

Streuungsmaße

Die angegebenen Maßzahlen sind empirisch, d.h. sie sind Schätzungen für die wahre Varianz

(empirische) Varianz (Streuung)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$s^2 \rightarrow_{n \rightarrow \infty} \text{var}(X)$$

Warum Division durch $(n - 1)$: Erwartungstreue (ÜA)

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

Streuungsmaße

Die angegebenen Maßzahlen sind empirisch, d.h. sie sind Schätzungen für die wahre Varianz

(empirische) Varianz (Streuung)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$s^2 \rightarrow_{n \rightarrow \infty} \text{var}(X)$$

Warum Division durch $(n - 1)$: Erwartungstreue (ÜA)

Standardabweichung

$$s = \sqrt{s^2}$$

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

Streuungsmaße (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)

Spannweite (Range)

$$X_{(n)} - X_{(1)}$$

Beschreibende

Streuungsmaße (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)

Spannweite (Range)

$$X_{(n)} - X_{(1)}$$

(Inter-)Quartilsabstand, IR

$$IR = x_{0.75} - x_{0.25}$$

Beschreibende

Streuungsmaße (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

Spannweite (Range)

$$X_{(n)} - X_{(1)}$$

(Inter-)Quartilsabstand, IR

$$IR = x_{0.75} - x_{0.25}$$

Mittlere absolute Abweichung vom Median

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - x_{0.5}|$$

Streuungsmaße (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 absolute deviation, MAD

$$MAD = med(|X_i - x_{0.5}|)$$

Wenn $X \sim \mathcal{N}$ so $\mathbf{E}(1.4826 \cdot MAD) = \sigma$

Variationskoeffizient

$$CV = \frac{s \cdot 100}{\bar{X}}$$

Gini's Mean Difference

$$G = \frac{1}{\binom{n}{2}} \sum_{i < j} |x_i - x_j|$$

$$X \sim \mathcal{N} \Rightarrow \mathbf{E}\left(\frac{\sqrt{\pi}}{2} G\right) = \sigma$$

Streuungsmaße (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

S_n und Q_n (Croux, Rousseuw 1992, 1993)

$$S_n = 1.1926 \cdot \text{med}_i(\text{med}_j |x_i - x_j|)$$

$$Q_n = 2.219 \cdot \{ |x_i - x_j|, i < j \}_{(k)}$$

$$k = h^2, h = \lfloor \frac{n}{2} \rfloor + 1$$

SAS verwendet einen modifizierten Schätzer
(Korrekturfaktor) für kleine Umfänge.

Die konstanten Faktoren sichern Erwartungstreue
bei Normalverteilung, $X \sim \mathcal{N} \Rightarrow \mathbf{E}(S_n) = \mathbf{E}(Q_n) = \sigma$

Streuungsmaße (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

Eigenschaften:

- Varianz und Standardabweichung und Spannweite sind nicht “robust”.
- IR und MAD sind robust.
(MAD etwas besser da höherer “Bruchpunkt”)
- G ist bedingt robust, effizient bei F normal.
- MAD ist wenig effizient.
(0.37 bei Normal)
- S_n oder Q_n sind geeignetste Schätzungen.

Streuungsmaße (6)

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

Nicht-Robuste Skalenschätzungen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

$$\text{Range} = X_{(n)} - X_{(1)}$$

$$CV = \frac{s \cdot 100}{\bar{X}}$$

Streuungsmaße (7)

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

Robuste Skalenschätzungen

$$IR = x_{0.75} - x_{0.25}$$

$$MAD = med(|x_i - x_{0.5}|)$$

$$G = \frac{1}{\binom{n}{2}} \sum_{i < j} |x_i - x_j|$$

$$S_n = 1.1926 \cdot med_i(med_j |x_i - x_j|)$$

$$Q_n = 2.219 \cdot \{|x_i - x_j|, i < j\}_{(k)}$$

$$k = h^2, h = \lfloor \frac{n}{2} \rfloor + 1$$

Lage- und Streuungsmaße in SAS (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

PROC MEANS;

VAR Zeit;

RUN;

Standardausgabe:

N, Mean, Std Dev, Minimum, Maximum

Vorteil: übersichtliche Darstellung

Nachteil: nur wenige Statistiken

Es können aber zusätzliche Statistiken durch
Optionen angefordert werden, z.B.

PROC MEANS Median Sum CL;

Descr1.sas

Lage- und Streuungsmaße in SAS (2)

Die Prozedur Univariate

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

```
PROC UNIVARIATE;  
    VAR Zeit;  
RUN;
```

N, Mean, Std Deviation, Variance
Sum Observations, Median, Mode
Range, Interquartile Range
Lokationstests (später)
Quantile
Extreme Beobachtungen

Lage- und Streuungsmaße in SAS (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

Getrimmte Mittel und robuste Skalenschätzer können einfach berechnet werden durch:

```
PROC UNIVARIATE ROBUSTSCALE TRIMMED=10  
WINSORISED=10;  
VAR ...;  
RUN;
```

TRIMMED: getrimmte Mittel

TRIMMED=10: die je 10 kleinsten und größten Beobachtungen werden weggelassen.

WINSORIZED: winsorisierte Mittel

ROBUSTSCALE: robuste Skalenschätzer

Formmaße

(Theoretische) Schiefe

$$\beta_1 = E\left(\frac{X - EX}{\sqrt{\text{var}(X)}}\right)^3$$

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

Formmaße

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

(Theoretische) Schiefe

$$\beta_1 = E\left(\frac{X - EX}{\sqrt{\text{var}(X)}}\right)^3$$

(Empirische) Schiefe

$$\hat{\beta}_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s}\right)^3$$

$$\hat{\beta}_{1,SAS} = \hat{\beta}_1 \frac{n^2}{(n-1)(n-2)}$$

$\beta_1 = 0$ falls F symmetrisch

$\beta_1 < 0$ falls F linksschief

$\beta_1 > 0$ falls F rechtsschief

ÜA: Berechnen Sie die (theor.) Schiefe von

$$X : \begin{pmatrix} \frac{1}{2}(-4 - \sqrt{6}) & -1 & \frac{1}{2}(-4 + \sqrt{6}) & 2 & 3 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}$$

und von

$$Y : \begin{pmatrix} -9 & -7 & 2 & 4 & 10 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}$$

PROC MEANS skewness;

PROC MEANS skewness vardef=n; (ohne Faktor)

Formmaße (2)

(Theoretische) Wölbung, Kurtosis

$$\beta_2 = E\left(\frac{X - EX}{\sqrt{\text{var}(X)}}\right)^4 - 3$$

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

Formmaße (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

(Theoretische) Wölbung, Kurtosis

$$\beta_2 = E\left(\frac{X - EX}{\sqrt{\text{var}(X)}}\right)^4 - 3$$

(Empirische) Wölbung, Kurtosis

$$\hat{\beta}_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s}\right)^4 - 3$$

$$\hat{\beta}_{2,SAS} = \hat{\beta}_2 \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} - 3 \frac{4n^2 - 3n + 1}{(n-1)(n-2)(n-3)}$$

Exzeß

$$\beta_2 + 3 \quad \hat{\beta}_2 + 3$$

$\beta_2 = 0$ bei Normalverteilung

$\beta_2 > 0$ Tails “dicker, länger, stärker” als bei NV

$\beta_2 < 0$ Tails “dünner, kürzer, schwächer” als
bei NV

PROC MEANS kurtosis;

PROC MEANS kurtosis vardef=n; (ohne Faktor)

$\beta_2 = 0$ heißt nicht notwendig: $F \sim \text{Normal}$.

Kurtosis

Werkzeuge der
empirischen
Forschung

W. Kössler

Einleitung

Datenbehandlung

- Syntax
- Tastatur
- Transformationen
- Externes File
- Input-Anweisung
- SAS-Files
- Zusammenfügen
- Output-Anweisung
- DO-Schleifen

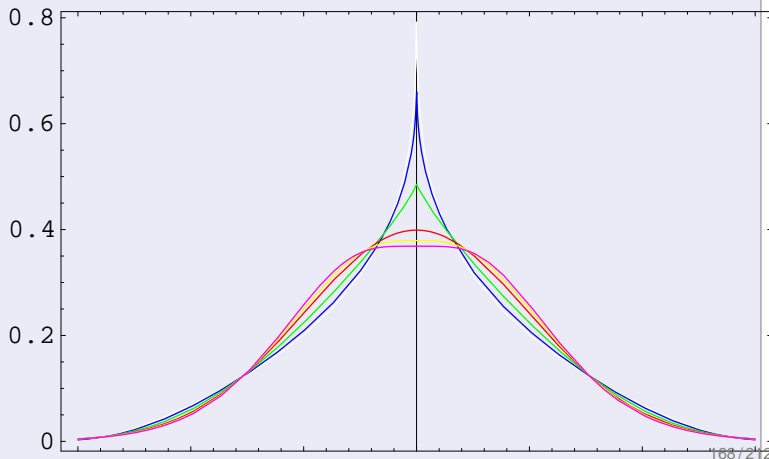
Wkt.rechnung

- Population
- Wahrscheinlichkeit
- Zufallsvariablen
- Diskrete Zufallsvariablen
- Stetige Zufallsvariablen
- Normalverteilung (1)
- Erwartungswert
- Varianz
- Normalverteilung (2)

Beschreibende

Dichten mit $E(X) = 0$, $var(X) = 1$, $\beta_1 = 0$, $\beta_2 = 0$

Dichte



Formmaße (3)

Theoretische Schiefe und Kurtosis verschiedener Verteilungen

Verteilung	Schiefe	Kurtosis
normal	0	0
gleich	0	-1.2
Doppelexp	0	3
Exponential	2	6
Bi(n,p)	$\frac{1-2p}{\sqrt{np(1-p)}}$	$-\frac{6}{n} + \frac{1}{np(1-p)}$
Poi(λ)	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$
Geo(p)	$\frac{2-p}{\sqrt{1-p}}$	$6 + \frac{p^2}{1-p}$

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

Einschub: GPLOT (vgl. ÜA 9)

Darstellung zweidimensionaler Zusammenhänge

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

```
SYMBOL1 i=spline c=green v=point;  
SYMBOL2 i=needle c=blue v=plus;  
PROC GPLOT;  
    PLOT y1*x=1 y2*x=2 /overlay;  
RUN;
```

Die darzustellenden Paare (x,y) sind vorher in einem DATA-Step zu erzeugen oder einzulesen.

Nach dem Gleichheitszeichen im Plot-Kommando steht die Nummer der zugehörigen SYMBOL-Anweisung.

Prozedur GLOT (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

Die Symbol-Anweisung beschreibt die Art, den Stil
des Plot

i=needle: Nadelplot (für diskrete
Wahrscheinlichkeiten praktisch)

i=join: (nach x) aufeinander folgende Punkte werden
verbunden

i=spline: Punkte werden durch einen Spline
verbunden

c=<Farbe>

v=<Zeichen>

overlay: alles in ein Plot.