1 Mathe Formeln Statistik und Wahrscheinlichkeitsrechnung

Jörn Horstmann, 26.10.2003. Alle Angaben ohne Gewähr.
http://www.ba-stuttgart.de/~wi02172/

1.1 Grundlagen

- Einzelklassen $[a_i; b_i[$
- Klassenweite $b_i - a_i = w_i$
- Klassenmitte $\tilde{x}_i = \frac{a_i+b_i}{2}$
- Mittelwert

$$\bar{x} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot n_i}_{\text{gruppierte Werte}} = \sum_{i=1}^{n} x_i \cdot h_i \approx \underbrace{\sum_{i=1}^{n} \tilde{x}_i \cdot h_i}_{\text{klassierte Daten}}$$

- Absolute Häufigkeit n_i
- Relative Häufigkeit $h_i = \frac{n_i}{n}$
- Kumulierte Häufigkeit $H_i = \sum h_j$
- Median, analog für Q_1 $(0,25)$ und Q_3 $(0,75)$

$$z = a_i + \frac{0,5 - H_{i-1}}{h_i} \cdot w_i$$

- Modalklasse (klassierte Daten) $\frac{h_i}{w_i} = max$
- Geometrisches Mittel (bei Wachstumsraten)

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n}$$

- Histogramm: Fläche des Rechtecks proportional zu h_i $h_i^* = \frac{h_i}{w_i}$

1.2 Streuungsmaße

$$\begin{aligned}
R &= \text{Spannbreite} = x_{max} - x_{min} \\
Q &= \text{Quartilsabstand} = Q_3 - Q_1 \\
d_m &= \frac{1}{n}\sum |x_i - m| \\
d_z &= \frac{1}{n}\sum |x_i - z| \\
d_{\bar{x}} &= \frac{1}{n}\sum |x_i - \bar{x}| \\
\sigma^2 &= \text{Varianz} = \frac{1}{n}\sum (x_i - \bar{x})^2 \\
&= \frac{1}{n}\sum x_i^2 - \bar{x}^2 = \sum x_i^2 h_i - \bar{x}^2 \\
\sigma &= \text{Standardabweichung} = \sqrt{\sigma^2} \\
V &= \text{Variationskoeffizient} = \frac{\sigma}{x}
\end{aligned}$$

- Der Gini-Koeffizient

$$\begin{aligned}
D_G &= \frac{\sum (H_{k-1} + H_k) \cdot n_k \cdot \bar{x}_k}{n \cdot \bar{x}} - 1 \quad \text{(grup. Daten)} \\
&\approx \frac{\sum (H_{i-1} + H_i) \cdot n_i \cdot \tilde{x}_i}{n \cdot \tilde{x}} - 1 \quad \text{(klass. Daten)}
\end{aligned}$$

$$\begin{aligned} & D_G = 0 && \text{absolute Gleichverteilung} \\ 0 < \; & D_G < 0,4 && \text{geringe Konzentration} \\ 0,4 \leq \; & D_G < 0,6 && \text{mittlere Konzentration} \\ 0,6 \leq \; & D_G < 1 && \text{hohe Konzentration} \\ & D_G = 1 && \text{absolute Konzentration} \end{aligned}$$

- Die Lorenz-Kurve

$$\begin{aligned} c_i &= \frac{n_i \cdot \bar{x}_i}{n \cdot \bar{x}} \approx \frac{n_i \cdot \tilde{x}_i}{n \cdot \tilde{x}} \qquad \text{Anteile der Merkmalssumme} \\ C_i &= \sum_{j=1}^{i} c_j \end{aligned}$$

$$\text{Schutzkoeffizient} \quad S = max\,(S_i) \quad S_i = H_i - C_i$$

$$P\,(H_i; C_i) \text{ ist ein Punkt auf der Lorenzkurve}$$

$$D_G = 2 \cdot \text{Fläche zw. Winkelhalbierender und Lorenzkurve}$$

1.3 Korrelation und Regression

- Bedingte Häufigkeiten

$$h\,(x_i|y_j \text{ fest}) = h_{i|j} = \frac{h_{ij}}{h_{\cdot j}} = \frac{n_{ij}}{n_{\cdot j}}$$

- Kovarianz $Cov(x, y) = \sigma_{xy}$

$$\begin{aligned} \sigma_{xy} &= \frac{1}{n} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) \\ &= \frac{1}{n} \sum x_i \cdot y_i - \bar{x} \cdot \bar{y} \end{aligned}$$

  x, y sind unabhängig $\Rightarrow Cov(x, y) = 0$
  $Cov(x, y) \neq 0 \Rightarrow x, y$ sind abhängig
  x, y sind unahängig wenn für alle i, j gilt: $n_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$

- Korrelationskoeffizient

$$\begin{aligned} r &= \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \\ \sigma_{xy} &= r \cdot \sigma_x \cdot \sigma_y \end{aligned}$$

- Regressionsgerade
  Linearer Zusammenhang zw. x und y, Suche einer Geraden $\tilde{y} = a + bx$ nach der Methode der kleinsten Quadrate

$$f = \sum (a + bx_i - y_i)^2$$

  f wird am kleinsten wenn

$$\begin{aligned} b &= \frac{\sum (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{\sigma_{xy}}{\sigma_x^2} \\ a &= \bar{y} - b\bar{x} \end{aligned}$$

1.4 Indizes

- Index nach Laspeyres

$$\begin{aligned} \text{Preisindex } I_{0,t}^L(p) &= \frac{\sum p_t^i \cdot q_0^i}{\sum p_0^i \cdot q_0^i} \cdot 100 \\ \text{Mengenindex } I_{0,t}^L(q) &= \frac{\sum p_0^i \cdot q_t^i}{\sum p_0^i \cdot q_0^i} \cdot 100 \end{aligned}$$

- Index nach Paasche

$$\begin{aligned}\text{Preisindex } I^P_{0,t}(p) &= \frac{\sum p^i_t \cdot q^i_t}{\sum p^i_0 \cdot q^i_t} \cdot 100 \\ \text{Mengenindex } I^P_{0,t}(q) &= \frac{\sum p^i_t \cdot q^i_t}{\sum p^i_t \cdot q^i_0} \cdot 100\end{aligned}$$

- Index nach Fischer

$$\begin{aligned}\text{Preisindex } I^F_{0,t}(p) &= \sqrt{I^L_{0,t}(p) \cdot I^P_{0,t}(p)} \\ \text{Mengenindex } I^F_{0,t}(q) &= \sqrt{I^L_{0,t}(q) \cdot I^P_{0,t}(q)}\end{aligned}$$

1.5 Wahrscheinlichkeiten

- Klassische Definition der Wahrscheinlichkeit (Laplace)
  nur bei Zufallsexperimenten, deren Ergebnisse gleiche Wahrscheinlichkeit haben

$$P(A) = \frac{\text{Anzahl der zu } A \text{ gehörenden Ereignisse}}{\text{Anzahl aller Ereignisse}} = \frac{|A|}{|G|}$$

- Bedinge Wahrscheinlichkeit (Wahrscheinlichkeit von A wenn B eingetreten ist)

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

- Additionssatz

$$\begin{aligned}P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A \setminus B) + P(A \cap B) + P(B \setminus A)\end{aligned}$$

- Spezieller Additionssatz wenn $A \cap B = \{\}$

$$P(A \cup B) = P(A) + P(B)$$

- Multiplikationssatz

$$\begin{aligned}P(A \cap B) &= P(A|B) \cdot P(B) \\ P(A \cap B) &= P(A) \cdot P(B|A)\end{aligned}$$

- Satz der totalen Wahrscheinlichkeit
  $A_1, A_2, A_3, \ldots, A_n$ seien disjunkte Ereignisse mit $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n = G$

$$\begin{aligned}B &= (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup \ldots \cup (B \cap A_n) \\ P(B) &= \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)\end{aligned}$$

- Satz von Bayes
  $A_1, A_2, A_3, \ldots, A_n$ seien disjunkte Ereignisse mit $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n = G$

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i) \cdot P(B|A_i)}{P(B)}$$

1.6 Kombinatorik

	ohne Wiederholung	mit Wiederholung
Permutation	verschiedene Elemente $$P(n) = n!$$	gruppenweise identische Elemente $n_1, n_2, n_3, \ldots$ $$P(n_1, n_2, \ldots) = \frac{(\sum n_i)!}{n_1! \cdot n_2! \cdot \ldots}$$
Variation (Reihenfolge wichtig)	m Elemente nacheinander aus n Elementen ziehen. $$V(m,n) = \frac{n!}{(n-m)!}$$	n^m
Kombination (Reihenfolge unwichtig)	gleichzeitiges Ziehen (mit einem Griff) m aus n $$C(m,n) = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$	$\binom{n+m-1}{m}$

1.7 Zufallsverteilungen

1.7.1 Diskrete ZV

$$\begin{aligned} E(X) &= \sum f(x_i) \cdot x_i \\ Var(x) &= \sum f(x_i) \cdot x_i^2 - \mu^2 \end{aligned}$$

- Diskrete Gleichverteilung (Laplace)
  x nimmt n verschiedene Werte an
  $$f(x = x_i) = \frac{1}{n}$$
- Binomialverteilung (Bernoulli - Experiment)
  Ein Zufallsexperiment mit zwei möglichen Ausgängen wird n-mal wiederholt, p ist Wahrscheinlichkeit des günstigen Ausgangs und k ist die Anzahl der günstigen Ausgänge
  $$\begin{aligned} P(A) &= p \\ P(\bar{A}) &= 1 - p = q \\ P(X = k) &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ E(X) &= n \cdot p \\ Var(X) &= n \cdot p \cdot (1-p) \end{aligned}$$
- Poisson-Verteilung
  Approximation der Binomialverteilung $B(n,p)$ wenn $n > 50$ und $p < 0,05$
  $$\begin{aligned} P(X = k) &= \frac{\mu^k \cdot e^{-\mu}}{k!} \\ &= P(k-1) \cdot \frac{\mu}{k} \end{aligned}$$
- Hypergeometrische Verteilung
  Eine Urne enthält N Elemente, davon M günstige. Es werden n Elemente ohne zurücklegen (mit

einem Griff) gezogen. X = Anzahl der dabei gezogenen günstigen Elemente.

$$\begin{aligned} P(X=k) &= \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \\ E(X) &= n \cdot \frac{M}{N} \\ Var(X) &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1} \end{aligned}$$

1.7.2 Stetige ZV

$$\begin{aligned} E(X) &= \int f(x)\, x\, dx \\ Var(x) &= \int f(x)\, x^2\, dx - \mu^2 \end{aligned}$$

- Normalverteilung

$$P(X=k) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Standardnormalverteilung, Verteilungsfunktion Φ

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} \\ E(Z) &= 0 \\ Var(Z) &= 1 \\ P(X \leq a) &= P\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right) \\ \Phi(-Z) &= 1 - \Phi(Z) \\ P(\mu - c \leq X \leq \mu + c) &= 2 \cdot \Phi\left(\frac{c}{\sigma}\right) - 1 \end{aligned}$$

- Lineare Transformation

$$\begin{aligned} Y &= aX + b \\ E(Y) &= a + E(X) \\ Var(Y) &= b^2 \cdot Var(X) \end{aligned}$$

- X, Y unabhängig

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) \\ Var(X+Y) &= Var(X) + Var(Y) \end{aligned}$$

- x_i ist $N(\mu, \sigma^2)$ für $i = 1 \ldots n$

$$\begin{aligned} \bar{x} &= \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad \text{ist ebenfalls normalverteilt} \\ E(\bar{x}) &= \mu \\ Var(\bar{x}) &= \frac{\sigma^2}{n} \end{aligned}$$

1.8 Hypothesentest

- Fehler 1. Art: Die richtige Hypothese wird abgelehnt
- Fehler 2. Art: Die falsche Hypothese wird angenommen