Formelsammlung – Statistik

1 Empirische Verteilungsfunktion

Empirische Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \textit{für}\, x < a_1 \\ \sum_{i=1}^{j} f_i & \textit{für}\, a_j \leq x \,\textit{und}\, a_{j+1} > x \\ 1 & \textit{für}\, x \geq a_k \end{cases}$$

Stetige empirische Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \textit{für} \quad x \leq g_0 \\ F(g_{i-1}) + \frac{x - g_{i-1}}{d_i} * f_i & \textit{für}\, g_{i-1} < x \leq g_i \\ 1 & \textit{für} \quad x \geq g_k \end{cases}$$

2 Statistische Lagemaße

Arithmetisches Mittel

a) bei unklassierten Daten

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

b) bei klassierten Daten

$$\bar{x}_g = \sum_{i=1}^{k} m_i * f_i$$

Median

a) bei einer ungeraden Anzahl von Werten

$$x_{med} = x_{(\frac{n+1}{2})}$$

b) bei einer geraden Anzahl von Werten

$$x_{med} = \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)})$$

Perzentilwerte

a) ergibt (n * p) keinen ganzzahligen Wert, ist k die auf (n * p) folgende ganze Zahl

$$x_p = x_{(k)}$$

b) ergibt (n * p) einen ganzzahligen Wert, entspricht k dem Ergebnis von (n * p)

$$x_p = \frac{1}{2}(x_{(k)} + x_{(k+1)})$$

Modus

$$x_{\text{mod}} = a_{x\,\text{max}}$$

(nur interpretierbar, wenn die Häufigkeitsverteilung ein eindeutiges Maximum besitzt)

Geometrisches Mittel

$$\bar{x}_{geom} = \sqrt[n]{x_1 * ... * x_n}$$

Harmonisches Mittel

$$\bar{x}_{har} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

3 Streuungsmaße / Dispersionsparameter

Spannweite

$$d_s = x_{\max} - x_{\min}$$

Interquartilsabstand (IQR)

$$IQR = x_{(0,75)} - x_{(0,25)}$$

Empirische Varianz

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

Standardabweichung

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

Variationskoeffizient

$$v = \frac{s}{\bar{x}}$$

(nur berechenbar, wenn das arithmetische Mittel positiv ausfällt)

Fünf-Werte-Zusammenfassung

$$\left[x_{\min}; x_{0,25}; x_{med}; x_{0,75}; x_{\max}\right]$$

4 Verteilungsmaße

Momentenkoeffizient der Schiefe

$$g_m = \frac{m_3}{s^3}$$

$$m_3 = \frac{1}{n} \sum_{i=1}^{n} (x_n - \bar{x})^3$$

$$s^3 = \left(\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \right)^3$$

Quartilskoeffizient der Schiefe

$$g_{0,25} = \frac{(x_{0,75} - x_{med}) - (x_{med} - x_{0,25})}{x_{0,75} - x_{0,25}}$$

Kurtosis / Exzeß

$$g_k = \frac{m_4}{s^4} - 3$$

$$m_4 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^4$$

$$s^4 = \left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^4$$

5 Zusammenhangsmaße

Bravais-Pearson-Korrelationskoeffizient

$$r = \frac{\sum_{i=1}^{n}(x_i * y_i) - n * \bar{x} * \bar{y}}{\sqrt{\sum_{i=1}^{n}(x_i^2) - n * \bar{x}^2} * \sqrt{\sum_{i=1}^{n}(y_i^2) - n * \bar{y}^2}}$$

Konkordanzkoeffizient nach Kendall

$$tau = \frac{2 * (K - D)}{n * (n-1)}$$

Rangkorrelationskoeffizient nach Spearman

$$rho = 1 - \frac{6 * \sum d_i^2}{(n^2 - 1) * n}$$

6 Lineare Regression

Berechnung des Regressionskoeffizienten

$$b = \frac{\sum_{i=1}^{n}(x_i * y_i) - n * \bar{x} * \bar{y}}{\sum_{i=1}^{n}(x_i^2) - n * \bar{x}^2}$$

Berechnung des konstanten Glieds

$$a = \bar{y} - b * \bar{x}$$

Bestimmtheitsmaß der Regressionsfunktion

$$R^2 = \frac{ESS}{TSS}$$

TSS = Total Sum of Squares

ESS = Explained Sum of Squares

7 Regeln für das Rechnen mit Mengen

Kommutativgesetz

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

Assoziativgesetz

$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$

Distributivgesetz

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

De Morgansche Regel

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$
$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

8 Die drei Axiome von Kolmogoroff

Axiom I: Die Wahrscheinlichkeit des Eintretens eines Ereignisses A eines Zufallsvorgangs ist stets eine nichtnegative reelle Zahl.

$$P(A) \geq 0$$

Axiom II: Die Wahrscheinlichkeiten aller möglichen Elementar-ereignisse eines Zufallsvorgangs ergeben zusammen den Wert 1.

$$P(\Omega) = 1$$

Axiom III: Axiom 3: Die Wahrscheinlichkeit der Vereinigungsmenge zweier oder mehrerer Ereignisse eines Zufallsvorgangs berechnet sich aus der Summe der Einzelwahrscheinlichkeiten der Ereignisse, wenn diese paarweise disjunkt sind.

$$P(A \cup B) = P(A) + P(B) \text{ falls } P(A \cap B) = \varnothing$$

9 Rechnen mit Wahrscheinlichkeiten

Klassische Wahrscheinlichkeitsdefinition
(nach Piere-Simon Maquis de Laplace)

$$P(A) = \frac{\text{Summe für A günstiger Elementarereignisse}}{\text{Summe aller möglichen Elementarereignisse}}$$

Additionssatz für unvereinbare Ereignisse

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

Additionssatz für beliebige Ereignisse

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

Multiplikationssatz bei stochastischer Unabhängigkeit

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

Satz der totalen Wahrscheinlichkeit

$$P(B)=\sum_{i=1}^{k}P(B\mid A_i)*P(A_i)$$

Satz von Bayes

$$P(A_i\mid B)=\frac{P(B\mid A_i)*P(A_i)}{P(B)}$$

10 Kombinatorik

Wie viele Möglichkeiten existieren für die Auswahl von k aus n Elementen?

Variation (Reihenfolge spielt eine Rolle) im Modell ohne Zurücklegen

$\frac{n!}{(n-k)!}$ (im Sonderfall der Permutation - Auswahl aller Elemente: $n!$)

Variation (Reihenfolge spielt eine Rolle) im Modell mit Zurücklegen

n^k

Kombination (Reihenfolge spielt keine Rolle) im Modell ohne Zurücklegen

$$\frac{n!}{k!*(n-k)!}$$

Kombination (Reihenfolge spielt keine Rolle) im Modell mit Zurücklegen

$\frac{(n+k-1)!}{(n-1)!*k!}$,

11 Rechnen mit Zufallsvariablen

Wahrscheinlichkeitsfunktion

$$P(X=x)=f(x_i)$$

Verteilungsfunktion

$$P(X\le g)=\sum_{x_i\le g}f(x_i)=F(g)$$

Erwartungswert einer Zufallsvariablen

$$E(X)=\sum_{i\ge 1}x_i*p_i$$

Varianz einer Zufallsvariablen

$$Var(X)=\sum_{i\ge i}(x_i-E(X))^2*p_i$$

12 Konfidenzintervalle um den Erwartungswert

Konfidenzintervall um μ bei bekanntem σ

$$P(\bar{x}-z_{(1-\frac{\alpha}{2})}*\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{x}+z_{(1-\frac{\alpha}{2})}*\frac{\sigma}{\sqrt{n}})=1-\alpha$$

$z_{(1-\frac{\alpha}{2})}$ bei 1 – α = 0,95 und df = 1: 1,96

$z_{(1-\frac{\alpha}{2})}$ bei 1 – α = 0,99 und df = 1: 2,58

Konfidenzintervall um μ bei unbekanntem σ

$$P(\bar{x} - t_{(1-\frac{\alpha}{2};n-1)} * \frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{(1-\frac{\alpha}{2};n-1)} * \frac{s}{\sqrt{n-1}}) = 1 - \alpha$$

$t_{(1-\frac{\alpha}{2};n-1)}$ bei 1 – α = 0,95 und df = 19: 2,093

$t_{(1-\frac{\alpha}{2};n-1)}$ bei 1 – α = 0,99 und df = 19: 2,861

13 Chi²-Test auf stochastische Unabhängigkeit

Tabellierte Quantile der Chi-Quadrat-Verteilung nach ausgewählten Wahrscheinlichkeiten und Freiheitsgraden

df/p	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75

14 Bestimmung der minimalen Stichproben-Größe nach Cochran

$$n = \frac{\frac{Z^2 * p * q}{e^2}}{1 + \frac{\frac{Z^2 * p * q}{e^2} - 1}{N}}$$

mit:

n = Stichprobenumfang
N = Größe der Grundgesamtheit
e = Breite des Konfidenzintervalls
p = Stichprobenanteil (falls unbekannt: 0,5)

q = (1-p)
Z = Z-Wert aus der Standard-normalverteilung für die avisierte Sicherheit des Konfidenzintervalls