

# Statistik – Einführung

## Unterlagen zur Lehrveranstaltung

Institut für Statistik  
WU Wien

Gerhard Derflinger  
Michael Hauser  
Jörg Lenneis  
Josef Leydold  
Günter Tirlir  
Rosmarie Wakolbinger

23. April 2003

# Inhaltsverzeichnis

- 1 – Daten und Statistisches Denken
- 2 – Beschreibende Statistik
- 3 – Wahrscheinlichkeitstheorie
- 4 – Diskrete Zufallsvariable
- 5 – Stetige Zufallsvariable
- 6 – Stichprobenverteilung
- 7 – Konfidenzintervalle für einen Parameter
- 8 – Tests auf einen Parameter
- 9 – Lineare Regression
- 10 – Kategoriale Daten
- 11 – Tests für zwei Grundgesamtheiten

# Übersicht

## 1 – Daten und Statistisches Denken

1. Definieren Statistik
2. Beschreiben die Verwendung von Statistik
3. Unterscheiden zwischen deskriptiver und schließender (Inferenz-)Statistik
4. Definieren Grundgesamtheit, Stichprobe, Parameter und den technischen Begriff Statistik
5. Definieren Merkmal und Merkmalsausprägungen

## 2 – Beschreibende Statistik

1. Charakterisieren Merkmale nach ihren Skalenniveaus.
2. Beschreiben qualitative Daten graphisch.
3. Beschreiben quantitative Daten graphisch.
4. Erzeugen und interpretieren graphische Darstellungen.
5. Erklären numerische Dateneigenschaften.
6. Beschreiben zusammenfassende Maßzahlen.
7. Analysieren numerische Daten mit Hilfe dieser Maßzahlen.

## 3 – Wahrscheinlichkeitstheorie

1. Definieren Experimente, Elementarereignisse, Ereignisse, Ereignisraum und Wahrscheinlichkeit.
2. Erklären wie Wahrscheinlichkeiten Ereignissen zugeordnet werden.
3. Verwenden Kontingenztafel, Venn-Diagramme und Bäume zum Bestimmen von Wahrscheinlichkeiten.
4. Beschreiben und verwenden Regeln für Wahrscheinlichkeiten.
5. Definieren von Unabhängigkeit und bedingter Wahrscheinlichkeit.

#### **4 – Diskrete Zufallsvariable**

1. Definieren Zufallsvariable (ZV).
2. Erklären diskrete Zufallsvariable.
3. Berechnen Erwartungswert und Varianz einer diskreten ZV.
4. Beschreiben Binomial- und Poissonverteilung.
5. Berechnen Wahrscheinlichkeiten diskreter ZV.

#### **5 – Stetige Zufallsvariable**

1. Definieren stetige Zufallsvariablen.
2. Beschreiben (stetig) gleich-, normal- und exponential- verteilte Zufallsvariablen.
3. Berechnen Erwartungswert und Varianz von stetigen ZV.
4. Bestimmen Wahrscheinlichkeiten stetiger Zufallsvariabler.

#### **6 – Stichprobenverteilung**

1. Beschreiben Eigenschaften von Schätzern.
2. Beschreiben Stichprobenverteilungen.
3. Beschreiben den Zusammenhang zwischen Grundgesamtheit und Stichprobenverteilung.
4. Formulieren den zentralen Grenzwertsatz.

#### **7 – Konfidenzintervalle für einen Parameter**

1. Formulieren, was geschätzt wird.
2. Unterscheiden Punktschätzer und Intervallschätzer.
3. Erklären Intervallschätzer.
4. Berechnen Konfidenzintervalle für Mittelwerte und Anteile.
5. Berechnen notwendige Stichprobengröße.

## **8 – Tests auf einen Parameter**

1. Beschreiben Test und Hypothese.
2. Unterscheiden Arten von Hypothesen.
3. Beschreiben das Testen von Hypothesen.
4. Erklären den Ablehnungsbereich und den  $p$ -Wert.
5. Lösen das (Hypothesen-)Testproblem für einen Parameter.
6. Definieren den Fehler 1. und 2. Art.
7. Erklären die Macht eines Tests.

## **9 – Lineare Regression**

1. Beschreiben lineares Regressionsmodell.
2. Geben die einzelnen Modellierungsschritte an.
3. Erklären die (gewöhnliche) Methode der kleinsten Quadrate, (gewöhnliche Kleinste-Quadrate-Methode).
4. Berechnen Regressionskoeffizienten.
5. Prognostizieren die „Antwort“-Variable.  
Sie wird auch als „Response“- , abhängige, zu erklärende Variable, oder Regressand bezeichnet.
6. Interpretieren einen Computerausdruck.

## **10 – Kategoriale Daten**

1. Erklären den  $\chi^2$ -Test auf Anteile. (Anpassungstest)
2. Erklären den  $\chi^2$ -Test auf Unabhängigkeit.

## **11 – Tests für zwei Grundgesamtheiten**

1. Testen Hypothesen über Parameter zweier Grundgesamtheiten:
  - Hypothesen über deren Mittel (Erwartungswerte)
  - Hypothesen über deren Anteile
2. Unterscheiden unabhängige Gesamtheiten und Gesamtheiten, die miteinander in Beziehung stehen.

# Formelübersicht

## 2 – Beschreibende Statistik

- $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$
- $\bar{x} = \sum_{i=1}^k x_i \frac{h_i}{n} = \frac{1}{n} \sum_{i=1}^k x_i h_i$
- Spannweite =  $x_{\max} - x_{\min}$
- $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$
- $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$
- Interquartilsabstand =  $Q_3 - Q_1$

## 3 – Wahrscheinlichkeitstheorie

- $P(\text{Ereignis}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{G}{M}$
- $P(\text{Ereignis}) = \frac{\text{Anzahl mit Eigenschaft}}{\text{Anzahl der Wiederholungen}} = \frac{X}{N}$
- $P(A \text{ oder } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \text{ oder } B) = P(A \cup B) = P(A) + P(B)$
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A|B) = P(A)$  und  $P(A \cap B) = P(A) P(B)$
- $P(A \text{ und } B) = P(A \cap B) = P(A) P(B|A) = P(B) P(A|B)$
- $P(A \text{ und } B) = P(A \cap B) = P(A) P(B)$

## 4 – Diskrete Zufallsvariable

- $\mu = E(X) = \sum x f(x)$

- $\sigma^2 = E[(X - \mu)^2] = \sum (x - \mu)^2 f(x)$
- $f(x) = \binom{n}{x} p^x (1 - p)^{n-x} = \frac{n!}{x! (n-x)!} p^x (1 - p)^{n-x}$
- $\mu = E(X) = n p$
- $\sigma = \sqrt{n p (1 - p)}$
- $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$
- $\mu = E(X) = \lambda$
- $\sigma = \sqrt{\lambda}$

## 5 – Stetige Zufallsvariable

- $f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right]$
- $Z = \frac{X - \mu}{\sigma}$
- $f(x) = \lambda \exp(-\lambda x)$

## 6 – Stichprobenverteilung

- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- $\mu_{\bar{x}} = \mu$
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- $Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- $\mu_{\bar{x}} = \mu$
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

## 7 – Konfidenzintervalle für einen Parameter

- $\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
- $\bar{X} - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}$
- $p - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \leq \theta \leq p + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}$

## 8 – Tests auf einen Parameter

- $H_0: \mu = \mu_0$  (Zahlenwert)
- $Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
- $Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
- $t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$
- $t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$
- $p = \frac{x}{n} = \frac{\text{Anzahl mit Eigenschaft}}{\text{Stichprobenumfang}}$
- $n p (1 - p) \geq 9$
- $Z = \frac{P - \mu_p}{\sigma_p} = \frac{P - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$

## 9 – Lineare Regression

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- $\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- $r^2 = 1 - \frac{\sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i}{\sum y_i^2 - (1/n)(\sum y_i)^2}$

$$\begin{aligned}
\cdot t_{n-2} &= \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \\
\cdot s_{\hat{\beta}_1}^2 &= \frac{s_\epsilon^2}{\sum (x_i - \bar{x})^2} \quad \text{und} \quad s_\epsilon^2 = \frac{1}{n-2} \sum \hat{\epsilon}_i^2 \\
\cdot r &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}
\end{aligned}$$

## 10 – Kategoriale Daten

$$\begin{aligned}
\cdot \chi^2 &= \sum \frac{(n_i^o - n_i^e)^2}{n_i^e} \\
\cdot \chi^2 &= \sum \frac{(n_{ik}^o - n_{ik}^e)^2}{n_{ik}^e} \\
\cdot n_{ik}^e &= \frac{n_{i\star} n_{\star k}}{n}
\end{aligned}$$

## 11 – Tests für zwei Grundgesamtheiten

$$\begin{aligned}
\cdot Z &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\
\cdot t &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
\cdot t &= \frac{\bar{D} - \mu_D}{\frac{s_D}{\sqrt{n_D}}} \quad \nu = n_D - 1 \\
\cdot Z &= \frac{(p_1 - p_2) - (\theta_1 - \theta_2)}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{wobei} \quad p = \frac{X_1 + X_2}{n_1 + n_2}
\end{aligned}$$