

4

## Beschreibende Statistik

- Statistische Maßzahlen für quantitative Merkmale
- Box-Plots
- Probability Plots
- Häufigkeitsdiagramme
- Häufigkeitstabellen

# Beschreibende Statistik

## Übersicht

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

- 4.1 Statistische Maßzahlen für quantitative Merkmale
- 4.2 Box-Plots
- 4.3 Probability Plots
- 4.4 Häufigkeitsdiagramme
- 4.5 Häufigkeitstabellen
- 4.6 Zusammenhangsmaße
- 4.7 Das Regressionsproblem

# Statistische Maßzahlen für quantitative Merkmale

Werkzeuge der empirischen Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

## 4.1.1 Lagemaße

Mittelwert, Quantile, Median, Quartile, Modalwert

## 4.1.2 Eigenschaften von Schätzungen

## 4.1.3 Streuungsmaße

Varianz, Standardabweichung, Spannweite, Quartilsabstand, MAD, Variationskoeffizient

## 4.1.4 Formmaße

Schiefe, Exzess, Wölbung, Kurtosis

Beschreibende

# Lagemaße (Lokationsparameter)

Das arithmetische Mittel

Die angegebenen Maßzahlen sind empirisch, d.h. sie sind Schätzungen für die wahre (i.A. unbekannte) Lage.

## Mittelwert (MEAN)

$$\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\bar{X}_n \xrightarrow{n \rightarrow \infty} \mathbf{E}X$$

Gesetz der Großen Zahlen.

Voraussetzungen:

a)  $X_i$  i.i.d.,  $\mathbf{E}X_i < \infty$  (Chintchin) oder

b)  $X_i$  beliebig,  $\mathbf{E}X_i^2 < \infty$  (Tschebycheff)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Lagemaße (2)

## Quantile

Die Beobachtungen  $x_1, \dots, x_n$  werden der Größe nach geordnet:  $x_{(1)} \leq \dots \leq x_{(n)}$ .

Sei  $0 \leq \alpha \leq 1$ ,  $\alpha \cdot n = \lfloor \alpha \cdot n \rfloor + r =: j + r$ .

### Quantile (Perzentile)

$$x_\alpha = \begin{cases} x_{(j+1)} & \text{für } r > 0 \\ 1/2(x_{(j)} + x_{(j+1)}) & \text{für } r = 0 \end{cases}$$

(empirisches)  $\alpha$ -Quantil bzw.  $\alpha \cdot 100\%$  Perzentil

mindestens  $\lfloor \alpha \cdot n \rfloor$  der Werte  $(x_1, \dots, x_n)$  sind  $\leq x_\alpha$   
mindestens  $\lfloor (1 - \alpha) \cdot n \rfloor$  der Werte  $(x_1, \dots, x_n)$  sind  $\geq x_\alpha$

Vereinbarung:  $x_0 = x_{(1)}$   $x_1 = x_{(n)}$

Werkzeuge der empirischen Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Quantile

## Beispiel

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

$$\begin{array}{ccccccccc} x_{(1)} & < & x_{(2)} & < & x_{(3)} & < & x_{(4)} & < & x_{(5)} \\ 1.5 & < & 2.7 & < & 2.8 & < & 3.0 & < & 3.1 \end{array}$$

$$\alpha = 0.25 :$$

$$\alpha \cdot n = 0.25 \cdot 5 = 1.25 = 1 + 0.25$$

$$\rightarrow x_{\alpha} = x_{0.25} = x_{(2)} = 2.7$$

$$\alpha = 0.75 :$$

$$\alpha \cdot n = 0.75 \cdot 5 = 3.75 = 3 + 0.75$$

$$\rightarrow x_{\alpha} = x_{0.75} = x_{(4)} = 3.0$$

$$\alpha = 0.5 :$$

$$\alpha \cdot n = 0.5 \cdot 5 = 2.5 = 2 + 0.5$$

$$\rightarrow x_{\alpha} = x_{0.5} = x_{(3)} = 2.8$$

# Lagemaße (3)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

## Median

ist das 0.5-Quantil  $x_{0.5}$ .

Beschreibende

# Lagemaße (3)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

## Median

ist das 0.5-Quantil  $x_{0.5}$ .

## Quartile

heißen die 0.25- und 0.75-Quantile  $x_{0.25}$  und  $x_{0.75}$ .

# Lagemaße (3)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

## Median

ist das 0.5-Quantil  $x_{0.5}$ .

## Quartile

heißen die 0.25- und 0.75-Quantile  $x_{0.25}$  und  $x_{0.75}$ .

## Modalwert

häufigster Wert

theoretischer Modalwert:

diskrete Merkmale: der wahrscheinlichste Wert

stetige Merkmale: Wert mit der größten Dichte

# Lagemaße (4)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

- Der Mittelwert ist in vielen Fällen eine 'gute' Lageschätzung, aber nicht robust (gegen Ausreißer).
- Der Median ist robust, aber meist nicht so 'gut'.

getrimmte Mittel, ( $\alpha$ -)getrimmtes Mittel

$$\bar{X}_\alpha := \frac{X_{(\lfloor n \cdot \alpha \rfloor + 1)} + \dots + X_{(n - \lfloor n \cdot \alpha \rfloor)}}{n - 2 \lfloor n \cdot \alpha \rfloor}, \quad \alpha \in \left[0, \frac{1}{2}\right)$$

Die  $\lfloor n \cdot \alpha \rfloor$  kleinsten und  $\lfloor n \cdot \alpha \rfloor$  größten Werte werden weggelassen und dann das arithmetische Mittel gebildet.

# Lagemaße (4)

- Der Mittelwert ist in vielen Fällen eine 'gute' Lageschätzung, aber nicht robust (gegen Ausreißer).
- Der Median ist robust, aber meist nicht so 'gut'.

getrimmte Mittel, ( $\alpha$ -)getrimmtes Mittel

$$\bar{X}_\alpha := \frac{X_{(\lfloor n \cdot \alpha \rfloor + 1)} + \dots + X_{(n - \lfloor n \cdot \alpha \rfloor)}}{n - 2 \lfloor n \cdot \alpha \rfloor}, \quad \alpha \in \left[0, \frac{1}{2}\right)$$

Die  $\lfloor n \cdot \alpha \rfloor$  kleinsten und  $\lfloor n \cdot \alpha \rfloor$  größten Werte werden weggelassen und dann das arithmetische Mittel gebildet.

$\bar{X}_\alpha$  ist robuster als  $\bar{X}$  und effizienter als  $x_{0.5}$ .

# Lagemaße (5)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

winsorisiertes Mittel,  $(\alpha)$ -winsorisiertes Mittel

Sei  $\alpha \in [0, \frac{1}{2})$  und jetzt  $n_1 := \lfloor n \cdot \alpha \rfloor + 1$ .

$$\bar{X}_{\alpha,w} := \frac{n_1 X_{(n_1)} + X_{(n_1+1)} + \dots + X_{(n-n_1)} + n_1 X_{(n-n_1+1)}}{n}$$

Die  $\lfloor n \cdot \alpha \rfloor$  kleinsten und  $\lfloor n \cdot \alpha \rfloor$  größten Werte werden “herangeschoben” und dann das arithmetische Mittel gebildet.

# Lagemaße (5)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

winsorisiertes Mittel, ( $\alpha$ -)winsorisiertes Mittel

Sei  $\alpha \in [0, \frac{1}{2})$  und jetzt  $n_1 := \lfloor n \cdot \alpha \rfloor + 1$ .

$$\bar{X}_{\alpha,w} := \frac{n_1 X_{(n_1)} + X_{(n_1+1)} + \dots + X_{(n-n_1)} + n_1 X_{(n-n_1+1)}}{n}$$

Die  $\lfloor n \cdot \alpha \rfloor$  kleinsten und  $\lfloor n \cdot \alpha \rfloor$  größten Werte werden “herangeschoben” und dann das arithmetische Mittel gebildet.

- winsorisiertes Mittel ist robuster als  $\bar{X}$  und effizienter als  $x_{0.5}$ .

Empfehlung für  $\bar{X}_{\alpha}, \bar{X}_{\alpha,w}$ :  $\alpha : 0.1 \quad \dots \quad 0.2$ .

# Lageschätzungen mit SAS

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

**Mittelwert:**

**PROC MEANS;**

**Median:**

**PROC MEANS MEDIAN;**

**PROC UNIVARIATE;**

**getrimmte Mittel:**

**PROC UNIVARIATE**

**TRIMMED=Zahl;**

**winsorisierte Mittel:**

**PROC UNIVARIATE**

**WINSORIZED=Zahl;**

**Quartile:**

**PROC UNIVARIATE;**

**Modalwert:**

**PROC UNIVARIATE;**

**Quantile:**

**PROC UNIVARIATE;**

`Descr1.sas`

`Mean.sas`

# Beispiele (1)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

## Tödliche Unfälle durch Pferdetritte

14 Corps, 20 Jahre, insges. 280 Einheiten. Erfasst wurde für jede Einheit die Anzahl der tödlichen Unfälle durch Pferdetritte.

| Anzahl | Häufigkeit |
|--------|------------|
| 0      | 144        |
| 1      | 91         |
| 2      | 32         |
| 3      | 11         |
| 4      | 2          |
| 5      | 0          |

Beschreibende

# Beispiele (1)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

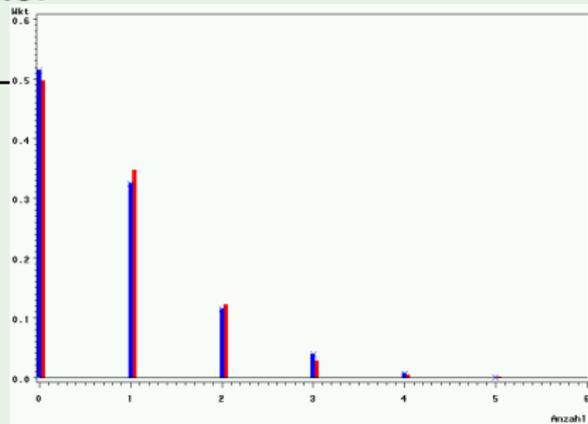
Normalverteilung (2)

Beschreibende

## Tödliche Unfälle durch Pferdetritte

14 Corps, 20 Jahre, insges. 280 Einheiten. Erfasst wurde für jede Einheit die Anzahl der tödlichen Unfälle durch Pferdetritte.

| Anzahl | Häufigkeit |
|--------|------------|
| 0      | 144        |
| 1      | 91         |
| 2      | 32         |
| 3      | 11         |
| 4      | 2          |
| 5      | 0          |



# Beispiele (1)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

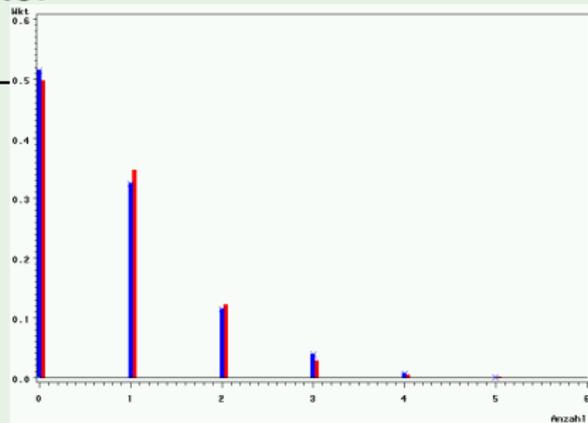
Normalverteilung (2)

Beschreibende

## Tödliche Unfälle durch Pferdetritte

14 Corps, 20 Jahre, insges. 280 Einheiten. Erfasst wurde für jede Einheit die Anzahl der tödlichen Unfälle durch Pferdetritte.

| Anzahl | Häufigkeit |
|--------|------------|
| 0      | 144        |
| 1      | 91         |
| 2      | 32         |
| 3      | 11         |
| 4      | 2          |
| 5      | 0          |



Poisson-Verteilung geeignet (?)

# Beispiele (2)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

## Anzahl von schwarzen Feldern

Ein Zufallszahlengenerator soll zufällige Bildpunkte erzeugen, weiß mit Wkt. 0.71 und schwarz mit Wkt. 0.29.

Dazu wurde ein großes Quadrat in 1000 Teilquadrate mit je 16 Bildpunkten zerlegt.

Gezählt wurde jeweils die Anzahl der schwarzen Bildpunkte.

# Beispiele (2)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

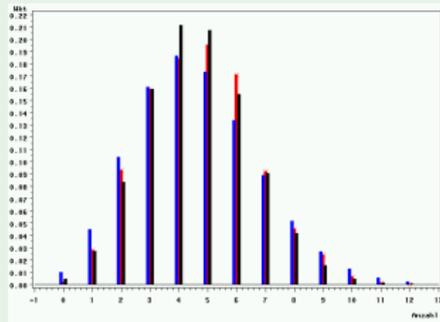
Beschreibende

## Anzahl von schwarzen Feldern

Ein Zufallszahlengenerator soll zufällige Bildpunkte erzeugen, weiß mit Wkt. 0.71 und schwarz mit Wkt. 0.29.

Dazu wurde ein großes Quadrat in 1000 Teilquadrate mit je 16 Bildpunkten zerlegt.

Gezählt wurde jeweils die Anzahl der schwarzen Bildpunkte.



# Beispiele (2)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

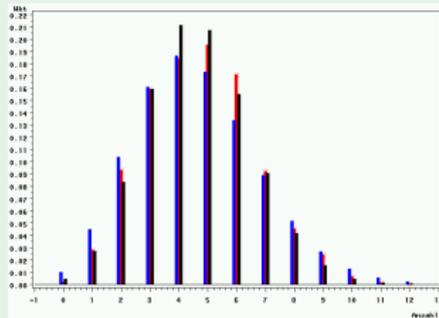
Beschreibende

## Anzahl von schwarzen Feldern

Ein Zufallszahlengenerator soll zufällige Bildpunkte erzeugen, weiß mit Wkt. 0.71 und schwarz mit Wkt. 0.29.

Dazu wurde ein großes Quadrat in 1000 Teilquadrate mit je 16 Bildpunkten zerlegt.

Gezählt wurde jeweils die Anzahl der schwarzen Bildpunkte.



| n        | 0 | 1  | 2  | 3   | 4   | 5   | 6   | 7  | 8  | 9  | 10 |
|----------|---|----|----|-----|-----|-----|-----|----|----|----|----|
| <b>h</b> | 2 | 28 | 93 | 159 | 184 | 195 | 171 | 92 | 45 | 24 | 6  |

Binomial-Verteilung (schwarz) geeignet (?)

# Eigenschaften von Schätzungen (1)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Sei  $\hat{\theta}_n$  eine Schätzung von  $\theta$ , die auf  $n$   
Beobachtungen beruht.

Konsistenz (Minimalforderung)

$$\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$$

Beschreibende

# Eigenschaften von Schätzungen (1)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Sei  $\hat{\theta}_n$  eine Schätzung von  $\theta$ , die auf  $n$  Beobachtungen beruht.

Konsistenz (Minimalforderung)

$$\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$$

Erwartungstreue, Asymptotische Erwartungstreue

$$\mathbf{E}\hat{\theta}_n = \theta$$

$$\mathbf{E}\hat{\theta}_n \rightarrow_{n \rightarrow \infty} \theta$$

Beschreibende

# Eigenschaften von Schätzungen (1)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

Sei  $\hat{\theta}_n$  eine Schätzung von  $\theta$ , die auf n Beobachtungen beruht.

Konsistenz (Minimalforderung)

$$\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$$

Erwartungstreue, Asymptotische Erwartungstreue

$$\mathbf{E}\hat{\theta}_n = \theta$$

$$\mathbf{E}\hat{\theta}_n \rightarrow_{n \rightarrow \infty} \theta$$

“gute”, “effiziente” Schätzung

$\text{var } \hat{\theta}_n$  möglichst klein

# Eigenschaften von Schätzungen (2)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

optimale Schätzung

wenn  $\text{var } \hat{\theta}_n$  den kleinstmöglichen Wert annimmt für  
alle e-treuen Schätzungen

Beschreibende

# Eigenschaften von Schätzungen (2)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

## optimale Schätzung

wenn  $\text{var } \hat{\theta}_n$  den kleinstmöglichen Wert annimmt für alle e-treuen Schätzungen

## Mean Square Error (MSE)

$$\begin{aligned}\text{MSE} &= \text{var } \hat{\theta}_n + \text{bias}^2 \hat{\theta}_n \\ &= \text{var } \hat{\theta}_n + (E\hat{\theta}_n - \theta)^2\end{aligned}$$

soll minimal oder möglichst klein sein.

# Eigenschaften von Schätzungen (2)

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

## optimale Schätzung

wenn  $\text{var } \hat{\theta}_n$  den kleinstmöglichen Wert annimmt für alle e-treuen Schätzungen

## Mean Square Error (MSE)

$$\begin{aligned}\text{MSE} &= \text{var } \hat{\theta}_n + \text{bias}^2 \hat{\theta}_n \\ &= \text{var } \hat{\theta}_n + (E\hat{\theta}_n - \theta)^2\end{aligned}$$

soll minimal oder möglichst klein sein.

## robuste Schätzung

Eigenschaften sollten “möglichst” auch bei (kleinen) Abweichungen von der (Normal-) Verteilungsannahme gelten

# Eigenschaften von Schätzungen (3)

## Cramer-Rao Ungleichung

$\theta$ : zu schätzender Parameter einer Population (Dichte  $f$ ).  
 $\hat{\theta} = \theta_n$ : eine erwartungstreue Schätzung von  $\theta$ .

### Cramer-Rao-Ungleichung

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)},$$

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (3)

## Cramer-Rao Ungleichung

$\theta$ : zu schätzender Parameter einer Population (Dichte  $f$ ).

$\hat{\theta} = \theta_n$ : eine erwartungstreue Schätzung von  $\theta$ .

### Cramer-Rao-Ungleichung

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)},$$

### Fisher-Information

$$\begin{aligned} I(f, \theta) &= \mathbf{E} \left( \frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2 \\ &= \int \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx \end{aligned}$$

Die Varianz einer Schätzung kann, bei gegebenem Stichprobenumfang, nicht beliebig klein werden.

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (4)

## Beispiele

$f$  normal

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\ln f(x, \mu) = -\ln(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x, \mu)}{\partial \mu} = \frac{x-\mu}{\sigma} \cdot \frac{1}{\sigma}$$

$$I(f, \mu) = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 \cdot f(x, \mu) dx = \frac{1}{\sigma^2}.$$

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (5)

## Beispiele (2)

Nach der Cramer-Rao-Ungleichung gilt also für jede Lageschätzung

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)} = \frac{\sigma^2}{n},$$

Werkzeuge der empirischen Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (5)

## Beispiele (2)

Nach der Cramer-Rao-Ungleichung gilt also für jede Lageschätzung

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)} = \frac{\sigma^2}{n},$$

insbesondere

$$\text{var}\bar{X} \geq \frac{\sigma^2}{n}.$$

Werkzeuge der empirischen Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (5)

## Beispiele (2)

Nach der Cramer-Rao-Ungleichung gilt also für jede Lageschätzung

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)} = \frac{\sigma^2}{n},$$

insbesondere

$$\text{var}\bar{X} \geq \frac{\sigma^2}{n}.$$

Vergleichen Sie das mit:

$$\text{var}\bar{X} = \frac{1}{n^2} \sum_{i=1}^n \text{var}X_i = \frac{\sigma^2}{n}.$$

Bei Normalverteilung ist also  $\bar{X}$  Lageschätzung mit minimaler Varianz.

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (6)

## Beispiele (3)

$f$  exponential

$$f(x, \lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

$$I(f, \lambda) = \frac{1}{\lambda^2} \quad (\text{ÜA, 2 P.})$$

Die Cramer-Rao-Schranke ist also:

$$\frac{1}{nI(\lambda)} = \frac{\lambda^2}{n}.$$

Vergleichen Sie mit  $\text{var}\bar{X} = \frac{\lambda^2}{n}$ .

Bei Exponentialverteilung ist also  $\bar{X}$  Parameterschätzung mit minimaler Varianz.

Werkzeuge der empirischen Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende

# Eigenschaften von Schätzungen (7)

## Beispiele (4)

$f$  Doppelexponential (=Laplace)

$$f(x, \lambda, \mu) = \frac{1}{2} \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x-\mu)} & \text{falls } x \geq \mu \\ \frac{1}{\lambda} e^{\frac{1}{\lambda}(x-\mu)} & \text{falls } x < \mu \end{cases}$$

Der hier interessierende (Lage-) Parameter ist  $\mu$ .

$$I(f, \mu) = \frac{1}{\lambda^2}. \quad (\text{ÜA, 5 P.}) \quad \text{var}(\bar{X}) = \frac{2\lambda^2}{n}. \quad (\text{ÜA, 2 P.})$$

Für den Median  $x_{0.5}$  gilt:

$$\text{var}(x_{0.5}) \sim \frac{\lambda^2}{n}. \quad (\text{ÜA, 10 P.})$$

Werkzeuge der  
empirischen  
Forschung

W. Kössler

Einleitung

Datenbehandlung

Syntax

Tastatur

Transformationen

Externes File

Input-Anweisung

SAS-Files

Zusammenfügen

Output-Anweisung

DO-Schleifen

Wkt.rechnung

Population

Wahrscheinlichkeit

Zufallsvariablen

Diskrete Zufallsvariablen

Stetige Zufallsvariablen

Normalverteilung (1)

Erwartungswert

Varianz

Normalverteilung (2)

Beschreibende