MAS3 Skript

SEbaBB 2024

10. Dezember 2024, Hagenberg

Inhaltsverzeichnis

1 Einführung

Folgende Teilbereiche der Statistik werden behandelt:

- Deskriptive Statistik
- Induktive Statistik
- Wahrscheinlichkeitsrechnung

1.1 Übersicht Deskriptive und Induktive Statistik

Die Deskriptive Statistik, auch beschreibende Statistik beschäftigt sich mit der Beschreibung von Daten. Man versucht dabei Daten mihilfe von Lage- und Streuungsmaßen in einer verständlichen Form darzustellen.

Manchmal hat man keinen Zugriff zu allen Daten, dann bedient man sich der Induktiven Statistik, auch schließende Statistik. Mithilfe von Stichproben versuchtman Schlussfolgerungen auf die Grundgesamtheit zu ziehen.

Folgende Bereiche werden behandelt:

- Lageparameter
- Streuungsparameter
- Lineare Transformationen
- Zusammenhangsmaße

1.2 Übersicht Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeitsrechnung beschäftigt sich mit der Quantifizierung von Zufäll. Zufälle sind Ereignisse, die nicht vollständig vorhersehbar sind. Die Wahrscheinlichkeitsrechnung modeliert diese Zufälle und ermöglicht es, Aussagen über die Wahrscheinlichkeit von Ereignissen zu treffen.

Folgende Bereiche werden behandelt:

- Laplace'sche Annahme
- Kombinatorik
- Axiome der Wahrscheinlichkeit
- Unaabhängigkeit
- Randwahrscheinlichkeit
- Diskrete und stetige Verteilungen
- Verteilungen

2 Statistik

2.1 Grundbegiffe

Begriff	Definition
Statistische Masse	Menge von statistischen Einheiten Die statistische Masse muss vor jeder statistischen Tätigkeit in räumlicher, zeitlicher und sachlicher Hinsicht präzise definiert werden.
Statistische Einheit	Objekt, das Merkmale besitzt.
Qualitatives Merkmal	Merkmal, das nicht gemessen werden kann. (z.B. Geschlecht, Farbe)
Quantitatives Merkmal	Merkmal, das gemessen werden kann und eine Einheit besitzt. (z.B. Größe, Gewicht)
Primärstatistik	Daten werden eigens für die statistische Zwecke erhoben.
Sekundärstatistik	Daten werden für andere Zwecke erhoben und für statistische Zwecke verwendet.
Voll/Totalerhebung	Alle statistischen Einheiten der Masse werden erfasst.
Teilerhebung/ Stichprobe	Nur ein Teil der s. Einheiten wird erfasst.
Zufallsstichprobe	s. Einheiten werden zufällig gewählt.
Repräsentative Stichprobe	s. Einheiten werden so gewählt, dass ihre Merkmale mit der gleicher relativen Häufigkeit (*Quoten*), wie in der Gesamtmasse vorkommen.
Urliste	Liste, die direkt bei der Datenerhebung entsteht.
Stab/Balkendiagramm	Diagramm, bei dem die Häufigkeit der Merkmalsausprägung durch die Länge der Stäbe/Balken dargestellt wird.
Klasse	Intervall, in dem die Merkmalsausprägungen zusammengefasst werden. (z.B. 0-10 Jahre, 11-20 Jahre...)
Kumulierte Häufigkeit	Summe der Häufigkeiten von einer Schranke bis zu einem bestimmten Wert. Nur bei qualitativ ordinalen oder quantitativen Merkmalen

Tabelle 1: Grundbegriffe 1

Begriff	Definition
Empirische Verteilung	Verteilung der Merkmalsausprägungen in einer Stichprobe. Gibt die kumulierte Häufigkeit der Merkmalsausprägungen an. Die dafür Formel ist: $F(x) = \frac{Anzahl mit Merkmal \leq x}{Anzahl alle Elemente}$
Lage	Lageparameter helfen das Zentrum der Verteilung zu bestimmen. (Modus, Median, Arithmetisches Mittel)
Streuung	Streuungsparameter helfen zu ermitteln wie stark die Werte um den Durchschnitt verteilt sind.
Schiefe	Schiefe der Verteilung gibt an, ob die Verteilung symmetrisch ist, oder ob sie nach links oder rechts neigt.
Wölbung	Wölbung der Verteilung gibt an, ob die Verteilung spitz oder flach ist, das heißt ob die Werte um den Durchschnitt geballt sind oder ob sie gleichmäßig verteilt sind
Zusammenhang	Zusammenhangsmaße geben an, ob und wie stark zwei Merkmale miteinander zusammenhängen.

Tabelle 2: Grundbegriffe 2

2.2 Notation

Zeichen	Bedeutung
N	Anzahl der Elemente der Grundgesamtheit
n	Anzahl der Elemente der Stichprobe
μ	Durchschnitt der Grundgesamtheit
$\bar{x}$	Durchschnitt der Stichprobe
σ	Standardabweichung der Grundgesamtheit
s	Standardabweichung der Stichprobe
q_n	n-tes Quantil

Tabelle 3: Notation

2.3 Lageparameter

2.4 Streuungsparameter

2.5 Lineare Transformationen

2.6 Zusammenhangsmaße

3 Wahrscheinlichkeitsrechnung

3.1 Laplace'sche Annahme

3.2 Kombinatorik

3.3 Axiome der Wahrscheinlichkeit

3.4 Unabhängigkeit

3.5 Bedingte und Unbedingte Wahrscheinlichkeiten

3.6 Diskrete und stetige Verteilungen

Es gibt zwei Hauptarten von Verteilungen:

- Diskrete Verteilungen
- Stetige Verteilungen

3.6.1 Diskrete Verteilungen

In einer diskreten Verteilung können nur bestimmte abzählbare Werte angenommen werden. Die Werte sind oft ganzzahlig, zum Beispiel die Anzahl von Würfen bei einem Würfelspiel oder die Anzahl von Menschen in einer Schlange.

Jedes Ereignis hat eine Wahrscheinlichkeit zwischen 0 und 1. Die Summe der Wahrscheinlichkeiten aller Ereignisse ist 1.

Die Wahrscheinlichkeitsfunktion $f(x)$ gibt die Wahrscheinlichkeit an, dass ein Ereignis x eintritt.

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

Die Verteilungsfunktion $F(x)$ gibt die Wahrscheinlichkeit an, dass ein Ereignis kleiner oder gleich x ist. Sie wird durch die Summe der Wahrscheinlichkeiten aller Ereignisse bis x berechnet.

$$F(x) = P(X \leq x) = \sum_{i=1}^{x} f(x_i)$$

Der Erwartungswert μ gibt den Durchschnitt der Verteilung an.

$$\mu = E(x) = \sum_{i=1}^{n} x_i * f(x_i)$$

Beispiele für diskrete Verteilungen sind die diskrete Gleichverteilung, die Hypergeometrische Verteilung, die Binomialverteilung und die Poissonverteilung.

3.6.2 Stetige Verteilungen

In einer stetigen Verteilung können unendlich viele Werte angenommen werden. Die Werte sind also nicht abzählbar. Das heißt aber nicht das die Werte nicht begrenzt sind, sondern dass es unendlich viele Werte zwischen zwei Werten gibt. Beispiele hierfür sind Körpergröße, Gewicht oder Temperatur.

Die Wahrscheinlichkeitsdichte $f(a, b)$ gibt die Wahrscheinlichkeit an, dass ein Ereignis in einem Intervall $[a, b]$ eintritt.

$$f(x) = P(a \leq X \leq b)$$

Die Wahrscheinlichkeit für ein genaues Ereignis ist 0, da es nur Wahrscheinlichkeiten für Intervalle gibt.

$$P(X = x) = 0$$

Die Verteilungsfunktion $F(x)$ gibt die Wahrscheinlichkeit an, dass ein Ereignis kleiner oder gleich x ist. Sie wird durch die Integration der Wahrscheinlichkeitsdichte von $-\infty$ bis x berechnet.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Das Integral der Wahrscheinlichkeitsdichte über den gesamten Wertebereich ist 1.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Der Erwartungswert μ gibt den Durchschnitt der Verteilung an.

$$\mu = E(x) = \int_{-\infty}^{\infty} x * f(x)dx$$

Die Varianz σ^2 gibt die Streuung der Verteilung an. Sie gibt die Verteilung der Werte um den Erwartungswert an.

$$\sigma^2 = Var(x) = E(x^2) - E(x)^2$$

Ein Beispiel für eine stetige Verteilung ist die Normalverteilung.

3.7 Verschiebungssatz

Der Verschiebungssatz besagt, dass der Erwartungswert einer linearen Transformation einer Zufallsvariablen gleich der linearen Transformation des Erwartungswerts ist.

$$E(aX + b) = aE(X) + b$$

3.8 Verteilungen

Verteilungen sind mathematische Modelle, die die Wahrscheinlichkeit von Ereignissen beschreiben.

3.8.1 Wie wähle ich die richtige Verteilung aus?

Verteilung	Anwendungsbereich
	Diskret
Diskrete Gleichverteilung	Zufällige Auswahl von Elementen
Hypergeometrische Verteilung	Ziehen ohne Zurücklegen
Binomialverteiltung	Ziehen mit Zurücklegen
Poissonverteilung	Anzahl von Ereignissen in einem Zeitintervall
	Stetig
Normalverteilung	Messwerte

Tabelle 4: Übersicht zu Verteilungen

Für die Klausur unbedingt einen Taschenrechner mitnehmen der die Verteilungen berechnen kann, da die Formeln sehr komplex sind.

3.8.2 Diskrete Gleichverteilung

Jedes Element hat die gleiche Wahrscheinlichkeit ausgewählt zu werden, daher eignet sich die Gleichverteilung für die Zufällige Auswahl von Elementen.

Bezeichnung	Formel
Untere Schranke	a
Obere Schranke	b
Wahrscheinlichkeit	$P(X = x) = \frac{1}{b-a+1}$
Erwartungswert	$\mu = \frac{a+b}{2}$
Varianz	$\sigma^2 = \frac{(b-a+1)^2-1}{12}$

Tabelle 5: Diskrete Gleichverteilung

3.8.3 Hypergeometrische Verteilung

Die Hypergeometrische Verteilung eignet sich für das Ziehen ohne Zurücklegen. Also wenn die Wahrscheinlichkeit für ein Ereignis nicht konstant ist und von vorherigen Ereignissen beeinflusst wird.

Bezeichnung	Formel
Anzahl Elemente	N
Anzahl gewünschter Erfolge	M
Anzahl Ziehungen	n
Anzahl gezogener Erfolge	m
Wahrscheinlichkeit	$P(X = x) = \frac{\binom{M}{m} * \binom{N-M}{n-m}}{\binom{N}{n}}$
Erwartungswert	$\mu = \frac{n*M}{N}$
Varianz	$\sigma^2 = n * \frac{M}{N} * (1 - \frac{M}{N}) * \frac{N-n}{N-1}$

Tabelle 6: Hypergeometrische Verteilung

3.8.4 Binomialverteilung

Die Binomialverteilung eignet sich für das Ziehen mit Zurücklegen. Also wenn die Wahrscheinlichkeit für ein Ereignis konstant und unabhängig ist.

Bezeichnung	Formel
Anzahl Ziehungen	n
Anzahl Erfolge	k
Wahrscheinlichkeit	p
Wahrscheinlichkeit	$P(X = x) = \binom{n}{k} * p^k * (1 - p)^{n-k}$
Erwartungswert	$\mu = n * p$
Varianz	$\sigma^2 = n * p * (1 - p)$

Tabelle 7: Binomialverteilung

3.8.5 Poissonverteilung

Die Poissonverteilung eignet sich für die Anzahl von Ereignissen in einem Zeitintervall. Die Ereignisse treten unabhängig voneinander auf und die Wahrscheinlichkeit für ein Ereignis ist konstant.

Bezeichnung	Formel
Anzahl Ereignisse	k
Ereignisrate	λ
Wahrscheinlichkeit	$P(X = k) = \frac{\lambda^k}{k!} * e^{-\lambda}$
Erwartungswert	$\mu = \lambda$
Varianz	$\sigma^2 = \lambda$

Tabelle 8: Poissonverteilung

3.8.6 Normalverteilung

Die Normalverteilung ist stetig und symmetrisch. Sie beschreibt die kontinuierliche Verteilung von Messwerten.

Die Wahrscheinlichkeit für ein genaues Ereignis ist 0, da es nur Wahrscheinlichkeiten für Intervalle gibt.

Bezeichnung	Formel
Erwartungswert	μ
Standardabweichung	σ
Wahrscheinlichkeit	$P(a \leq X \leq b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$
Varianz	σ^2

Tabelle 9: Normalverteilung