Empirische Verteilungen

Wilfried Mann,
Mettmann

Verteilungen dienen der übersichtlichen Darstellung und Beurteilung von statistischem Material sowie zum Vergleich mit theoretischen Verteilungen, um qualitative Aussagen zur Grundgesamtheit zu treffen. Basis ist eine Zufallsstichprobe für kategorial skalierte oder stetige Variablen, für die zuvor Intervalle gebildet wurden.

1 Arten von Verteilungen

Absolute Häufigkeiten

Ausgangsdaten für die Bestimmung absoluter Häufigkeiten ist die geordnete Urliste einer Zufallsstichprobe. Jede Zahl einer Ausprägung (Variable) wird gezählt, wobei gleiche Werte zusammengefasst werden. Somit erhält man eine Übersicht darüber, wie oft eine Zahl vorkommt. Absolute Häufigkeiten werden mit »H« bezeichnet. Bei stetigen Variablen mit sehr vielen unterschiedlichen Werten müssen Klassen gebildet werden, die sich aus dem Gesamtintervall aller Werte und den zu bildenden Intervallen der Klassen ergeben. Klassen sollten i.d.R. gleich lang und mindestens mit 5 Werten besetzt sein.

Eine Faustformel lautet: Klassenanzahl (m) = $\sqrt{n} + 3$; mit n = Anzahl der Versuche.

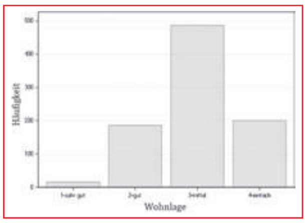
Relative Häufigkeiten

Relative Häufigkeiten (h) werden aus der absoluten Häufigkeit (H) einer Ausprägung (A) abgeleitet und allgemein berechnet nach

$$\text{relative Häufigkeit } h_n = \frac{\text{absolute Häufigkeit } H}{\text{Anzahl der Versuche } n}$$

Sie hat nur Werte zwischen 0 und 1.

Mit den relativen Häufigkeiten lassen sich Wahrscheinlichkeiten anhand von Beobachtungen bestimmen. Einem Ereignis A wird die Wahrscheinlichkeit P(A) zugeordnet. Absolute oder relative Häufigkeiten werden graphisch in Diagrammen dargestellt. Diskrete Variablen z.B. in Balken-, Linien- oder Kreisdiagrammen, stetige Variablen in Histogrammen oder auch als Polygonzug.



Balkendiagramm der Wohnlagen

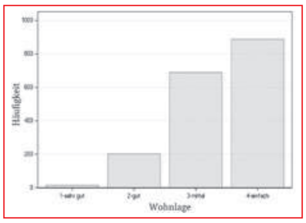


Diagramm zur Summenhäufigkeit der Wohnlagen

Summenhäufigkeiten

Absolute oder relative Häufigkeiten können aufsummiert werden (Summenhäufigkeiten oder kumulierte Verteilungen). Die Summe dieser relativen Häufigkeiten (F) ist immer gleich 1.

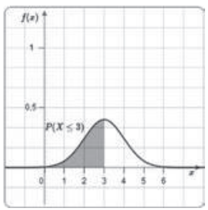
Häufigkeits- und Summenhäufigkeitsfunktion

Häufigkeiten und Summenhäufigkeiten lassen sich auch funktional darstellen. Hierbei wird die Dichtefunktion $f(x)$ aus relativen Häufigkeiten abgeleitet (auch Häufigkeitsfunktion, Zähldichte und Wahrscheinlichkeitsfunktion). Aus kumulierten Häufigkeiten F(x) ergibt sich die Summenhäufigkeitsfunktion bzw. Verteilungsfunktion.

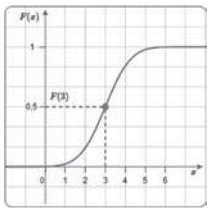
Eigenschaften der Dichtefunktion	Eigenschaften einer Verteilungsfunktion
• Die Dichtefunktion nimmt nur positive Werte an. $f(x) \geq 0$ für alle $x \in \mathbb{R}$ • Die Fläche unter der Dichtefunktion ist 1. $\int_{-\infty}^{\infty} f(x)\,dx = 1$	• $F(x)$ ist monoton steigend. • $F(x)$ ist rechtsseitig stetig. • $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to+\infty} F(x) = 1$

2 Wahrscheinlichkeitsverteilungen

Eine **Wahrscheinlichkeitsverteilung** gibt an, wie sich die Wahrscheinlichkeiten auf die möglichen Werte einer Zufallsvariable verteilen. Diese wird aus der Dichtefunktion oder der Verteilungsfunktion abgeleitet. Ein Beispiel: Die Wahrscheinlichkeit $P(X \leq 3)$ ist gleich der Fläche zwischen der Dichtefunktion f, der x-Achse und der Senkrechten bei x=3 und kann auch dem Funktionswert der Verteilungsfunktion an der Stelle x=3 entnommen werden.



Dichtefunktion



Verteilungsfunktion

3 Anwendung

Mit Hilfe von geeigneten Standard-Softwareprodukten lassen sich ohne vertiefte Kenntnisse von Formelzusammenhängen **Häufigkeiten** bestimmen und graphisch darstellen. Ergänzend neben Dichte- und Verteilungsfunktionen, mit den Kenngrößen arithmetischer Mittelwert (($\bar{x}$)) und der Standardabweichung (s), auch der Boxplot mit den Lagemerkmalen Median und den Quartilen. Bei metrischen Variablen werden die notwendigen Klassen (Intervalle) nach statistischen Gesichtspunkten automatisch gebildet.

Bei Kaufpreisanalysen kann es dennoch sinnvoll sein, Gruppen nach sachlogischen bzw. wertbestimmenden Kriterien zu bilden, um diese kategorial zu analysieren; z.B. Wohnflächen bei Wohnungseigentum mit 20 bis 40 m^2 (Appartements), 41 bis 80 m^2, 81 bis 120 m^2 und 121 bis 200 m^2 (Großwohnungen).

Neben der übersichtlichen Darstellung zur Beurteilung des statistischen Basismaterials dienen die empirischen Häufigkeitsverteilungen zur Ableitung von Wahrscheinlichkeiten und zum Vergleich mit theoretischen Verteilungen. Dieser Vergleich hat dann Bedeutung bei Signifikanz-Tests zur Bestimmung der Güte von statistischen Aussagen in Bezug auf die Grundgesamtheit.