Statistik

für Betriebswirtschaft, Internationales Management, Wirtschaftsinformatik und Informatik

12.04.2017:
Aufgabe 23e, 25d-e, 26c,
30, 31, 32a-d

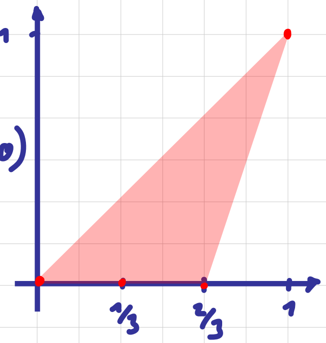
Sommersemester 2017

HSA Statistik SS 2017 Sessionlist		
Datum	Statistik für BW/IM/I/Winf	Nr.
15.03.2017	Einführung Statistik	1
22.03.2017	Differentialrechnung, 2-dim Diff.Rechnung	2
29.03.2017	univ. deskr. Stat., Quantile, Plots	3
05.04.2017	Streuung, Konzentrationsmaße	4
12.04.2017	Kontingenztabellen, Mosaikplots, Korrelation	5
19.04.2017	Preisindizes, lineare Regression	6
26.04.2017	Wahrscheinlichkeitsbegriff	7
03.05.2017	Bedingte Wahrscheinlichkeit, Bayes	8
10.05.2017	diskrete Zufallsvariablen	9
17.05.2017	Stetige ZV, Gleichverteilung	10
24.05.2017	Pyramid	
31.05.2017	Normalverteilung, Verteilungsparameter	11
07.06.2017	Schätzfunktionen und Punktschätzer	12
14.06.2017	Konfidenzintervalle	13
21.06.2017	Wiederholung, Besprechung Probeklausur	14
28.06.2017	Prüfungswoche	15

Stefan Etschberger

Beispiel (Gini-Koeffizienten)

Urliste: $x_i = 0, 0, 100$ € $\quad (\Sigma = 100)$



i	1	2	3
p_i	0	0	1

Gini-Koeffizient: $G = \frac{1}{3}\left(2\cdot(1\cdot 0 + 2\cdot 0 + 3\cdot 1) - (3+1)\right)$

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

$= \frac{1}{3}(6-4) = \frac{2}{3} < 1$ (trotz maximaler Konzentration)

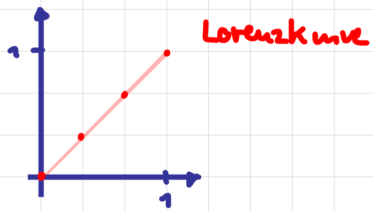
allgemein: Urliste: x_i: $\underbrace{0,0,0,0,\ldots,0}_{n-1}, 1$ ← n-te Wert

$\Rightarrow G = \frac{1}{n}\left(2\cdot(1\cdot 0 + 2\cdot 0 + \dots + (n-1)\cdot 0 + n\cdot 1) - (n+1)\right)$

$= \frac{1}{n}(2n - n - 1) = \frac{n-1}{n}$

Beispiel: x_i: 10 €, 10 €, 10 € $\quad (\Sigma = 30)$

i	1	2	3
p_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$



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

Gini-Koeffizient

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

- Numerisches Maß der Konzentration: Gini-Koeffizient G

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

- Aus den Daten:

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

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

➡ Normierter Gini-Koeffizient: $G_* = \frac{G}{G_{max}}$

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

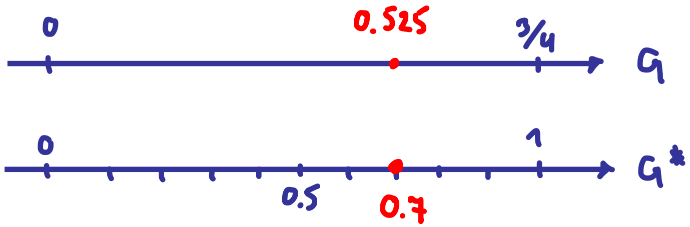
Gini-Koeffizient: Beispiel

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

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

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

Aufgabe 3 (zu Lorenzkurve und Gini-Koeffizient)

Stichprobe aus Umfragedaten: Ausgaben für Telekommunikation

5000 500 800 200 400

a) Zeichne Lorenzkurve (5 Minuten)
b) Berechne Gini-Koeffizient (3 Minuten)

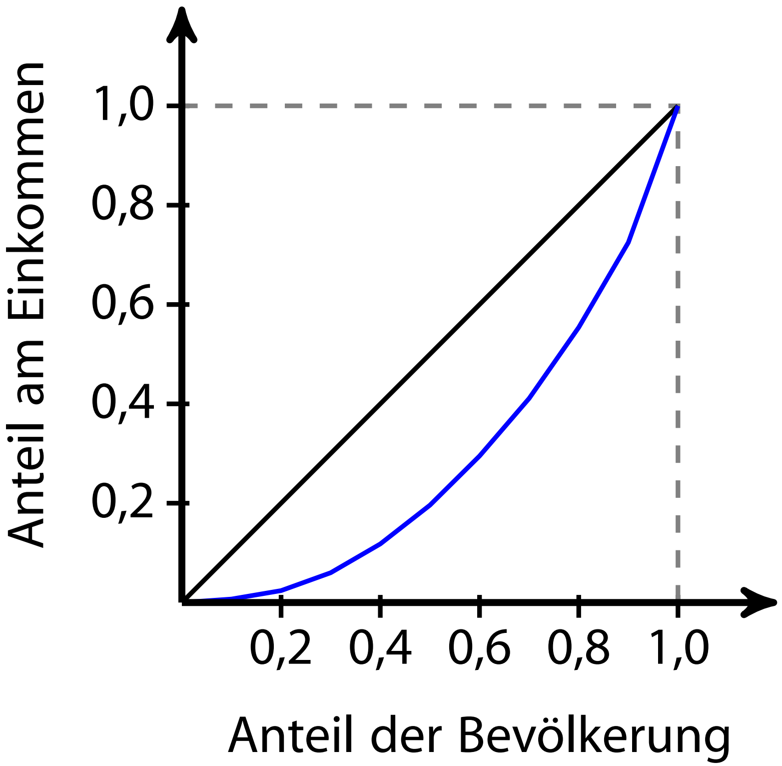
Umfrage: In der letzten Aufgabe hatte ich

A) Alles richtig
B) Alles bis auf die Zeichnung richtig
C) Einen Fehler in den Zahlen
D) Mehr als einen Fehler in den Zahlen
E) Ich wusste nicht, was zu tun ist

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

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

Einkommensverteilung (Nettoäquivalenzeinkommen)[1] in Deutschland

Soziodemographische Untergliederung	Erhebungsjahr							
	2008	2009	2010	2011	2012	2013	2014	2015
Median des Äquivalenzeinkommens in EUR je Jahr								
Insgesamt	18 309	18 586	18 797	19 043	19 595	19 582	19 733	20 668
Männer	18 777	18 927	19 186	19 389	20 074	20 081	20 228	21 194
Frauen	17 909	18 219	18 448	18 700	19 137	19 067	19 319	20 238
Durchschnittliches Äquivalenzeinkommen in EUR je Jahr								
Insgesamt	21 086	21 223	21 470	21 549	22 022	22 471	22 537	23 499
Männer	21 595	21 648	21 937	22 077	22 663	23 127	23 131	24 042
Frauen	20 595	20 814	21 018	21 037	21 401	21 840	21 964	22 973
Ungleichheit der Einkommensverteilung (S80 / S20 – Rate)[2]								
Insgesamt	4,8	4,5	4,5	4,5	4,3	4,6	5,1	4,8
Gini - Koeffizient								
Insgesamt	30,2	29,1	29,3	29,0	28,3	29,7	30,7	30,1

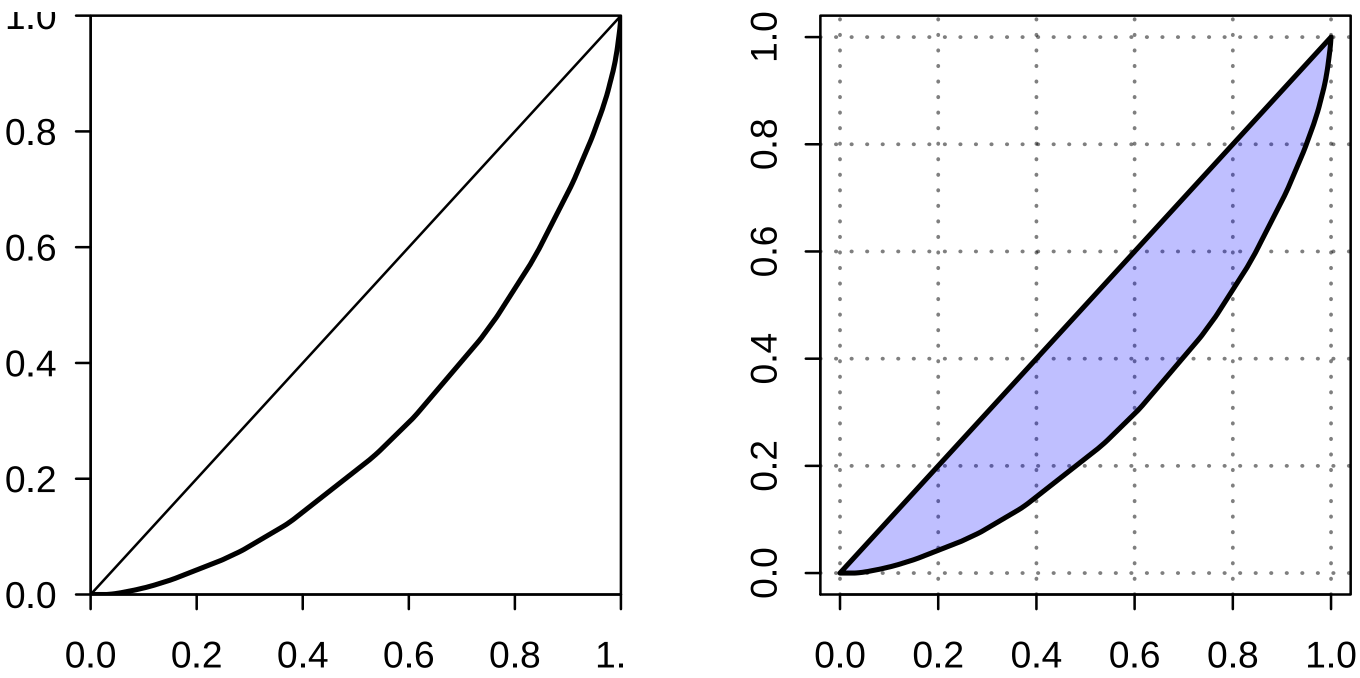
[1] Referenzjahr für die Ermittlung des Nettoäquivalenzeinkommens ist bei Leben in Europa jeweils das dem Erhebungsjahr vorangegangene Jahr.
[2] Der Quotient stellt das Verhältnis zwischen dem Gesamteinkommen des oberen Fünftels und dem des unteren Fünftels der Einkommensverteilung dar.

Quelle: Leben in Europa (EU-SILC).

Quelle: DeStatis,
https://goo.gl/tVlNLW
abgerufen: 5.4.2017

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] 0.4277039
```


1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

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:**

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

Beispiele: ① 0,0,100
p_i 0,0,1
$\Rightarrow H = 0^2 + 0^2 + 1^2 = 1$
② 10,10,10
p_i $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$
$\Rightarrow H = (\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2 = 3 \cdot \frac{1}{9} = \frac{1}{3}$

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

Variationskoeffizient $\frac{S}{\bar{x}}$

- **Exponentialindex:**

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

Beispiel ①: 0,0,100
p_i 0,0,1
$E = 0^0 \cdot 0^0 \cdot 1^1 = 1 \cdot 1 \cdot 1 = 1$
②: 10,10,10 $\quad p_i = \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$
$E = (\frac{1}{3})^{\frac{1}{3}} \cdot (\frac{1}{3})^{\frac{1}{3}} \cdot (\frac{1}{3})^{\frac{1}{3}} = \frac{1}{3}$

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

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

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

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

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

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

1. Einführung

2. Differenzieren 2

3. Deskriptive Statistik

Häufigkeiten

Lage und Streuung

Konzentration

Zwei Merkmale

Korrelation

Preisindizes

Lineare Regression

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Unterscheide:

- **Gemeinsame Häufigkeiten:**

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

- **Randhäufigkeiten:**

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

- **Bedingte (relative) Häufigkeiten:**

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

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
- Häufigkeiten
- Lage und Streuung
- Konzentration
- Zwei Merkmale
- Korrelation
- Preisindizes
- Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Beispiel: 400 unfallbeteiligte Autoinsassen:

Rand-häufigkeiten

	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 $(= n)$

Rand h.

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

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

Streuungsdiagramm

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Beispiel:

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

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

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

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Streuungsdiagramm

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

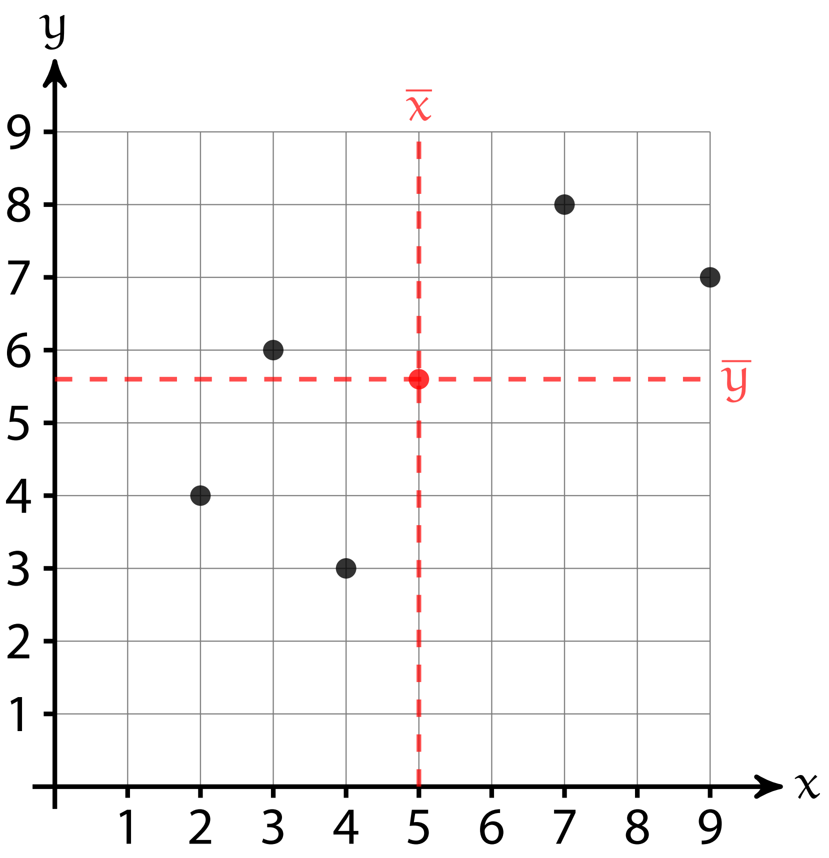
➟ Alle (x_i, y_i) sowie $(\bar{x}, \bar{y})$ in Koordinatensystem eintragen.

Beispiel:

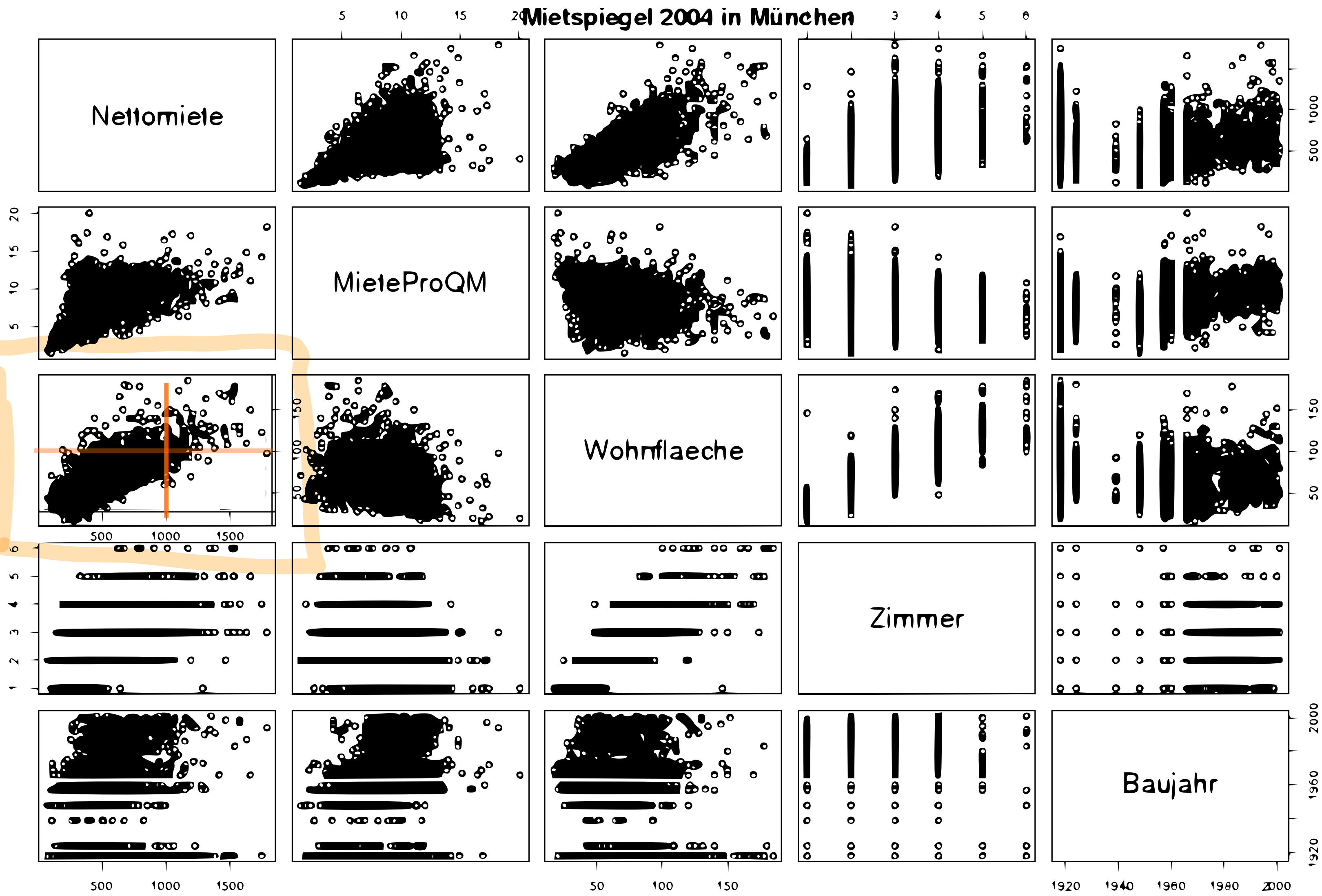
i	1	2	3	4	5	$\sum$
x_i	2	4	3	9	7	25
y_i	4	3	6	7	8	28

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

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



Beispiel Streuungsdiagramm



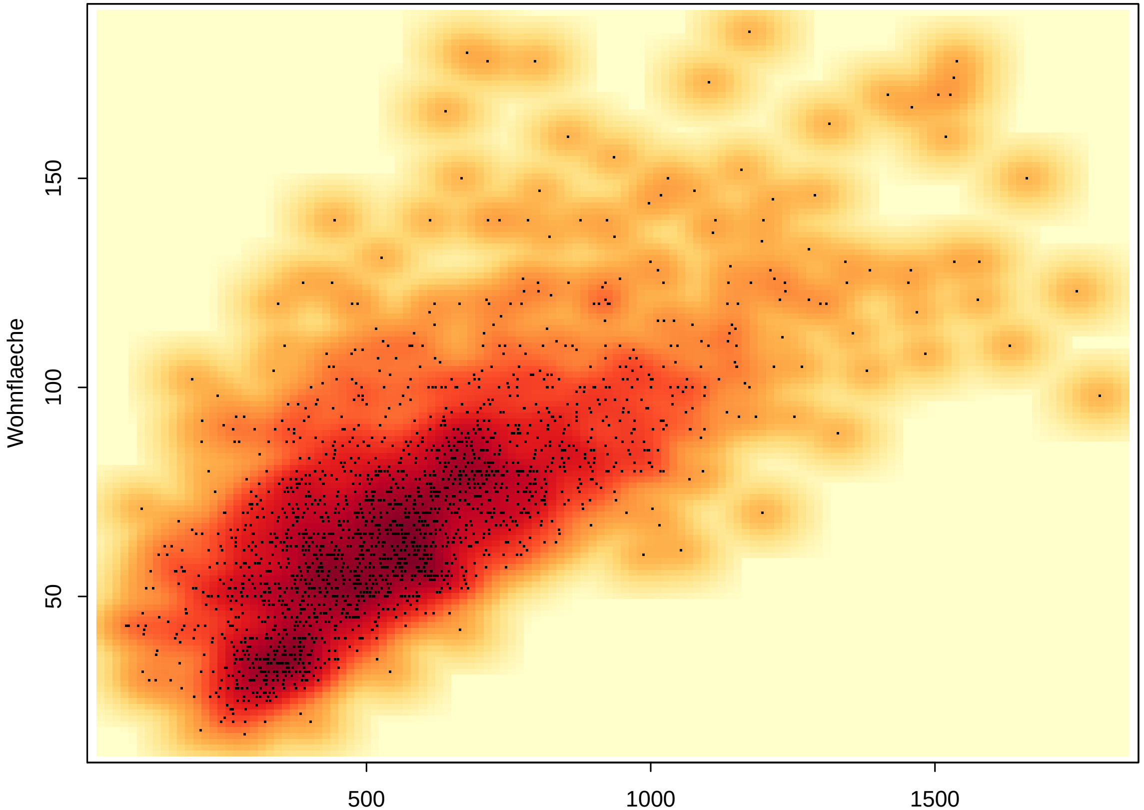
(Datenquelle: Fahrmeir u. a., (2009))

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Beispiel Streuungsdiagramm

```
if (!require("RColorBrewer")) {
install.packages("RColorBrewer")
library(RColorBrewer)
}
mieten <- read.table('https://goo.gl/yzJMJF', 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))
```

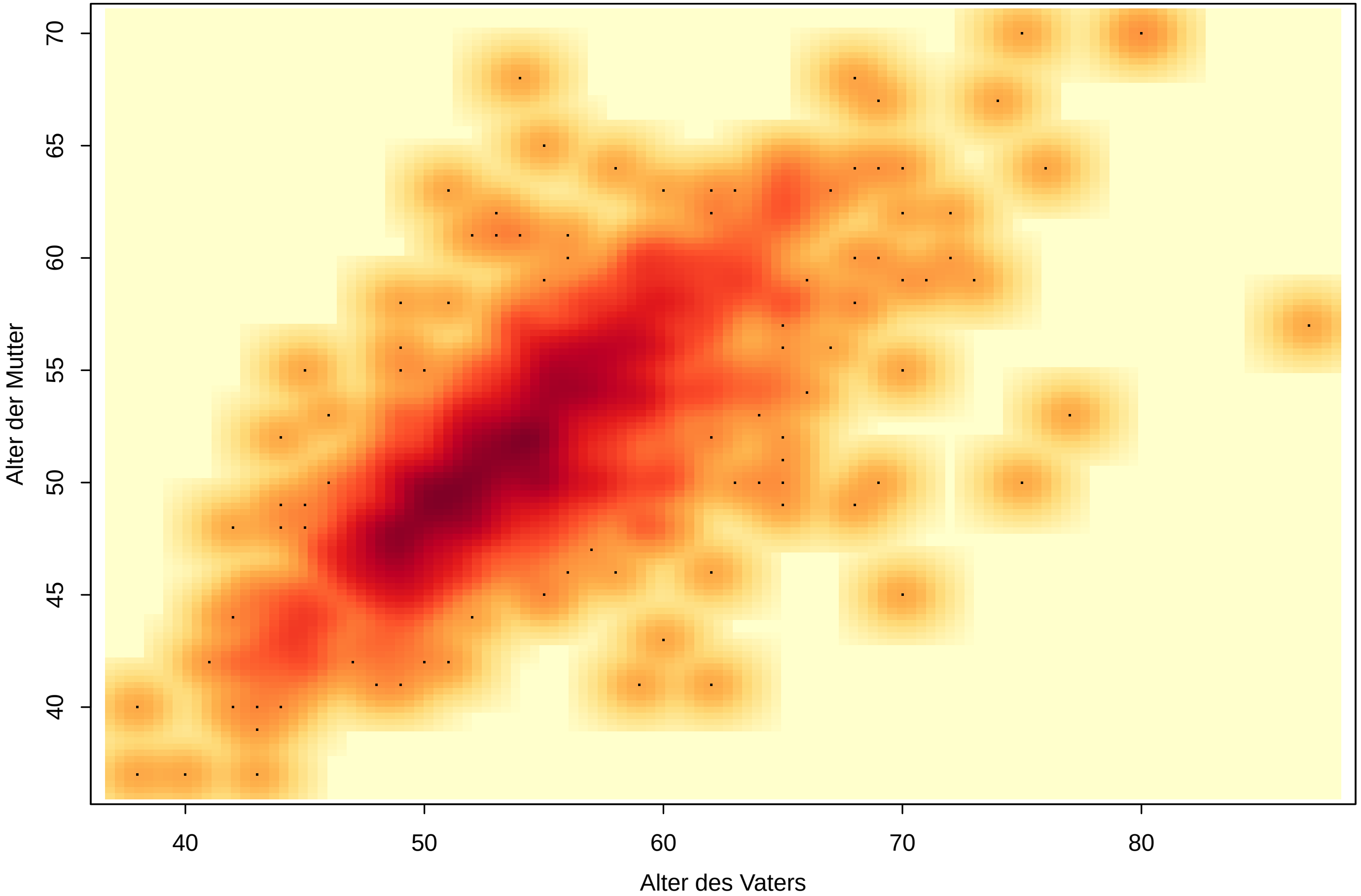


1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

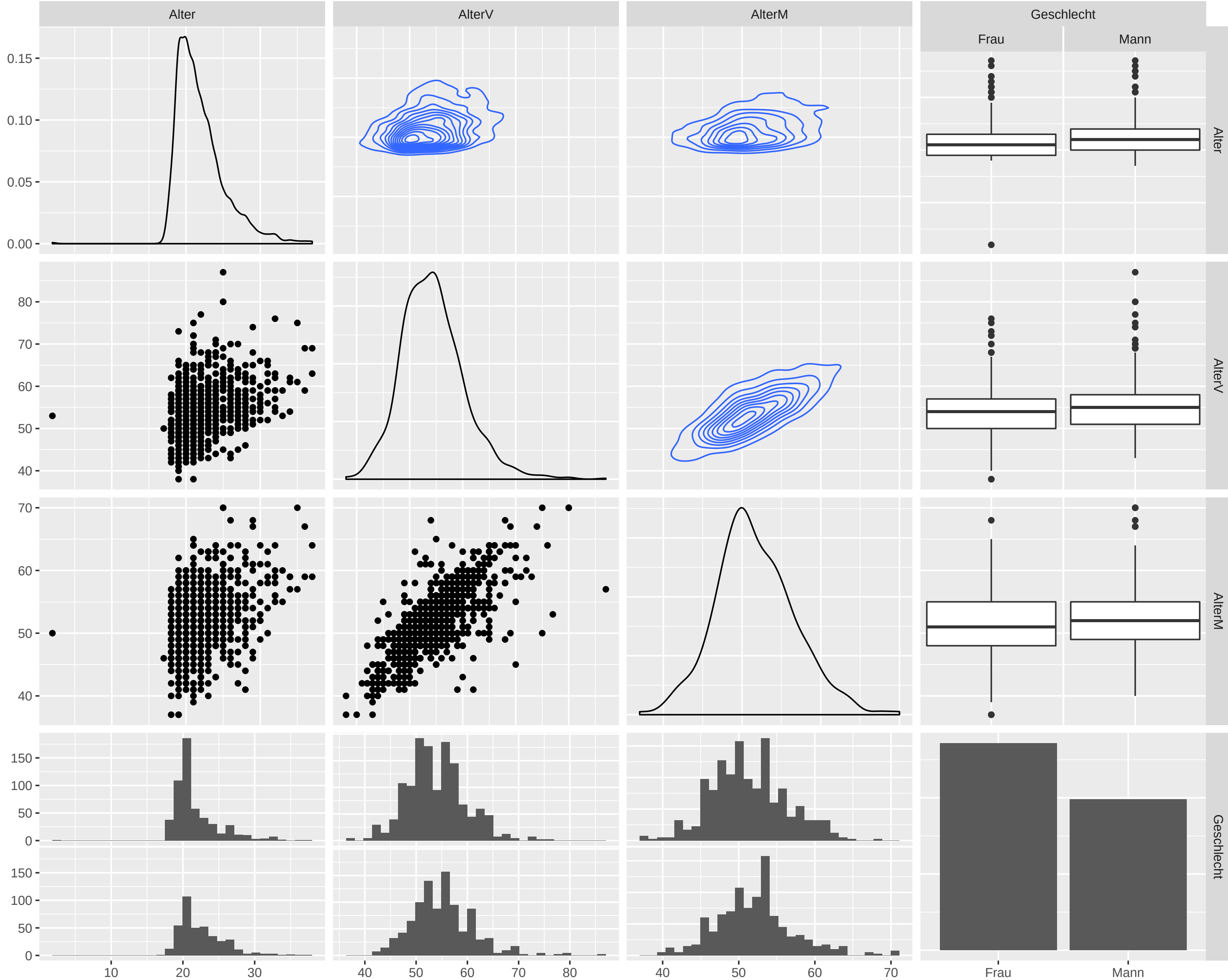
Beispiel Streuungsdiagramm

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

```
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)
```

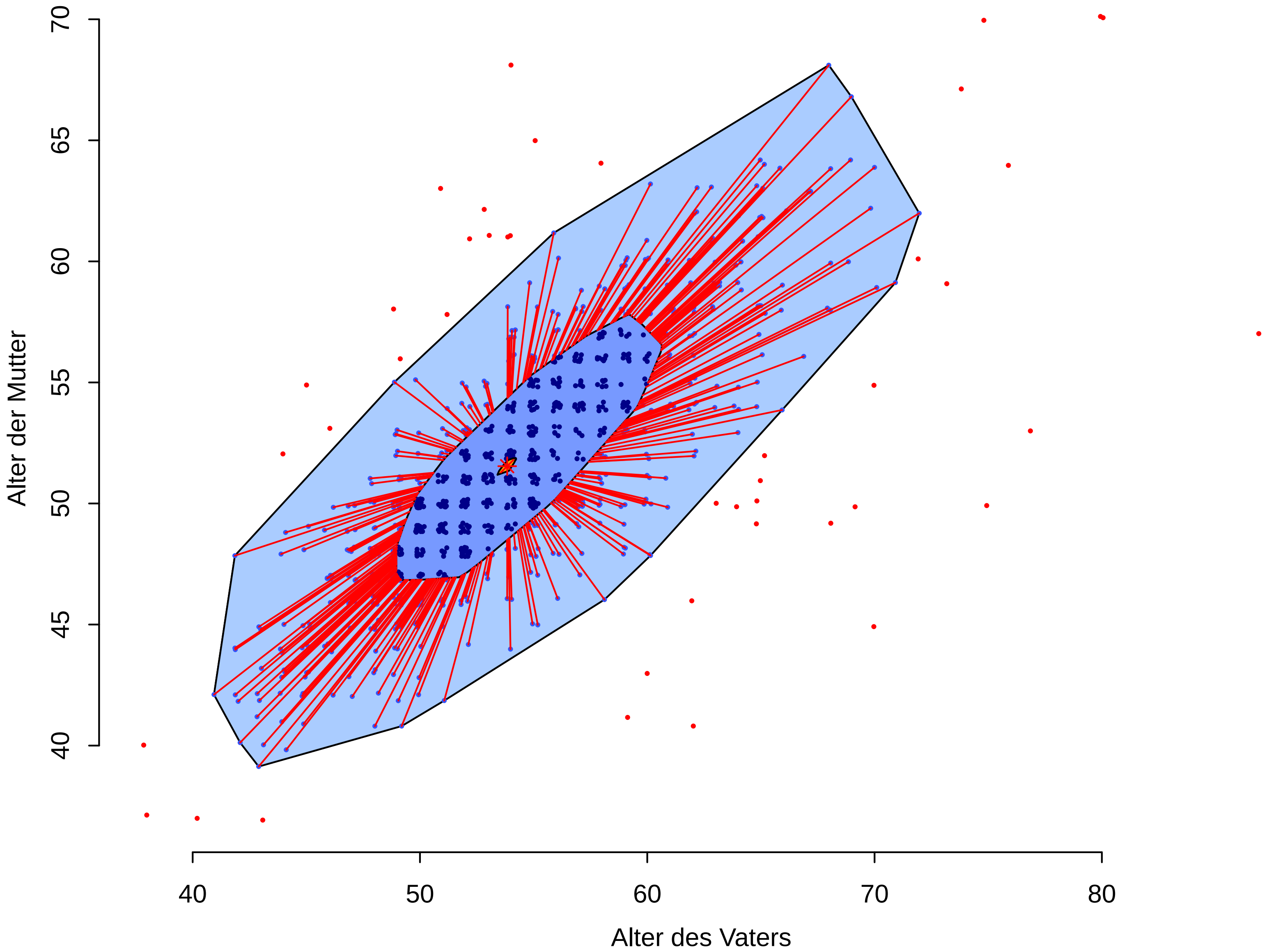


1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Bagplot: Boxplot in 2 Dimensionen

```
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!!"
```

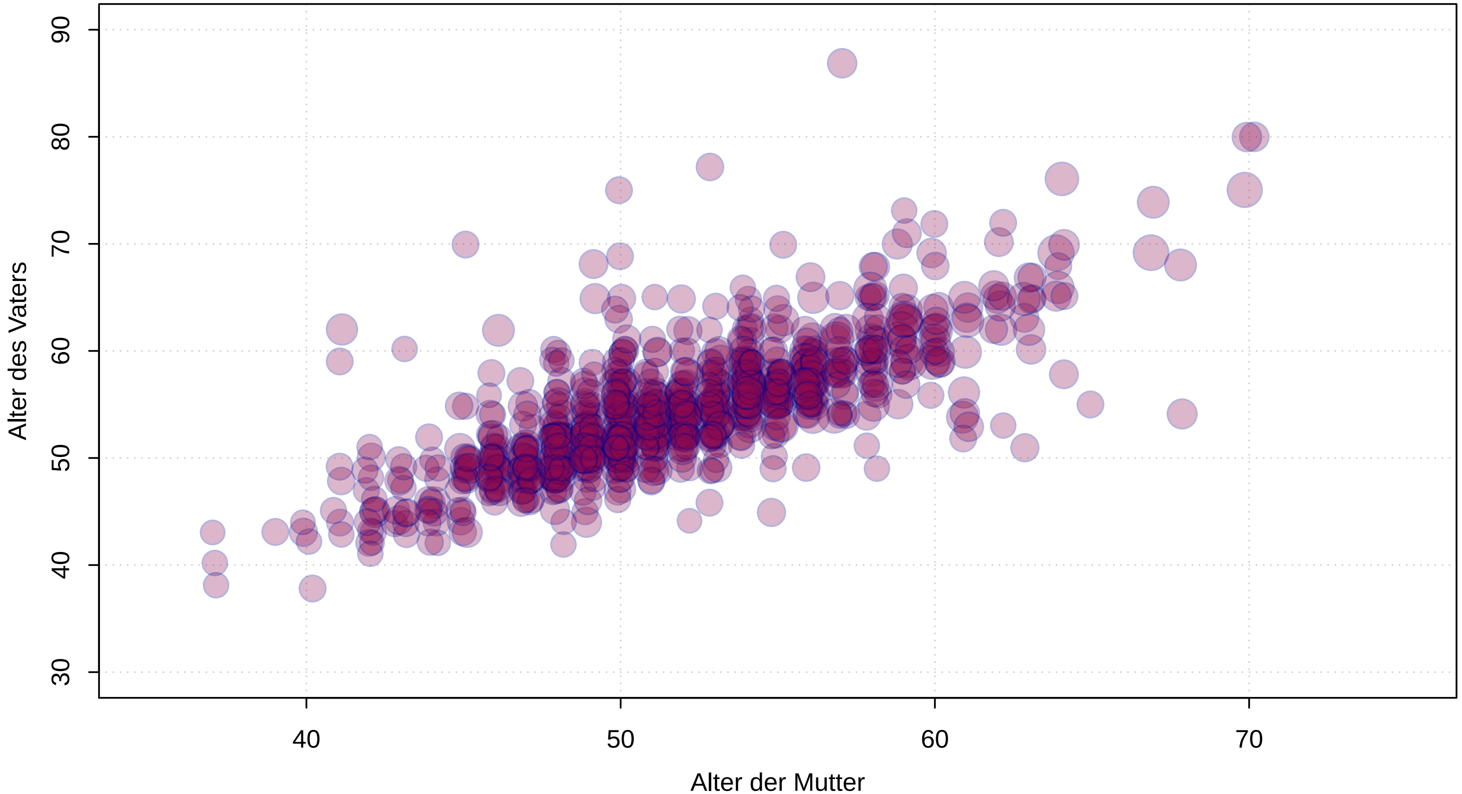


1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Bubbleplot: 3 metrische Variablen

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

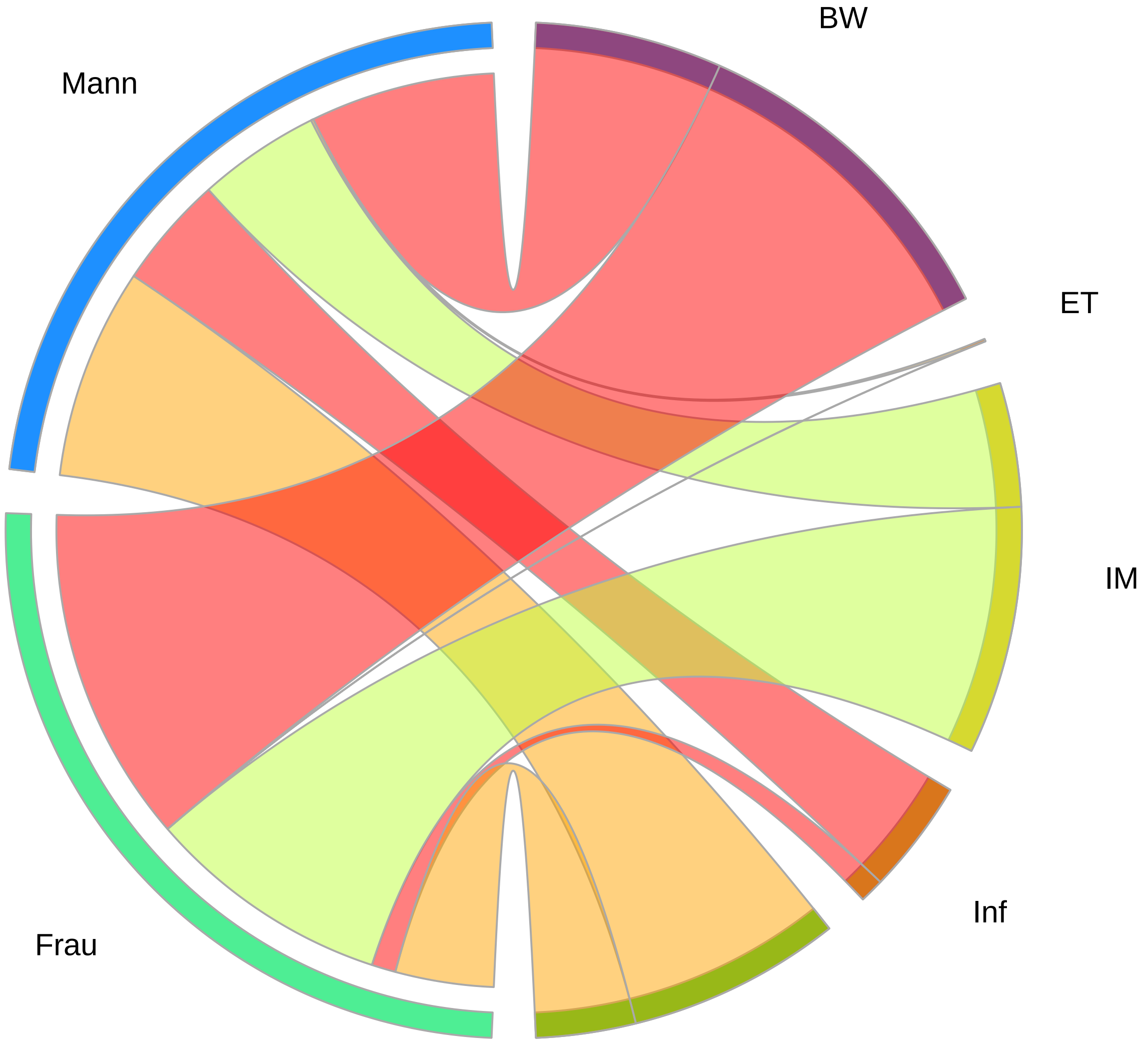
```
require(DescTools)
My.ohne.NA = na.exclude(MyData[,c("AlterM", "AlterV", "Alter")])
with(My.ohne.NA, {
  Alter.skaliert = (Alter-min(Alter))/(max(Alter)-min(Alter))
  PlotBubble(jitter(AlterM), jitter(AlterV), Alter.skaliert,
             col=SetAlpha("deeppink4",0.3),
             border=SetAlpha("darkblue",0.3),
             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


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



1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Korrelationsrechnung

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

- Frage: Wie stark ist der Zusammenhang zwischen X und Y?
- Dazu: Korrelationskoeffizienten
- Verschiedene Varianten: Wahl abhängig vom Skalenniveau von X und Y:

	Skalierung von Y		
Skalierung von X	kardinal	ordinal	nominal
kardinal	Bravais-Pearson-Korrelations-koeffizient		
ordinal		Rangkorrelations-koeffizient von Spearman	
nominal			Kontingenz-koeffizient

Korrelationsrechnung

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

- Frage: Wie stark ist der Zusammenhang zwischen X und Y?
- Dazu: Korrelationskoeffizienten
- Verschiedene Varianten: Wahl abhängig vom Skalenniveau von X und Y:

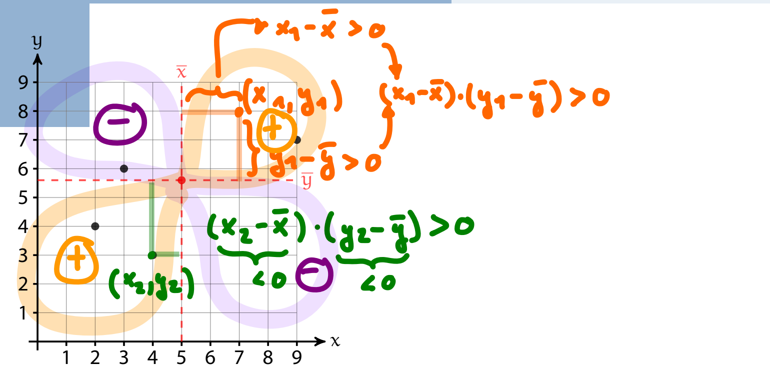
	Skalierung von Y		
Skalierung von X	kardinal	ordinal	nominal
kardinal	Bravais-Pearson-Korrelationskoeffizient		
ordinal		Rangkorrelationskoeffizient von Spearman	
nominal			Kontingenzkoeffizient

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Korrelationsrechnung

- Frage: Wie stark ist der Zusammenhang zwischen X und Y?
- Dazu: Korrelationskoeffizienten
- Verschiedene Varianten: Wahl abhängig vom Skalenniveau von X und Y:

	Skalierung von Y		
Skalierung von X	kardinal	ordinal	nominal
kardinal	Bravais-Pearson-Korrelationskoeffizient		
ordinal		Rangkorrelationskoeffizient von Spearman	
nominal			Kontingenzkoeffizient



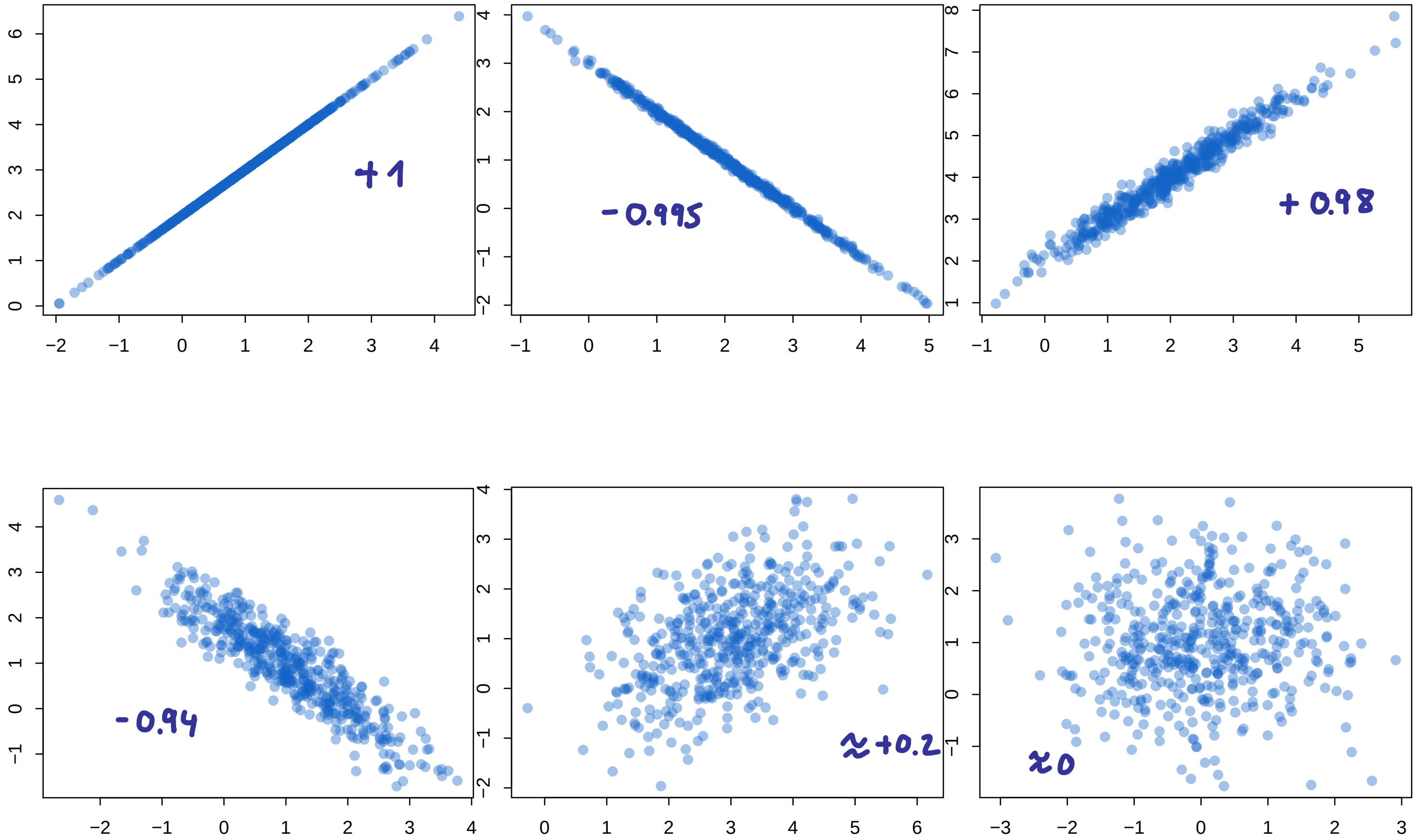
Bravais-Pearson-Korrelationskoeffizient

Voraussetzung: X, Y kardinalskaliert

Kovarianz von X,Y

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \, \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}} \in [-1;+1]$$

σ_x σ_y

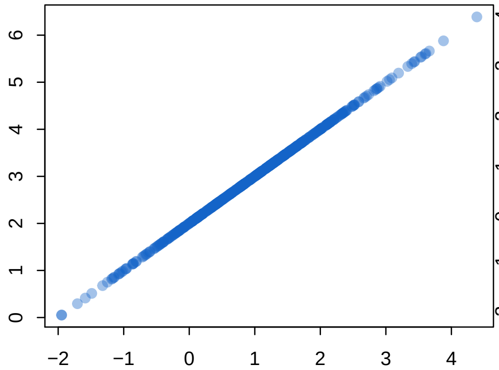


1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

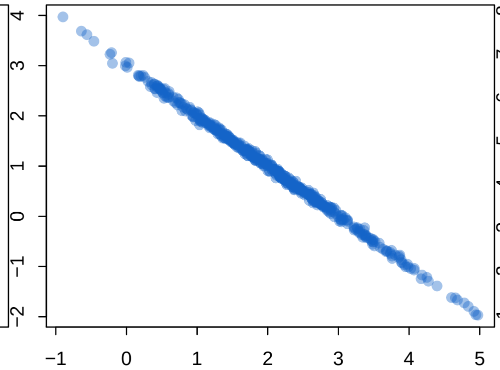
Bravais-Pearson-Korrelationskoeffizient

Voraussetzung: X, Y kardinalskaliert

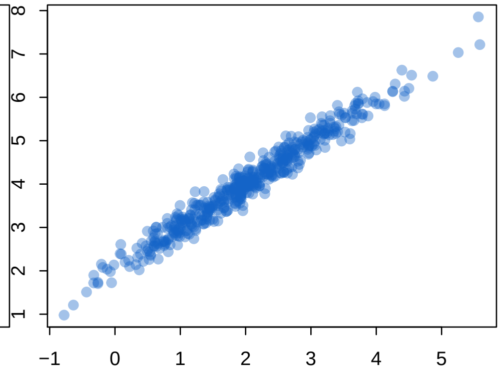
$$r = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2}\sqrt{\sum\limits_{i=1}^{n} y_i^2 - n\bar{y}^2}} \in [-1;+1]$$



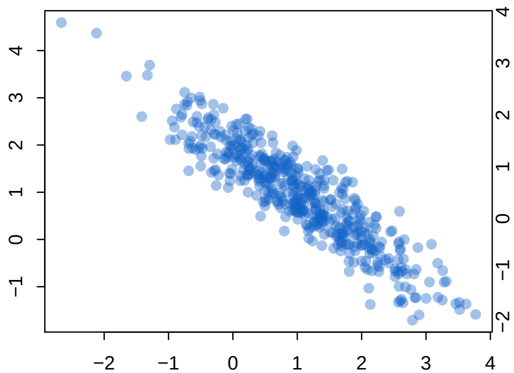
$r = 1$



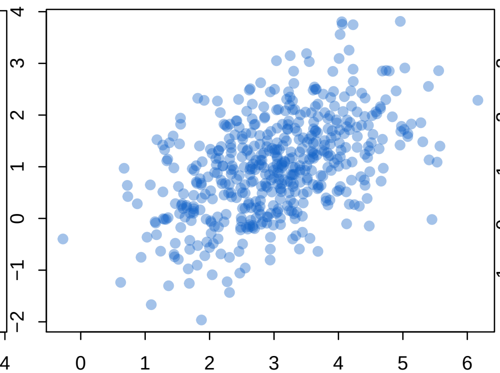
$r = -0{,}999$



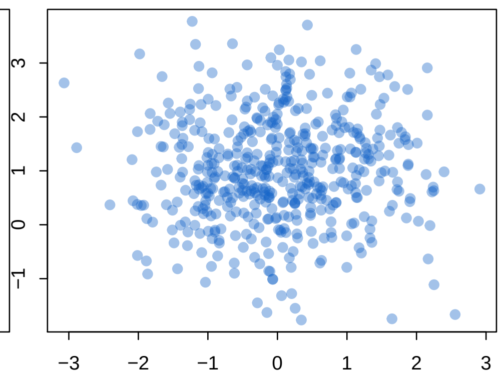
$r = 0{,}981$



$r = -0{,}91$



$r = 0{,}487$



$r = 0{,}031$

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Bravais-Pearson-Korrelationskoeffizient

TR: ① Mode → STAT → A+Bx
↳ 2-Var → LIN

② Eingabe:

x	y	Freq
2	4	1
4	3	1
3	6	1
9	7	1
7	8	1

③: $\boxed{\text{AC}}$ → Shift → STAT → REG → r $\boxed{=}$

Im Beispiel:

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	4	4	16	8
2	4	3	16	9	12
3	3	6	9	36	18
4	9	7	81	49	63
5	7	8	49	64	56
$\sum$	25	28	159	174	157

$$\Rightarrow \quad \bar{x} = 25/5 = 5$$

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

$$r = \frac{157 - 5 \cdot 5 \cdot 5{,}6}{\sqrt{159 - 5 \cdot 5^2}\sqrt{174 - 5 \cdot 5{,}6^2}} = 0{,}703$$

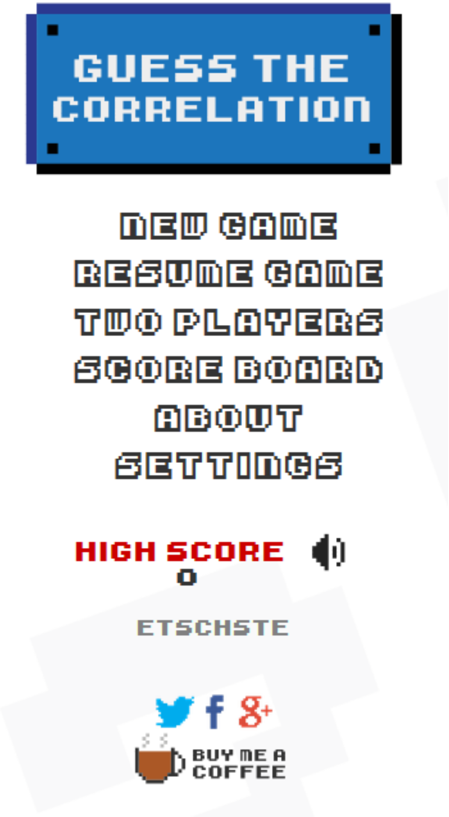
(deutliche positive Korrelation)

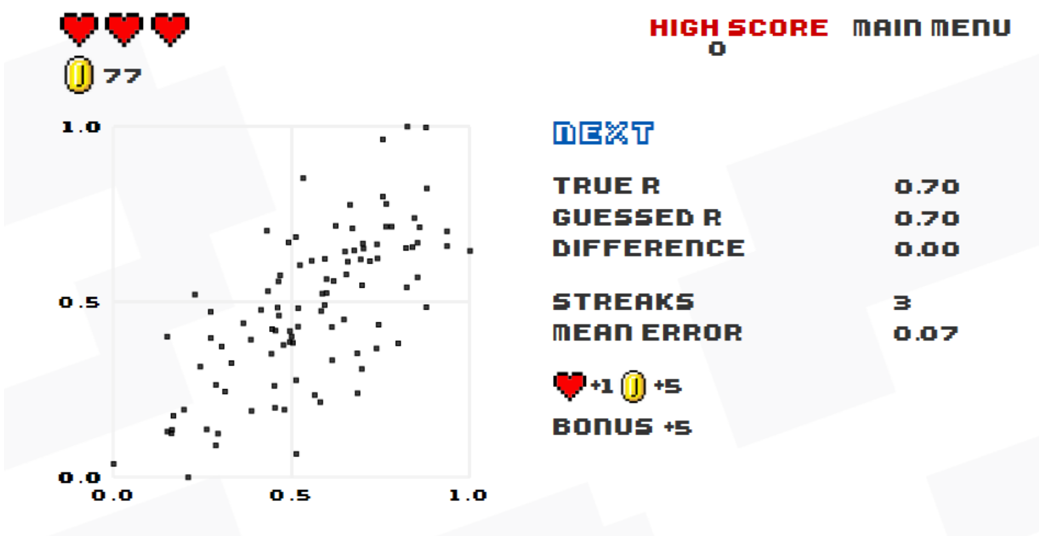
1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Guess The Correlation

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

guessthecorrelation.com





Go for the Highscore!

Rangkorrelationskoeffizient von Spearman

- Voraussetzungen: X, Y (mindestens) ordinalskaliert, Ränge eindeutig (keine Doppelbelegung von Rängen)
- Vorgehensweise:
  1. Rangnummern R_i (X) bzw. R_i' (Y) mit $R_i^{(\prime)} = 1$ bei größtem Wert usw.
  2. Berechne

$$r_{SP} = 1 - \frac{6 \sum\limits_{i=1}^{n} (R_i - R_i')^2}{(n-1)\, n\, (n+1)} \in [-1; +1]$$

- Hinweise:
  - $r_{SP} = +1$ wird erreicht bei $R_i = R_i'$ $\quad \forall i = 1, \ldots, n$
  - $r_{SP} = -1$ wird erreicht bei $R_i = n + 1 - R_i'$ $\quad \forall i = 1, \ldots, n$
- Falls Ränge nicht eindeutig: Bindungen, dann Berechnung von r_{SP} über Ränge und Formel des Korr.-Koeff. von Bravais-Pearson

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Beispiel (Rangkorrelationskoeffizient)

Person	Rang	Preis	Qualität	„Vorranking"	
1	6	199,-	sehr gut	1	(Mittel) 1.5
2	7	39,50	sehr gut	2	1.5
3	5	350,-	geht so	6	6
4	1	1200,-	schlecht	7	7
5	2	899,-	gut	3	4
6	4	400,-	gut	4	4
7	3	500,-	~~geht so~~ gut	5	4

Rangkorr. koeffizient r_{sp}

$\hat{=}$ Bravais-Pearson der Rangnummern

$r_{sp} \approx -0.730$

deutlich negative Korr., d.h.
steigender Preis bedeutet tendentielle
sinkendes Qualitätsempfinden

Rangkorrelationskoeffizient von Spearman

1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression
4. W-Theorie
5. Induktive Statistik
Quellen
Tabellen

Im Beispiel:

x_i	R_i	y_i	R'_i
2	5	4	4
4	3	3	5
3	4	6	3
9	1	7	2
7	2	8	1

$$r_{SP} = 1 - \frac{6 \cdot [(5-4)^2 + (3-5)^2 + (4-3)^2 + (1-2)^2 + (2-1)^2]}{(5-1) \cdot 5 \cdot (5+1)} = 0{,}6$$