

Scatterplots

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)

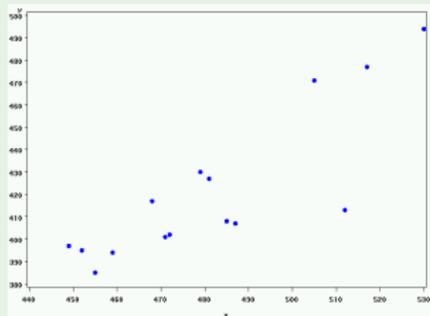
Beschreibende

Scatterplot

Zweidimensionale Stichproben können als Punkte in der Ebene dargestellt werden

Länge und Breite von Venusmuscheln

```
PROC GPLOT;  
PLOT y*x;  
RUN;
```



Descr_Scatter.sas

Descr_Scatter1.sas

Das Regressionsproblem

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

X, Y : Zufallsvariablen (auch mehrdimensional)

Modell:

$$Y = f(X, \underbrace{\theta_1, \dots, \theta_p}_{\text{Parameter}}) + \epsilon, \quad \epsilon \sim (0, \sigma^2).$$

Parameter zuf. Fehler

f linear, bekannt bis auf Parameter:
lineare Regression

f nichtlinear, bekannt bis auf Parameter:
nichtlineare Regression

f unbekannt: nichtparametrische Regression

Regression

f bekannt (bis auf Parameter)

Aufgabe:

$$\min_{\theta_1, \dots, \theta_p} \mathbf{E}(Y - f(\mathbf{X}, \theta_1, \dots, \theta_p))^2$$

$\theta_1, \dots, \theta_p$ unbekannt.

Beobachtungen: (Y_i, \mathbf{X}_i) .

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

Regression

f bekannt (bis auf Parameter)

Aufgabe:

$$\min_{\theta_1, \dots, \theta_p} \mathbf{E}(Y - f(\mathbf{X}, \theta_1, \dots, \theta_p))^2$$

$\theta_1, \dots, \theta_p$ unbekannt.

Beobachtungen: (Y_i, \mathbf{X}_i) .

Erwartungswert durch arithmetisches Mittel ersetzen

$$\min_{\theta_1, \dots, \theta_p} \frac{1}{n} \sum_{i=1}^n (Y_i - f(\mathbf{X}_i, \theta_1, \dots, \theta_p))^2$$

Kleinste Quadrat-Schätzung für $\theta_1, \dots, \theta_p$ (KQS)
Least-Squares-Estimation (LSE)

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

Regression

f bekannt (bis auf Parameter)

Lösung des Minimum-Problems

$$\min_{\theta_1, \dots, \theta_p} \frac{1}{n} \sum_{i=1}^n (Y_i - f(\mathbf{X}_i, \theta_1, \dots, \theta_p))^2$$

zu minimierende Funktion differenzieren und Null setzen:

$$\frac{2}{n} \cdot \sum_{i=1}^n (Y_i - f(\mathbf{X}_i, \theta_1, \dots, \theta_p)) \cdot \frac{\partial f(\mathbf{X}_i, \theta_1, \dots, \theta_p)}{\partial \theta_j} = 0$$

$j = 1, \dots, p, \Rightarrow$ Gleichungssystem mit p Gleichungen.

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

Regression

f linear: lineares Gleichungssystem (1)

$$f(X, \theta_1, \theta_2) = \theta_1 X + \theta_2$$

$$\frac{\partial f}{\partial \theta_1} = X \quad \frac{\partial f}{\partial \theta_2} = 1$$

$$\frac{1}{n} \sum_{i=1}^n (Y_i - (\theta_1 X_i + \theta_2)) \cdot X_i = 0$$

$$\frac{1}{n} \sum_{i=1}^n (Y_i - (\theta_1 X_i + \theta_2)) \cdot 1 = 0$$

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \theta_2 \sum_i X_i = 0$$

$$\sum_i Y_i - \theta_1 \sum_i X_i - \theta_2 \cdot n = 0$$

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

Regression

f linear: lineares Gleichungssystem (2)

Die zweite Gleichung nach θ_2 auflösen:

$$\theta_2 = \frac{1}{n} \sum_i Y_i - \theta_1 \frac{1}{n} \sum_i X_i$$

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

Regression

f linear: lineares Gleichungssystem (2)

Die zweite Gleichung nach θ_2 auflösen:

$$\theta_2 = \frac{1}{n} \sum_i Y_i - \theta_1 \frac{1}{n} \sum_i X_i$$

und in die erste einsetzen:

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \frac{1}{n} \sum_i Y_i \sum_i X_i + \theta_1 \frac{1}{n} \sum_i X_i \sum_i X_i = 0$$

$$\sum_i X_i Y_i - \frac{1}{n} \sum_i Y_i \sum_i X_i - \theta_1 \left(\sum_i X_i^2 - \frac{1}{n} \sum_i X_i \sum_i X_i \right) = 0$$

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

Regression

f linear: lineares Gleichungssystem (2)

Die zweite Gleichung nach θ_2 auflösen:

$$\theta_2 = \frac{1}{n} \sum_i Y_i - \theta_1 \frac{1}{n} \sum_i X_i$$

und in die erste einsetzen:

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \frac{1}{n} \sum_i Y_i \sum_i X_i + \theta_1 \frac{1}{n} \sum_i X_i \sum_i X_i = 0$$

$$\sum_i X_i Y_i - \frac{1}{n} \sum_i Y_i \sum_i X_i - \theta_1 \left(\sum_i X_i^2 - \frac{1}{n} \sum_i X_i \sum_i X_i \right) = 0$$

\Rightarrow

$$\hat{\theta}_1 = \frac{\sum_i X_i Y_i - \frac{1}{n} \sum_i X_i \sum_i Y_i}{\sum_i X_i^2 - \frac{1}{n} (\sum_i X_i)^2} = \frac{S_{XY}}{S_X^2}, \hat{\theta}_2 = \frac{1}{n} \left(\sum_i Y_i - \hat{\theta}_1 \sum_i X_i \right)$$

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

Regression

Zähler und Nenner in $\hat{\theta}_1$

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

$$\begin{aligned}S_{XY} &= \frac{1}{n-1} \sum_i (X_i - \bar{X})(Y_i - \bar{Y}) \\&= \frac{1}{n-1} \left(\sum_i X_i Y_i - \bar{X} \sum_i Y_i - \bar{Y} \sum_i X_i + n\bar{X}\bar{Y} \right) \\&= \frac{1}{n-1} \left(\sum_i X_i Y_i - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} \right) \\&= \frac{1}{n-1} \left(\sum_i X_i Y_i - n\bar{X}\bar{Y} \right) \\&= \frac{1}{n-1} \left(\sum_i X_i Y_i - \frac{1}{n} \sum_i X_i \sum_i Y_i \right) \\S_{X^2} &= \frac{1}{n-1} \left(\sum_i X_i X_i - \frac{1}{n} \sum_i X_i \sum_i X_i \right)\end{aligned}$$

Spezialfall $f(X, \theta) = \theta$ (konstant)

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)

$$Y_i = \theta + \epsilon_i, \quad \epsilon_i \sim (0, \sigma^2)$$

Minimierungsaufgabe:

$$\min_{\theta} \left(\sum_{i=1}^n (Y_i - \theta)^2 \right)$$

Lösung:

$$2 \sum_{i=1}^n (Y_i - \theta) = 0 \quad \sum_{i=1}^n Y_i - n\theta = 0$$

$$\hat{\theta} = \frac{1}{n} \sum Y_i = \bar{Y}$$

D.h. \bar{Y} ist auch KQS.

Beschreibende

Spezialfall $f(X, \theta) = \theta$

Schätzung des Schätzfehlers

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

$$\sigma_{Y_i}^2 = \sigma_{\theta + \epsilon_i}^2 = \sigma_{\epsilon_i}^2 = \sigma^2.$$

Schätzfehler:

$$\begin{aligned}\sigma_{\hat{\theta}}^2 &= \text{var}(\hat{\theta}) = \text{var}\left(\frac{1}{n} \cdot \sum Y_i\right) = \frac{1}{n^2} \cdot n \cdot \text{var}Y_i \\ &= \frac{1}{n} \cdot \sigma^2 \quad \rightarrow_{n \rightarrow \infty} 0\end{aligned}$$

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{\hat{\sigma}^2}{n}$$

Lineare und Nichtlineare Regression

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

f : linear, $f(X, \theta_1, \theta_2) = \theta_1 X + \theta_2$

θ_1 und θ_2 werden geschätzt.

Descr_Scatter_1.sas

Descr_Scatter_Heroin.sas

Lineare und Nichtlineare Regression

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

f : linear, $f(X, \theta_1, \theta_2) = \theta_1 X + \theta_2$

θ_1 und θ_2 werden geschätzt.

Descr_Scatter_1.sas

Descr_Scatter_Heroin.sas

f : nichtlinear, z.B. $f(X, \theta_1, \theta_2) = \ln(\theta_1 X + \theta_2)$

a) Lösung des nichtlinearen Gleichungssystems

b) wird auf den linearen Fall zurückgeführt

$$Y = \ln(\theta_1 X + \theta_2) + \epsilon$$

$$e^Y = \theta_1 X + \theta_2 + \tilde{\epsilon}$$

Modelle sind aber i.A. nicht äquivalent!

Weitere nichtlineare Regressionsfunktionen

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)

$$f(t) = a + bt + ct^2 \quad \text{Parabel}$$

$$f(t) = at^b \quad \text{Potenzfunktion}$$

$$f(t) = ae^t \quad \text{Exponentialfunktion}$$

$$f(t) = k - ae^{-t}$$

$$f(t) = \frac{k}{1 + be^{-ct}} \quad \text{logistische Funktion}$$

$$\ln f(t) = k - \frac{a}{b + t} \quad \text{Johnson-Funktion}$$

$$\ln f(t) = k - \lambda e^{-t} \quad \text{Gompertz-Funktion}$$

Beschreibende

Nichtparametrische Regression

f unbekannt, aber "glatt"

z.B. f 2x stetig differenzierbar, $f \in C_2$, $\lambda \geq 0$

Glättender Kubischer Spline ist Lösung von

$$\min_{f \in C_2} \sum_{i=1}^n (Y_i - f(X_i))^2 + \lambda \cdot \int (f''(x))^2 dx$$

Descr_Scatter.sas

SYMBOL I=SMnnS;

SM: Smoothing Spline

nn: Glättungsparameter

nn=00: Interpolierender Spline

nn=99: Gerade

S: Punktepaare werden vor der Auswertung nach dem Argument sortiert.

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

Nichtparametrische Regression

Kernschätzung, Motivation

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

geg.: Kernfunktion K , standardisierte Dichte, z.B. Normaldichte, Epanechnikov-Kern.

Regressionsmodell:

$$\begin{aligned} Y &= f(X) + \epsilon, \quad \epsilon \sim (0, \sigma^2) \quad \text{also} \\ \mathbf{E}(Y|X = x) &= f(x) \\ f(x) &= \mathbf{E}(Y|X = x) \\ &= \int y f_{Y|X}(y|x) dy \\ &= \int y \frac{g(x, y)}{f_0(x)} dy \\ &= \frac{\int y g(x, y) dy}{f_0(x)} \end{aligned}$$

Regression

Kernschätzung

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)

$$f(x) = \frac{\int yg(x, y)dy}{f_0(x)}$$

$g(x, y)$: gemeinsame Dichte von (X, Y)

$f_0(x)$: Randdichte von X

$f_{Y|X}$: bedingte Dichte von Y

Beschreibende

Regression

Kernschätzung

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

$$f(x) = \frac{\int yg(x, y)dy}{f_0(x)}$$

$g(x, y)$: gemeinsame Dichte von (X, Y)

$f_0(x)$: Randdichte von X

$f_{Y|X}$: bedingte Dichte von Y

Der Nenner wird geschätzt durch

$$\hat{f}_0(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \cdot K\left(\frac{x - X_i}{h}\right)$$

und der Zähler durch

$$\frac{1}{n} \sum_{i=1}^n Y_i \hat{g}(X_i, Y_i) = \frac{1}{n} \sum_{i=1}^n Y_i \cdot \frac{1}{h} \cdot K\left(\frac{x - X_i}{h}\right)$$

Regression

Kernschätzung

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

Beide zusammen ergeben die

Kernschätzung

$$\hat{f}(x) = \frac{\sum_{i=1}^n Y_i \cdot \frac{1}{h} \cdot K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n \frac{1}{h} \cdot K\left(\frac{x-X_i}{h}\right)}$$

K: Kernfunktion

h: Glättungsparameter

Nichtparametrische Kurvenschätzung

Spline- und Kernschätzung

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

Illustration: SAS-INSIGHT.

Analyse

Fit(Y X)

Output

Nonparametric Curves

Smoothing Spline

Normal kernel smoother

- Venus-Muschel Daten (WORK/Descr_Scatter)
- Heroin-Daten (SASUSER/heroin) (TIME-DOSE)

Glättende Splines können auch mit Hilfe der
Prozedur GPLOT erzeugt werden.

Beschreibende Statistik

Zusammenfassung (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)

Verteilungsfunktion

$$F(x) = P(X \leq x)$$

diskrete Verteilung

$$F(x) = \sum_{i:i \leq x} p_i \quad p_i = P(X = x_i)$$

stetige Verteilung

$$F(x) = \int_{-\infty}^x f(t) dt, \quad f(t) : \text{Dichte.}$$

Bsp: diskrete Verteilung: Binomial, Poisson
stetige Verteilung: Normal, Gleich, Exp

Beschreibende Statistik

Zusammenfassung (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

Erwartungswert

$$\mathbf{E}(X) = \begin{cases} \sum x_i p_i & X \text{ diskret} \\ \int x f(x) dx & X \text{ stetig} \end{cases}$$

Varianz

$$\text{var}(X) = \mathbf{E}(X - \mathbf{E}X)^2$$

Normalverteilung, Dichte

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}} \quad \text{Standard}$$

$$f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}$$

Beschreibende Statistik

Zusammenfassung (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)

Gesetz der Großen Zahlen ($\mathbf{E}(X) < \infty$)

$$\bar{X} \longrightarrow \mathbf{E}X.$$

Zentraler Grenzwertsatz (X_i iid)

$$\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \longrightarrow Z \sim \mathcal{N}(0, 1)$$

$$\sqrt{n} \cdot \frac{\bar{X} - \mu}{s} \longrightarrow Z \sim \mathcal{N}(0, 1)$$

$$\bar{X} = \frac{1}{n} \sum X_i$$

$$s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 \rightarrow \sigma^2$$

Beschreibende

Beschreibende Statistik

Zusammenfassung (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)

Statistische Maßzahlen

Lagemaße: \bar{X} , $x_{0.5}$, x_{α} , $x_{0.25}$, $x_{0.75}$, \bar{x}_{α} , $\bar{x}_{\alpha,w}$

Skalenmaße: s^2 , s , R , IR , MAD , S_n , Q_n

Formmaße: β_1 , β_2

PROC UNIVARIATE
PROC UNIVARIATE ROBUSTSCALE
PROC UNIVARIATE TRIMMED=
PROC UNIVARIATE WINSORIZED=
PROC MEANS MEDIAN STD

Beschreibende Statistik

Zusammenfassung (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)

Boxplots

PROC BOXPLOT

PROC GPLOT

Häufigkeitsdiagramme

PROC GCHART

PROC UNIVARIATE

HISTOGRAM

Häufigkeitstabellen:

PROC FREQ

Zusammenhangsmaße:

PROC CORR

Pearson, Spearman, Kendall-Korrelationskoeff.

Scatterplots, Regression, Schätzung der

Regressionskoeffizienten: **PROC GPLOT**