

# Statistik

für Betriebswirtschaft, Internationales Management,  
Wirtschaftsinformatik und Informatik

Sommersemester 2017

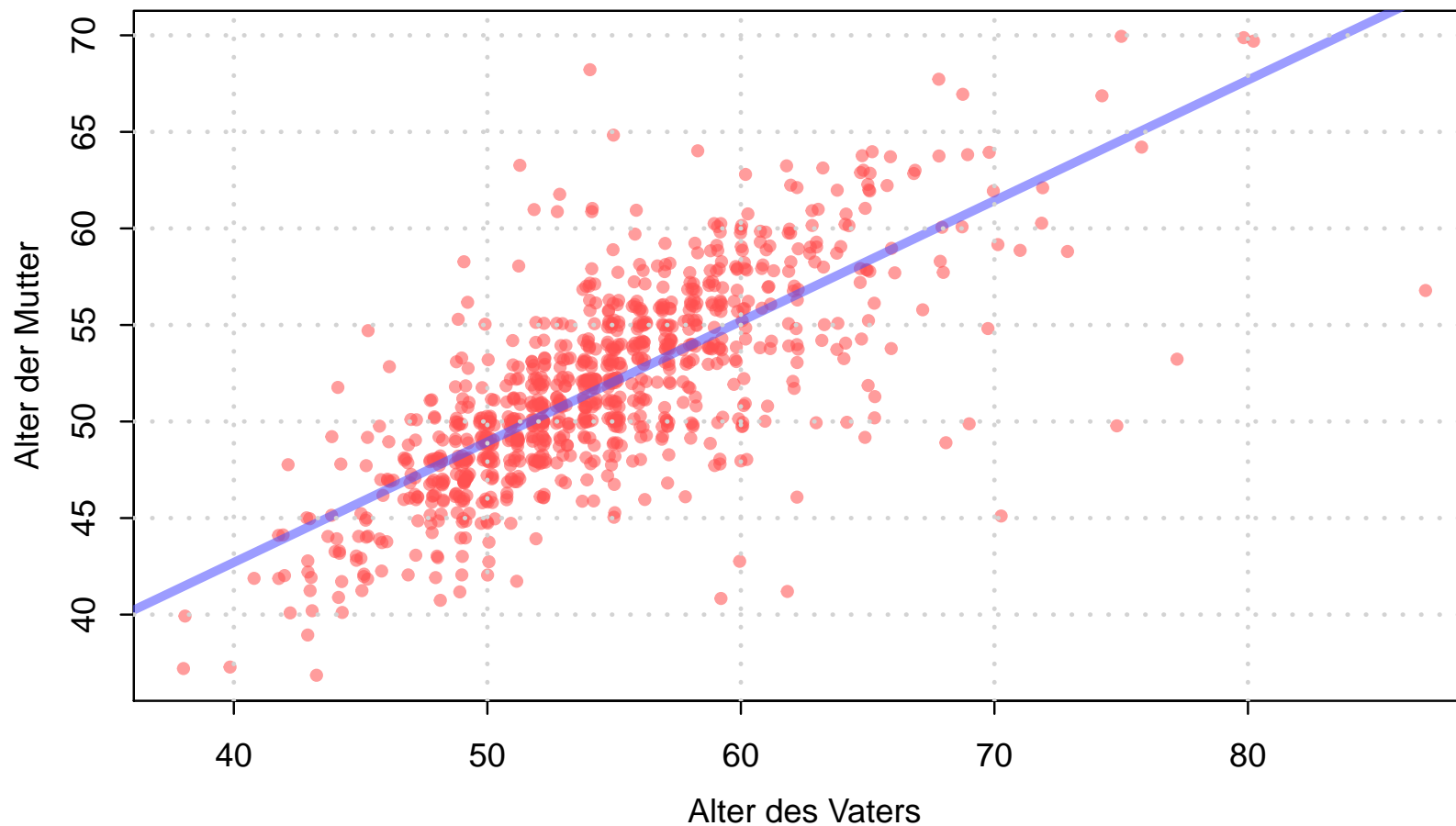
HSA Statistik SS 2017 Sessionlist		
Datum	Statistik für BW/IM/I/Winf	Nr.
15.03.2017	Einführung Statistik	1
22.03.2017	Differentialrechnung, 2-dim Diff.Rechnung	2
29.03.2017	univ. deskr. Stat., Quantile, Plots	3
05.04.2017	Streuung, Konzentrationsmaße	4
12.04.2017	Kontingenztabellen, Mosaikplots, Korrelation	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



```
meineRegression = lm(AlterM ~ AlterV)  
meineRegression
```

```
plot(AlterV, AlterM,  
     xlab="Alter des Vaters",  
     ylab="Alter der Mutter")  
abline(meineRegression)
```

```
##  
## Call:  
## lm(formula = AlterM ~ AlterV)  
##  
## Coefficients:  
## (Intercept)      AlterV  
##      17.7196      0.6245
```



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

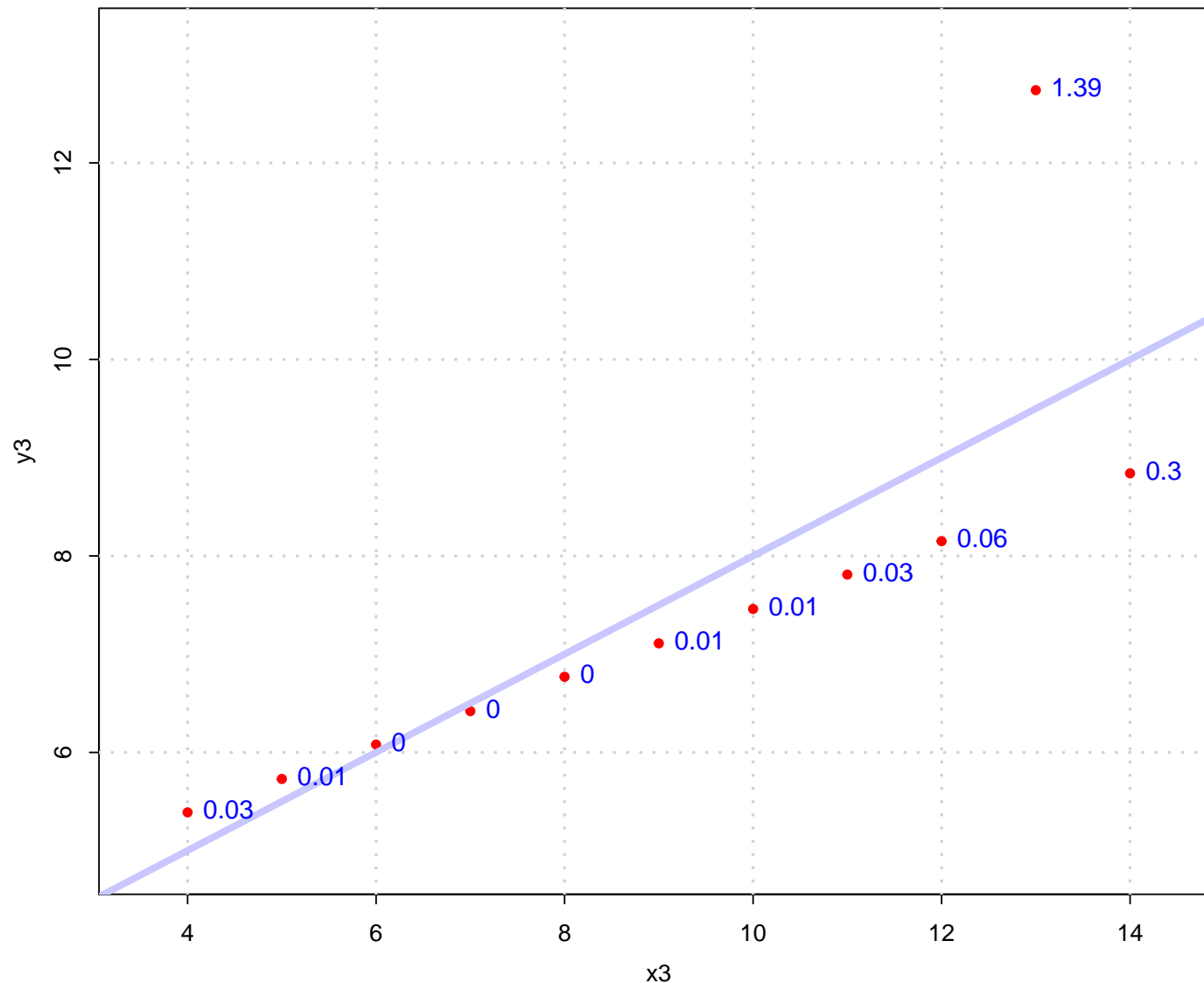
- ▶ Oft Kritisch: Einzelne Punkte, die Modell stark beeinflussen
- ▶ Idee: Was würde sich ändern, wenn solche Punkte weggelassen würden?
- ▶ **Cook-Distanz**: Misst den Effekt eines gelöschten Objekts
- ▶ Formel für ein lineares Modell mit einem unabh. Merkmal:

$$D_i = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{y}_{j(\text{ohne } i)})^2}{\text{MSE}}$$

- ▶ Dabei bedeutet:
  - $\hat{y}_j$ : Prognosewert des kompletten Modells für das j-te Objekt
  - $\hat{y}_{j(\text{ohne } i)}$ : Prognosewert des Modells ohne Objekt i für das j-te Objekt
  - $\text{MSE} = \frac{1}{n} \cdot \sum (\hat{y}_i - y_i)^2$ : Normierender Term (Schätzwert für Fehlerstreuung)



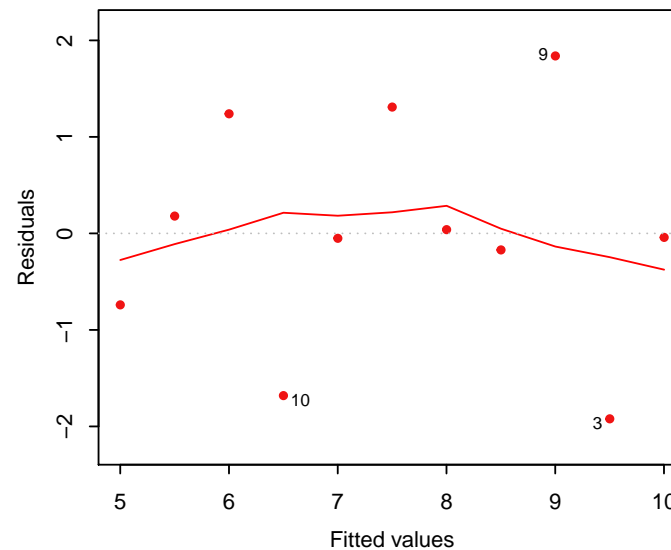
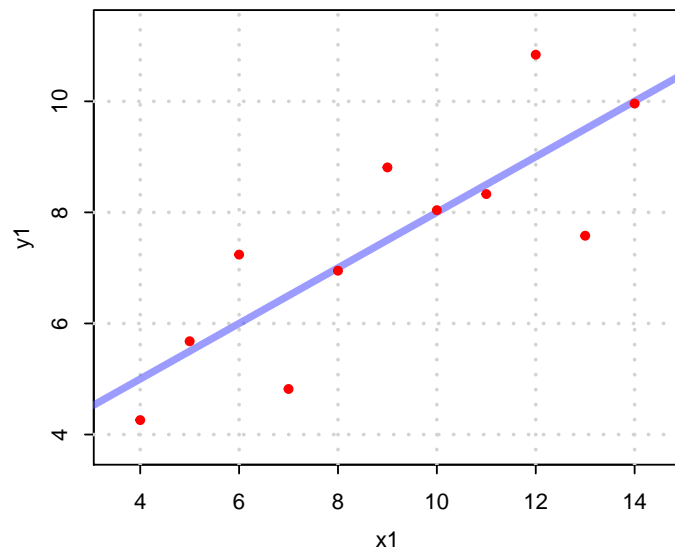
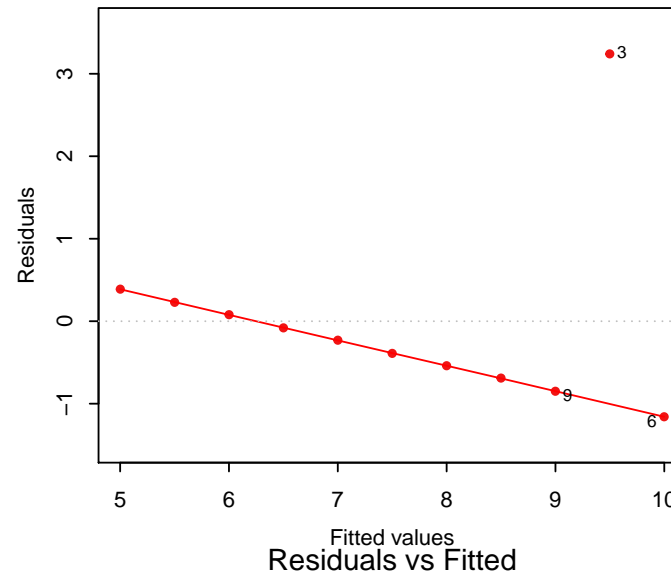
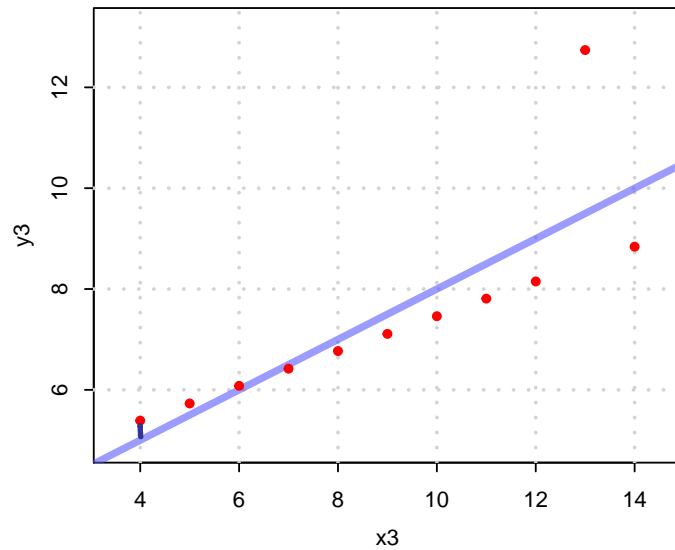
- ▶ Anscombe-Daten: Regressionsmodell Nr. 3
- ▶ Darstellung der Cook-Distanz neben Punkten
- ▶ Faustformel: Werte über 1 sollten genau untersucht werden



- 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



- ▶ Oft aufschlussreich: Verteilung der **Residuen**  $e_i$
- ▶ Verbreitet: Graphische Darstellungen der Residuen
- ▶ Z.B.:  $e_i$  über  $\hat{y}_i$

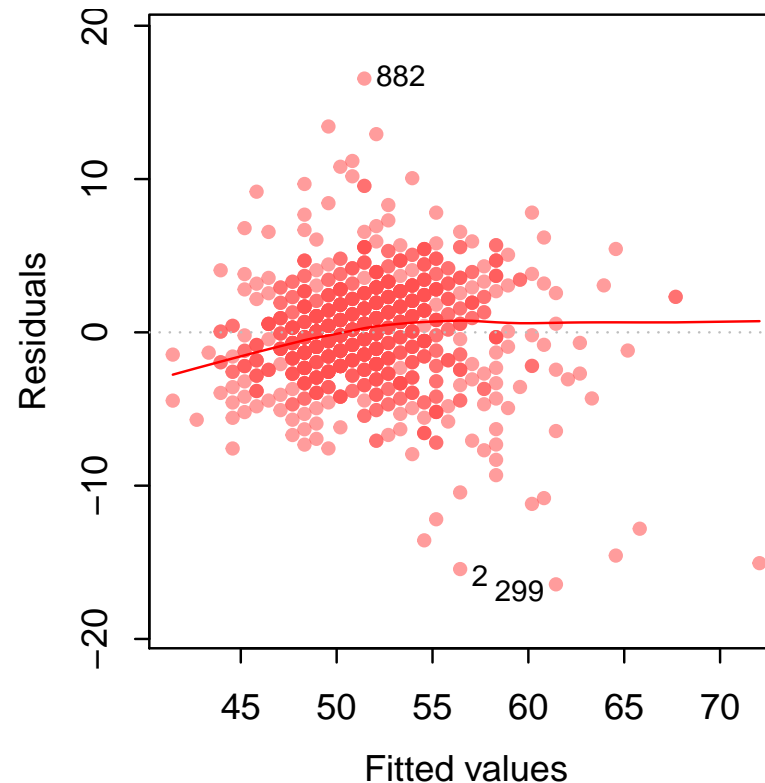
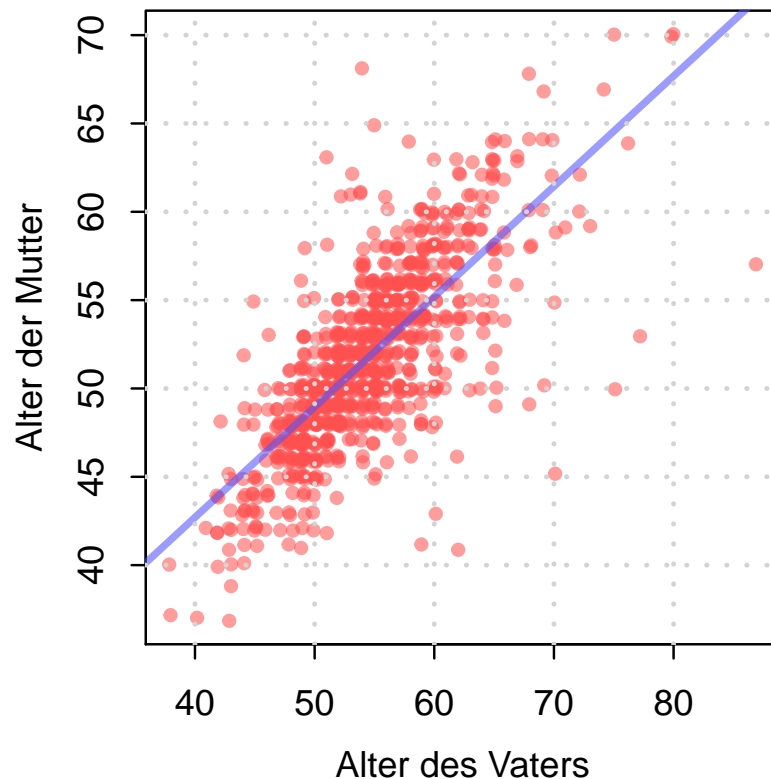
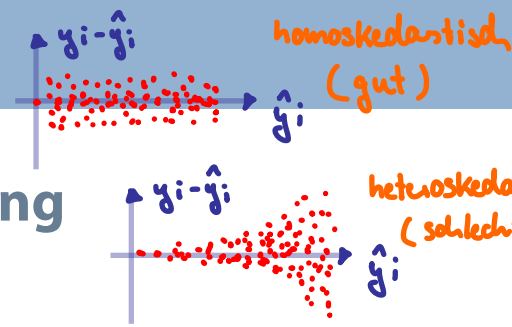


- 1. Einführung
- 2. Differenzieren 2
- 3. Deskriptive Statistik
  - Häufigkeiten
  - Lage und Streuung
  - Konzentration
  - Zwei Merkmale
  - Korrelation
  - Preisindizes
  - Lineare Regression
- 4. W-Theorie
- 5. Induktive Statistik
- Quellen
- Tabellen

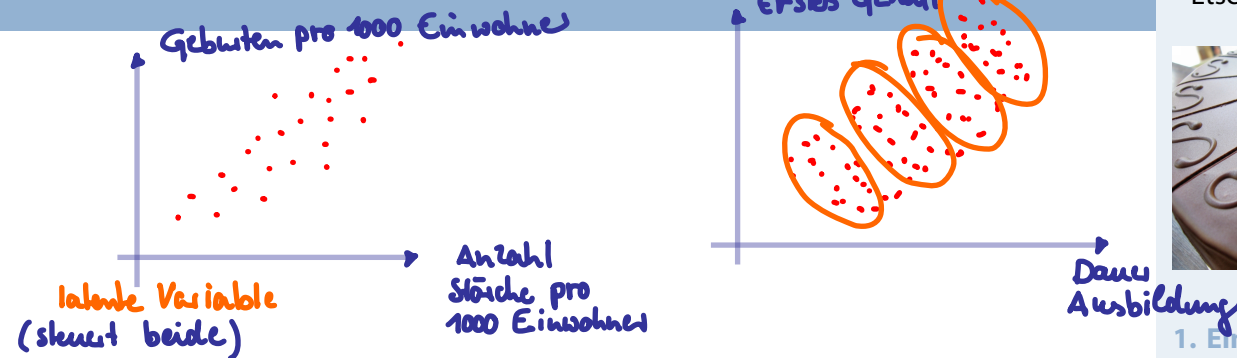


## Wichtige Eigenschaften der Residuenverteilung

- ▶ Möglichst **keine systematischen Muster**
- ▶ Keine Änderung der Varianz in Abhängigkeit von  $\hat{y}_i$  (**Homoskedastizität**)
- ▶ Nötig für inferentielle Analysen: Näherungsweise **Normalverteilung** der Residuen (q-q-plots)



- 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



## Exkurs: Kausalität vs. Korrelation

- ▶ Meist wichtig für sinnvolle Regressionsanalysen:
- ▶ **Kausale Verbindung** zwischen unabhängigem und abhängigem Merkmal
- ▶ Sonst bei Änderung der unabhängigen Variablen keine sinnvollen Prognosen möglich
- ▶ Oft: **Latente Variablen** im Hintergrund



- 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 Statistik: Einführung
- 2 Differenzieren 2
- 3 Deskriptive Statistik
- 4 Wahrscheinlichkeitstheorie
- 5 Induktive Statistik



- 4 Wahrscheinlichkeitstheorie
  - Kombinatorik
  - Zufall und Wahrscheinlichkeit
  - Zufallsvariablen und Verteilungen
  - Verteilungsparameter





A A A | B B B B | C C C  
 A A A A A A A A | |  
 | B | C C C C C C C C

·	·	·		·	·	·	·		·	·	·
·	·	·	·	·	·	·	·	·	·		
	·		·	·	·	·	·	·	·	·	·

$$\frac{12 \cdot 11}{1 \cdot 2} = \binom{12}{2} = \binom{12}{10}$$

$\nearrow n+k-1$   
 $\searrow k$

allgemein :  $\binom{n+k-1}{R}$  mit Wiederholung ohne Reihenfolge



2-mal Würfeln, das heißt Auswahl von  $k = 2$  aus  $n = 6$  Zahlen.



(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- ▶ mit WH, mit RF: alle Möglichkeiten,  $6^2 = 36$
- ▶ ohne WH, ohne RF: Hälfte des letzten Ergebnisses:  $\frac{30}{2} = 15 = \frac{6!}{4!2!} = \binom{6}{2}$
- ▶ ohne WH, mit RF: Diagonale entfällt,  $36 - 6 = 30 = 6 \cdot 5 = \frac{6!}{(6-2)!}$
- ▶ mit WH, ohne RF: Letztes Ergebnis plus Diagonale,  $15 + 6 = 21 = \binom{7}{2}$

## Auswahl von $k$ aus $n$ Dingen

	mit Wiederholung	ohne Wiederholung
mit Reihenfolge	$n^k$	$\frac{n!}{(n-k)!}$
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k}$



1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik

### 4. W-Theorie

- Kombinatorik
- Zufall und Wahrscheinlichkeit
- Zufallsvariablen und Verteilungen
- Verteilungsparameter

### 5. Induktive Statistik

#### Quellen

#### Tabellen