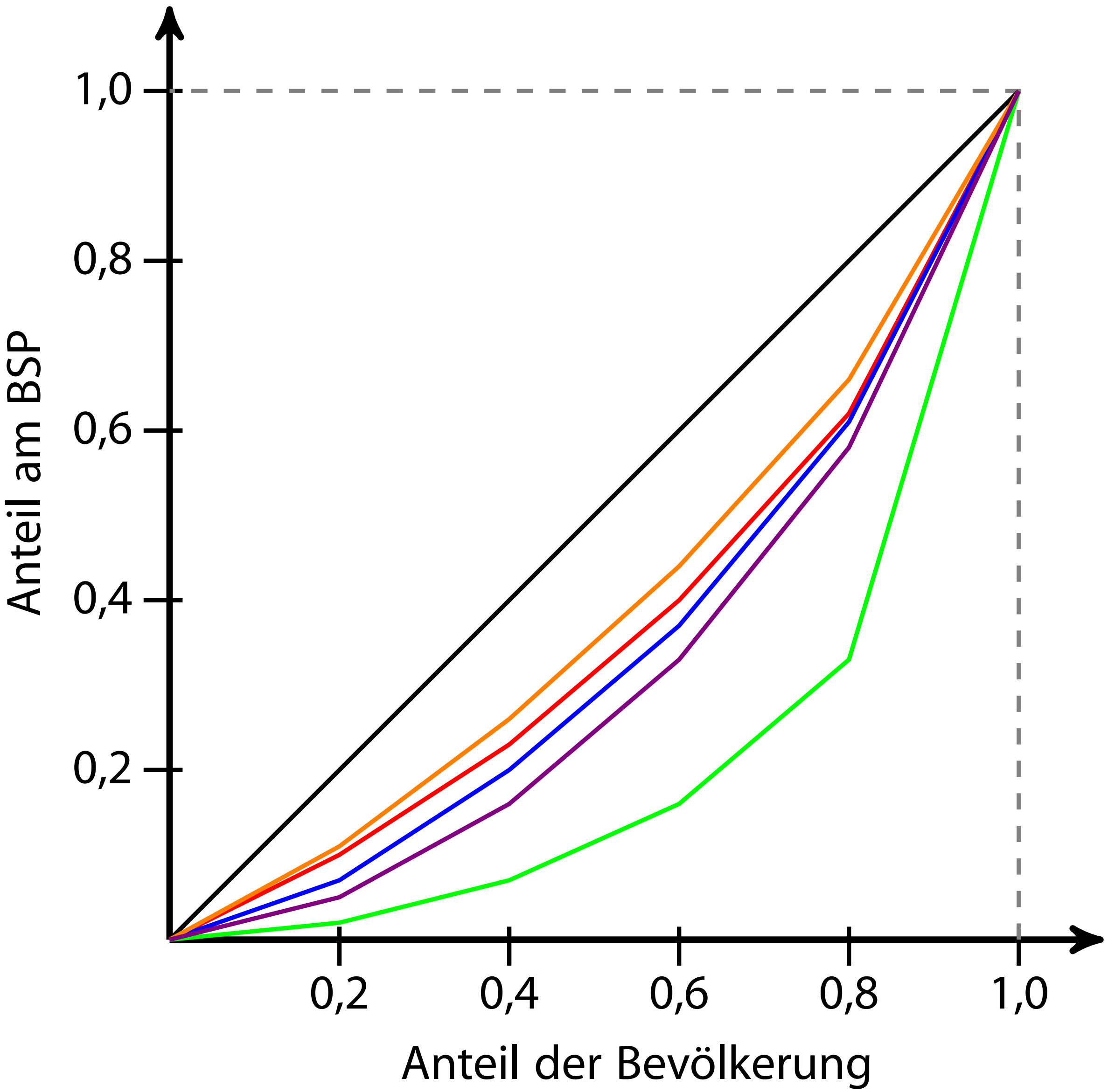
(Stand 2000)



Lorenzkurve: Beispiel Bevölkerungsanteil gegen BSP

Bangladesch
Brasilien
Deutschland
Ungarn
USA

(Stand 2000)



Gini-Koeffizient

- Numerisches Maß der Konzentration: Gini-Koeffizient G

$$G = \frac{\text{Fläche zwischen } 45^\circ\text{-Linie und } L}{\text{Fläche unter } 45^\circ\text{-Linie}} =$$

$= \cdot 2$

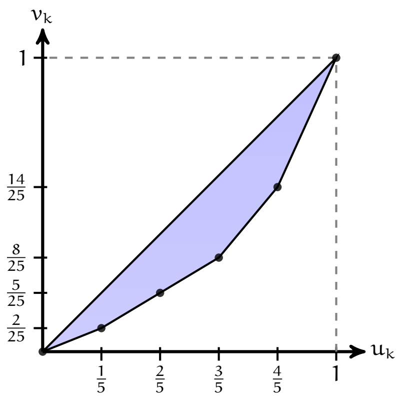
- Aus den Daten:

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

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

➠ Normierter Gini-Koeffizient:

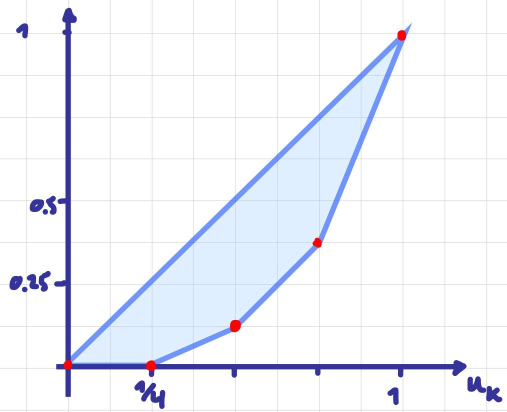
$$G_* = \frac{n}{n-1} \cdot G \in [0; 1]$$



Beispiel: Lorenzkurve und Gini-Koeffizient

x_i : 0, 50, 100, 250

i	1	2	3	4
x_i	0	50	100	250
p_i	0	0.125	0.25	0.625
v_i	0	0.125	0.375	1
u_K	1/4	2/4	3/4	4/4



$\sum x_i = 400$

Gini: $G = \frac{1}{n} \cdot \left[2 \cdot \sum i p_i - (n+1)\right]$

$$= \frac{1}{4} \cdot \left[2 \cdot \left(\underbrace{1 \cdot 0}_{1 \cdot p_1} + \underbrace{2 \cdot 0.125}_{2 \cdot p_2} + 3 \cdot 0.25 + 4 \cdot 0.625\right) - (4+1)\right]$$

$$= \frac{1}{4} \cdot \left[2 \cdot (0.25 + 0.75 + 2.5) - 5\right] = 0.5$$

normiert: $G^* = \frac{4}{3} \cdot 0.5 \approx 0.667$

Gini-Koeffizient: Beispiel

Beispiel:

i	1	2	3	4	$\sum$
x_i	1	2	2	15	20
p_i	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{15}{20}$	1

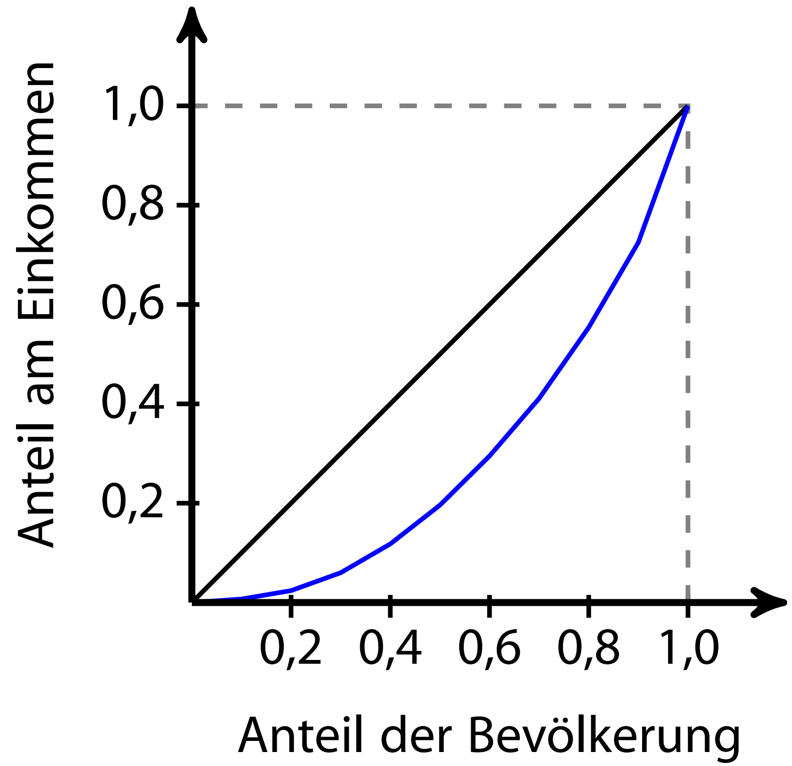
$$G = \frac{2 \cdot \left(1 \cdot \frac{1}{20} + 2 \cdot \frac{2}{20} + 3 \cdot \frac{2}{20} + 4 \cdot \frac{15}{20}\right) - (4+1)}{4} = 0{,}525$$

Mit $G_{\text{max}} = \frac{4-1}{4} = 0{,}75$ folgt

$$G_* = \frac{4}{4-1} \cdot 0{,}525 = 0{,}7$$

Armutsbericht der Bundesregierung 2008

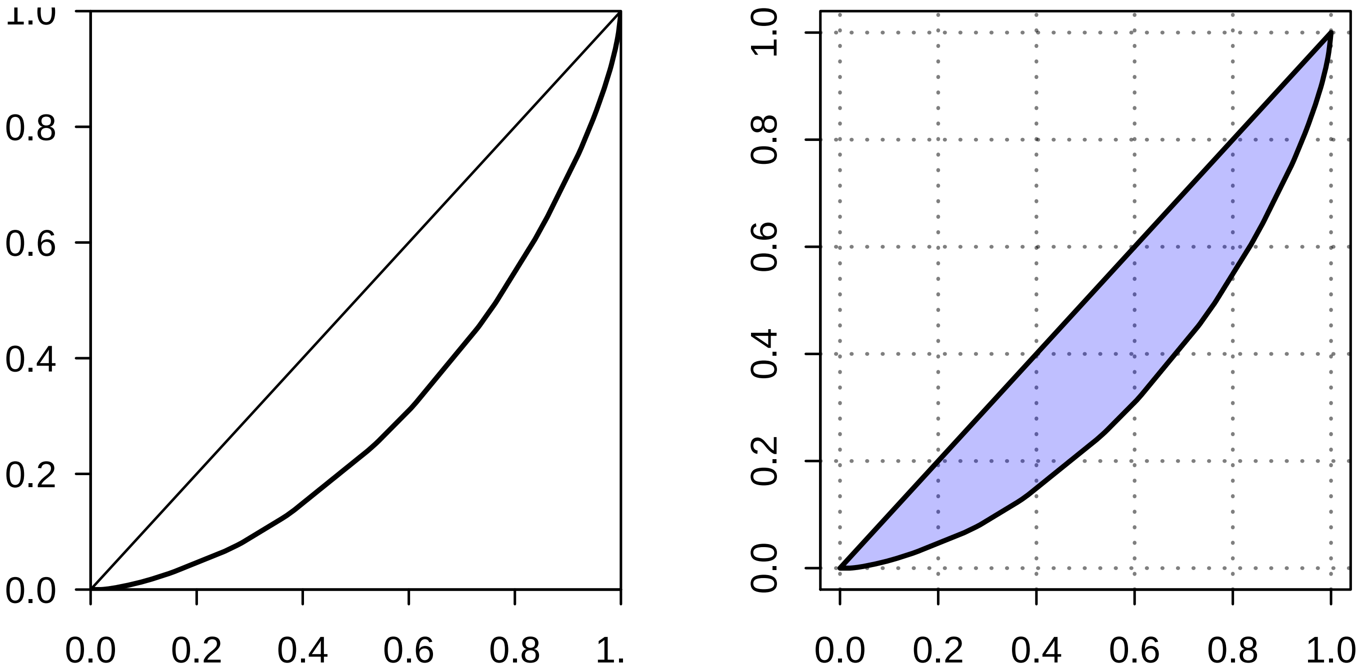
- Verteilung der Bruttoeinkommen in Preisen von 2000
- aus unselbständiger Arbeit der Arbeitnehmer/-innen insgesamt



	2002	2003	2004	2005
Arithmetisches Mittel	24.873	24.563	23.987	23.648
Median	21.857	21.531	20.438	20.089
Gini-Koeffizient	0,433	0,441	0,448	0,453

Lorenzkurve mit R

```r
require(ineq) # inequality Paket
Lorenz = Lc(na.exclude(MyData$AusgSchuhe))
   plot(Lorenz, xlab="", ylab="", main="") # Standard plot
   plot(c(0,1), c(0,1), type="n", # bisschen netter
       panel.first=grid(lwd=1.5, col=rgb(0,0,0,1/2)),
       xlab="", main="", ylab="")
   polygon(Lorenz$p, Lorenz$L, density=-1, col=rgb(0,0,1,1/4), lwd=2)
```




```r
Gini(na.exclude(AusgSchuhe)) # Gini-Koeffizient

## [1] 0.4069336
```

Weitere Konzentrationsmaße

- Konzentrationskoeffizient:

$$CR_g = \text{Anteil, der auf die } g \text{ größten entfällt} = \sum_{i=n-g+1}^{n} p_i = 1 - v_{n-g}$$

- Herfindahl-Index:

$$H = \sum_{i=1}^{n} p_i^2 \qquad (\in [\tfrac{1}{n}; 1])$$

Es gilt: $H = \frac{1}{n}(V^2 + 1)$ bzw. $V = \sqrt{n \cdot H - 1}$

- Exponentialindex:

$$E = \prod_{i=1}^{n} p_i^{p_i} \qquad (\in [\tfrac{1}{n}; 1]) \qquad \text{wobei} \qquad 0^0 = 1$$

- Im Beispiel mit $x = (1, 2, 2, 15)$:

$$CR_2 = \frac{17}{20} = 0{,}85$$

$$H = \left(\frac{1}{20}\right)^2 + \dots + \left(\frac{15}{20}\right)^2 = 0{,}59$$

$$E = \left(\frac{1}{20}\right)^{\frac{1}{20}} \cdots \left(\frac{15}{20}\right)^{\frac{15}{20}} = 0{,}44$$

Zweidimensionale Urliste

Urliste vom Umfang n zu **zwei** Merkmalen X und Y:

$$(x_1, y_1),\quad (x_2, y_2),\quad \ldots,\quad (x_n, y_n)$$

Kontingenztabelle:

Sinnvoll bei wenigen Ausprägungen bzw. bei klassierten Daten.

	Ausprägungen von Y			
Ausprägungen von X	b_1	b_2	$\ldots$	b_l
a_1	h_{11}	h_{12}	$\ldots$	h_{1l}
a_2	h_{21}	h_{22}	$\ldots$	h_{2l}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a_k	h_{k1}	h_{k2}	$\ldots$	h_{kl}

Kontingenztabelle

Unterscheide:

- **Gemeinsame Häufigkeiten:**

$$h_{ij} = h(a_i, b_j)$$

- **Randhäufigkeiten:**

$$h_{i\cdot} = \sum_{j=1}^{l} h_{ij} \quad \text{und} \quad h_{\cdot j} = \sum_{i=1}^{k} h_{ij}$$

- **Bedingte (relative) Häufigkeiten:**

$$f_1(a_i \mid b_j) = \frac{h_{ij}}{h_{\cdot j}} \quad \text{und} \quad f_2(b_j \mid a_i) = \frac{h_{ij}}{h_{i\cdot}}$$

Beispiel: 400 unfallbeteiligte Autoinsassen:

	leicht verletzt $(= b_1)$	schwer verletzt $(= b_2)$	tot $(= b_3)$	
angegurtet $(= a_1)$	264 $(= h_{11})$	90 $(= h_{12})$	6 $(= h_{13})$	360 $(= h_{1\cdot})$
nicht angegurtet $(= a_2)$	2 $(= h_{21})$	34 $(= h_{22})$	4 $(= h_{23})$	40 $(= h_{2\cdot})$
	266 $(= h_{\cdot 1})$	124 $(= h_{\cdot 2})$	10 $(= h_{\cdot 3})$	400 $(= n)$

$f_2(b_3 \mid a_2) = \frac{4}{40} = 0{,}1$ (10 % der nicht angegurteten starben.)

$f_1(a_2 \mid b_3) = \frac{4}{10} = 0{,}4$ (40 % der Todesopfer waren nicht angegurtet.)

Beispiel:

i	1	2	3	4	5	$\sum$
x_i	2	4	3	9	7	25
y_i	4	3	6	7	8	28

$$\Rightarrow \bar{x} = \frac{25}{5} = 5$$

$$\bar{y} = \frac{28}{5} = 5{,}6$$

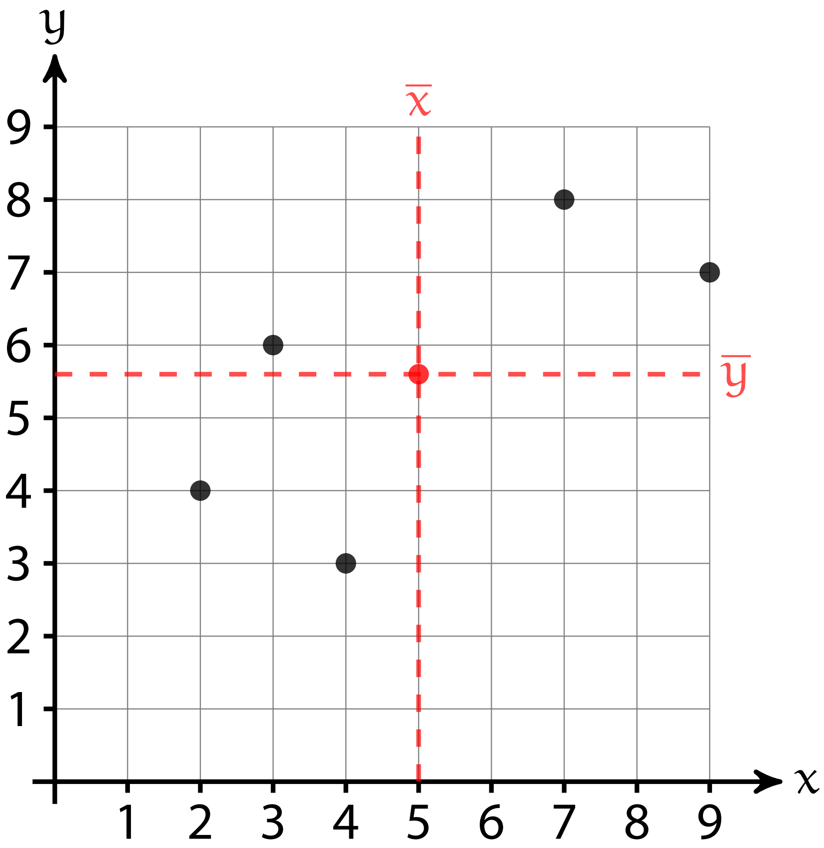
Streuungsdiagramm sinnvoll bei vielen verschiedenen Ausprägungen (z.B. stetige Merkmale)

➠ Alle (x_i, y_i) sowie $(\bar{x}, \bar{y})$ in Koordinatensystem eintragen.

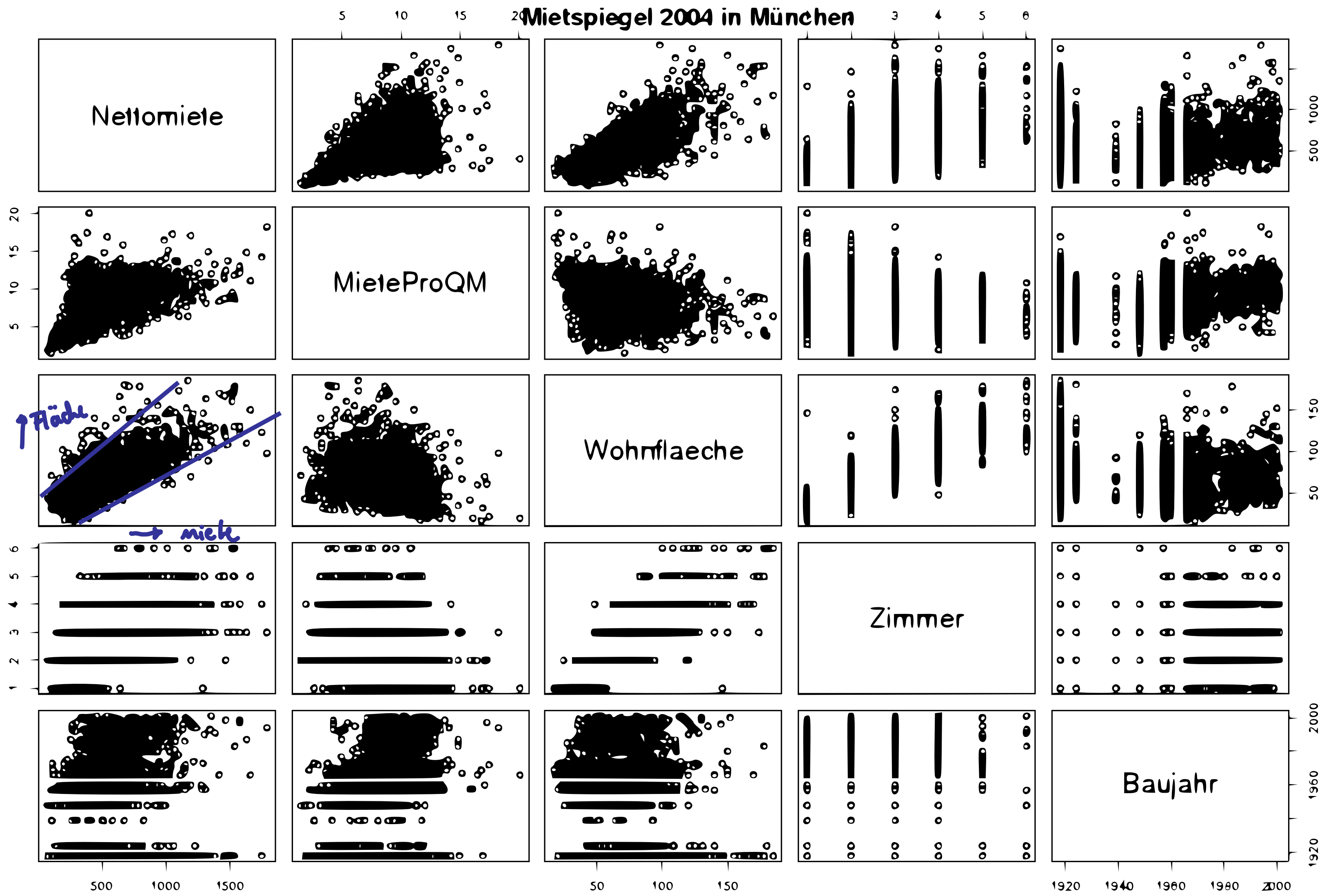
Beispiel:

i	1	2	3	4	5	$\sum$
x_i	2	4	3	9	7	25
y_i	4	3	6	7	8	28

$$\Rightarrow \bar{x} = \frac{25}{5} = 5$$
$$\bar{y} = \frac{28}{5} = 5{,}6$$



Beispiel Streuungsdiagramm

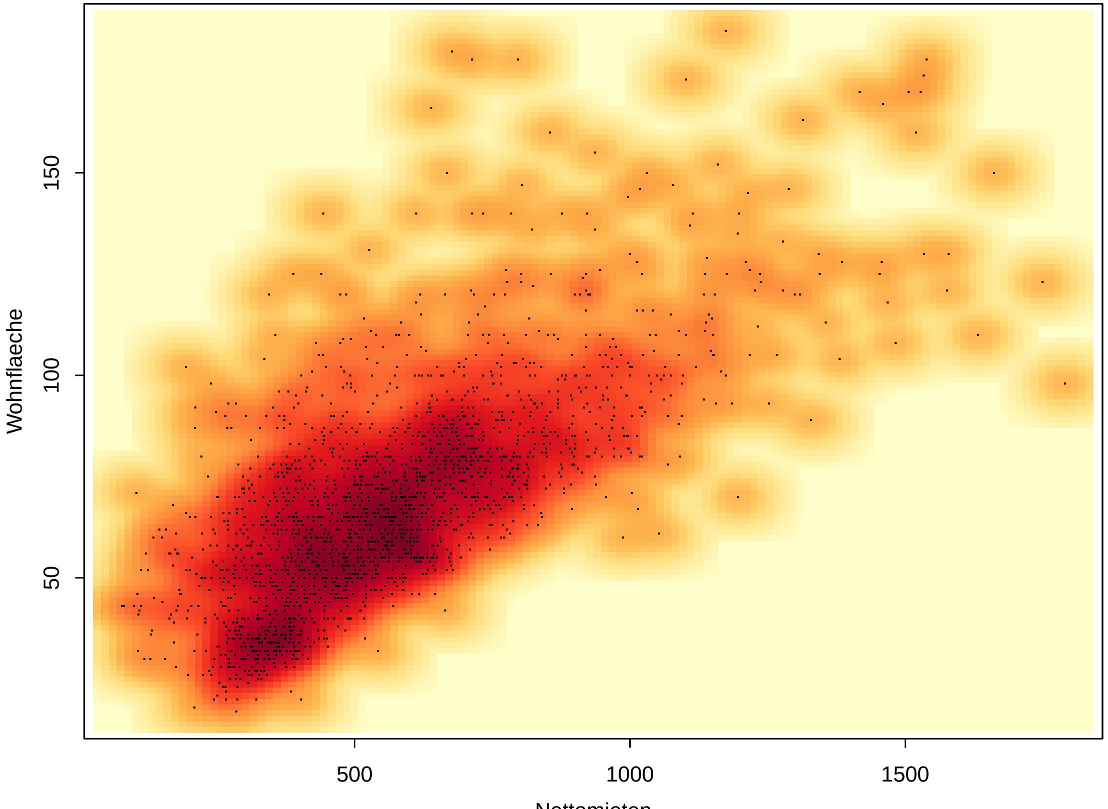


(Datenquelle: Fahrmeir u. a., (2009))

Beispiel Streuungsdiagramm

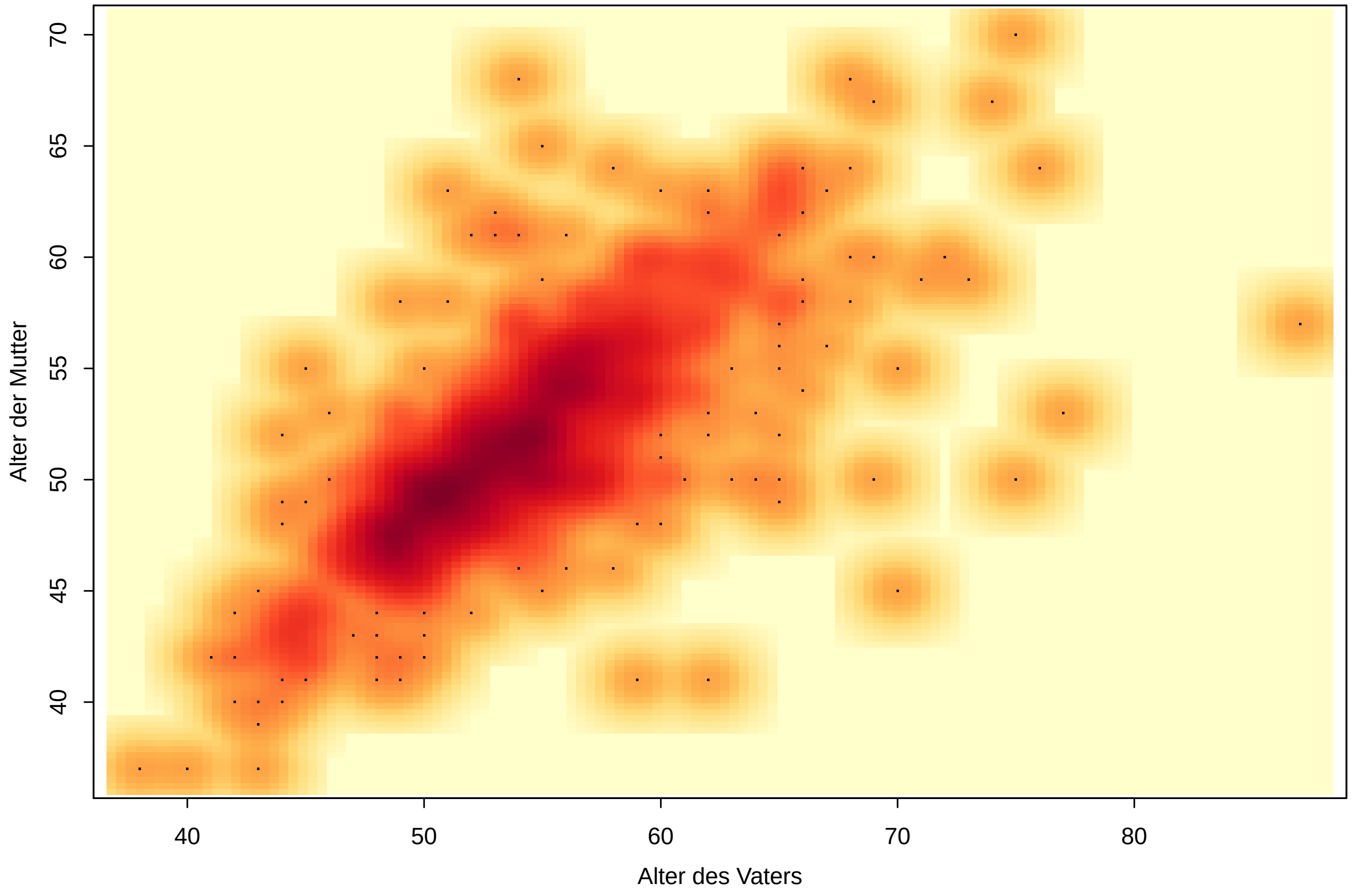

```r
if (!require("RColorBrewer")) {
install.packages("RColorBrewer")
library(RColorBrewer)
}
mieten <- read.table('http://goo.gl/jhpJW4', header=TRUE, sep='\t',
                     check.names=TRUE, fill=TRUE, na.strings=c('',''))
x <- cbind(Nettomieten=mieten$nm, Wohnflaeche=mieten$wfl)

library("geneplotter")  ## from BioConductor
smoothScatter(x, nrpoints=Inf,
              colramp=colorRampPalette(brewer.pal(9,"YlOrRd")),
              bandwidth=c(30,3))
```

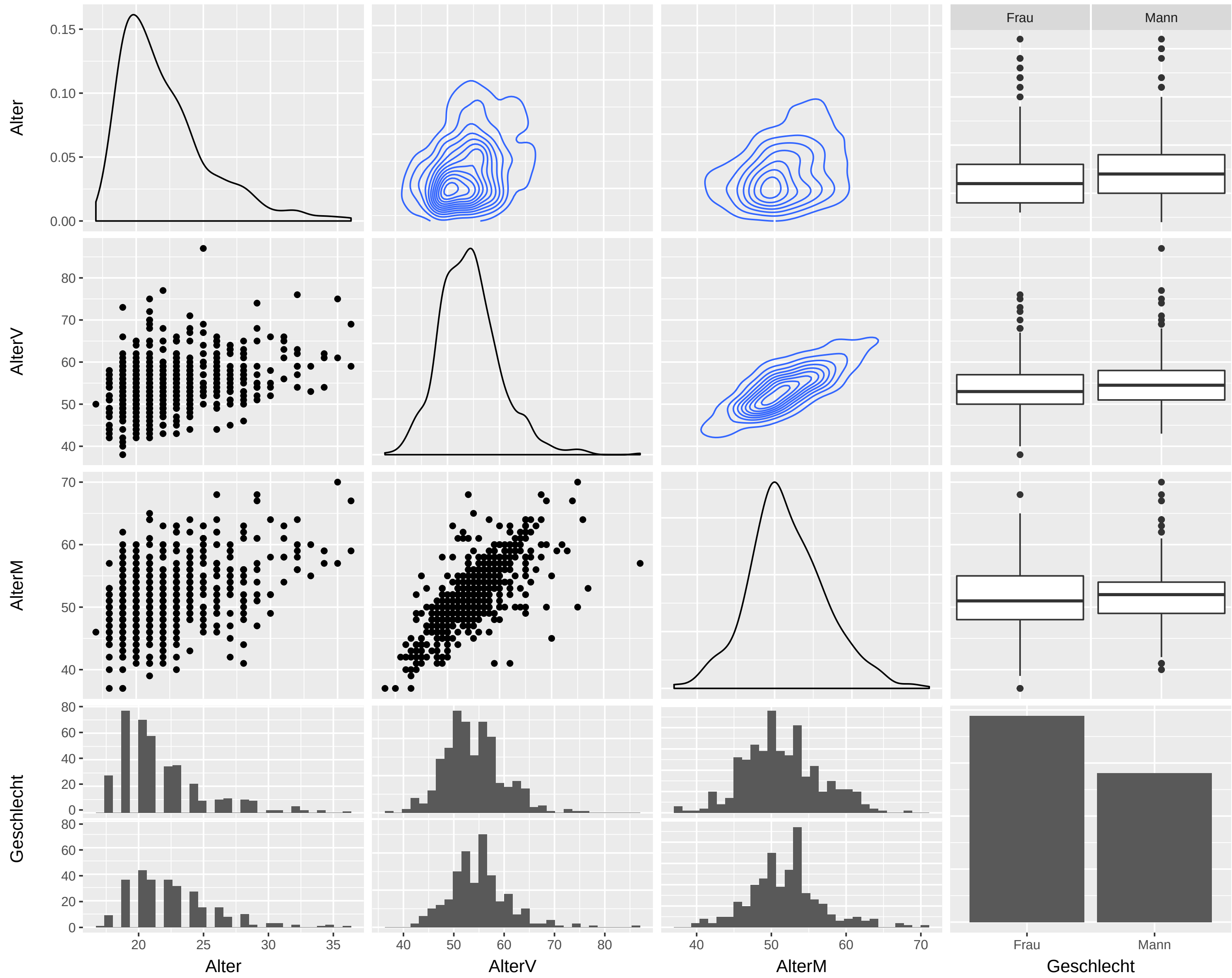


Beispiel Streuungsdiagramm

```
x = cbind("Alter des Vaters"=AlterV, "Alter der Mutter"=AlterM)
require("geneplotter") ## from BioConductor
smoothScatter(x, colramp=colorRampPalette(brewer.pal(9,"YlOrRd")) )
```



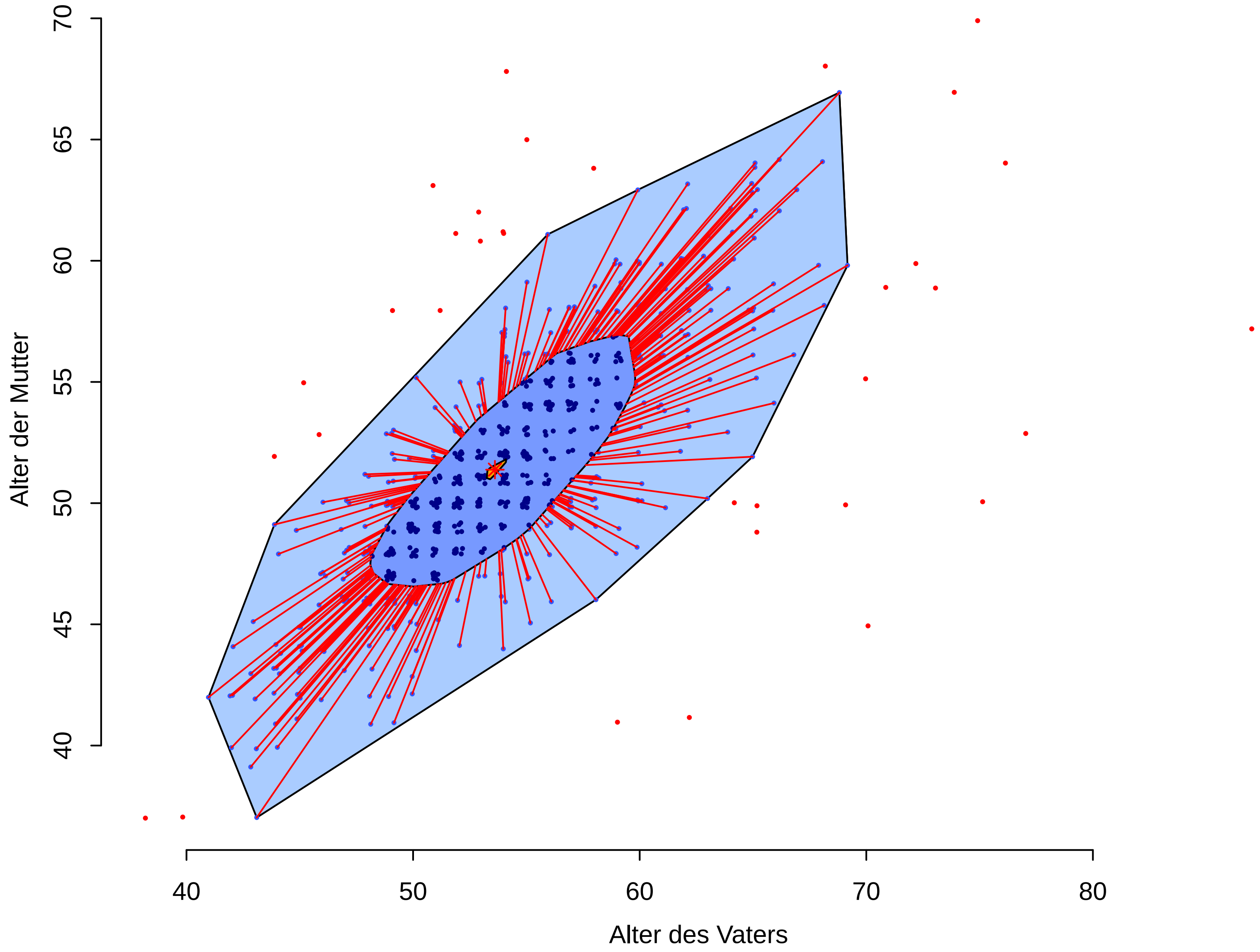

```r
require(GGally)
ggpairs(MyData[, c("Alter", "AlterV", "AlterM", "Geschlecht")],
        upper = list(continuous = "density", combo = "box"),
       color='Geschlecht', alpha=0.5)
```



Bagplot: Boxplot in 2 Dimensionen

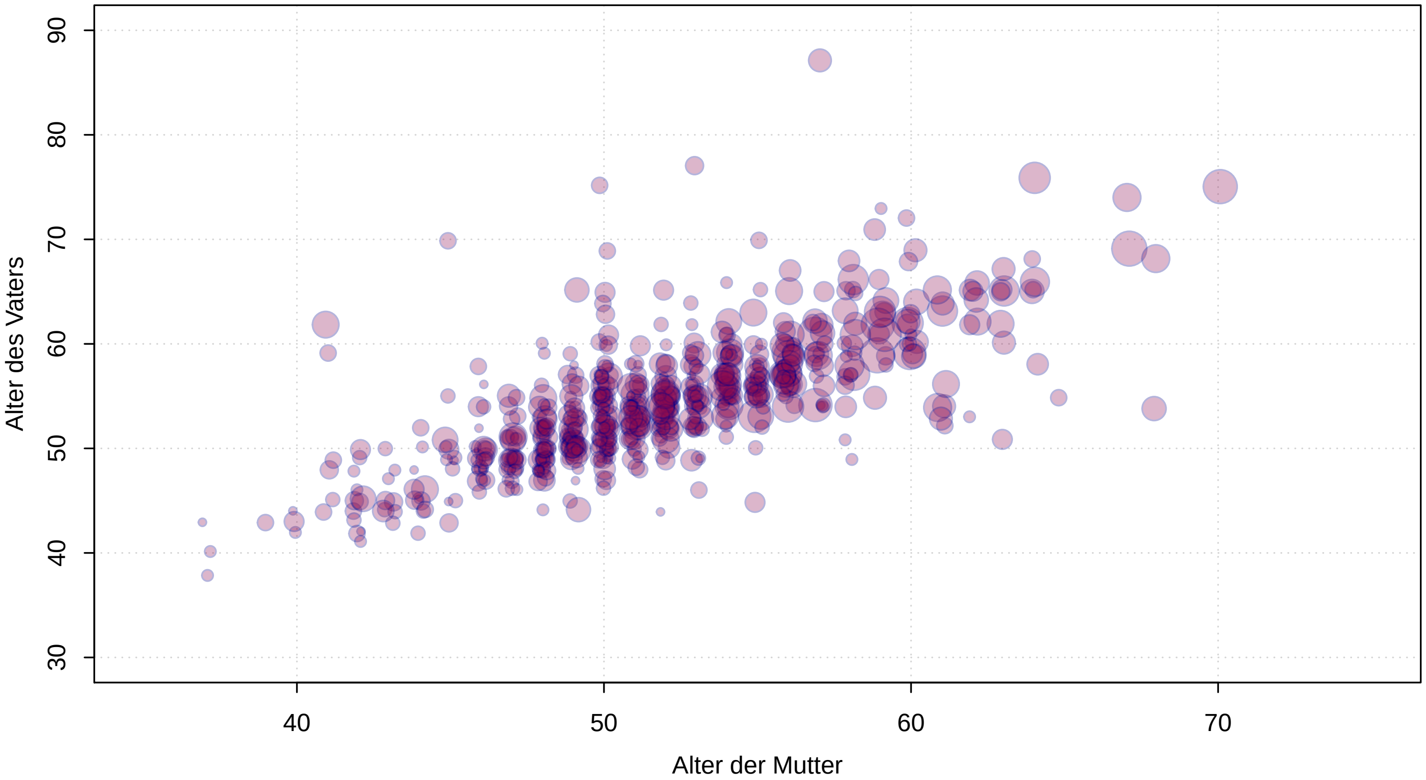

```r
require(aplpack)
bagplot(jitter(AlterV), jitter(AlterM), xlab="Alter des Vaters", ylab="Alter der Mutter")

## [1] "Warning: NA elements have been exchanged by median values!!"
```



Bubbleplot: 3 metrische Variablen

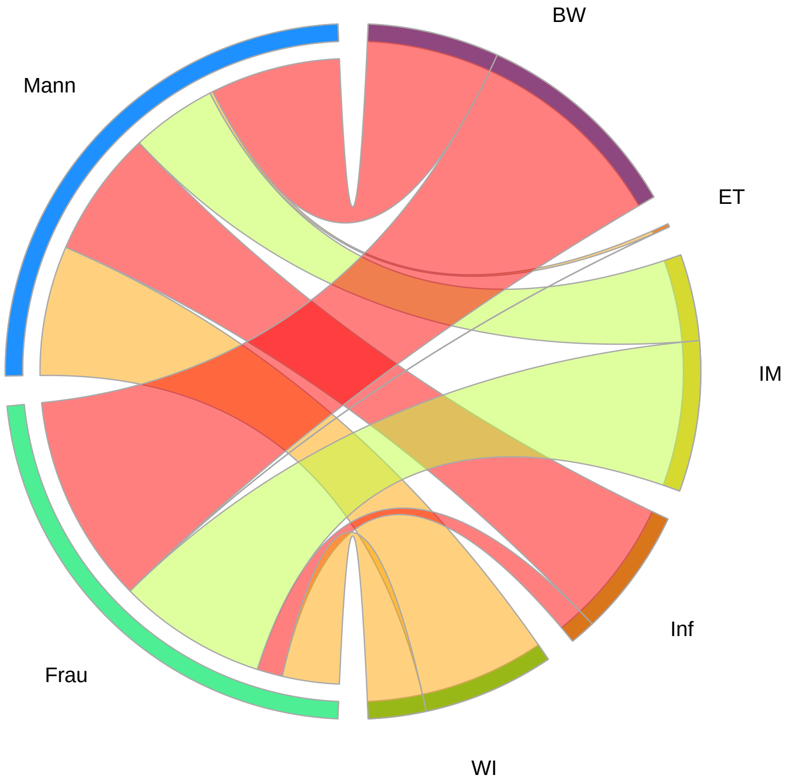
```
require(DescTools)
My.ohne.NA = na.exclude(MyData[,c("AlterM", "AlterV", "Alter")])
with(My.ohne.NA, {
  Alter.skaliert = (Alter-min(Alter))/(max(Alter)-min(Alter))
  PlotBubble(jitter(AlterM), jitter(AlterV), Alter.skaliert,
             col=SetAlpha("deeppink4",0.3),
             border=SetAlpha("darkblue",0.3),
             xlab="Alter der Mutter", ylab="Alter des Vaters",
             panel.first=grid(),
             main="")
  })
```



Größe der Blasen: Alter zwischen 0 (Jüngster) und 1 (Ältester)

Circular Plots: Assoziationen


```r
require(DescTools)
with(MyData, {
PlotCirc(table(Studiengang, Geschlecht),
acol=c("dodgerblue","seagreen2","limegreen","olivedrab2","goldenrod2","tomato2"),
rcol=SetAlpha(c("red","orange","olivedrab1"), 0.5)
)})
```



Gute Idee: Noch Experimentell