

Häufigkeitsverteilungen

1.1 Empirische Verteilungsfunktion

Die empirische Verteilungsfunktion beschreibt anhand von Werten einer Stichprobe, welcher Anteil dieser Werte kleiner oder gleich einem bestimmten Wert x ist.

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

1.2 Stetige empirische Verteilungsfunktion

Die stetige empirische Verteilungsfunktion ist eine kontinuierliche Variante der empirischen Verteilungsfunktion, bei der die kumulierten relativen Häufigkeiten zwischen den geordneten Stichprobenwerten linear interpoliert werden.

$$F(x) = \begin{cases} 0 & \text{für } x \leq g_0, \\ F(g_{i-1}) + \frac{x - g_{i-1}}{d_i} f_i & \text{für } g_{i-1} < x \leq g_i, \\ 1 & \text{für } x \geq g_k. \end{cases} \quad (2)$$

Lagemaße

Statistische Lagemaße sind Kennzahlen, die die zentrale Lage bzw. den typischen Wert („Mittelwert“, „Durchschnitt“) einer Verteilung beschreiben.

2.1 Arithmetisches Mittel

2.1.1 Arithm. Mittel bei unklassierten Daten

Das arithmetische Mittel ist der Durchschnitt einer Menge von Zahlen. Es wird berechnet, indem alle Werte addiert und die Summe durch die Anzahl der Werte geteilt werden.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

2.1.2 Arithm. Mittel bei klassierten Daten

Das arithmetische Mittel bei klassierten Daten ist ein Näherungswert für den Durchschnitt, bei dem jede Klasse durch ihre Klassenmitte und den relativen Anteil der Werte dieser Klasse an der Gesamtverteilung repräsentiert wird.

$$\bar{x}_g = \sum_{i=1}^k (m_i f_i) \quad (4)$$

2.2 Median

Der Median ist derjenige Wert, der genau in der Mitte einer geordneten Verteilung liegt. 50% der Werte einer Verteilung sind kleiner oder gleich dem Median, 50% der Werte einer Verteilung sind größer oder gleich dem Median.

2.2.1 Median bei ungerader Anzahl von Werten

$$x_{med} = x_{(\frac{n+1}{2})} \quad (5)$$

2.2.2 Median bei gerader Anzahl von Werten

$$x_{med} = \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) \quad (6)$$

2.3 Perzentilwerte

Perzentilwerte sind Lagekennzahlen, die eine geordnete Datenverteilung in beliebig definierbare Segmente unterteilen. Das p -Perzentil ist der Wert, unterhalb dessen oder auf dem p % aller Werte liegen.

2.3.1 Perzentil bei nicht ganzzahligem np

(hier ist k die nächste auf np folgende ganze Zahl)

$$x_p = x_k \quad (7)$$

2.3.2 Perzentil bei ganzzahligem np

(hier entspricht k dem Ergebnis von np)

$$x_p = \frac{1}{2} (x_k + x_{(k+1)}) \quad (8)$$

2.4 Modus

Der Modus ist der Wert, der in einer Datenmenge am häufigsten vorkommt. Er ist nur interpretierbar, wenn die Verteilung ein eindeutiges Maximum (unimodal) besitzt.

$$x_{mod} = a_{x\ max} \quad (9)$$

2.5 Geometrisches Mittel

Das geometrische Mittel wird berechnet, indem man das Produkt aller Werte bildet und daraus die n -te Wurzel zieht. Es eignet sich besonders für Wachstumsraten oder multiplikative Zusammenhänge.

$$\bar{x}_{geom} = \sqrt[n]{x_1 \cdot \dots \cdot x_n} \quad (10)$$

2.6 Harmonisches Mittel

Das harmonische Mittel eignet sich vor allem zur Mittelung von Verhältniszahlen wie Geschwindigkeiten oder Raten.

$$\bar{x}_{har} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (11)$$

3 Streuungsmaße

Streuungsmaße sind statistische Kennzahlen, die beschreiben, wie stark (oder schwach) die Werte einer Datenreihe um ein Lagemaß variieren (streuen).

3.1 Spannweite

Die Spannweite ist definiert als die Differenz zwischen dem größten und dem kleinsten Wert einer Datenreihe.

$$d_s = x_{max} - x_{min} \quad (12)$$

3.2 Interquartilsabstand

Der Interquartilsabstand entspricht der Spannweite der mittleren 50% einer Verteilung. Er wird als Differenz zwischen dem oberen und dem unteren Quartilswert berechnet.

$$IQR = x_{0,75} - x_{0,25} \quad (13)$$

3.3 Empirische Varianz

Die empirische Varianz ist die durchschnittliche quadratische Abweichung der Werte einer Verteilung vom arithmetischen Mittel dieser Verteilung. Bei der empirischen Varianz wird durch n , bei der Stichprobenvarianz durch $n - 1$ geteilt.

$$s^2 = \left\{ \frac{1}{n} \sum_{i=1}^n x_i^2 \right\} - \bar{x}^2 \quad (14)$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (15)$$

3.4 Stichprobenvarianz

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (16)$$

3.5 Standardabweichung

Die Standardabweichung ist die positive Wurzel der Varianz. Sie liefert einen nicht-quadrierten Wert für die Streuung, der einfacher als die Varianz zu interpretieren ist.

$$s = +\sqrt{s^2} \quad (17)$$

3.6 Variationskoeffizient

Der Variationskoeffizient ist ein relatives Streuungsmaß und gibt das Verhältnis der Standardabweichung zum arithmetischen Mittel an. Er ermöglicht den Vergleich der Streuung von Verteilungen, die in unterschiedlichen Einheiten erfasst wurden, und kann nur berechnet werden, wenn ein positives arithmetisches Mittel vorliegt.

$$v = \frac{s}{\bar{x}} \quad (18)$$

3.7 Fünf-Werte-Zusammenfassung

Die Fünf-Werte-Zusammenfassung besteht aus dem Minimum und dem Maximum einer Verteilung sowie den drei Quartilswerten. Es kann als Sechs-Werte-Zusammenfassung durch das arithmetische Mittel ergänzt werden.

$$[x_{min}; x_{0,25}; x_{med}; x_{0,75}; x_{max}] \quad (19)$$

4 Verteilungsmaße

Verteilungsmaße geben Auskunft darüber, wie eine Verteilung optisch aufgebaut ist: flach oder steil bzw. symmetrisch oder asymmetrisch.

4.1 Momentenkoeffizient der Schiefe

Der Momentenkoeffizient der Schiefe ist ein dimensionsloses Maß für die Symmetrie einer Verteilung. Ein positiver Wert zeigt Rechtsschiefe, ein negativer Wert Linksschiefe an, Verteilungen mit einem Wert nahe 0 sind symmetrisch aufgebaut.

$$g_m = \frac{m_3}{s^3} \quad (20)$$

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (21)$$

$$s^3 = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^3 \quad (22)$$

4.2 Quartilkoeffizient der Schiefe

Der Quartilkoeffizient der Schiefe ist ein robustes Maß für die Asymmetrie einer Verteilung, das auf den Quartilen basiert. Er vergleicht die Abstände zwischen Median und oberem bzw. unterem Quartil und ist weniger empfindlich gegenüber Ausreißern als momentbasierte Schiefe Maße.

$$g_{0,25} = \frac{(x_{0,75} - x_{med}) - (x_{med} - x_{0,25})}{x_{0,75} - x_{0,25}} \quad (23)$$

4.3 Kurtosis / Exzeß

Die Kurtosis ist ein dimensionsloses Maß für die Form einer Verteilung im Vergleich zur Normalverteilung. Bei einer Kurtosis nahe 0 entspricht die Wölbung der Verteilung der einer Normalverteilung, positive Werte zeigen eine steilere, negative Werte eine flachere Wölbung an. Eine hohe Kurtosis weist zudem auf eine stärkere Ausreißerneigung hin.

$$g_k = \frac{m_4}{s^4} - 3 \quad (24)$$

$$m_4 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^4 \quad (25)$$

$$s^4 = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^4 \quad (26)$$

5 Zusammenhangsmaße

Zusammenhangsmaße bzw. Korrelationskoeffizienten sind statistische Kennzahlen, die Stärke und Richtung des Zusammenhangs zwischen zwei Variablen beschreiben. Je nach Skalenniveau und Art des gesuchten Zusammenhangs werden unterschiedliche Maße verwendet.

5.1 Bravais-Pearson-Korrelationskoeffizient

Der Bravais-Pearson-Korrelationskoeffizient misst die Stärke der linearen Korrelation zwischen zwei metrischen Variablen. Er nimmt Werte zwischen -1 und $+1$ an, wobei -1 für einen perfekten negativen, $+1$ für einen perfekten positiven und Werte nahe 0 für keinen linearen Zusammenhang stehen.

$$r = \frac{\sum_{i=1}^n (x_i y_i) - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n (x_i^2) - n \bar{x}^2} \sqrt{\sum_{i=1}^n (y_i^2) - n \bar{y}^2}} \quad (27)$$

5.2 Konkordanzkoeffizient nach Kendall

Der Konkordanzkoeffizient nach Kendall misst die Stärke einer monotonen Korrelation zwischen zwei Variablen anhand der Übereinstimmung der Rangreihen. Er nimmt Werte zwischen -1 und $+1$ an und ist auch für ordinale Daten geeignet.

$$\tau = \frac{2(K - D)}{n(n - 1)} \quad (28)$$

5.3 Rangkorrelationskoeffizient nach Spearman

Der Rangkorrelationskoeffizient nach Spearman misst die Stärke einer monotonen Korrelation zwischen zwei Variablen anhand ihrer Rangplätze. Auch er eignet sich für ordinale Daten und nimmt Werte zwischen -1 und $+1$ an.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (29)$$

6 Lineare Regression

Die univariate lineare Regressionsanalyse untersucht den linearen Zusammenhang zwischen einer abhängigen und einer unabhängigen Variablen. Ziel ist es, diejenige Gerade zu ermitteln, die die abhängige Variable möglichst gut durch die unabhängige Variable erklärt bzw. vorhersagt.

6.1 Aufstellung der Regressionsfunktion

6.1.1 Regressionskoeffizient

Der Regressionskoeffizient gibt die geschätzte Änderung der abhängigen Variable an, wenn die unabhängige Variable um eine Einheit steigt. Er entspricht der Steigung der Regressionsgeraden.

$$b = \frac{\sum_{i=1}^n (x_i y_i) - n \bar{x} \bar{y}}{\sum_{i=1}^n (x_i^2) - n \bar{x}^2} \quad (30)$$

6.1.2 Konstantes Glied

Das konstante Glied ist der geschätzte Wert der abhängigen Variablen, wenn die unabhängige Variable den Wert 0 annimmt. Es entspricht dem y-Achsenschnittpunkt der Regressionsgeraden.

$$a = \bar{y} - b \bar{x} \quad (31)$$

6.1.3 Regressionsfunktion

$$y = a + bx \quad (32)$$

6.2 Überprüfung der Modellgüte

Das Bestimmtheitsmaß R^2 gibt an, welcher Anteil der Streuung der abhängigen Variable durch das lineare Regressionsmodell erklärt wird. Es liegt zwischen 0 und 1 ; je höher der Wert, desto besser passt das Modell zu den Daten. R^2 entspricht dem quadrierten Bravais-Pearson-Koeffizienten.

$$R^2 = \frac{ESS}{TSS} \quad (33)$$

7 Mengenlehre

7.1 Kommutativgesetz

Das Kommutativgesetz besagt, dass bei bestimmten Rechenoperationen die Reihenfolge der Operanden vertauscht werden kann, ohne dass sich das Ergebnis ändert.

$$A \cap B = B \cap A \quad (34)$$

$$A \cup B = B \cup A \quad (35)$$

7.2 Assoziativgesetz

Das Assoziativgesetz besagt, dass bei bestimmten Rechenoperationen die Klammerung der Operanden verändert werden kann, ohne dass sich das Ergebnis ändert.

$$(A \cap B) \cap C = A \cap (B \cap C) \quad (36)$$

$$(A \cup B) \cup C = A \cup (B \cup C) \quad (37)$$

7.3 Distributivgesetz

Das Distributivgesetz beschreibt, wie Klammern aufgelöst werden können, indem eine Operation über eine andere Operation verteilt wird.

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad (38)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C) \quad (39)$$

7.4 De Morgansche Regeln

Die De Morganschen Regeln besagen, dass das Komplement der Vereinigungsmenge der Mengen A und B der Schnittmenge aus dem Komplement der Menge A und dem Komplement der Menge B entspricht - und umgekehrt. Sie ermöglichen damit das „Auflösen“ von Negationen.

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B} \quad (40)$$

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B} \quad (41)$$

8 Axiome von Kolmogoroff

Die Axiome von Kolmogoroff sind die grundlegenden Regeln der Wahrscheinlichkeitstheorie. Sie besagen, dass Wahrscheinlichkeiten nichtnegativ sind, die Wahrscheinlichkeit des sicheren Ereignisses 1 beträgt und sich die Wahrscheinlichkeiten paarweise disjunkter Ereignisse addieren.

1. Erstes Axiom von Kolmogoroff
 $P(A) \geq 0$
2. Zweites Axiom von Kolmogoroff
 $P(\Omega) = 1$
3. Drittes Axiom von Kolmogoroff
 $P(A \cup B) = P(A) + P(B)$
falls $P(A \cap B) = \emptyset$

9 Wahrscheinlichkeitslehre

9.1 Laplace-Wahrscheinlichkeit

Die Laplace-Wahrscheinlichkeit beschreibt die Wahrscheinlichkeit eines Ereignisses bei einem Zufallsexperiment mit endlich vielen und gleich wahrscheinlichen Ergebnissen. Sie wird als Anzahl der günstigen Ereignisse geteilt durch die Anzahl aller möglichen Ereignisse berechnet.

$$P(A) = \frac{\sum \text{zu } A \text{ passende Elementarereignisse}}{\sum \text{denkbare Elementarereignisse}} \quad (42)$$

9.2 Additionssatz für unvereinbare Ereignisse

Der Additionssatz für unvereinbare Ereignisse besagt, dass sich die Wahrscheinlichkeit der Vereinigung zweier disjunkter Ereignisse durch Addition ihrer Einzelwahrscheinlichkeiten ergibt.

$$P(A \cup B) = P(A) + P(B) \quad (43)$$

9.3 Additionssatz für beliebige Ereignisse

Der Additionssatz für beliebige Ereignisse besagt, dass die Wahrscheinlichkeit der Vereinigung zweier Ereignisse der Summe ihrer Einzelwahrscheinlichkeiten minus der Wahrscheinlichkeit ihres gemeinsamen Auftretens entspricht.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (44)$$

9.4 Multiplikationssatz bei stochastischer Unabhängigkeit

Der Multiplikationssatz bei stochastischer Unabhängigkeit besagt, dass die Wahrscheinlichkeit des gemeinsamen Eintretens unabhängiger Ereignisse dem Produkt ihrer Einzelwahrscheinlichkeiten entspricht.

$$P(A \cap B) = P(A)P(B) \quad (45)$$

9.5 Satz der totalen Wahrscheinlichkeit

Mit dem Satz der totalen Wahrscheinlichkeit lässt sich die Wahrscheinlichkeit des Eintretens eines Ereignisses aus bedingten Wahrscheinlichkeiten und den Wahrscheinlichkeiten bedingender Ereignisse bestimmen.

$$P(B) = \sum_{i=1}^k P(B|A_i)P(A_i) \quad (46)$$

9.6 Satz von Bayes

Der Satz von Bayes ermöglicht die Berechnung der Eintrittswahrscheinlichkeit eines Ereignisses A unter der Bedingung, dass zuvor ein anderes Ereignis B eingetreten ist.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} \quad (47)$$

10 Kombinatorik

Die Kombinatorik befasst sich mit dem Anordnen und Auswählen von Elementen nach bestimmten Regeln. Sie untersucht u.a., wie viele Möglichkeiten es gibt, Elemente miteinander zu kombinieren oder in eine Reihenfolge zu bringen.

10.1 Variation ohne Zurücklegen

Die Variation ohne Zurücklegen beschreibt die Anzahl geordneter Auswahlen von k Elementen aus n verschiedenen Elementen, wobei jedes Element höchstens einmal gewählt werden darf. Die Reihenfolge der ausgewählten Elemente spielt dabei eine Rolle.

$$\frac{n!}{(n-k)!} \quad (48)$$

10.2 Permutation

Die Permutation ist ein Sonderfall der Variation ohne Zurücklegen, der dann eintritt, wenn k und n identisch sind.

$$n! \quad (49)$$

10.3 Variation mit Zurücklegen

Die Variation mit Zurücklegen beschreibt die Anzahl geordneter Auswahlen von k Elementen aus n Elementen, wobei jedes Element mehrfach gewählt werden kann. Die Reihenfolge der ausgewählten Elemente spielt dabei eine Rolle.

$$n^k \quad (50)$$

10.4 Kombination ohne Zurücklegen

Eine Kombination ohne Zurücklegen ist eine ungeordnete Auswahl von k Elementen aus n verschiedenen Elementen, bei der jedes Element höchstens einmal gewählt werden darf. Die Reihenfolge der Auswahl spielt dabei keine Rolle.

$$\frac{n!}{k!(n-k)!} \quad (51)$$

10.5 Kombination mit Zurücklegen

Eine Kombination mit Zurücklegen ist eine ungeordnete Auswahl von k Elementen aus n verschiedenen Elementen, bei der Elemente mehrfach ausgewählt werden dürfen. Die Reihenfolge spielt dabei keine Rolle.

$$\frac{(n+k-1)!}{(n-1)!k!} \quad (52)$$

11 Rechnen mit Zufallsvariablen

11.1 Wahrscheinlichkeitsfunktion

$$P(X = x_i) = f(x_i) \quad (53)$$

11.2 Verteilungsfunktion

$$P(X \leq g) = \sum_{x_i \leq g} f(x_i) = F(g) \quad (54)$$

11.3 Erwartungswert einer Zufallsvariablen

$$E(X) = \sum_{i \geq 1} x_i p_i \quad (55)$$

11.4 Varianz einer Zufallsvariablen

$$Var(X) = \sum_{i \geq 1} (x_i - E(X))^2 p_i \quad (56)$$

12 Konfidenzintervalle

Ein Konfidenzintervall ist ein aus Stichprobendaten berechneter Wertebereich (Vertrauensbereich), der einen unbekannten Populationsparameter mit einer festgelegten Vertrauenswahrscheinlichkeit einschließt.

12.1 Konfidenzintervall um μ bei bekanntem σ

$$P(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha \quad (57)$$

$z_{1-\frac{\alpha}{2}}$ bei $1 - \alpha = 0,95$ und $df = 1 : 1,96$
 $z_{1-\frac{\alpha}{2}}$ bei $1 - \alpha = 0,99$ und $df = 1 : 2,58$

12.2 Konfidenzintervall um μ bei unbekanntem σ

$$P(\bar{x} - t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n-1}}) = 1 - \alpha \quad (58)$$

$t_{1-\frac{\alpha}{2}; n-1}$ bei $1 - \alpha = 0,95$ und $df = 1 : 2,093$
 $t_{1-\frac{\alpha}{2}; n-1}$ bei $1 - \alpha = 0,99$ und $df = 1 : 2,861$

13 Chi-Quadrat-Test auf stochastische Unabhängigkeit

Der Chi-Quadrat-Test auf stochastische Unabhängigkeit prüft, ob zwei kategoriale Variablen voneinander abhängig sind. Er vergleicht hierfür die tatsächlich beobachteten Häufigkeiten mit den Werten, die man bei völliger Unabhängigkeit der Variablen voneinander erwarten würde.

Tabellierte Quantile der Chi-Quadrat-Verteilung (Auszug):

df/p	0,9	0,95	0,975	0,99	0,995
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,81	9,35	11,34	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75

14 Minimale Stichprobengröße nach Cochran

Die Formel nach Cochran dient zur Bestimmung der minimal erforderlichen Stichprobengröße, um einen Anteilswert mit vorgegebener Genauigkeit und Sicherheit zu schätzen. Dabei ist z der z -Wert des Konfidenzniveaus, p der erwartete Anteilswert und e die gewünschte Fehlertoleranz.

$$n = \frac{\frac{Z^2 pq}{e^2}}{1 + \frac{\frac{Z^2 pq}{e^2} - 1}{N}} \quad (59)$$

15 Übersicht der Verfahrensrobustheiten

Robust
Modus, Median, Quartile, Quantile, Perzentile, Interquartilsabstand, Quartilkoeffizient der Schiefe, Konkordanzkoeffizient nach Kendall, Rangkorrelationskoeffizient nach Spearman
Nicht robust
Arithmetisches Mittel, Geometrisches Mittel, Harmonisches Mittel, Spannweite, Varianz, Standardabweichung, Variationskoeffizient, Fünf-Werte-Zusammenfassung, Kurtosis/Exzeß, Momentenkoeffizient der Schiefe, χ^2 -Test, Bravais-Pearson-Korrelationskoeffizient, Univariate Lineare Regression

16 Übersicht der Mindestskalenniveaus

Nominalskalenniveau
Modus, Venn-Diagramme, χ^2 -Test
Ordinalskalenniveau
Median, Quartile, Quantile, Perzentile, Interquartilsabstand, Fünf-Werte-Zusammenfassung, Quartilkoeffizient der Schiefe, Stem-and-Leaf-Diagramme, Box-Plots, Rangkorrelationskoeffizient nach Spearman, Konkordanzkoeffizient nach Kendall
Kardinalskalenniveau
Arithmetisches Mittel, Geometrisches Mittel, Harmonisches Mittel, Spannweite, Varianz, Standardabweichung, Variationskoeffizient, Momentenkoeffizient der Schiefe, Kurtosis/Exzeß, Bravais-Pearson-Korrelationskoeffizient, Univariate Lineare Regression

Diese Formelsammlung wurde von Christian Reinboth mit L^AT_EX auf Basis eines Templates von Jean Amadeus Elsner für die Grundlagenvorlesung Statistik im berufsbegleitenden Bachelorstudiengang Betriebswirtschaftslehre am Fachbereich Wirtschaftswissenschaften der Hochschule Harz erstellt. Sie ist unter der Lizenz CC BY 4.0 zur weiteren Nutzung freigegeben.

<https://creativecommons.org/licenses/by/4.0/deed.de>

