

Inhalt

Einführung

Datenerhebung und Messung

**Wahrscheinlichkeitsrechnung: Grundlagen
und Definitionen**

Zufallsvariablen, Verteilungen & Häufigkeiten

**Stochastische Unabhängigkeit und
Zusammenhangsmaße für diskrete Merkmale**

Statistische Grafiken

Kennwerte & Verteilungseigenschaften

Wichtige parametrische Verteilungen

Schätzung & Grenzwertsätze

**Zufallsvektoren und multivariate
Verteilungen**

**Zusammenhangsmaße für metrische
Merkmale**

Korrelation und Kausalität

Statistik	Aufgaben	Techniken
Deskriptive Statistik	Beschreibung, graphische Darstellung und Validierung von Daten. ! Keine Rückschlüsse auf Grundgesamtheit möglich.	Grafiken, Tabellen, Kennzahlen
Explorative Statistik	Suche nach Struktur in den Daten (ohne stochastische Methoden). Formulierung von Hypothesen für das den Daten zugrunde liegende stochastische Modell.	Iterative und interaktive Anwendung von Techniken aus der deskriptiven und induktiven Statistik.
Induktive Statistik	Ziehung von Schlüssen von den Daten (Stichprobe) auf Grundgesamtheit. Basierend auf stochastischen Modellen.	Statistische Modellierung, statistische Tests, Konfidenzintervalle, Schätzer

Datenerhebung & Messung

Matrikelnummer	Name	Vorname	Geburtsdatum	Hauptfach	Nebenfach
xxxxx 234	Muster	Peter	01.01.2001	Statistik	Informatik
xxxxx 556	Schmid	Lena	31.10.2002	Informatik	Statistik
xxxxx 123	Müller	Jonas	23.08.1999	Mathematik	NA
xxxxx 767	Nguyen	Cho	24.12.2000	Medizin	Soziologie
xxxxx 111	Nagel	Cosima	26.10.1996	Jura	Ethik

- Grundgesamtheit:** Studenten der LMU (über welche "Objekte" erhebe ich Daten?)
- Stichprobe:** z.B. Alle Statistik Studenten **!** Stichproben müssen nicht per Definition zufällig gewählt sein.
- Statistische Einheit/ Untersuchungseinheit (UE):** Ein Student bzw. ein Element der Grundgesamtheit
- Merkmal:** Messbare Eigenschaft einer statistischen Einheit. In der Tabelle quasi der (sinnvolle) Spaltenname. z.B. Hauptfach ist ein Merkmal
- Merkmalsausprägung:** Der tatsächliche Wert des Merkmals bei einer statistischen Einheit. In der Tabelle ist das ein Wert in einer Zelle.
- Beobachtung:** Alle Merkmalsausprägungen einer statistischen Einheit zu einem Zeitpunkt. In der Tabelle sind das die Werte in einer Zeile.

Unterscheidung nach		
... Quantifizierbarkeit der Ausprägungen	Qualitative Merkmale: <ul style="list-style-type: none"> nur zuordenbar (einstufig) Beispiele: Wohnort, Name 	Quantitative Merkmale: <ul style="list-style-type: none"> mess- oder zählbar Beispiele: Alter, Körpergröße
... Anzahl der Ausprägungen*	Diskrete Merkmale: <ul style="list-style-type: none"> höchstens abzählbar unendlich viele mögliche Ausprägungen Beispiele: Gehaltsklassen, Kaufverhalten 	Stetige Merkmale: <ul style="list-style-type: none"> überabzählbar unendlich viele mögliche Ausprägungen Beispiele: Geschwindigkeit, Gewicht
... Direktheit der Informationsgewinnung	Beobachtbare Merkmale: <ul style="list-style-type: none"> können direkt erhoben werden Beispiel: Abiturnote 	Latente Merkmale: <ul style="list-style-type: none"> Operationalisierung über Indikatoren/Items notwendig Beispiele: Bildungsgrad, Kreativität, Nutzen

- (*) Merkmale die eigentlich diskret sind, aber so viele Ausprägungen haben, dass sie wie stetige Merkmale behandelt werden können, nennt man auch quasi-stetig (z.B. Einkommen)
- (*) Stetige Merkmale können durch Klassenbildung in diskrete Merkmale umgewandelt werden.

Skalenniveaus

Skalenniveau	Beispiele	Erlaubte Transformationen um Strukturen zu erhalten	natürliche Ordnung	sinnvolle Abstände	natürliche Null	natürliche Einheit	Berechenbare Kennzahlen
Nominalskala	Wohnort, Farbe	Bijektionen	×	×	×	×	Mode
Ordinal - Rangskala	Noten, Michelin-Sterne Platzierung bei Sportevent	str. monoton steig. Abb.	✓	×	×	×	Median
Intervallskala	Temperatur in °C Jahreszahlen	affin lin. str. mon. steig. Abb.	✓	✓	×	×	Arithm. Mittel
Verhältnisskala	Preis, Länge, Gewicht, Temp. in K°	lineare str. mon. steig. Abb.	✓	✓	✓	×	Geom. Mittel Harm. Mittel
Absolutskala	Häufigkeit, Anzahl, Prozentpunkte	Identität	✓	✓	✓	✓	Alle

Datenerhebung

Methoden:

Beobachtung

Datengewinnung durch Erfassen von ungesteuerten Sachverhalten

Befragung

Fragebögen für mündliche/schriftliche/online Umfrage.

Experiment

Erzeugung der Daten durch Simulation von Situationen.

Umfang:

Vollerhebung

Alle stat. Einheiten einer GG werden untersucht.

Stichprobe (Teilerhebung)

Ein Teil der UE in einer GG wird untersucht.

Datenform:

Querschnittsdaten

Eine Beobachtung pro UE.

- Noten, Aktivitäten, Geschlecht, können zu bestimmtem Zeitpunkt von UE erhoben werden und z.B. mittels Regression auf Zusammenhänge untersucht werden.

Zeitreihe

Mehrere Beobachtungen einer UE.

- Temperatur, Wind & Luftfeuchtigkeit werden in regelmäßigen Abständen gemessen um Prognosen über die zeitliche Entwicklung der UE 'Wetter' zu machen

Längsschnittsdaten

Mehrere Beobachtungen mehrerer UE.

- Kohortenstudien in Medizin
- Mikrozensus

Wahrscheinlichkeitsrechnung

Elementarereignisse

Für eine Grundmenge Ω wird die ein-elementige Teilmenge $\{\omega\} \subseteq \Omega$ als Elementarereignis bezeichnet

Ereignisse

Für eine Grundmenge Ω wird $A \subseteq \Omega$ als Ereignis bezeichnet.

Laplace - Wahrscheinlichkeit

Für eine abzählbare Grundmenge Ω und ein Ereignis $A \subseteq \Omega$ ist die Laplace-Wahrscheinlichkeit $P(A) := \frac{|A|}{|\Omega|}$

Wahrscheinlichkeitsverteilung (Axiome von Kolmogorov) (vereinfacht)

Sei Ω eine Grundmenge und P ein Funktion auf $P(\Omega)$. P heißt Wahrscheinlichkeitsverteilung oder Wahrscheinlichkeitsmaß auf Ω , wenn sie folgende Eigenschaft erfüllt: 1) $P(\Omega) = 1$ 2) $\forall A \subseteq \Omega: P(A) \geq 0$ 3) $\forall A, B \subseteq \Omega: A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit von A gegeben B für Ereignisse $A, B \subseteq \Omega$ mit $P(B) > 0$ ist $P(A|B) := \frac{P(A \cap B)}{P(B)}$

Satz

$$P\left[\bigcap_{i=1}^n A_i\right] = \prod_{i=1}^n P[A_i | \bigcap_{j=1}^{i-1} A_j] = P[A_1] \cdot P[A_2 | A_1] \cdot P[A_3 | A_1 \cap A_2] \cdot \dots \cdot P[A_n | A_1 \cap \dots \cap A_{n-1}] \quad (\text{Multiplikationssatz})$$

Sei $(A_i)_{i \in I}$ eine disjunkte Zerlegung von Ω . D.h. $\Omega = \bigcup_{i \in I} A_i$. Dann gilt für beliebiges B

$$P[B] = \sum_{i: P(A_i) > 0} P[B|A_i] \cdot P[A_i] \quad (\text{Satz von totaler Wahrscheinlichkeit})$$

$$\text{Spezialfall: } P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$$

Satz von Bayes

$(A_i)_{i \in I}$ sei so, dass $\Omega = \bigcup_{i \in I} A_i$. B sei so, dass $P[B] \neq 0$.

$$\text{Dann ist } P[A_i | B] = \frac{P[B|A_i] \cdot P[A_i]}{\sum_j P[B|A_j] \cdot P[A_j]} \quad \text{Spezialfall: } P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}$$

Folgerungen

Korollar

$$P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

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

$$B \subseteq A \Rightarrow P(B) \leq P(A)$$

$$P\left[\bigcup_{i=1}^n A_i\right] = \sum_{i=1}^n P[A_i] - \sum_{1 \leq i < j \leq n} P[A_i \cap A_j] + \sum_{1 \leq i < j < k \leq n} P[A_i \cap A_j \cap A_k] - \dots + (-1)^{n+1} P\left[\bigcap_{i=1}^n A_i\right]$$

← Siebformel von Sylvester-Poincaré

$$\text{Spezialfall: } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Kombinatorik	mit Wiederholung/ mit Zurücklegen	ohne Wiederholung/ ohne Zurücklegen
Anzahl Kombinationen ohne Reihenfolge	$\binom{n}{m}$	$\binom{n+m-1}{m}$
Anzahl Kombinationen mit Reihenfolge	$\frac{n!}{(n-m)!}$	n^m
Anzahl Permutationen	$n!$	$\frac{n!}{n_1! \cdot \dots \cdot n_k!}$

Wettverhältnis (Odds - Update)

$$\underbrace{\frac{P(B|A)}{P(B|A^c)}}_{\text{a-posteriori Verhältnis}} = \underbrace{\frac{P(A|B)}{P(A|B^c)}}_{\text{Faktor der neuen Information}} \cdot \underbrace{\frac{P(B)}{P(B^c)}}_{\text{n-priori-Verhältnis}}$$