### 0.1 Grundbegriffe

Grundgesamtheit, Population
(gesamte Gruppe, Bsp. Schüler einer Schule)

1. Umfang
   - Stichprobe/Teilerhebung
     (nur kleiner Teil der gesamten Gruppe untersucht)
2. Datenform
   - Querrschnittsdaten
     (Momentaufnahme einer Stichprobe)
   - Zeitreihe
     (quantitative Auswertung über einen Zeitverlauf)
   - Längsschnittdaten
     (Datenerhebung zu versch. Zeitpunkten)
3. Methoden operationaler Messungen
   - Beobachtung:
     verdeckt oder teilnehmend; systematisch mit Beobachtungsprotokoll
   - Befragung:
     mündlich mit/ohne Interviewer, schriftlich/online, Fragebogen
   - Experiment:
     kontrollierte Situation, evtl. Erhebung durch Beobachtung oder Befragung

## 1 Grundlagen: Wahrscheinlichkeit

**Wichtiges zur Wahrscheinlichkeitsrechnung**

- $P(\cup_{i=1}^n) = \sum_{i=1}^n P(A_i)$ für paarweise disjunkte Ereignisse $A_1, A_2, ..., A_n \subset \Omega$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(\overline{A}) = 1 - P(A)$
- $P(A|\overline{B}) = 1 - P(\overline{A}|\overline{B})$
- $P(\overline{B} \cap A) = P(A|\overline{B}) \cdot P(\overline{B})$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ für beliebige A, B $\subset \Omega$
- $P(A \cup B) = P(A) + P(B)$ für disjunkte Ereignisse $A, B \subseteq \Omega$ (Additivität)

### 1.1 Bedingte Wahrscheinlichkeiten

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

**Satz der totalen Wahrscheinlichkeit**

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$

**Wichtiger Spezialfall: Multiplikationssatz**

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

### 1.2 Stochastische Unabhängigkeit

$$P(A|B) = P(A)\ bzw.\ P(B|A) = P(B)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

### 1.3 Bedingte Unabhängigkeit

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

### 1.4 Satz von Bayes

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)P(B)$$

$$\Rightarrow P(A|B)P(B) = P(B|A)P(A)$$

### 1.5 Satz von Bayes II

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})}$$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}$$

## 2 Zusammenhangsmaße diskreter Merkmale

X = Zeilen
Y = Spalten

### 2.1 Kontigenztafeln

- Absolute Häufigkeit

| | $b_1$ | $\cdots$ | $b_m$ | $h_{1\cdot}$ |
|---|---|---|---|---|
| $a_1$ | $h_{11}$ | $\cdots$ | $h_{1m}$ | $h_{1\cdot}$ |
| $a_2$ | $h_{21}$ | $\cdots$ | $h_{2m}$ | $h_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $a_k$ | $h_{k1}$ | $\cdots$ | $h_{km}$ | $h_{k\cdot}$ |
| | $h_{\cdot 1}$ | $\cdots$ | $h_{\cdot m}$ | $n$ |

- Relative Häufigkeiten

| | $b_1$ | $\cdots$ | $b_m$ | $h_{1\cdot}$ |
|---|---|---|---|---|
| $a_1$ | $f_{11}$ | $\cdots$ | $f_{1m}$ | $f_{1\cdot}$ |
| $a_2$ | $f_{21}$ | $\cdots$ | $f_{2m}$ | $f_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $a_k$ | $f_{k1}$ | $\cdots$ | $f_{km}$ | $f_{k\cdot}$ |
| | $f_{\cdot 1}$ | $\cdots$ | $f_{\cdot m}$ | 1 |

**Bedingte Häufigkeiten**

$$f_{Y|X}(b_1|a_1) = \frac{h_{i1}}{h_{i\cdot}}, ..., f_{Y|X}(b_m|a_i) = \frac{h_{im}}{hi\cdot}$$

$$f_{X|Y}(a_1|b_1) = \frac{h_{1j}}{h_{\cdot j}}, ..., f_{X|Y}(a_k|b_j) = \frac{h_{kj}}{h\cdot j}$$

### 2.2 Empirische Unabhängigkeit

$$f_{Y|X}(b_j|a_1) = f_Y(b_j|a_2) = ... = f_{Y|X}(b_j|a_k) \forall j = 1, ..., m$$

oder halt:

$$\forall i,j: \ h_{ij} = \tilde{h}_{ij}$$

### 2.3 $\chi^2$- und Kontigenzkoeffizient

**Berechnung von $\tilde{h}$**

ist erwartete Häufigkeit in einer Kontingenztafel

$$\text{mit } \tilde{h}_{ij} = \frac{h_{i\cdot} h_{\cdot j}}{n}$$

**$\chi^2$-Koeffizient**

$$\chi^2 := \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(h_{ij} - \frac{h_{i\cdot} h_{\cdot j}}{n})^2}{\frac{h_{i\cdot} h_{\cdot j}}{n}}$$

$$= n \sum_i \sum_j \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}}$$

$$\text{mit } \tilde{h}_{ij} = \frac{h_{i\cdot} h_{\cdot j}}{n}$$

- $\chi^2 \in [0, n(min(k,m) - 1)]$
- $\chi^2 = 0 \Leftrightarrow$ X und Y empirisch unabhängig
- $\chi^2$ groß, starker Zusammenhang
- $\chi^2$ klein, schwacher Zusammenhang

**Spezialfall für 2x2-Tafel**

| | | |
|---|---|---|
| $a$ | $b$ | $a+b$ |
| $c$ | $d$ | $c+d$ |
| $a+c$ | $b+d$ | |

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

**Kontigenzkoeffizient**

$$K := \sqrt{\frac{\chi^2}{n + \chi^2}},\ K \in \left[0, \sqrt{\frac{M-1}{M}}\right],\ M = min\{k, m\}$$

**Korrigierter Koeffizient**

$$K^* := \frac{K}{\sqrt{(M-1)/M}},\ K^* \in [0,1]$$

## 3 ZVs, Verteilungen und Häufigkeiten

### 3.1 Zufallsvariable

$$X: \Omega \to T_X,\ T_X \subseteq \mathbb{R}$$

- ordnet jedem $\omega \in \Omega$ genau ein $x \in T_X$ zu: $X(\omega) = x$
- Mehrere Elementarereignisse können selben Zahlenwert zugeordnet sein
- Ausprägungen/Relisierungen: $x = X(\omega)$ von X

### 3.2 Diskreter Erwartungswert

$$E(X) := \sum_{x \in T_X} x \cdot P(X = x) = \sum_{\omega \in \Omega} P(\{\omega\}) X(\omega)$$
$$= \sum_{x \in T_X} x \cdot f_X(x)$$

### 3.3 Stetiger Erwartungswert

$$E(X) := \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

### 3.4 Eigenschaften Erwartungswert

- X=a mit Wahrscheinlichkeit 1 (determinische ZV)

$$E(X) = a$$

- Linearität

$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$$
$$E(aX + b) = a \cdot E(X) + b$$

- symmetrisch um Punkt c: $f(c-x) = f(c+x) \forall x \in T_X$, dann $E(X) = c$
- für bel. $a_1, ..., a_n \in \mathbb{R}$ und bel. ZV $X_1, ..., X_n$

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \cdot E(X_i)$$

### 3.5 Transformationsregel Erwartungswert

Für reelle Funktion Y=g(X):

$$E(Y) = E[g(X)] = \begin{cases} \sum_{x \in T} g(x) f(x) & \text{X diskret} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{X stetig} \end{cases}$$

### 3.6 Varianz ZV

$$Var(X) := E[(X - E(X))^2]$$
$$= \begin{cases} \sum_{x \in T_X} (x - E(X))^2 f_X(x), & \text{X diskret} \\ \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx, & \text{X stetig} \end{cases}$$

#### Eigenschaften Varianz

- Einfachere Berechnung mit Varianz:

$$Var(X) = E(X^2) - (E(X))^2$$

- $Var(aX + b) = a^2 Var(X) \forall a, b \in \mathbb{R}$
- unabhängige ZV X, Y: $Var(X + Y) = Var(X) + Var(Y)$

# 4 Wichtige parametrische Verteilungen

### 4.0.1 Binomialverteilung

Modelliert Anzahl Erfolge mit fester Anzahl Versuchen

$$P(X = x) = f(x) = \binom{n}{k} \cdot \pi^x (1-\pi)^{n-x}, \ x \in T$$

$$X \sim \mathscr{B}(n, \pi), \ n \in \mathbb{N}, \ \pi \in [0,1]$$

### 4.0.2 Normalverteilung

modelliert kontinuierliche Werte um Mittelwert mit sym. Verteilung (Körpergröße, Gewicht)
Träger $T = \mathbb{R}$ mit Parametern $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), x \in \mathbb{R}$$

$\mu = 0$ und $\sigma^2 = 1$ ist standardnormalverteilt.

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

- Verschieben und skalieren

$$X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow (a + bX) \sim \mathcal{N}(a + b\mu, b^2\sigma^2)$$

- Summen und Differezen

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \Rightarrow \sum_i X_i \sim \mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$$

- Integral der Normalverteilungsdichte nicht analytisch machbar

# 5 Schätzung und Grenzwertsätze

## 5.1 Standadisierte ZV

$$\tilde{X} = \frac{X - \mu_X}{\sigma_X}$$
$$E(\tilde{X}) = (E(X) - \mu_X) = 0$$
$$Var(\tilde{X}) = \frac{1}{\sigma_X^2} Var(X) = 1$$

## 5.2 Standadisierung summierter ZVn

$$E(Y_n) = n \cdot \mu_X, \ Var(Y_n) = n \cdot \sigma_X^2$$
$$Z_n = \frac{Y_n - n\mu_X}{\sqrt{n}\sigma_X} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu_X}{\sigma_X}$$

## 5.3 Zentraler Grenzwertsatz

$$Fn(z) \xrightarrow{n \to \infty} F_{N(0,1)}(z) \ \forall z \in \mathscr{R}$$

$$Z_n \sim \mathcal{N}(\mu = 0, \sigma^2 = 1) \text{ a } :\cong \text{ asymptotisch}$$

Standadisierung:

$$Y_n \sim \mathcal{N}(\mu = n\mu_X, \sigma^2 = n\sigma_X^2)$$

Speziell:

$$\overline{X}n \sim \mathcal{N}(\mu = \mu_X, \sigma^2 = \frac{\sigma_X^2}{n}$$