Rheinische Hochschule Köln

Plattformmodul Statistische Grundlagen V 3.0

Prof. Dr. Jürgen Krob

Rheinische Hochschule Köln gGmbH

Sommersemester 2025

Rheinische
Hochschule
Köln

Einleitung

Lernergebnisse / Kompetenzen

Rheinische Hochschule Köln

Die Studierenden

- erhalten einen ganzheitlichen und breiten Einblick in die Methoden der deskriptiven Statistik.
- verstehen die grundlegenden Prinzipien der Wahrscheinlichkeitsrechnung.
- erlangen erste Kenntnisse über Methoden der Statistik für Anwendungen in Unternehmen, dem Qualitätsmanagement, der Marktforschung und der empirischen Datenanalyse.
- sind in der Lage, diese Methoden sachgemäß auf betriebswirtschaftliche bzw. medizinökonomische bzw. psychologische Fragestellungen anzuwenden.

Vorgehensweise

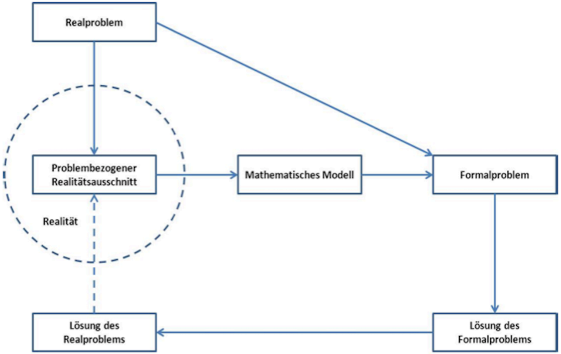

Abbildung: Vorgehensweise

Literatur

Rheinische Hochschule Köln

- Backhaus, K./ Erichson, B./ Plinke, W./ Weiber R. (2021). Multivariate Analysemethoden. Berlin.
- Eckstein, P. P. (2019): Statistik für Wirtschaftswissenschaftler, 6. Aufl., Wiesbaden.
- Sibbertsen, P./ Lehne, H. (2015): Statistik: Einführung für Wirtschafts- und Sozialwissenschaftler, 2. Aufl., Berlin.

ILIAS

Rheinische
Hochschule
Köln

Die seminaristische Lehrveranstaltung wird durch Einsatz der E-Learningplattform ILIAS der RH begleitet und unterstützt.

Abbildung: E-Learningplattform ILIAS

Einführung

Einführung

Rheinische Hochschule Köln

Statistik ist die Bezeichnung für die Gesamtheit von Verfahren und Methoden zur Gewinnung, Erfassung, Aufbereitung, Analyse, Abbildung, Nachbildung und Vorhersage von (möglichst) massenhaften, zähl-, mess- und/ oder systematisch beobachtbaren Daten über reale Sachverhalte zum Zwecke der Erkenntnisgewinnung und Entscheidungsfindung (meist unter Ungewissheit).

Rheinische Hochschule Köln

5 Stufen in statistischen Projekten/empirischen Studien:

1. Übertragung der materiellen Frage in eine statistische Frage/ Untersuchungsplanung
2. Gewinnung des Beobachtungsmaterials/ Datenerhebung
3. Aufbereitung und Darstellung des Beobachtungsmaterials/ Datenaufbereitung
   Die Aufbereitung und Darstellung kann in verschiedener Weise erfolgen:
   1. durch Gruppierung des Materials nach Merkmalsklassen und graphisch oder tabellarische Darstellung der daraus gewonnenen Häufigkeitsverteilung,
   2. durch Berechnung von beschreibenden Maßzahlen wie Mittelwert, Streuung, Abhängigkeitsmaße
4. Schluß von der Stichprobe auf die Gesamtheit/ Datenanalyse
   Hierbei benutzt man insbesondere Methoden zur
   1. Schätzung unbekannter Parameter,
   2. Angabe von Schätzbereichen für Parameter,
   3. Prüfung von Hypothesen über Parameter
5. Sachgerechte Interpretation der Ergebnisse/ Dateninterpretation

Deskriptive Statistik - Grundbegriffe

Grundbegriffe

Rheinische Hochschule Köln

Definition
Statistische Einheit, Merkmalsträger Eine statistische Einheit (Merkmalsträger) ist das kleinste Element in der Statistik. Eine statistische Einheit ist Träger von Informationen bzw. Eigenschaften, die für eine statistische Untersuchung von Interesse sind.

Definition
Statistische Gesamtheit Eine endliche Menge mit n Elementen wohl unterschiedener, sachlich, örtlich und zeitlich gleich abgegrenzter statistischer Einheiten heißt statistische Gesamtheit vom Umfang n.

Grundbegriffe

Rheinische Hochschule Köln

Definition

Statistisches Merkmal
Eine Eigenschaft einer statistischen Einheit, die Grundlage bzw. Gegenstand einer statistischen Untersuchung ist, heißt statistisches Merkmal.

Definition

Merkmalsausprägung
Eine Aussage über ein Merkmal bzw. über eine Eigenschaft einer statistischen Einheit heißt Merkmalsausprägung.

Grundbegriffe

Rheinische Hochschule Köln

Definition
Statistische Skala Eine relationstreue Abbildung von Merkmalsausprägungen eines Erhebungsmerkmals auf eine Zeichen- bzw. Zahlenmenge heißt statistische Skala.

Grundbegriffe

Skala					
Typ	Kategorial-		Kardinal-		
Name	Nominal-	Ordinal-	Absolut-	Verhältnis-	Intervall-
Operationen	Bildung von Häufigkeiten	Median, Quantile	= ≠ > < + - * /		= ≠ > < + -
Beispiel	Geschlecht	Prädikat	Anzahl	Umsatz	Temperatur
Merkmal					
Art	qualitativ		quantitativ		
Skalierung	nominal	ordinal	kardinal, metrisch		
Ausprägung	Kategorie		Wert		
	Begriff	Intensität	diskret	quasi-stetig	stetig
Beispiel	weiblich	sehr gut	20 Stück	1,2 Mio. €	$20{,}2^0$ Grad

Abbildung: Skalen und Merkmalsklassifikationen

Grundbegriffe

Rheinische Hochschule Köln

Definition

Datenerhebung
Für eine statistische Gesamtheit vom Umfang n heißt der Vorgang der Ermittlung von Ausprägungen A eines statistischen Merkmals Datenerhebung.

Definition

Urliste
Wird ein Merkmal statistisch erhoben, dann heißt die Zusammenstellung der Merkmalsausprägungen in der Reihenfolge ihrer statistischen Erhebung Urliste.

Deskriptive Statistik - Eindimensionale Verteilungen

Rheinische Hochschule Köln

Nominale Merkmale

Beispiel

Eine Befragung des Deutschen Instituts für Wirtschaftsforschung über Auswirkungen von Umweltschutzauflagen bei Kölner Unternehmen ergab folgende Tabelle:

Auflage, die zur stärksten Kostensteigerung führte	Zahl der Nennungen n_i
Luftreinhaltung (A_1)	2
Abfallbeseitigung (A_2)	5
Lärmschutz (A_3)	5
Gewässerschutz (A_4)	3

Eindimensionale Verteilungen

Rheinische Hochschule Köln

Häufigkeitstabelle:

Ausprägung A_i des Merkmals „Auflage, die zur stärksten Kosten-steigerung führte“	absolute Häufigkeit $n(A_i) = n_i$	relative Häufigkeit $h(A_i) = \frac{n_i}{n}$
A_1		
A_2		
A_3		
A_4		

Eindimensionale Verteilungen

Rheinische Hochschule Köln

Ordinal und quantitativ-diskrete Merkmale

Als quantitativ-diskretes Merkmal betrachten wir die Haushaltsgröße (= Anzahl der im Haushalt lebenden Personen) von privaten Haushalten.
In einer Untersuchung über die Einkaufsgewohnheiten im Internet wurden für 25 Haushalte neben dem verfügbaren Einkommen auch die Haushaltsgröße gemessen an
der Personenanzahl x_i, die die Merkmalsausprägungen 1, 2, 3, 4 usw. annehmen kann, erhoben (vgl. Beispiel 2.1 in Sibbertsen, P./ Lehne, H. (2015)).

Wie im Falle des qualitativen Merkmals erhält man durch Auszählen die absoluten Häufigkeiten für jede Merkmalsausprägung x_i, aus denen dann die relativen Häufigkeiten zu errechnen sind.

Eindimensionale Verteilungen

Beispiel 2.1. Haushaltseinkommen und -größe

In einer Untersuchung über die Konsumgewohnheiten wurden 25 Haushalte u.a. nach dem verfügbaren Einkommen des Haushalts (in €) und nach der Haushaltsgröße (= Anzahl der im Haushalt lebenden Personen) befragt. Die Tabelle 2.1 gibt die Urliste für die beiden Merkmale an.

i	Einkommen	HH-Größe	i	Einkommen	HH-Größe
1	1600	1	14	2150	3
2	2900	4	15	3200	4
3	3200	2	16	2500	3
4	4200	5	17	1800	2
5	2700	5	18	1600	1
6	2050	1	19	2700	3
7	3500	3	20	5100	2
8	2050	2	21	2600	3
9	2100	2	22	6250	4
10	2700	1	23	2500	2
11	2150	3	24	2500	3
12	1550	2	25	2200	2
13	1920	1			
$\sum$				67720	64

Tabelle 2.1: Urliste von Haushaltseinkommen und -größe

Abbildung: Beispiel 2.1. Haushaltseinkommen und -größe

Eindimensionale Verteilungen

Rheinische Hochschule Köln

Tabelle: Haushaltsgröße

Merkmalsausprägung	absolute Häufigkeit	relative Häufigkeit	Verteilungs-funktion
x_i = Anzahl der Personen	$n(x_i) = n_i$ (= Anzahl der Haushalte)	$h_i = \frac{n_i}{n}$	$F(x_i)$
$x_1 = 1$	5		
$x_2 = 2$	8		
$x_3 = 3$	7		
$x_4 = 4$	3		
$x_5 = 5$	2		
Insgesamt	25		

Eindimensionale Verteilungen

Eine Verteilungsfunktion hat stets die folgenden drei Eigenschaften:

- Sie ist für alle reelle Zahlen definiert.
- Sie ist monoton steigend.
- Ihre Werte liegen zwischen 0 und 1.

Im Falle von ordinalen bzw. quantitativ-diskreten Merkmalen ist es eine Treppenfunktion, im Falle von quantitativ-stetigen Merkmalen eine stetige Funktion.

Rheinische Hochschule Köln

Quantitativ-stetige Merkmale

Klasseneinteilung

Die Berechnung einer Häufigkeitsverteilung setzt bei quantitativ-stetigen Merkmalen eine Klassenbildung voraus, da ein einzelner Merkmalswert in der Regel nur einmal beobachtet wird, wenn eine beliebige Messgenauigkeit erzielt werden kann.

$\Longrightarrow$ Die Häufigkeitsverteilung ist in erheblichen Maße abhängig von der Klassenbildung.

Die Klassenbildung wird durch das Ziel der Untersuchung bestimmt.

Eindimensionale Verteilungen

Rheinische Hochschule Köln

1. Bei einem Vergleich mit anderen Verteilungen wählt man zweckmäßigerweise gleiche Klassen für alle Verteilungen.

2. Sollen die vorliegenden Beobachtungen bestmöglich dargestellt werden, so ist Gleichartiges in einer Klasse zusammenzufassen und Verschiedenartiges zu trennen.

   Hierbei sind möglichst sachliche Kriterien bei der Definition der Gleichartigkeit zu verwenden. Wenn keine sachlichen Kriterien existieren,

   können folgende formale Kriterien helfen:

   1. Rundung der Meßergebnisse beachten, da die Urliste meistens Aussagen über bestimmte Intervalle, wie z.B. 0 - unter 1 Euro, 1 - unter 2 Euro, usw. enthält. Man sollte die Klassengrenzen so wählen, dass sie mit Grenzen der Urlistenintervalle zusammenfallen.
   2. Möglichkeit absolut oder relativ konstante Klassenbreiten wählen
   3. Die Anzahl der Klassen (k) wegen der Übersichtlichkeit nicht zu groß wählen, z.B. $k \leq 20$.
   4. Die Häufigkeiten der Randklassen dürfen nicht zu klein werden.
   5. Der häufigste Wert der Urliste sollte die Klassenmitte der Klasse mit der größten Häufigkeit bilden, damit die Werte um den häufigsten Wert in eine Klasse fallen.

Eindimensionale Verteilungen

Rheinische Hochschule Köln

In der bereits erwähnten Untersuchung über die Einkaufsgewohnheiten im Internet wurden für die 25 betrachteten Haushalte das verfügbaren monatlichen Brutto-Haushaltseinkommen in Euro erhoben (vgl. Beispiel 2.1 in Sibbertsen, P./ Lehne, H. (2015)).

Nach der Bildung von Klassen ergibt sich folgende Tabelle.

Eindimensionale Verteilungen

Rheinische Hochschule Köln

Tabelle: monatliches Brutto-Haushaltseinkommen

Einkommens-klassse $x_i^u \leq x < x_i^o$ (in Euro)	absolute Häufigkeit n_i	relative Häufigkeit $h_i = h(x_i^u \leq x < x_i^o) = \frac{n_i}{n}$	Verteilungs-funktion an den Klassen-obergrenze $F(x_i^o)$	Klasse-breite Δi	Dichte-funktion $f(i) = \frac{n_i}{n * \Delta i} = \frac{h_i}{\Delta i}$	Klasse-mitte x_i^*
$1.500 - < 2.000$	5					
$2.000 - < 2.500$	6					
$2.500 - < 3.000$	8					
$3.000 - < 4.000$	3					
$4.000 - < 5.000$	1					
$5.000 - < 6.500$	2					
Insgesamt	25					

Deskriptive Statistik -
Lage- und Streuparameter eindimensionaler Verteilungen

Lage- und Streuparameter

Lageparameter

Lageparameter sind Maßzahlen, die die durchschnittliche Lage der Merkmalsausprägungen einer Häufigkeitsverteilung wiedergeben.

Zu Kennzeichnung der ”typischen” Merkmalsgröße kann man am einfachsten den häufigsten Wert verwenden.

Definition

Modus/Modalwert

Die Merkmalsausprägung eines beliebig skalierten Merkmals, die in einer statistischen Gesamtheit am häufigsten beobachtet wird, heißt Modus oder Modalwert.

Lage- und Streuparameter

Rheinische Hochschule Köln

Definition

Median
Ist X ein mindestens ordinales, zahlenmäßig erfasstes und aufsteigend geordnetes Merkmal einer statistischen Gesamtheit, dann heißt der kleinste Wert x, für den $F(x) \geq 0,5$ gilt, Median, kurz $x_{0,5}$.

Definition

Quantile (Verallgemeinerung des Begriffs Median)
Ist X ein mindestens ordinales, zahlenmäßig erfasstes und aufsteigend geordnetes Merkmal einer statistischen Gesamtheit, dann heißt der kleinste Wert x, für den $F(x) \geq p$ und $0 < p < 1$ gilt, Quantil der Ordnung p, kurz x_p.

Lage- und Streuparameter

Definition

Box-and-Whisker-Plot
Die grafische Darstellung

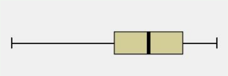

Abbildung: Box-and-Whisker-Plot

auf der Basis der 5 Verteilungsmaßzahlen:

- kleinster Merkmalswert x_{min}
- unteres Quartil $x_{0,25}$
- mittleres Quartil $x_{0,50}$ = Median
- oberes Quartil $x_{0,75}$ und
- größter Merkmalswert x_{max}

heißt Box-and-Whisker-Plot, kurz Boxplot

Rheinische Hochschule Köln

Definition

Arithmetisches Mittel
Ist X ein quantitatives Merkmal einer statistischen Gesamtheit vom Umfang n, dann heißt der Wert, der sich ergibt, wenn man die Summe aller beobachteten Merkmalswerte gleichmässig auf alle Merkmalsträger verteilt, arithmetisches Mittel $\bar{x}$.

Werden bei statistischen Erhebungen nicht die einzelnen Merkmalswerte erhoben (siehe Beispiel Haushaltseinkommen) und sind auch nicht die arithmetischen Mittel der einzelnen Klassen bekannt, so läßt sich das arithmetische Mittel i. a. nur mit einem Informationsverlust auf Basis der Klassenmitten x_i^* berechnen. In diesem Falle gilt

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i^* .$$

Rheinische Hochschule Köln

Streuparameter

Die Mittelwerte allein liefern nur eine unvollständige Beschreibung einer Verteilung, da sie nichts über die Größe der Abweichung der einzelnen Merkmalswerte von diesem Mittelwert aussagen.

Definition

Spannweite (range)

Ist X ein quantitatives Merkmal, dann heißt die Differenz $R = x_{max} - x_{min}$ aus dem größten Merkmalswert x_{max} und dem kleinsten Merkmalswert x_{min} Spannweite Sp.

Rheinische Hochschule Köln

Definition

Interquartilsabstand

Ist X ein quantitatives Merkmal, dann heißt die Differenz $Q_{0,5} = x_{0,75} - x_{0,25}$ aus oberem und unterem Quartil Interquartilsabstand.

Lage- und Streuparameter

Bei einem Streuungsmaß, welches zum Konzept des arithmetischen Mittels passt, sollten die Abweichungen aller Merkmalswerte vom arithmetischen Mittel eingehen.
Da die durchschnittliche Abweichung definitionsgemäß gleich Null ist,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = 0,$$

liegt es nahe, entweder die **durchschnittliche absoluten Abweichungen** (mean deviation)

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

Lage- und Streuparameter

oder die **durchschnittliche quadratische Abweichung** (Standardabweichung)

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

zu verwenden.
Das Quadrat der Standardabweichung (s^2) wird als Varianz bezeichnet.

Lage- und Streuparameter

Rheinische Hochschule Köln

Es gilt bei einer Urliste:

$$
\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2) \\
&= \frac{1}{n}(\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2) \\
&= \frac{1}{n}(\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2) \\
&= (\frac{1}{n}\sum_{i=1}^{n}x_i^2)-\bar{x}^2
\end{aligned}
$$

Lage- und Streuparameter

Werden bei statistischen Erhebungen nicht die einzelnen Merkmalswerte erhoben (siehe Beispiel zum monatlichen Brutto-Haushaltseinkommen) und sind auch nicht die arithmetischen Mittel der einzelnen Klassen bekannt, so läßt sich die Streuung i. a. nur mit einem Informationsverlust auf Basis der Klassenmitten x_i^* berechnen. In diesem Falle gilt

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

Lage- und Streuparameter

Will man in der Deskriptiven Statistik die Streuung gleicher, allerdings unterschiedlich dimensionierter bzw. verschiedener quantitativer Merkmale vergleichen, verwendet man den Variationskoeffizienten.

Definition

Variationskoeffizient
Ist X ein quantitatives Merkmal einer statistischen Gesamtheit vom Umfang n mit nur positiven Merkmalswerten $x_i > 0$ und einer von Null verschiedenen Standardabweichung $s_X > 0$, dann heißt der Quotient

$$v_X = \frac{s_X}{\bar{x}}$$

Variationskoeffizient v_X.

Rheinische Hochschule Köln

Deskriptive Statistik - Zweidimensionale Verteilungen/ Zusammenhangsanalysen

Rheinische Hochschule Köln

Überblick über die weiteren Abschnitte:

2. Merkmal	1. Merkmal: nominal	1. Merkmal: ordinal	1. Merkmal: metrisch
nominal-skaliert	Cramer's V	Cramer's V	Cramer's V
ordinal-skaliert	Cramer's V	Rangkorrelations-koeffizient	Rangkorrelations-koeffizient
metrisch skaliert	Cramer's V	Rangkorrelations-koeffizient	Maßkorrelations-koeffizient

Zusammenhangsanalysen

Kontingenzanalyse

Beispiel

Im Rahmen seiner Abschlussarbeit an der RH im letzten Semester befragte ein Studierender des Studiengangs Wirtschaft seine Kommilitoninnen und Kommilitonen u. a. danach, ob sie einem Nebenjob nachgehen und warum. Ein Teilergebnis seiner Befragung ist in der nachfolgenden Tabelle zusammengefasst:

Neben-job X	Finanzielle Situation Y: y_1 = unbefriedigend	Finanzielle Situation Y: y_2 = befriedigend	ins-gesamt
x_1 = ja	10	190	200
x_2 = nein	65	17	82
insgesamt	75	207	282

Zur Vermeidung von Fehlinterpretationen klären Sie bitte zunächst die statistischen Grundbegriffe, wie Merkmalsträger, etc.

Zusammenhangsanalysen

Definition

Statistische Unabhängigkeit
Stimmen die zeilenweisen Konditionalverteilungen mit der zeilenweisen Randverteilung überein, so sind die beiden Merkmale statistisch unabhängig, ansonsten statistische abhängig.

Analog gilt dies auch spaltenweise.

Zusammenhangsanalysen

Wenn eine Kreuztabelle I Zeilen und J Spalten bei der gemeinsamen Verteilung hat, ist es eine (IxJ)-Kreuztabelle und χ^2 ergibt sich aus folgender Formel:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}.$$

χ^2 ist genau dann 0, wenn die beiden Merkmale statistisch unabhängig sind. Ansonsten ist χ^2 größer als 0.

Rheinische Hochschule Köln

Indikatoren für die Stärke des Zusammenhangs, die auf χ^2 beruhen:

- Phi-Koeffizient

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

Je größer der Wert von Φ ist, desto stärker ist der Zusammen- hang.
Als Faustregel gilt, dass ein Wert größer als $0,3$ eine Stärke der Abhängigkeit anzeigt, die relevant ist. Diese Faustregel gilt auch für die beiden nachfolgenden Indikatoren.

Ein Nachteil von Φ ist, dass Φ auch Werte über 1 annehmen kann und damit nicht beschränkt ist.

Zusammenhangsanalysen

Rheinische Hochschule Köln

- Kontingenzkoeffizient

$$CC = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Dieser Koeffizient nimmt nur Werte zwischen 0 und 1 an, aber nur selten den Maximalwert 1, denn es gilt:

$CC_{max} = \sqrt{(R-1)/R}$, wobei $R = min(Zeilenanzahl\ I, Spaltenanzahl\ J)$

- Cramer's V

$$V = \sqrt{\frac{\chi^2}{n(R-1)}},$$

wobei $R = min(Zeilenanzahl\ I, Spaltenanzahl\ J)$

V nimmt nur Werte zwischen 0 und 1 an, wobei der Maximalwert 1 unabhängig von der Anzahl der Zeilen bzw. Spalten ist.

Die oben genannten Indikatoren nehmen den Wert 0 an, falls kein Zusammenhang existiert und den maximalen Wert bei vollständiger Abhängigkeit.

Rheinische Hochschule Köln

Maßkorrelationsanalyse

Beispiel

Für $n = 10$ vergleichbare Zwei-Zimmer-Mietwohnungen, die auf dem Berliner Mietwohnungsmarkt angeboten wurden, soll statistisch untersucht werden, ob zwischen der Wohnfläche X (Angaben in m^2) und der monatlichen Warmmiete Y (Angabe in Euro) ein statistischer Zusammenhang besteht, wie stark er ausgeprägt ist und welche Richtung er besitzt.

i	x_i	y_i
1	68	538
2	72	590
3	47	429
4	61	518
5	55	484
6	63	486
7	51	456
8	65	510
9	54	445
10	64	544

Zusammenhangsanalysen

i	xi	yi
1	68	538
2	72	590
3	47	429
4	61	518
5	55	484
6	63	486
7	51	456
8	65	510
9	54	445
10	64	544

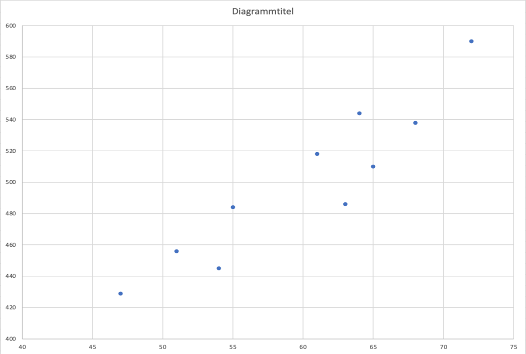

Abbildung: Streudiagramm

Rheinische Hochschule Köln

Interpretation Kovarianz:

Um die Kovarianz zu verstehen, hilft das Streudiagramm. Man zeichne durch $\bar{x}$ und $\bar{y}$ parallel zu den Koordinatenachsen zwei Geraden. Dadurch wird das Streudiagramm in vier Quadrante eingeteilt.

- **Quadrant I** mit $x_i > \bar{x}$ und $y_i > \bar{y}$, rechts oben
  Dieser Quadrant erzeugt nur positive Terme.
- **Quadrant II** mit $x_i < \bar{x}$ und $y_i > \bar{y}$, links oben
  Dieser Quadrant erzeugt nur negative Terme.
- **Quadrant III** mit $x_i < \bar{x}$ und $y_i < \bar{y}$, links unten
  Dieser Quadrant erzeugt nur positive Terme.
- **Quadrant IV** mit $x_i > \bar{x}$ imd $y_i < \bar{y}$, rechts unten
  Dieser Quadrant erzeugt nur negative Terme.

Zusammenhangsanalysen

Rheinische Hochschule Köln

Die Kovarianz ist ein Maß, das einen ersten Anhaltspunkt für den Zusammenhang zweier Merkmale liefern kann. Je nach Lage der Punktwolke im Streudiagramm lassen sich unterschiedliche Wechselbeziehungen (Korrelationen) zwischen den Merkmalen feststellen.

1. **Positive Korrelation**
   Die Punktwolke verläuft von links unten nach rechts oben. Kleine x_i-Werte sind mit kleinen y_i-Werten gekoppelt, große x_i mit großen y_i. In der Kovarianzrechnung überwiegen die Werte, die einen positiven Beitrag leisten: Die Kovarianz ist positiv.

2. **Negative Korrelation**
   Die Punktwolke verläuft von links oben nach rechts unten. Kleine x_i-Werte sind mit großen y_i-Werten gekoppelt, große x_i mit kleinen y_i. In der Kovarianzrechnung überwiegen die Werte, die einen negativen Beitrag leisten: Die Kovarianz ist negativ.

Rheinische Hochschule Köln

3 **Keine Korrelation**
Die Punktwolke ist diffus über alle Quadranten verteilt. Eine eindeutige Richtung lässt sich nicht erkennen. In die Kovarianzrechnung gehen positive wie negative Werte ein: Die Kovarianz ist nahe Null.

Die Kovarianz hat den entscheidenden Nachteil, dass sie von den in sie eingehenden Dimensionen abhängt. Eine dimensionslose Größe ist zu Interpretationszwecken und für die Vergleichbarkeit sinnvoller.

Rheinische Hochschule Köln

Interpretation Maßkorrelationskoeffizient:

- Der Koeffizient kann nur bestimmt werden, wenn sich im Nenner keine Null ergibt. Daher muss der Fall, dass alle x_i gleich sind bzw. alle y_i gleich sind, ausgeschlossen werden.
- Er ist ein normiertes und symmetrisches Zusammenhangsmaß, für das stets $-1 \leq r_{xy} \leq 1$ gilt. Ein r nahe 1 kennzeichnet einen starken gleichläufigen, ein r nahe -1 einen starken gegenläufigen linearen Zusammenhang.

$$r_{xy} = +1 \iff y_i = ax_i + b \text{ mit } a > 0$$

$$r_{xy} = -1 \iff y_i = ax_i + b \text{ mit } a < 0$$

Ein r um 0 ist ein Indiz dafür, dass die Merkmale X und Y (linear) voneinander unabhängig sind.

Zusammenhangsanalysen

Der Maßkorrelationskoeffizient ist nicht nur ein Maß für die Richtung des Zusammenhangs zwischen den Merkmalen X und Y wie die Kovarianz, sondern gibt gleichzeitig auch die Stärke des linearen Zusammenhangs zwischen den Merkmalen an.

1. **Positive Korreliertheit**
   Die Punktwolke verläuft von links unten nach rechts oben. Je näher die Punktewolke an einer (imaginären) Geraden durch den I. und III. Quadranten liegt, desto stärker die positive Korreliertheit. Um so näher ist auch der Wert von r_{xy} an +1 dran. Für $r_{xy} = +1$ schließlich liegen alle Wertepaare (x_i, y_i) tatsächlich auf einer Geraden mit positiver Steigung ($a > 0$).

Zusammenhangsanalysen

2 Negative Korreliertheit
Die Punktwolke verläuft von links oben nach rechts unten. Je näher die Punktewolke an einer (imaginären) Geraden durch den II. und VI. Quadranten liegt, desto stärker die negative Korreliertheit. Um so näher ist auch der Wert von r_{xy} an -1 dran. Für $r_{xy} = -1$ schließlich liegen alle Wertepaare (x_i, y_i) tatsächlich auf einer Geraden mit negativer Steigung ($a < 0$).

3 Unkorreliertheit
Die Punktwolke ist diffus über alle Quadranten verteilt, damit sicherlich nicht auf oder nahe einer Geraden. Kein linearer Zusammenhang feststellbar, r_{xy} nahe Null.

Rheinische Hochschule Köln

Rangkorrelationsanalyse

Sind zwei Merkmale mindestens ordinal skaliert, kann der Rangkorrelationskoeffizient ρ (Rho) von Charles Edward Spearman (1863-1945) berechnet werden. Die beiden Merkmale X und Y werden dazu in ihre natürliche Rangordnung (aufsteigend oder absteigend) gebracht und bei der Analyse des Zusammenhangs werden die beiden Rangordnungen dann auf Übereinstimmungen geprüft. Dabei ist zu beachten, dass die Art der Anordnung für die spätere Interpretation von großer Wichtigkeit ist. Dem Rangkorrelationskoeffizienten von Spearmann liegt letztendlich der Korrelationskoeffizient r von Bravais-Pearson zugrunde, denn nachdem für die Merkmale X und Y jeweils eine Rangordnung erstellt wurde und diesen dann Rangziffern zugeordnet wurden, wird zur Ermittlung der Übereinstimmung der Korrelationskoeffizient r berechnet.

Unter dem Begriff einer Rangkorrelationsanalyse subsummiert man in der statistischen Methodenlehre eine sachlogisch begründete Analyse eines Zusammenhanges zwischen zwei ordinalen Merkmalen. Aus der Vielzahl der in der Statistik bekannten und applizierten ordinalen Zusammenhangsmaße kommt wegen seiner einfachen Berechnung dem Rangkorrelationskoeffizienten nach Spearman eine besondere praktische Bedeutung zu.

Rheinische Hochschule Köln

Definition

Rangzahl

Ist X ein mindestens ordinales Merkmal, dessen n beobachtete Merkmalsausprägungen x_i $(l = 1, 2, \cdots, n)$ aufsteigend $x_i \leq x_{i+1}$ oder absteigend $x_i \geq x_{i+1}$ geordnet sind, dann heißt die der geordneten Merkmalsausprägungen $x_{(i)}$ zugewiesene Platznummer i Rangzahl R_i^X der Merkmalsausprägung x_i

Bindung:

Gleiche Merkmalsausprägungen in einer geordneten Folge heißen Bindungen (engl. ties). Treten Bindungen auf, so ordnet man in der Regel allen gleichen Merkmalsausprägungen das arithmetische Mittel ihrer Rangzahlen zu, die sie im Fall ihrer Unterscheidbarkeit erhalten hätten.

Zusammenhangsanalysen

Rheinische Hochschule Köln

Definition

Rangkorrelationskoeffizient nach Spearman

Für eine statistische Gesamtheit, an deren n Merkmalsträger die n Ausprägungspaare $\{(x_i; y_i), i = 1, 2, \cdots, n\}$ der beiden (mindestens) ordinalen Merkmale X und Y beobachtet wurden, wobei dem Ausprägungspaar $(x_i; y_i)$ das Rangzahlenpaar $(R_i^X; R_i^Y)$ zugeordnet wird, heißt für den Fall einer Urliste die Größe, die sich ergibt, wenn man den Maßkorrelationskoeffiziten für die Rangzahlenpaare $(R_i^X; R_i^Y)$ berechnet **Rangkorrelationskoeffizient nach Spearman.**

Zusammenhangsanalysen

Der Rangkorrelationskoeffizient nach Spearman ist ein normiertes Zusammenhangmaß zur Ermittlung der Richtung und der Stärke eines ordinal statistischen Zusammenhangs, für das stets $-1 \leq r_S \leq 1$ gilt. Ein r_S nahe 1 kennzeichnet einen gleichläufigen oder positiven Zusammenhang, ein r_S nahe -1 einen gegenläufigen oder negativen Zusammenhang. Ein r_S um 0 deutet man als Indiz dafür, dass statistisch zwischen den Merkmalen X und Y kein Zusammenhang nachweisbar ist bzw. dass sie empirisch als voneinander unabhängig angesehen werden können.

Zusammenhangsanalysen

Rheinische Hochschule Köln

Beispiel

Im 100m-Sprint laufen 6 Sportler um die Medaillen.
Frage: Gibt es einen Zusammenhang zwischen Größe und Platzierung (nicht Geschwindigkeit!)?
Die folgende Tabelle gibt die Größe (Merkmal X) und die Zeiten (Merkmal Z) der Läufer wieder:

Läufer i	Größe x_i	Zeit z_i	Rang gem. X R_i^X	Rang gem. Zeit R_i^Z
1	1,80	10,46		
2	1,72	10,54		
3	1,76	10,46		
4	1,85	10,4		
5	1,69	10,52		
6	1,88	10,31		

Bedeutung und Interpretation des Rangkorrelationskoeffizienten

Auch hier informiert das Vorzeichen des Koeffizienten über die Richtung des Zusammenhangs. Es handelt sich also wegen des positiven Vorzeichens um einen gleichläufigen Zusammenhang, d.h. wird das eine Merkmal größer (die Körpergröße), so wächst auch das andere (die Geschwindigkeit). Wichtig hierbei ist allerdings, dass der Zusammenhang nur mittelbar gemessen wird, nämlich in Form des Zusammenhangs der Ränge.
Der Betrag informiert über die Tendenz, dass bei steigender Körpergröße (höherer Rang für X) die Geschwindigkeit ebenfalls ansteigt (höherer Rang für Y). Nachteil: Da nur die Rangordnungen verglichen werden, bleiben die Abstufungen zwischen den Merkmalswerten unberücksichtigt. Hätte ein Läufer das Rennen z.B. nicht beendet oder wäre er weit hinter den anderen ins Ziel gekommen, so hätte sich an seinem Rangplatz und der Berechnung nichts verändert.

Zusammenhangsanalysen

Rheinische Hochschule Köln

Definition

Rangkorrelationskoeffizient nach Spearman **näherungsweise**

Für eine statistische Gesamtheit, an deren n Merkmalsträger die n Ausprägungspaare $\{(x_i; y_i), i = 1, 2, \cdots, n\}$ der beiden (mindestens) ordinalen Merkmale X und Y beobachtet wurden, wobei dem Ausprägungspaar $(x_i; y_i)$ das Rangzahlenpaar $(R_i^X; R_i^Y)$ zugeordnet wird, heißt für den Fall einer Urliste die Größe

$$r_S = 1 - \frac{6 * \sum_{i=1}^{n} (R_i^x - R_i^y)^2}{n * (n^2 - 1)}$$

Rangkorrelationskoeffizient nach Spearman.

Zusammenhangsanalysen

Die näherungsweise Berechnung des Korrelationskoeffizienten nach Spearman ist streng genommen an die Bedingung gebunden, dass keine Rangbindungen in einem oder in beiden ordinalen Merkmalen auftreten. Da in praktischen Anwendungen Rangbindungen häufig vorkommen, sollte in jeder der beiden Rangreihen höchstens ein Fünftel der Rangzahlen als Bindungen auftreten, wenn der Korrelationskoeffizient aussagefähig sein soll.

Rheinische Hochschule Köln

Ein neues Messinstrument soll die Risikobereitschaft von Studienabgängern messen (Skala 0-100). Nun soll untersucht werden, ob die Selbsteinschätzung der Risikobereitschaft bei den Studienabgängern mit der Fremdeinschätzung durch deren Partner/Partnerin (Skala 0-10) zusammenhängt.
Die folgende Tabelle gibt die Variablen Selbsteinschätzung (X) und Fremdeinschätzung (Y) für 12 Probanden an:

Proband	X	Y
1	43	3,0
2	47	1,2
3	52	4,2
4	54	4,0
5	56	4,3
6	58	6,6
7	60	2,8
8	68	5,9
9	69	4,1
10	72	4,8
11	83	9,2
12	84	5,0

Berechnen Sie den zugehörigen Rangkorrelationskoeffizienten.
Welche Schlüsse ziehen Sie daraus?

Abbildung: Beispiel

Rheinische Hochschule Köln

Deskriptive Statistik - Regressionsanalyse

Beispiel

Für $n = 10$ vergleichbare Zwei-Zimmer-Mietwohnungen, die auf dem Berliner Mietwohnungsmarkt angeboten wurden, soll statistisch untersucht werden, ob zwischen der Wohnfläche X (Angaben in m^2) und der monatlichen Warmmiete Y (Angabe in Euro) ein statistischer Zusammenhang besteht, wie stark er ausgeprägt ist und welche Richtung er besitzt.

i	x_i	y_i	$f(x_i) = \hat{y}_i$	$y_i - \hat{y}_i = \epsilon_i$	ϵ_i^2
1	68	538			
2	72	590			
3	47	429			
4	61	518			
5	55	484			
6	63	486			
7	51	456			
8	65	510			
9	54	445			
10	64	544			

Regressionsanalyse

Rheinische
Hochschule
Köln

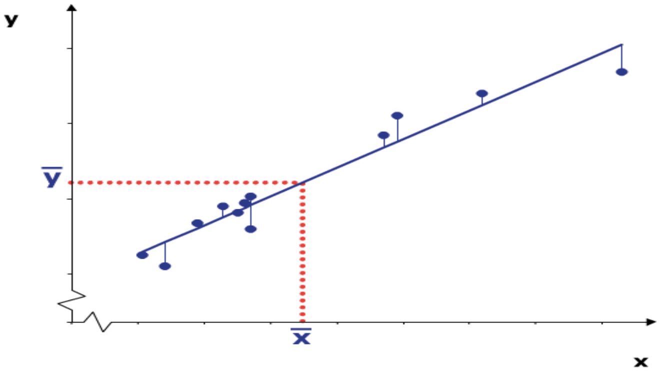

Abbildung: Regressionsgerade

Rheinische Hochschule Köln

Anmerkungen zum Bestimmtheitsmaß:

- Für die einfache lineare Regression $f(x) = \hat{a} + \hat{b}x$ (und nur für diese!) gilt die folgende wichtige Beziehung:
$$R^2 = r_{XY}^2 = r_{YX}^2$$
Demnach ist das Quadrat des einfachen linearen Maßkorrelations-koeffizienten r_{XY} gleich dem Bestimmtheitsmaß R^2 einer einfachen linearen Regression Y auf X.
- Das Bestimmtheitsmaß ist ein normiertes Maß, für das stets $0 \leq R^2 \leq 1$ gilt.
- Das zum Bestimmtheitsmaß R^2 komplementäre Maß $1 - R^2$ wird auch als Unbestimmtheitsma”bezeichnet.
- Beide Maße können gleichermaßen als Gütemaße für eine Regression betrachtet werden. Je näher das Bestimmtheitsmaß am Wert eins liegt, um so höher ist die Bestimmtheit und damit die statistische Erklärungsfähigkeit einer Regression.

Grundlagen Wahrscheinlichkeitstheorie

Grundlagen Wahrscheinlichkeitstheorie

Ein Grundanliegen der Stochastik ist die mathematische Beschreibung und Nachbildung von zufallsabhängigen Vorgängen. Für die inhaltliche Bestimmung und für das Verständnis stochastischer Grundbegriffe erweisen sich vor allem Zufallsexperimente, die Glücksspielen entlehnt sind, als sehr anschaulich und vorteilhaft.

Definition

Zufallsexperiment
Ein unter gleichen Rahmenbedingungen zumindest gedanklich beliebig oft wiederholbarer Versuch mit unbestimmtem Ausgang heißt Zufallsexperiment.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Entscheiden Sie im Folgenden, ob die angegebenen Vorgänge als Zufallsexperimente gelten.

1. Anzahl der Zuschauer der nächsten ”heute”-Nachrichten
2. Ihre Punktzahl in der nächsten Statistik-Klausur
3. Das Herauskommen durch das Werfen einer Sechs beim Mensch-ärgere-dich-nicht-Spiel

Definition

Ergebnis und Ergebnismenge

Der Ausgang eines Zufallsexperimentes heißt Ergebnis ω. Die Menge Ω aller möglichen Ergebnisse $\omega \in \Omega$ eines Zufallsexperimentes heißt Ergebnismenge.

Definition

Ereignis

Eine Teilmenge einer Ergebnismenge Ω eines Zufallsexperimentes heißt Ereignis.

- Elementarereignis
- Sicheres Ereignis
- Unmögliches Ereignis

Grundlagen Wahrscheinlichkeitstheorie

Welche Darstellungsmöglichkeiten für Ereignisse (Mengen) kennen Sie?

Beispiel

Bestimmen und benennen Sie alle Ereignisse des Zufallsexperimentes "Einmaliges Werfen eines idealen Spielwürfels".

Rheinische Hochschule Köln

Ereignisrelationen und -operationen

Definition

Ereignisrelation $A \subset B$

Sind $A, B \subset \Omega$ zufällige Ereignisse einer Ergebnismenge Ω, dann beschreibt die Ereignisrelation $A \subset B$, dass das Ereignis A das Ereignis B nach sich zieht.

Definition

Ereignisoperation $A \cup B$

Sind $A, B \subset \Omega$ zufällige Ereignisse einer Ergebnismenge Ω, dann beschreibt die Ereignisoperation $A \cup B$ das zufällige Ereignis, dass mindestens eines der beiden zufälligen Ereignisse A oder B eintritt.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Definition

Ereignisoperation $A \cap B$
Sind $A, B \subset \Omega$ zufällige Ereignisse einer Ergebnismenge Ω, dann beschreibt die Ereignisoperation $A \cap B$ das zufällige Ereignis, dass sowohl das zufällige Ereignis A als auch das zufällige Ereignis B eintritt.

Definition

Disjunkte Ereignisse
Sind $A, B \subset \Omega$ zufällige Ereignisse einer Ergebnismenge Ω und stellt die Ereignisoperation $A \cap B$ unmögliches Ereignis dar, dann heißen die zufällige Ereignisse A und B disjunkte zufällige Ereignisse.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Definition

Ereignisoperation $A \setminus B$

Sind $A, B \subset \Omega$ zufällige Ereignisse einer Ergebnismenge Ω, dann beschreibt die Ereignisoperation $A \setminus B$ das zufällige Ereignis, dass das zufällige Ereignis A, aber nicht zufällige Ereignis B eintritt.

Definition

Rheinische Hochschule Köln

Komplementäres Ereignis
Ist $B \subset \Omega$ ein zufälliges Ereignisse einer Ergebnismenge Ω, dann beschreibt die Ereignisoperation $\bar{B} = \Omega \setminus B$ das zum zufälligen Ereignis B komplementäre zufällige Ereignis $\bar{B}$, das darin besteht, dass das sichere Ereignis Ω, aber nicht das zufällige Ereignis B eintritt.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Rechenregeln für Mengen

1. $A \cup B = B \cup A$ (Kommutativgesetz)
2. $A \cap B = B \cap A$ (Kommutativgesetz)
3. $A \cup (B \cup C) = (A \cup B) \cup C$ (Assoziativgesetz)
4. $A \cap (B \cap C) = (A \cap B) \cap C$ (Assoziativgesetz)
5. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (Distributivgesetz)
6. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivgesetz)
7. $A \cup (A \cap B) = A$ (Absorptionsgesetz)
8. $A \cap (A \cup B) = A$ (Absorptionsgesetz)
9. $(A \backslash B) \cap B = \{\}$ (Satz vom Widerspruch)
10. $(A \backslash B) \cup B = A \cup B$ (Satz vom ausgeschlossenen Dritten)
11. $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$ (DeMorgan'sche Regel)
12. $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$ (DeMorgan'sche Regel)

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Definition

Mächtigkeit/Kardinalität
Die Anzahl der Elemente eines Ereignisses wird mit dem aus der Mengenlehre stammenden Begriff der Mächtigkeit erfaßt, geschrieben $|A|$.

Es gilt:
Die Anzahl aller Ereignisse zu einer Ergebnismenge Ω ergibt sich aus

$$2^{|\Omega|}.$$

Rheinische Hochschule Köln

Beispiel

Sei $A = \{1, 2, 3, 4, 5\}$ und $B = \{2, 3, 7\}$.

1. $A \subset B$?
2. $A \supset B$?
3. $A \cap B$?
4. $A \cup B$?
5. $A \cap A$?
6. $B \cup B$?
7. $A \setminus B$?
8. $B \setminus A$?
9. Welche Teilmengen besitzt B?

Rheinische Hochschule Köln

Beispiel

Zur Beschreibung des Kölner Mietwohnungsmarktes wurde für das Marktsegment von 2-Zimmer-Mietwohnungen unter anderem auch erfasst, ob eine derartige Wohnung einen Balkon besitzt (Ereignis A), ob eine Einbauküche vorhanden ist (Ereignis B) bzw. ob die Wohnung mit einer Zentralheizung (Ereignis C) ausgestattet ist.
Stellen Sie die folgenden Ereignisse durch geeignete Verknüpfungen der Ereignisse A, B, C dar:
Eine Kölner 2-Zimmer-Mietwohnung besitzt

1. einen Balkon und Zentralheizung.
2. zwar Zentralheizung, aber keinen Balkon.
3. weder einen Balkon noch eine Einbauküche
4. Welche Berliner 2-Zimmer-Mietwohnungen sind durch die folgenden Ereignisse gekennzeichnet:
   1. $B \setminus A$
   2. $\bar{B} \cap \bar{C}$
   3. $\bar{A} \cup \bar{B}$
   4. $C \cap (\overline{A \cup B})$?

Grundlagen Wahrscheinlichkeitstheorie

Um Zufallsvorgänge besser einschätzen zu können, wir versucht, den möglichen Ausgängen, sprich den Ereignissen, zahlen zuzuordnen. Geschieht diese Zuordnung (*Ereignis* $\rightarrow$ *Zahl*) unter Einhaltung gewisser mathematischer Grundsätze/ Axiomen, so werden die zugeordneten Zahlen auch Wahrscheinlichkeiten genannt.

$$\mathcal{P} : \mathcal{A} \longmapsto \mathcal{P}(\mathcal{A})$$

Bezeichnung: $P(A)$ gibt die Wahrscheinlichkeit für das Eintreten des Ereignisses A an. Der Buchstabe P ist von der englischen Übersetzung für Wahrscheinlichkeiten probability abgeleitet.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Definition

Axiome von Kolmogorov

Sind $\Omega \neq \{\}$ eine (nichtleere) Ergebnismenge und $A, B \subset \Omega$ zufällige Ereignisse.

(K1) $P(A) \geq 0$ für alle Ereignisse A von Ω (Nichtnegativitätsbedingung)

(K2) $P(\Omega) = 1$ (Normierungsbedingung)

(K3) Falls $A \cap B = \{\}$, so gilt: $P(A \cup B) = P(A) + P(B)$. (Additivitätsbedingung)

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Aus diesen Grundsätzen lassen sich nun eine Menge Rechenregeln für Wahrscheinlichkeiten ableiten.

Theorem

Rechenregeln für Wahrscheinlichkeiten
Sei Ω die Ergebnismenge eines Zufallsexperimentes, A und B zwei Ereignisse von Ω. Dann gilt:
(R1) $0 \leq P(A) \leq 1$,
(R2) $P(\{\}) = 0$,
(R3) Falls $A \subset B$, so gilt: $P(A) \leq P(B)$,
(R4) $P(A) = 1 - P(\bar{A})$ *mit* $\bar{A} = \Omega \backslash A$,
(R5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
(R6) $P(A \backslash B) = P(A) - P(A \cap B)$.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Beispiel

In einer Kleinstadt erscheinen die Lokalblätter ”Abendzeitung” und ”Bildpost”. Die Wahrscheinlichkeit, dass eine Bewohnerin/ein Bewohner

- die Abendzeitung liest, sei 0,6,
- die Bildpost liest, sei 0,5 und
- die Abendzeitung oder die Bildpost (oder beide) liest, sei 0,9.

Wie groß ist dann die Wahrscheinlichkeit, dass eine Bewohnerin/ein Bewohner

1. beide Blätter liest?
2. keines der beiden Blätter liest?
3. eine der beiden Lokalzeitungen, aber nicht beide liest?

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Definition

Laplace-Experiment

Ein Zufallsexperiment heißt **Laplace-Experiment**, wenn die Ergebnismenge Ω eine endliche Menge ist mit $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ und alle Elementarereignisse $\{\omega_1\}$ bis $\{\omega_n\}$ gleich wahrscheinlich sind.

Rheinische Hochschule Köln

Definition

Laplace-Wahrscheinlichkeit oder klassische Wahrscheinlichkeit
Für die Elementarereignisse $\{\omega_i\}$, $i = 1, \cdots, n$, eines Laplace-Experimentes ermittelt sich die (Laplace-)Wahrscheinlichkeit oder klassische Wahrscheinlichkeit wie folgt: Für alle $\omega_i \in \Omega$ gilt

$$P(\{\omega_i\}) = \frac{1}{n}$$

Die Wahrscheinlichkeit für ein Ereignis A ergibt sich dann zu:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n}$$

Diese Zuordnung von Zahlen zu Ereignissen erfüllt die Axiome (K1)-(K3) und darf somit als Wahrscheinlichkeit im Sinne von Kolmogorov bezeichnet werden.

Rheinische Hochschule Köln

Beispiel

Das Werfen eines ”fairen” Würfels stellt ein Laplace-Experiment dar. Es gilt:

$$\Omega = \{1, 2, 3, 4, 5, 6\}, |\Omega| = 6, P(\{\omega_i\}) = \frac{1}{6}.$$

Wie lauten die Wahrscheinlichkeiten für folgende Ereignisse?

- A: Es wird eine Sechs gewürfelt.
- B: Eine ungerade Augenzahl erscheint.
- C: Eine ungerade Augenzahl oder die Sechs erscheint.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Definition

Subjektive Wahrscheinlichkeit
Die wissensbasierte Bestimmung der Wahrscheinlichkeit $P(A)$ eines zufälligen Ereignisses A durch eine mit dem jeweiligen Zufallsexperiment vertraute Person heißt subjektive Wahrscheinlichkeit.

Subjektive Wahrscheinlichkeiten spielen in der Entscheidungstheorie für Lösungsansätze von Entscheidungsproblemen unter Ungewissheit eine besondere Rolle.

Relative Häufigkeit und Wahrscheinlichkeit

Von praktischer Bedeutung ist der Zusammenhang zwischen relativer Häufigkeit $h(A)$ und Wahrscheinlichkeit $P(A)$ eines zufälligen Experimentes A. Es kann gezeigt werden, dass es gemäß dem Gesetz der großen Zahlen gerechtfertigt ist, eine relative Häufigkeit als einen Schätzwert für eine unbekannte Wahrscheinlichkeit zu verwenden. Diese Herangehensweise ist vor allem dann von Vorteil, wenn Wahrscheinlichkeiten für interessierende Fragestellungen nicht ohne weiteres berechnet werden können. Man verwendet dann die statistisch beobachteten relativen Häufigkeiten als Näherung für die nicht unmittelbar berechenbaren Wahrscheinlichkeiten.

Grundlagen Wahrscheinlichkeitstheorie

Beispiel

In einer durch Studierende der RH Köln durchgeführte Befragung von Fahrgästen zu aktuellen Fragen des Öffentlichen Personennahverkehrs erbrachte unter anderem das folgende Ergebnis: siehe Tabelle 1.

	vorrangig	benutztes	Verkehrsmittel	
Wohnort	U-Bahn	S-Bahn	Bus	gesamt
Köln	145	181	242	568
Düsseldorf	200	96	152	448
außerhalb des Rheinlandes	14	57	10	81
gesamt	359	334	404	1097

Gehen Sie von folgendem Zufallsexperiment aus: Aus dem Kreis der befragten Fahrgäste wird zufällig ein Fahrgast ausgewählt. Vereinbaren Sie für die folgenden Ereignisse geeignete Symbole und geben Sie die jeweiligen Ereigniswahrscheinlichkeiten an:

1. Der Fahrgast ist aus Düsseldorf.
2. Der Fahrgast ist ein S-Bahn-Nutzer.
3. Der Fahrgast ist ein S-Bahn-Nutzer und aus Düsseldorf.
4. Der Fahrgast ist Rheinländer.
5. Der Fahrgast ist kein Rheinländer.

Beispiel

Eine Befragung von Reisenden auf dem Flughafen Köln-Bonn erbrachte u. a. das folgende Ergebnis: Von den insgesamt 340 befragten Fluggästen gaben 177 Fluggäste an, privat unterwegs zu sein. Von den 164 Fluggästen, die mit einem Taxi zum Flughafen fuhren, waren 121 Fluggäste geschäftlich unterwegs. Von den 128 Fluggästen, die mit dem Bus anreisten, waren 94 Fluggäste privat unterwegs.

1. Erstellen Sie für die Erhebungsmerkmale eine Kontingenztabelle/Kreuztabelle.
2. Von Interesse sind die folgenden Ereignisse: Ein zufällig ausgewählter Fluggast, der vom Flughafen Köln-Bonn abreist, ist
   1. mit dem Bus zum Flughafen gefahren (Ereignis B),
   2. geschäftlich unterwegs (Ereignis G).

   Geben Sie anhand der Kontingenztabelle die folgenden Wahrscheinlichkeiten an: $P(B)$, $P(G)$, $P(B \cap G)$. Benennen Sie den theoretischen Sachverhalt, auf dessen Grundlage Sie die Wahrscheinlichkeiten bestimmt haben.
3. Benennen Sie die folgenden Beziehungen und überprüfen Sie unter Verwendung der Kontingenztabelle deren Gültigkeit:
   1. $P(B \cup G) = P(B) + P(G)$,
   2. $P(B \cup G) = P(B) + P(G) - P(B \cap G)$

Rheinische Hochschule Köln

Definition

Bedingte Wahrscheinlichkeit

Sind $A, B \subset \Omega$ zwei zufällige Ereignisse einer Ergebnismenge Ω, dann heißt die Wahrscheinlichkeit

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

falls $P(B) > 0$, für das Eintreten des Ereignisses A unter der Bedingung, dass das Ereignis B bereits eingetreten ist, bedingte Wahrscheinlichkeit $P(A|B)$.

Rheinische Hochschule Köln

Definition

Stochastische Unabhängigkeit

Sind $C, D \subset \Omega$ zwei zufällige Ereignisse einer Ergebnismenge Ω, deren unbedingte Wahrscheinlichkeiten $P(C)$ und $P(D)$ sowie deren bedingte Wahrscheinlichkeiten $P(C|D)$ und $P(D|C)$ bekannt sind, dann heißen die Ereignisse C und D stochastisch unabhängig, wenn $P(C|D) = P(C)$ bzw. $P(D|C) = P(D)$ gilt.

Rheinische Hochschule Köln

Beispiel

In einer Filiale eines Kölner Kreditinstituts besitzen 80% der Kunden/Kundinnen ein Gehaltskonto und 50% der Kunden/Kundinnen ein Sparkonto. Alle Kunden/Kundinnen der Filiale verfügen über mindestens eine der beiden Anlageformen.
Wie groß ist die Wahrscheinlichkeit, dass ein/e zufällig ausgewählte/r Kunde/Kundin dieser Bankfiliale

1. ein Gehaltskonto und ein Sparkonto besitzt?
2. ein Sparkonto besitzt, wenn bereits bekannt ist, dass der Kunde/die Kundin ein Gehaltskonto hat?
3. ein Gehaltskonto besitzt, wenn bereits bekannt ist, dass der Kunde/die Kundin ein Sparkonto hat?
4. ein Sparkonto hat, aber kein Gehaltskonto?
5. höchstens eines von beiden Konten besitzt?

Grundlagen Wahrscheinlichkeitstheorie

Beispiel

Eine Umfrage unter Studierenden ergab, dass 70% aller Studierenden regelmäßig in der Mensa essen und dass 40% aller Studierenden eine längere Öffnungszeit der Mensa wünschen. 20% aller Studierenden gehen regelmäßig in der Mensa essen und wünschen eine längere Öffnungszeit.

1. Wie groß ist die Wahrscheinlichkeit, dass Studierende, die längere Öffnungszeiten der Mensa wünschen, regelmäßig dort essen?
2. Wie groß ist die Wahrscheinlichkeit, dass Studierende, die nicht regelmäßig in der Mensa essen, längere Öffnungszeiten wünschen?

Rheinische Hochschule Köln

Theorem

Es bezeichnen A und B zwei zufällige Ereignisse mit der Wahrscheinlichkeit $P(A) > 0$ und $P(B) > 0$. Dann gilt:

1. *Sind A und B disjunkt, so sind sie voneinander abhängig.*
2. *Sind A und B unabhängig, so sind sie nicht disjunkt.*
3. *Wenn die Ereignisse A und B unabhängig sind, so sind auch die Ereignisse $\bar{A}$ und $\bar{B}$ unabhängig.*

Rheinische Hochschule Köln

Theorem

Multiplikationsregel
Sind $A, B \subset \Omega$ zwei zufällige Ereignisse einer Ergebnismenge Ω, deren unbedingte Wahrscheinlichkeiten $P(A)$ und $P(B)$ sowie deren bedingte Wahrscheinlichkeiten $P(A|B)$ und $P(B|A)$ bekannt sind, dann heißen die Gleichung

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

allgemeine Multiplikationsregel für zwei zufällige Ereignisse.

Spezialfall Sind $A, B \subset \Omega$ zwei stochastisch unabhängige zufällige Ereignisse einer Ergebnismenge Ω, deren unbedingte Wahrscheinlichkeiten $P(A)$ und $P(B)$ bekannt sind, dann heißt die Gleichung

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

Multiplikationsregel für zwei stochastisch unabhängige Ereignisse.

Rheinische Hochschule Köln

Theorem

Formel von Bayes (1. Version)
Sind $A, B \subset \Omega$ zufällige Ereignisse mit $P(B) > 0$, dann heißt die Gleichung

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Formel von Bayes.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Theorem

Totale Wahrscheinlichkeit
Sind $A_i \subset \Omega$ $(i = 1, 2, \cdots, n)$ paarweise disjunkte Ereignisse einer Ergebnismenge Ω mit den Wahrscheinlichkeiten $P(A_i)$, wobei $P(A_1) + \cdots + P(A_n) = 1$ gilt, und ist $B \subset \Omega$ ein zufälliges Ereignis, dessen bedingte Wahrscheinlichkeiten $P(B|A_i)$ bezüglich der zufälligen Ereignisse A_i bekannt sind, dann heißt die Gleichung

$$P(B) = \sum_{i=1}^{n} P(B|A_i) * P(A_i)$$

Formel der totalen Wahrscheinlichkeit für das zufällige Ereignis B.

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Theorem

Formel von Bayes (2. Version)
Sind $A_i \subset \Omega$ *(*$i = 1, 2, \cdots, n$*) paarweise disjunkte Ereignisse einer Ergebnismenge* Ω *mit den Wahrscheinlichkeiten* $P(A_i)$*, wobei* $P(A_1) + \cdots + P(A_n) = 1$ *gilt, und* $A, B \subset \Omega$ *zufällige Ereignisse, dessen bedingte Wahrscheinlichkeiten* $P(B|A_i)$ *bezüglich der zufälligen Ereignisse* A_i *bekannt sind, dann heißt die Gleichung*

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(B|A) * P(A)}{\sum_{i=1}^{n} P(B|A_i) * P(A_i)}$$

Formel von Bayes.

Rheinische Hochschule Köln

Beispiel

Eine Offsetdruckerei eines Zeitungsverlages benötigt innerhalb eines Monats 10.000 Rollen Papier. Die entsprechenden Rollen werden von den Papiermühlen A, B und C geliefert. Die Papiermühle A liefert 5.000 Rollen, B und C jeweils 2.500 Rollen. Erfahrungsgemäß haben 5% der von A gelieferten Rollen Qualitätsmängel, 2% der von B gelieferten und 4% der von C gelieferten.
Mit welcher Wahrscheinlichkeit ist eine zufällig ausgewählte Rolle mit Qualitätsmängel aus dem Papierlager der Offsetdruckerei von der

1. Papiermühle A,
2. Papiermühle B und
3. Papiermühle C?

Grundlagen Wahrscheinlichkeitstheorie

Rheinische Hochschule Köln

Beispiel

Erfahrungsgemäß gehen vier Fünftel der Klausurteilnehmenden im Fach Statistik einem Nebenjob nach. Während sich erfahrungsgemäß für Klausurteilnehmende, die einem Nebenjob nachgehen, die Chance, die Klausur im ersten Anlauf zu bestehen, auf sechs zu vier beläuft, beträgt für Klausurteilnehmende, die keinem Nebenjob nachgehen, die Chance, die Klausur im ersten Anlauf zu bestehen, drei zu eins.

1. Geben Sie die Wahrscheinlichkeiten dafür an, dass zufällig ausgewählte Klausurteilnehmende
   1. einem Nebenjob nachgehen.
   2. keinem Nebenjob nachgehen.
   3. bedingt dadurch, dass sie einem Nebenjob nachgehen, im ersten Anlauf die Klausur nicht bestehen.
   4. bedingt dadurch, dass sie keinem Nebenjob nachgehen, im ersten Anlauf die Klausur nicht bestehen.
2. Geben Sie die Wahrscheinlichkeit dafür an, dass zufällig ausgewählte Klausurteilnehmende im ersten Anlauf die Klausur nicht bestehen. Benennen Sie die angewandte Rechenregel.
3. Alle Klausurteilnehmenden, welche die Klausur im ersten Anlauf nicht bestanden haben, nehmen am zusätzlich angebotenen Klausurtraining teil. Geben Sie die Wahrscheinlichkeit dafür an, das zufällig ausgewählte Teilnehmer am Klausurtraining Studierende sind, die einem Nebenjob nachgehten Benennen Sie die angewandte Rechenregel.
4. Geben Sie unter Verwendung des Ergebnisses aus der letzten Teilaufgabe die Ereigniswahrscheinlichkeit dafür an, dass zufällig ausgewählte Teilnehmer am Klausurtraining Studierende sind, die entweder einem Nebenjob oder keinem Nebenjob nachgeht. Wie nennt man das interessierende Ereignis?

Grundlagen Zufallsvariablen/ Wahrscheinlichkeitsverteilungen

Grundlagen Zufallsvariablen

Beispiel

Werfen mit drei unterschiedlichen Münzen

Es werden drei unterschiedliche Münzen geworfen. Die daraus abgeleitete Zufallsvariable X sei Anzahl Kopf.

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Beispiel

Für die stetige Zufallsvariable X mit der Dichtefunktion

$$f(x) = \begin{cases} 2x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Bestimmen Sie die zugehörige Verteilungsfunktion, den Erwartungswert und die Varianz der obigen Zufallsvariablen.

Grundlagen Zufallsvariablen

Definition

Zufallsvariable

Ist Ω die Ergebnismenge eines Zufallsexperimentes, so heißt eine Funktion X, die jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ derart zuordnet, dass $X : \Omega \longrightarrow \mathbb{R}$ und $\omega \longmapsto X(\omega)$ gilt, eine Zufallsvariable X (auf Ω).

Eine Zufallsvariable ist eine Abbildung bzw. Funktion, deren Argumente bzw. Werte auf der Ergebnismenge Ω variieren und in diesem Sinne variabel sind. Der reelle Funktionswert $X(\omega) = a \in \mathbb{R}$ bezeichnet eine Realisation bzw. eine Realisierung einer Zufallsvariablen X. Nach Art des Wertebereiches unterscheidet man zwischen diskreten und stetigen Zufallsvariablen.

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Definition

Zufallsvariable (alternative Definition)
Eine Funktion X heißt eine Zufallsvariable oder stochastische Variable, wenn sie einem Zufallsexperiment zugeordnet ist und folgende Eigenschaften hat:

1. Die Werte von X sind reelle Zahlen
2. Für jede Zahl a und für jedes Intervall I auf der Zahlengeraden ist die Wahrscheinlichkeit des Ereignisses "X hat den Wert a" bzw. "X liegt in dem Intervall I" im Einklang mit den Axiomen von Kolmogorov (Nichtnegativitäts-, Normierungs-, Additivitätsbedingung).

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Definition

Wahrscheinlichkeitsfunktion (diskrete Zufallsvariablen)

Es seien $x_1, x_2, \ldots$ die Werte der Zufallsvariable X und $p_1 = P(X = x_1), p_2 = P(X = x_2), \ldots$ die zugehörige Wahrscheinlichkeiten. Dann heißt die Funktion

$$f(x) = \begin{cases} p_i & \textit{für } x = x_i, i = 1, 2, \ldots \\ 0 & \textit{für alle übrigen } x \end{cases}$$

die Wahrscheinlichkeitsfunktion der Zufallsvariablen X.

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Definition

Verteilungsfunktion

Ist X eine Zufallsvariable, so heißt die für jede reelle Zahl $a \in \mathbb{R}$ definierte Funktion $F_X : \mathbb{R} \longrightarrow [0;1]$ und $a \longmapsto F_X(a) = P(X \leq a) = P(\{\omega \in \Omega : X(\omega) \leq a\})$ Verteilungsfunktion F_X der Zufallsvariable X an der Stelle $a \in \mathbb{R}$.

Grundlagen Zufallsvariablen

Definition

Dichtefunktion (stetige Zufallsvariable)
Existiert eine reellwertige, nichtnegative Funktion mit $f_X : t \longrightarrow f_X(t) >= 0$ und $t \in \mathbb{R}$ in der Form
$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(t)dt$ zu einer Verteilungsfunktion F_X, so heißt f_X die zugehörige Dichtefunktion.

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Wahrscheinlichkeiten

Es gilt im diskreten Fall:

$$P(a < X \leq b) = \Sigma_{i:a<x_i \leq b} f(x_i)$$

Es gilt im stetigen Fall:

$$P(a < X \leq b) = \int_a^b f_X(t)dt = F(b) - F(a)$$

Erwartungswert einer diskreten Zufallsvariablen

Der Erwartungswert einer Zufallsvariablen X entspricht dem arithmetischen Mittel aus der Deskriptiven Statistik und dient wie dieser der Charakterisierung der Verteilung durch einen einzigen Wert.

Definition

Erwartungswert einer diskreten Zufallsvariablen
Für eine diskrete Zufallsvariable X mit den Werten $x_1, x_2, \ldots$ $(x_i \in W)$ und der Wahrscheinlichkeitsfunktion $f(x)$ gilt:

$$\mu = E(X) = x_1 f(x_1) + x_2 f(x_2) + \ldots = \Sigma_{x_i \in W} x_i f(x_i)$$

$E(X)$ wird der Erwartungswert der Zufallsvariable X genannt.

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Varianz einer diskreten Zufallsvariablen

Definition

Varianz/ Standardabweichung einer diskreten Zufallsvariablen
Für eine diskrete Zufallsvariable X mit den Werten $x_1, x_2, \ldots$ ($x_i \in W$) und der Wahrscheinlichkeitsfunktion $f(x)$ gilt:

$$\sigma^2 = Var(X) = (x_1 - E(X))^2 f(x_1) + (x_2 - E(X))^2 f(x_2) + \ldots$$

$$= \Sigma_{x_i \in W}(x_i - E(X))^2 f(x_i)$$

$Var(X)$ wird die Varianz der Zufallsvariable X genannt. Die positive Quadratwurzel

$$\sigma = Std(X) = \sqrt{(\sigma^2)}$$

heißt Standardabweichung der Zufallsvariable X

Grundlagen Zufallsvariablen

Ähnlich wie auch in der Deskriptiven Statistik lässt sich die Varianz mittels folgender Formel etwas einfacher berechnen:

$$\sigma^2 = Var(X) = x_1^2 f(x_1) + x_2^2 f(x_2) + \ldots - (x_1 f(x_1) + x_2 f(x_2) + \ldots)^2$$

$$= E(X^2) - [E(X)]^2$$

Rheinische Hochschule Köln

Erwartungswert einer stetigen Zufallsvariablen

Definition

Erwartungswert einer stetigen Zufallsvariablen
Für eine stetige Zufallsvariable X mit Dichtefunktion $f(x)$ gilt:

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

$E(X)$ wird der Erwartungswert der Zufallsvariable X genannt.

Rheinische Hochschule Köln

Varianz einer stetigen Zufallsvariablen

Definition

Varianz/ Standardabweichung einer stetigen Zufallsvariablen
Für eine stetige Zufallsvariable X mit Dichtefunktion $f(x)$ gilt:

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

$Var(X)$ wird die Varianz der Zufallsvariable X genannt. Die positive Quadratwurzel

$$\sigma = Std(X) = \sqrt{(\sigma^2)}$$

ist Standardabweichung der Zufallsvariable X.

Auch hier gilt: $\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$

Rheinische Hochschule Köln

Theorem

Ungleichung von Tschebyscheff
Die Ungleichung von Tschebyscheff gibt eine Schätzung ab, mit welcher Wahrscheinlichkeit eine (kardinal skalierte) Zufallsgröße X in einem bestimmten Intervall um den Erwartungswert mindestens zu liegen kommt:

$$P(\mu - 2\sigma < X < \mu + 2\sigma) > \frac{3}{4}$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) > \frac{8}{9}$$

Allgemein für $c \in \mathbb{R}$

$$P(\mu - c\sigma < X < \mu + c\sigma) > 1 - \frac{1}{c^2}$$

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

- Exkurs: Binomialverteilung (diskrete Verteilung)
- Exkurs: Normalverteilung (stetige Verteilung)

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Binomialverteilung

Definition

Wahrscheinlichkeitsfunktion einer Binomialverteilung Bin(n;p)
Es seien $0, 1, \ldots, n$ die Werte einer Zufallsvariablen X. Dann heißt die diskrete Wahrscheinlichkeitsverteilung

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x = 0, 1, 2, \ldots, n \\ 0 & \text{für alle übrigen } x \end{cases}$$

die Binomialverteilung zu den Parametern n (Anzahl der Versuche) und $p \in [0; 1]$ (der Erfolgs- oder Trefferwahrscheinlichkeit).

Rheinische Hochschule Köln

Binomialverteilung

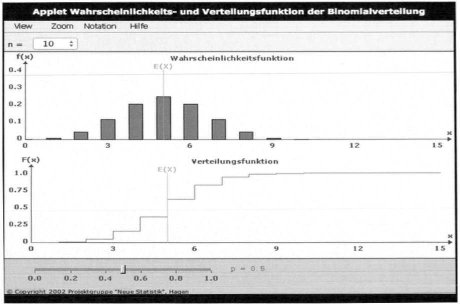

Abbildung: Binomialverteilung

Rheinische Hochschule Köln

Normalverteilung

Definition

Dichtefunktion einer Normalverteilung
Eine Zufallsvariable X heißt N(μ;σ)-verteilt,
falls ihre Dichtefunktion wie folgt definiert ist:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Definition

Standardnormalverteilung

Eine Standardnormalverteilung ist der Sonderfall einer Normalverteilung mit $\mu = 0$ und $\sigma = 1$.

Standardnormalverteilung

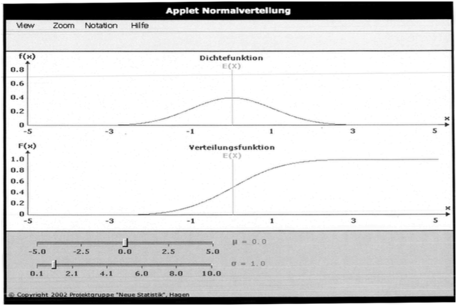

Abbildung: Standardnormalverteilung

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Typische Eigenschaften:

- Glockenförmiger Verlauf der Dichtefunktion
- Symmetrische Dichtefunktion, die ihren zugehörigen Erwartungwert bei $x = \mu$ hat, also dem einen Parameter der Normalverteilung.
- Median und Erwartungswert liegen identisch in der "Mitte"der Dichtefunktion
- Die Dichtefunktion nähert sich asymptotisch der Abzisse (x-Achse).
- Die Standardabweichung ist identisch mit σ, dem zweiten Parameter der Normalverteilung.
- Die Verteilungsfunktion einer Standardnormalverteilung wird oft auch mit Φ bezeichnet.
  Ist Z eine standardnormalverteilte Zufallsvariable, dann gilt
  $\Phi(a) = F_Z(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Theorem

Ist Z eine standardnormalverteilte Zufallsvariable (N(0;1)), dann gilt

$$\Phi(-a) = 1 - \Phi(a).$$

Beispiel

Sei Z eine standardnormalverteilte Zufallsvariable, d.h. $\mu = 0$ und $\sigma = 1$.
Gesucht ist die Wahrscheinlichkeit

$$P(-1 < Z \leq 2,56).$$

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Theorem

Z-Transformation
Ist X eine N(μ;σ)-verteilte Zufallsvariable, dann ist die Zufallsvariable

$$Z = \frac{X - \mu}{\sigma}$$

standardnormalverteilt (N(0;1)).

Beispiel

Die Zufallsvariable X sei N(2;3)-verteilt.
Gesucht ist die Wahrscheinlichkeit

$$P(X \leq 5).$$

Rheinische Hochschule Köln

Beispiel

Die Zufallsvariable X sei N(3;4)-verteilt.
Gesucht ist die Wahrscheinlichkeit

$$P(2 < X \leq 7).$$

Grundlagen Zufallsvariablen

Umgekehrt gilt auch:

Theorem

Ist Z eine N(0;1)-verteilte Zufallsvariable, dann ist die Zufallsvariable

$$X = \mu + Z * \sigma$$

eine N(μ;σ)-verteilte Zufallsvariable.

Theorem

p-Quantile einer Normalverteilung
Das p-Quantil einer normalverteilten Zufallsvariablen X mit Erwartungswert μ und Standardabweichung σ ergibt sich aus dem zugehörigen p-Quantil z_p der Standardanormalverteilung wie folgt:

$$x_p = \mu + z_p * \sigma.$$

Um die Gültigkeit dieses Theorems zu beweisen, muss gezeigt werden, dass

$$F_X(x_p) = p.$$

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Theorem

„Sigma-Regeln“
Für eine Zufallsvariable X, die normalverteilt ist mit Erwartungswert μ und Standardabweichung σ gilt:

$$P(\mu - 1\sigma < X \leq \mu + 1\sigma) = 0,6826$$

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) = 0,9544$$

$$P(\mu - 3\sigma < X \leq \mu + 3\sigma) = 0,9974$$

Grundlagen Zufallsvariablen

Rheinische Hochschule Köln

Beispiel

Das Gewicht beim Abfüllen einer Zuckertüte soll normalverteilt sein mit Erwartungswert 1000 g. und Standardabweichung 4 g.
Wie viel Prozent der Zuckertüten wiegen mehr als 1008 g?

Rheinische Hochschule Köln

Vielen Dank

Kontaktdaten:

Prof. Dr. Jürgen Krob
Dr. rer. nat., Dipl.-Math. oec.

E-Mail: juergen.krob@rh-koeln.de

Rheinische Hochschule Köln gGmbH
University of Applied Sciences
Schaevenstraße 1 a/b
50676 Köln
URL: www.rh-koeln.de