Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

Kapitel VIII - Mehrdimensionale Merkmale

Deskriptive Statistik

Prof. Dr. W.-D. Heller
Hartwig Senska
Carlo Siebenschuh

Agenda

1. **Einleitung**
2. Absolute bzw. relative Häufigkeiten
3. Graphische Darstellung
4. Zusammenhang zweier Merkmale
5. Überprüfung auf Unabhängigkeit

Einleitung

Häufig werden bei einer statistischen Untersuchung mehrere Merkmale auf einer statistischen Masse erhoben, d.h. es werden die Merkmalsausprägungen zu mehreren Merkmalen einer statistischen Einheit festgehalten.

Beispiele sind etwa

(1) bei der Untersuchung des Bremsverhaltens eines PKW die Geschwindigkeit bei Beginn des Bremsvorgangs, die Länge des Bremsweges, die Reifenarten, die Profiltiefe der Reifen etc.,

(2) bei der Umsatzstatistik eines Automobilkonzerns neben regionalen Aspekten auch Modelltyp, Ausstattungsvariante, Farbe, Eigenschaften des Käufers etc.

Einleitung

Häufig werden bei einer statistischen Untersuchung mehrere Merkmale auf einer statistischen Masse erhoben, d.h. es werden die Merkmalsausprägungen zu mehreren Merkmalen einer statistischen Einheit festgehalten.

Beispiele sind etwa

(1) bei der Untersuchung des Bremsverhaltens eines PKW die Geschwindigkeit bei Beginn des Bremsvorgangs, die Länge des Bremsweges, die Reifenarten, die Profiltiefe der Reifen etc.,

(2) bei der Umsatzstatistik eines Automobilkonzerns neben regionalen Aspekten auch Modelltyp, Ausstattungsvariante, Farbe, Eigenschaften des Käufers etc.

Einleitung

Häufig werden bei einer statistischen Untersuchung mehrere Merkmale auf einer statistischen Masse erhoben, d.h. es werden die Merkmalsausprägungen zu mehreren Merkmalen einer statistischen Einheit festgehalten.

Beispiele sind etwa

(1) bei der Untersuchung des Bremsverhaltens eines PKW die Geschwindigkeit bei Beginn des Bremsvorgangs, die Länge des Bremsweges, die Reifenarten, die Profiltiefe der Reifen etc.,

(2) bei der Umsatzstatistik eines Automobilkonzerns neben regionalen Aspekten auch Modelltyp, Ausstattungsvariante, Farbe, Eigenschaften des Käufers etc.

Einleitung

Das Ergebnis der Untersuchung einer statistischen Masse S auf mehrere (etwa $k \geq 2$) Merkmale besteht also zunächst in einer Tabelle, bei der für jede statistische Einheit die Merkmalswerte der einzelnen Merkmale aufgelistet sind.

		Merkmale				
		1	2	3		k
	s_1	x_{11}	x_{12}	x_{13}	$\dots$	x_{1k}
	s_2	x_{21}	x_{22}	x_{23}	$\dots$	x_{2k}
stat.	.	.				.
Einheiten	.	.				.
	.	.				.
	s_n	x_{n1}	x_{n2}	x_{n3}	$\dots$	x_{nk}

Dabei ist z.B. x_{23} Merkmalswert der statistischen Einheit s_2 bei Merkmal 3.

Einleitung

Das Ergebnis der Untersuchung einer statistischen Masse S auf mehrere (etwa $k \geq 2$) Merkmale besteht also zunächst in einer Tabelle, bei der für jede statistische Einheit die Merkmalswerte der einzelnen Merkmale aufgelistet sind.

		Merkmale				
		1	2	3		k
	s_1	x_{11}	x_{12}	x_{13}	...	x_{1k}
	s_2	x_{21}	x_{22}	x_{23}	...	x_{2k}
stat.	.	.				.
Einheiten	.	.				.
	.	.				.
	s_n	x_{n1}	x_{n2}	x_{n3}	...	x_{nk}

Dabei ist z.B. x_{23} Merkmalswert der statistischen Einheit s_2 bei Merkmal 3.

Einleitung

Sei also S eine statistische Masse, so liegt nun nicht mehr ein einzelnes Merkmal vor:

$$b : S \mapsto M,$$

sondern $k \geq 2$ Merkmale

$$b_1 : S \mapsto M, \quad \dots \quad , b_k : S \mapsto M.$$

Die Merkmalswerte $b_1(s), \dots, b_k(s)$ einer statistischen Einheit s kann man zusammenfassen zu einem k-Tupel (Merkmals-k-Tupel)

$$(b_1(s), \dots, b_k(s)).$$

Man spricht daher auch von einem **k-dimensionalen Merkmal.**

Einleitung

Sei also S eine statistische Masse, so liegt nun nicht mehr ein einzelnes Merkmal vor:

$$b : S \mapsto M,$$

sondern $k \geq 2$ Merkmale

$$b_1 : S \mapsto M, \quad \ldots \quad , b_k : S \mapsto M.$$

Die Merkmalswerte $b_1(s), \ldots, b_k(s)$ einer statistischen Einheit s kann man zusammenfassen zu einem k-Tupel (Merkmals-k-Tupel)

$$(b_1(s), \ldots, b_k(s)).$$

Man spricht daher auch von einem **k-dimensionalen Merkmal.**

Einleitung

Sei also S eine statistische Masse, so liegt nun nicht mehr ein einzelnes Merkmal vor:

$$b : S \mapsto M,$$

sondern $k \geq 2$ Merkmale

$$b_1 : S \mapsto M, \quad ... \quad , b_k : S \mapsto M.$$

Die Merkmalswerte $b_1(s), ..., b_k(s)$ einer statistischen Einheit s kann man zusammenfassen zu einem k-Tupel (Merkmals-k-Tupel)

$$(b_1(s), ..., b_k(s)).$$

Man spricht daher auch von einem **k-dimensionalen Merkmal.**

Einleitung

Zweidimensionale Merkmale

Bei zwei Merkmalen erhält man für jede statistische Einheit ein Paar von Beobachtungswerten. Die Urliste besteht daher aus n Paaren von Beobachtungswerten (Beobachtungspaaren), wobei n die Anzahl der statistischen Einheiten ist:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n).$$

Einleitung

Beispiel 8.1
In der Situation von Beispiel 4.2 ist neben der Punktzahl im Leistungskurs Mathematik noch die Punktzahl im Grundkurs Deutsch beobachtet worden. Man erhält folgende Tabelle:

Schüler	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
Punktzahl in: Mathematik	8	14	9	13	8	12	9	11	12	9
Deutsch	5	9	9	8	12	8	10	10	9	10
	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}
Mathematik	12	14	10	12	9	7	11	12	13	9
Deutsch	7	13	9	11	6	14	12	8	6	5

Einleitung

Beispiel 8.1
In der Situation von Beispiel 4.2 ist neben der Punktzahl im Leistungskurs Mathematik noch die Punktzahl im Grundkurs Deutsch beobachtet worden. Man erhält folgende Tabelle:

Schüler	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
Punktzahl in: Mathematik	8	14	9	13	8	12	9	11	12	9
Deutsch	5	9	9	8	12	8	10	10	9	10
	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}
Mathematik	12	14	10	12	9	7	11	12	13	9
Deutsch	7	13	9	11	6	14	12	8	6	5

Die Urliste lautet damit:

$(8,5),(14,9),(9,9),(13,8),(8,12),(12,8),(9,10),(11,10),(12,9),(9,10),$

$(12,7),(14,13),(10,9),(12,11),(9,6),(7,14),(11,12),(12,8),(13,6),(9,5)$

Einleitung

Lexikographische Ordnung
Ordnungsprinzip für mehrdimensionale Merkmale. Zunächst werden die Wertepaare (bei zweidimensionalen Merkmalen) nach dem ersten Beobachtungswert ordnet. Stimmen bei zwei Merkmalspaaren die ersten "Komponenten" überein, so richtet sich die Reihenfolge nach dem zweiten Merkmal.

In Beispiel 8.1 ergibt sich somit folgende geordnete Urliste:

(7,14), (8,5), (8,12), (9,5), (9,6), (9,9), (9,10), (9,10), (10,9)
(11,10), (11,12), (12,7), (12,8), (12,8), (12,9), (12,11), (13,6)
(13,8), (14,9), (14,13)

Agenda

1. Einleitung
2. **Absolute bzw. relative Häufigkeiten**
3. Graphische Darstellung
4. Zusammenhang zweier Merkmale
5. Überprüfung auf Unabhängigkeit

Absolute bzw. relative Häufigkeiten

Absolute Häufigkeit der Merkmalskombination (a_i,b_j)

$$h(a_i, b_j)$$

Anzahl des Auftretens der Merkmalskombination (a_i, b_j) in der Urliste, wobei

a_i die Merkmalsausprägung i des ersten Merkmals,
b_j die Merkmalsausprägung j des zweiten Merkmals ist.

Die Summe aller absoluten Häufigkeiten ergibt damit die Anzahl der Beobachtungen:

$$n = \sum_{i=1}^{r} \sum_{j=1}^{s} h(a_i, b_j)$$

Absolute bzw. relative Häufigkeiten

Absolute Häufigkeit der Merkmalskombination (a_i,b_j)

$$h(a_i, b_j)$$

Anzahl des Auftretens der Merkmalskombination (a_i, b_j) in der Urliste, wobei

a_i die Merkmalsausprägung i des ersten Merkmals,
b_j die Merkmalsausprägung j des zweiten Merkmals ist.

Die Summe aller absoluten Häufigkeiten ergibt damit die Anzahl der Beobachtungen:

$$n = \sum_{i=1}^{r} \sum_{j=1}^{s} h(a_i, b_j)$$

Absolute bzw. relative Häufigkeiten

Relative Häufigkeit der Merkmalskombination (a_i,b_j)

$$p(a_i, b_j) = \frac{1}{n} \cdot h(a_i, b_j)$$

Absolute bzw. relative Häufigkeiten

Die Gesamtheit aller Merkmalskombinationen zusammen mit ihren absoluten Häufigkeiten heißt **zweidimensionale Häufigkeitsverteilung.**

Die tabellarische Darstellung einer zweidimensionalen Häufigkeitsverteilung von nominalskalierten Merkmalen heißt **Kontingenztabelle**. Sind beide Merkmale mindestens ordinalskaliert (Rangmerkmal oder quantitatives Merkmal), so nennt man sie die **Korrelationstabelle.**

Absolute bzw. relative Häufigkeiten

Die Gesamtheit aller Merkmalskombinationen zusammen mit ihren absoluten Häufigkeiten heißt **zweidimensionale Häufigkeitsverteilung.**

Die tabellarische Darstellung einer zweidimensionalen Häufigkeitsverteilung von nominalskalierten Merkmalen heißt **Kontingenztabelle**. Sind beide Merkmale mindestens ordinalskaliert (Rangmerkmal oder quantitatives Merkmal), so nennt man sie die **Korrelationstabelle.**

	b_1	b_2	b_3	$\cdots$	b_s
a_1	$h(a_1, b_1)$	$h(a_1, b_2)$	$h(a_1, b_3)$	$\cdots$	$h(a_1, b_s)$
a_2	$h(a_2, b_1)$	$h(a_2, b_2)$	$h(a_2, b_3)$	$\cdots$	$h(a_2, b_s)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_r	$h(a_r, b_1)$	$h(a_r, b_2)$	$h(a_r, b_3)$	$\cdots$	$h(a_r, b_s)$

Abbildung 8.2 - Zweidimenionale Häufigkeitstabelle

Absolute bzw. relative Häufigkeiten

Beispiel 8.2
Zu Beispiel 8.1 ergibt sich folgende Korrelationstabelle:

		Punkte in Mathematik									
		6	7	8	9	10	11	12	13	14	15
	5	-	-	1	1	-	-	-	-	-	-
	6	-	-	-	1	-	-	-	1	-	-
	7	-	-	-	-	-	-	1	-	-	-
	8	-	-	-	-	-	-	2	1	-	-
Punkte	9	-	-	-	1	1	-	1	-	1	-
in	10	-	-	-	2	-	1	-	-	-	-
Deutsch	11	-	-	-	-	-	-	1	-	-	-
	12	-	-	1	-	-	1	-	-	-	-
	13	-	-	-	-	-	-	-	-	1	-
	14	-	1	-	-	-	-	-	-	-	-
	15	-	-	-	-	-	-	-	-	-	-

Absolute bzw. relative Häufigkeiten

Beispiel 8.3

Eine Untersuchung bei 50 Familien hat folgendes Ergebnis über die Anzahl der im Haushalt lebenden Kinder und der Religionszugehörigkeit der Mutter erbracht:

		Kinderzahl					
		0	1	2	3	4	
Religions-	rk	1	5	7	2	2	17
zu-	ev	2	4	10	5	-	21
gehörigkeit	sonst	2	1	7	2	-	12
		5	10	24	9	2	50

Die Tabelle der relativen Häufigkeiten ist dann:

		Kinderzahl					
		0	1	2	3	4	
Religions-	rk	.02	.10	.14	.04	.04	.34
zu-	ev	.04	.08	.20	.10	-	.42
gehörigkeit	sonst	.04	.02	.14	.04	-	.24
		.1	.2	.48	.18	.04	1

Absolute bzw. relative Häufigkeiten

Beispiel 8.3
Eine Untersuchung bei 50 Familien hat folgendes Ergebnis über die Anzahl der im Haushalt lebenden Kinder und der Religionszugehörigkeit der Mutter erbracht:

		Kinderzahl					
		0	1	2	3	4	
Religions-	rk	1	5	7	2	2	17
zu-	ev	2	4	10	5	-	21
gehörigkeit	sonst	2	1	7	2	-	12
		5	10	24	9	2	50

Die Tabelle der relativen Häufigkeiten ist dann:

		Kinderzahl					
		0	1	2	3	4	
Religions-	rk	.02	.10	.14	.04	.04	.34
zu-	ev	.04	.08	.20	.10	-	.42
gehörigkeit	sonst	.04	.02	.14	.04	-	.24
		.1	.2	.48	.18	.04	1

Absolute bzw. relative Häufigkeiten

Klassierte Merkmale

(1) Sind beide Merkmale klassiert mit den Klassen $I_1, ..., I_k$ für das erste Merkmal und $J_1, ..., J_m$ für das zweite Merkmal, so ist

$$\begin{aligned} h(I_j, J_t) &= \#\{(x_i, y_i) \mid x_i \in I_j, y_i \in J_t\} \\ p(I_j, J_t) &= \tfrac{1}{n} \cdot h(I_j, J_t), j = 1, ..., k, t = 1, ..., m. \end{aligned}$$

(2) Ist das erste Merkmal klassiert in den Klassen $I_1, ..., I_k$, das zweite nicht, so ist

$$\begin{aligned} h(I_j, b_t) &= \#\{(x_i, y_i) \mid x_i \in I_j, y_i = b_t\} \\ p(I_j, b_t) &= \tfrac{1}{n} \cdot h(I_j, b_t), j = 1, ..., k, t = 1, ..., s. \end{aligned}$$

Umgekehrt analog.

Absolute bzw. relative Häufigkeiten

Klassierte Merkmale

(1) Sind beide Merkmale klassiert mit den Klassen $I_1, ..., I_k$ für das erste Merkmal und $J_1, ..., J_m$ für das zweite Merkmal, so ist

$$\begin{aligned} h(I_j, J_t) &= \#\{(x_i, y_i) \mid x_i \in I_j, y_i \in J_t\} \\ p(I_j, J_t) &= \tfrac{1}{n} \cdot h(I_j, J_t), j = 1, ..., k, t = 1, ..., m. \end{aligned}$$

(2) Ist das erste Merkmal klassiert in den Klassen $I_1, ..., I_k$, das zweite nicht, so ist

$$\begin{aligned} h(I_j, b_t) &= \#\{(x_i, y_i) \mid x_i \in I_j, y_i = b_t\} \\ p(I_j, b_t) &= \tfrac{1}{n} \cdot h(I_j, b_t), j = 1, ..., k, t = 1, ..., s. \end{aligned}$$

Umgekehrt analog.

Absolute bzw. relative Häufigkeiten

Klassierte Merkmale

(1) Sind beide Merkmale klassiert mit den Klassen $I_1, ..., I_k$ für das erste Merkmal und $J_1, ..., J_m$ für das zweite Merkmal, so ist

$$\begin{aligned} h(I_j, J_t) &= \#\{(x_i, y_i) \mid x_i \in I_j, y_i \in J_t\} \\ p(I_j, J_t) &= \tfrac{1}{n} \cdot h(I_j, J_t), j = 1, ..., k, t = 1, ..., m. \end{aligned}$$

(2) Ist das erste Merkmal klassiert in den Klassen $I_1, ..., I_k$, das zweite nicht, so ist

$$\begin{aligned} h(I_j, b_t) &= \#\{(x_i, y_i) \mid x_i \in I_j, y_i = b_t\} \\ p(I_j, b_t) &= \tfrac{1}{n} \cdot h(I_j, b_t), j = 1, ..., k, t = 1, ..., s. \end{aligned}$$

Umgekehrt analog.

Agenda

1. Einleitung
2. Absolute bzw. relative Häufigkeiten
3. **Graphische Darstellung**
4. Zusammenhang zweier Merkmale
5. Überprüfung auf Unabhängigkeit

Graphische Darstellung

Bei der graphischen Darstellung zwei- und mehrdimensionaler Merkmale ergeben sich dieselben Prinzipien wie im eindimensionalen Fall, wobei im Unterschied zum eindimensionalen Fall die absoluten bzw. relativen Häufigkeiten jetzt Kombinationen von Merkmalsausprägungen zugeordnet sind.

Graphische Darstellung

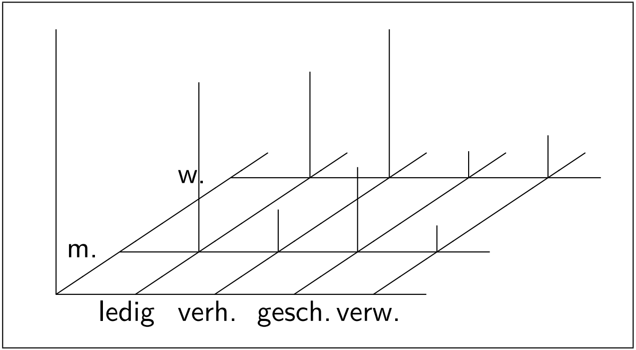


Abbildung 8.3 - Stabdiagramm eines zweidimensionalen Merkmals

Graphische Darstellung

Anstelle der dritten Dimension besteht noch die Möglichkeit der Differenzierung mit Farben, unterschiedlich starken Grautönen oder verschiedenartigen Schraffierungen.

Abbildung 8.4 - Graphische Darstellung mit unterschiedlicher Schattierung

Graphische Darstellung

Anstelle der dritten Dimension besteht noch die Möglichkeit der Differenzierung mit Farben, unterschiedlich starken Grautönen oder verschiedenartigen Schraffierungen.

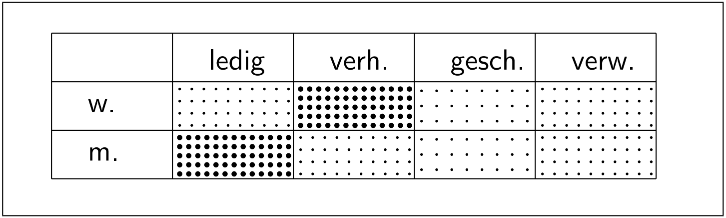


Abbildung 8.4 - Graphische Darstellung mit unterschiedlicher Schattierung

Graphische Darstellung

Histogramm
Bei der graphischen Darstellung von klassierten Daten eines zweidimensionalen Merkmals erhält man anstelle des zweidimensionalen Histogramms eine dreidimensionale Figur, wobei bei Abbildungen wieder eine perspektivische Darstellung verwendet wird.

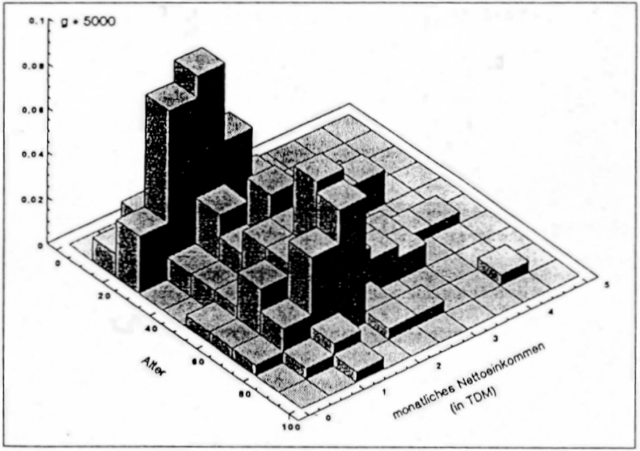
Beachte: Es handelt sich hierbei um eine **volumenproportionale** Darstellung. Bei unterschiedlichen Klassenbreiten bei beiden Merkmalen ist die unterschiedliche Grundfläche jeder Säule zu berücksichtigen.

Graphische Darstellung

Histogramm
Bei der graphischen Darstellung von klassierten Daten eines zweidimensionalen Merkmals erhält man anstelle des zweidimensionalen Histogramms eine dreidimensionale Figur, wobei bei Abbildungen wieder eine perspektivische Darstellung verwendet wird.

Beachte: Es handelt sich hierbei um eine **volumenproportionale** Darstellung. Bei unterschiedlichen Klassenbreiten bei beiden Merkmalen ist die unterschiedliche Grundfläche jeder Säule zu berücksichtigen.

Graphische Darstellung



Graphische Darstellung

Streuungsdiagramm
Trägt man die Beobachtungspaare bei quantitativen Merkmalen in einem Koordinatensystem ein, so erhält man ein sogenanntes Streuungsdiagramm. Durch das Streuungsdiagramm kann man einen Anhaltspunkt dafür erhalten, ob (und welcher Art) möglicherweise ein Zusammenhang zwischen den beiden Merkmalen besteht.

Graphische Darstellung

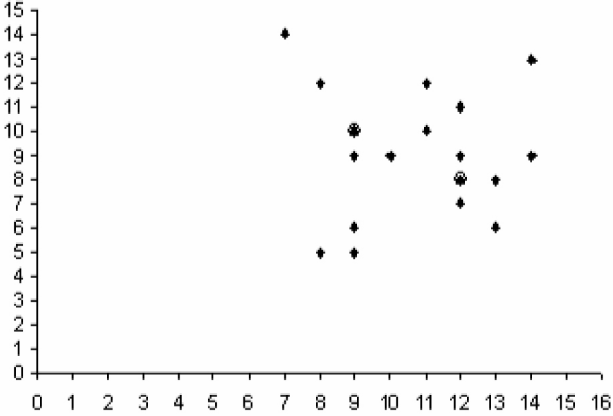


Abbildung: Streudiagramm zu Beispiel 8.1

Agenda

1. Einleitung
2. Absolute bzw. relative Häufigkeiten
3. Graphische Darstellung
4. **Zusammenhang zweier Merkmale**
5. Überprüfung auf Unabhängigkeit

Zusammenhang zweier Merkmale

Randverteilung
Die Häufigkeit einer Merkmalsausprägung a_i bzw. b_j unter Vernachlässigung des jeweils anderen Merkmals erhält man durch Summation der Häufigkeiten der entsprechenden Zeile bzw. Spalte.
Da diese Häufigkeiten üblicherweise in einer Randspalte bzw. Randzeile eingetragen werden, spricht man von der Randverteilung.

Die Summe der Randspalte und der Randzeile muss bei den absoluten Häufigkeiten jeweils die Gesamtzahl n der Beobachtungen ergeben. Bei den relativen Häufigkeiten ist diese Summe jeweils 1.

Zusammenhang zweier Merkmale

Randverteilung
Die Häufigkeit einer Merkmalsausprägung a_i bzw. b_j unter Vernachlässigung des jeweils anderen Merkmals erhält man durch Summation der Häufigkeiten der entsprechenden Zeile bzw. Spalte.
Da diese Häufigkeiten üblicherweise in einer Randspalte bzw. Randzeile eingetragen werden, spricht man von der Randverteilung.

Die Summe der Randspalte und der Randzeile muss bei den absoluten Häufigkeiten jeweils die Gesamtzahl n der Beobachtungen ergeben. Bei den relativen Häufigkeiten ist diese Summe jeweils 1.

Zusammenhang zweier Merkmale

	b_1	b_2	b_3	$\cdots$	b_s	
a_1	$h(a_1, b_1)$	$h(a_1, b_2)$	$h(a_1, b_3)$	$\cdots$	$h(a_1, b_s)$	$h(a_1)$
a_2	$h(a_2, b_1)$	$h(a_2, b_2)$	$h(a_2, b_3)$	$\cdots$	$h(a_2, b_s)$	$h(a_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
a_r	$h(a_r, b_1)$	$h(a_r, b_2)$	$h(a_r, b_3)$	$\cdots$	$h(a_r, b_s)$	$h(a_r)$
	$h(b_1)$	$h(b_2)$	$h(b_3)$	$\cdots$	$h(b_s)$	n

Abb. 8.7 - Zweidimensionale Häufigkeitsverteilung mit Randverteilung

Zusammenhang zweier Merkmale

Randverteilung zu Beispiel 8.2 in absoluten Häufigkeiten:

		Punkte in Mathematik										
		6	7	8	9	10	11	12	13	14	15	
	5	-	-	1	1	-	-	-	-	-	-	2
	6	-	-	-	1	-	-	-	1	-	-	2
	7	-	-	-	-	-	-	1	-	-	-	1
	8	-	-	-	-	-	-	2	1	-	-	3
Punkte	9	-	-	-	1	1	-	1	-	1	-	4
in	10	-	-	-	2	-	1	-	-	-	-	3
Deutsch	11	-	-	-	-	-	-	1	-	-	-	1
	12	-	-	1	-	-	1	-	-	-	-	2
	13	-	-	-	-	-	-	-	-	1	-	1
	14	-	1	-	-	-	-	-	-	-	-	1
	15	-	-	-	-	-	-	-	-	-	-	0
	0	-	1	2	5	1	2	5	2	2	0	20

Zusammenhang zweier Merkmale

Beispiel 8.5

Randverteilung zu Beispiel 8.3 in relativen Häufigkeiten:

		Kinderzahl					
		0	1	2	3	4	
Religions-	rk	.02	.10	.14	.04	.04	.34
zu-	ev	.04	.08	.20	.10	-	.42
gehörigkeit	sonst	.04	.02	.14	.04	-	.24
		.10	.20	.48	.18	.04	1.0

Zusammenhang zweier Merkmale

Gegenseitige Beeinflussung zweier Merkmale
Um nun den Einfluss (falls vorhanden) des einen Merkmals auf das andere feststellen zu können, fixiert man zunächst eine bestimmte Ausprägung dieses Merkmals (a sei diese Merkmalsausprägung) und betrachtet nur die statistischen Einheiten, die diese Merkmalsausprägung tragen.

Zusammenhang zweier Merkmale

Beispiel 8.6
Betrachtet man die Merkmale Geschlecht und Körpergröße bei 20 Schüler(inne)n einer Abiturklasse erhält man die folgende Häufigkeitsverteilung:

	Körpergröße (über bis einschließlich)						
Geschlecht	140-150	150-160	160-170	170-180	180-190	190-200	$\sum$
weiblich	1	2	5	1	0	0	9
männlich	0	0	1	5	3	2	11

Man erkennt deutlich die Unterschiede des Merkmals Körpergröße bei den Schülerinnen gegenüber den Schülern.

Zusammenhang zweier Merkmale

Beispiel 8.6

Relative Häufigkeiten der Körpergrößen der Schülerinnen:

140-150	150-160	160-170	170-180	180-190	190-200
$\frac{1}{9}$	$\frac{2}{9}$	$\frac{5}{9}$	$\frac{1}{9}$	0	0

Relative Häufigkeiten der Schüler:

140-150	150-160	160-170	170-180	180-190	190-200
0	0	$\frac{1}{11}$	$\frac{5}{11}$	$\frac{3}{11}$	$\frac{2}{11}$

Dieser Vergleich der Häufigkeitsverteilungen zeigt klare Unterschiede. Also hat das Merkmal Geschlecht Einfluss auf das Merkmal Körpergröße. Nur wenn sie identisch wären, könnte man einen Einfluss ausschließen.

Zusammenhang zweier Merkmale

Beispiel 8.6

Relative Häufigkeiten der Körpergrößen der Schülerinnen:

140-150	150-160	160-170	170-180	180-190	190-200
$\frac{1}{9}$	$\frac{2}{9}$	$\frac{5}{9}$	$\frac{1}{9}$	0	0

Relative Häufigkeiten der Schüler:

140-150	150-160	160-170	170-180	180-190	190-200
0	0	$\frac{1}{11}$	$\frac{5}{11}$	$\frac{3}{11}$	$\frac{2}{11}$

Dieser Vergleich der Häufigkeitsverteilungen zeigt klare Unterschiede. Also hat das Merkmal Geschlecht Einfluss auf das Merkmal Körpergröße. Nur wenn sie identisch wären, könnte man einen Einfluss ausschließen.

Zusammenhang zweier Merkmale

Beispiel 8.6

Relative Häufigkeiten der Körpergrößen der Schülerinnen:

140-150	150-160	160-170	170-180	180-190	190-200
$\frac{1}{9}$	$\frac{2}{9}$	$\frac{5}{9}$	$\frac{1}{9}$	0	0

Relative Häufigkeiten der Schüler:

140-150	150-160	160-170	170-180	180-190	190-200
0	0	$\frac{1}{11}$	$\frac{5}{11}$	$\frac{3}{11}$	$\frac{2}{11}$

Dieser Vergleich der Häufigkeitsverteilungen zeigt klare Unterschiede. Also hat das Merkmal Geschlecht Einfluss auf das Merkmal Körpergröße. Nur wenn sie identisch wären, könnte man einen Einfluss ausschließen.

Zusammenhang zweier Merkmale

Bedingte Häufigkeit
Sei $S_a = \{s \in S \mid b_1(s) = a\}$ die Menge der statistischen Einheiten, die bei Merkmal 1 die Merkmalsausprägung a tragen. S_a nennen wir **reduzierte statistische Masse unter der Bedingung $\boldsymbol{a}$**.

Zu Bsp. 8.1: Alle Schüler, die in Mathematik=Merkmal die Merkmalsausprägung a=8 Punkte haben: $S_a = S_8 = \{s_1, s_5\}$
Die Anzahl der statistischen Einheiten in S_a (die absolute Häufigkeit) ist $h(a)$ und ergibt sich aus der Randverteilung.
($h(a)$=h(8 Punkte in Mathematik)=2)
Absolute Häufigkeit einer Merkmalsausprägung b von Merkmal 2 in der reduzierten statistischen Masse S_a unter der Bedingung a ist $h(a, b)$.
($b = 5$ Punkte in Deutsch, somit $h(a, b)$=$h(8, 5)$=1)

Zu Bsp. 8.1: Alle Schüler, die in Mathematik=Merkmal die Merkmalsausprägung a=8 Punkte haben: $S_a = S_8 = \{s_1, s_5\}$
Die Anzahl der statistischen Einheiten in S_a (die absolute Häufigkeit) ist $h(a)$ und ergibt sich aus der Randverteilung.
($h(a)$=h(8 Punkte in Mathematik)=2)
Absolute Häufigkeit einer Merkmalsausprägung b von Merkmal 2 in der reduzierten statistischen Masse S_a unter der Bedingung a ist $h(a, b)$.
($b = 5$ Punkte in Deutsch, somit $h(a, b)$=$h(8, 5)$=1)

Zu Bsp. 8.1: Alle Schüler, die in Mathematik=Merkmal die Merkmalsausprägung a=8 Punkte haben: $S_a = S_8 = \{s_1, s_5\}$
Die Anzahl der statistischen Einheiten in S_a (die absolute Häufigkeit) ist $h(a)$ und ergibt sich aus der Randverteilung.
($h(a)$=h(8 Punkte in Mathematik)=2)
Absolute Häufigkeit einer Merkmalsausprägung b von Merkmal 2 in der reduzierten statistischen Masse S_a unter der Bedingung a ist $h(a, b)$.
($b = 5$ Punkte in Deutsch, somit $h(a, b)$=$h(8, 5)$=1)

Zu Bsp. 8.1: Alle Schüler, die in Mathematik=Merkmal die Merkmalsausprägung a=8 Punkte haben: $S_a = S_8 = \{s_1, s_5\}$
Die Anzahl der statistischen Einheiten in S_a (die absolute Häufigkeit) ist $h(a)$ und ergibt sich aus der Randverteilung.
($h(a)$=h(8 Punkte in Mathematik)=2)
Absolute Häufigkeit einer Merkmalsausprägung b von Merkmal 2 in der reduzierten statistischen Masse S_a unter der Bedingung a ist $h(a, b)$.
($b = 5$ Punkte in Deutsch, somit $h(a, b)$=$h(8, 5)$=1)

Zu Bsp. 8.1: Alle Schüler, die in Mathematik=Merkmal die Merkmalsausprägung a=8 Punkte haben: $S_a = S_8 = \{s_1, s_5\}$
Die Anzahl der statistischen Einheiten in S_a (die absolute Häufigkeit) ist $h(a)$ und ergibt sich aus der Randverteilung.
($h(a)$=h(8 Punkte in Mathematik)=2)
Absolute Häufigkeit einer Merkmalsausprägung b von Merkmal 2 in der reduzierten statistischen Masse S_a unter der Bedingung a ist $h(a, b)$.
($b = 5$ Punkte in Deutsch, somit $h(a, b)$=$h(8, 5)$=1)

Damit erhält man die relative Häufigkeit der Merkmalsausprägung b in der unter der Bedingung a reduzierten statistischen Masse S_a

$$p(b \mid a) := \frac{h(a,b)}{h(a)} = \frac{p(a,b)}{p(a)} \quad \text{für} \quad h(a), p(a) \neq 0$$

Zusammenhang zweier Merkmale

Bedingte relative Häufigkeit
$p(b \mid a)$ wird wie folgt gelesen: "p von b unter der Bedingung a", und "bedingte relative Häufigkeit der Merkmalsausprägung b unter der Bedingung a" genannt.

Die Rollen der Merkmale 1 und 2 bzw. von a und b lassen sich natürlich vertauschen, so dass

$$p(a \mid b) := \frac{h(a,b)}{h(b)} = \frac{p(a,b)}{p(b)} \quad \text{für} \quad h(b), p(b) \neq 0$$

die bedingte relative Häufigkeit von a unter der Bedingung b heißt.

Allgemein spricht man von den **bedingten Häufigkeitsverteilungen.**

Zusammenhang zweier Merkmale

Bedingte relative Häufigkeit
$p(b \mid a)$ wird wie folgt gelesen: "p von b unter der Bedingung a", und "bedingte relative Häufigkeit der Merkmalsausprägung b unter der Bedingung a" genannt.

Die Rollen der Merkmale 1 und 2 bzw. von a und b lassen sich natürlich vertauschen, so dass

$$p(a \mid b) := \frac{h(a,b)}{h(b)} = \frac{p(a,b)}{p(b)} \quad \text{für} \quad h(b), p(b) \neq 0$$

die bedingte relative Häufigkeit von a unter der Bedingung b heißt.

Allgemein spricht man von den **bedingten Häufigkeitsverteilungen.**

Zusammenhang zweier Merkmale

Bedingte relative Häufigkeit
$p(b \mid a)$ wird wie folgt gelesen: "p von b unter der Bedingung a", und "bedingte relative Häufigkeit der Merkmalsausprägung b unter der Bedingung a" genannt.

Die Rollen der Merkmale 1 und 2 bzw. von a und b lassen sich natürlich vertauschen, so dass

$$p(a \mid b) := \frac{h(a,b)}{h(b)} = \frac{p(a,b)}{p(b)} \quad \text{für} \quad h(b), p(b) \neq 0$$

die bedingte relative Häufigkeit von a unter der Bedingung b heißt.

Allgemein spricht man von den **bedingten Häufigkeitsverteilungen.**

Zusammenhang zweier Merkmale

Beispiel 8.7

Zu der Untersuchung in Beispiel 8.3 möchte man angeben, welcher Anteil der Mütter mit rk., ev. oder sonstiger Religionszugehörigkeit 0,1,2,3 oder 4 Kinder haben. Dazu muss die bedingte Häufigkeitsverteilung der Kinderzahl unter der Bedingung der Religionszugehörigkeit bestimmt werden. (a=Anzahl der Kinder)

$$\begin{aligned}
p(a \mid rk) &= \frac{p(a, rk)}{p(rk)} = \frac{h(a, rk)}{h(rk)} \\
p(a \mid ev) &= \frac{p(a, ev)}{p(ev)} = \frac{h(a, ev)}{h(ev)} \\
p(a \mid \text{sonst.}) &= \frac{p(a, \text{sonst.})}{p(\text{sonst.})} = \frac{h(a, sonst.)}{h(sonst.)}.
\end{aligned}$$

Zusammenhang zweier Merkmale

Beispiel 8.7
Zu der Untersuchung in Beispiel 8.3 möchte man angeben, welcher Anteil der Mütter mit rk., ev. oder sonstiger Religionszugehörigkeit 0,1,2,3 oder 4 Kinder haben. Dazu muss die bedingte Häufigkeitsverteilung der Kinderzahl unter der Bedingung der Religionszugehörigkeit bestimmt werden. (a=Anzahl der Kinder)

$$
\begin{aligned}
p(a \mid rk) &= \frac{p(a, rk)}{p(rk)} = \frac{h(a, rk)}{h(rk)} \\
p(a \mid ev) &= \frac{p(a, ev)}{p(ev)} = \frac{h(a, ev)}{h(ev)} \\
p(a \mid \text{sonst.}) &= \frac{p(a, \text{sonst.})}{p(\text{sonst.})} = \frac{h(a, sonst.)}{h(sonst.)}.
\end{aligned}
$$

Zusammenhang zweier Merkmale

Beispiel 8.7

Zu der Untersuchung in Beispiel 8.3 möchte man angeben, welcher Anteil der Mütter mit rk., ev. oder sonstiger Religionszugehörigkeit 0,1,2,3 oder 4 Kinder haben. Dazu muss die bedingte Häufigkeitsverteilung der Kinderzahl unter der Bedingung der Religionszugehörigkeit bestimmt werden. (a=Anzahl der Kinder)

$$
\begin{aligned}
p(a \mid rk) &= \frac{p(a, rk)}{p(rk)} = \frac{h(a, rk)}{h(rk)} \\
p(a \mid ev) &= \frac{p(a, ev)}{p(ev)} = \frac{h(a, ev)}{h(ev)} \\
p(a \mid \text{sonst.}) &= \frac{p(a, \text{sonst.})}{p(\text{sonst.})} = \frac{h(a, sonst.)}{h(sonst.)}.
\end{aligned}
$$

Zusammenhang zweier Merkmale

Beispiel 8.7
Zu der Untersuchung in Beispiel 8.3 möchte man angeben, welcher Anteil der Mütter mit rk., ev. oder sonstiger Religionszugehörigkeit 0,1,2,3 oder 4 Kinder haben. Dazu muss die bedingte Häufigkeitsverteilung der Kinderzahl unter der Bedingung der Religionszugehörigkeit bestimmt werden. (a=Anzahl der Kinder)

$$\begin{aligned}
p(a \mid rk) &= \frac{p(a, rk)}{p(rk)} = \frac{h(a, rk)}{h(rk)} \\
p(a \mid ev) &= \frac{p(a, ev)}{p(ev)} = \frac{h(a, ev)}{h(ev)} \\
p(a \mid \text{sonst.}) &= \frac{p(a, \text{sonst.})}{p(\text{sonst.})} = \frac{h(a, sonst.)}{h(sonst.)}.
\end{aligned}$$

Zusammenhang zweier Merkmale

Beispiel 8.7

Bedingte Häufigkeitsverteilung:

		0	1	2	3	4	$\sum$
	rk.	0.059	0.294	0.412	0.118	0.118	1
Religions-	ev.	0.095	0.190	0.476	0.238	0	1
zugehörigkeit	sonst.	$0.1\bar{6}$	$0.08\bar{3}$	$0.58\bar{3}$	$0.1\bar{6}$	0	1

Zusammenhang zweier Merkmale

Bedingte Lage- und Streuungsparameter
Zu den bedingten Häufigkeitsverteilungen kann man die Lage- und Streuungsparameter wie bei gewöhnlichen Häufigkeitsverteilungen bestimmen.

Beispiel 8.8
Bedingte Lage- und Streuungsparameter aus Beispiel 8.7:

- Arithmetische Mittel der Kinderzahl für Mütter mit Religionszugehörigkeit rk

$$\begin{aligned}\bar{x} &= \sum_{a=0}^{4} a\ p(a \mid rk) \\ &= 0 + 0.294 + 2 \cdot 0.412 + 3 \cdot 0.118 + 4 \cdot 0.118 = 1.944\end{aligned}$$

Zusammenhang zweier Merkmale

Bedingte Lage- und Streuungsparameter
Zu den bedingten Häufigkeitsverteilungen kann man die Lage- und Streuungsparameter wie bei gewöhnlichen Häufigkeitsverteilungen bestimmen.

Beispiel 8.8
Bedingte Lage- und Streuungsparameter aus Beispiel 8.7:

- Arithmetische Mittel der Kinderzahl für Mütter mit Religionszugehörigkeit rk

$$\begin{aligned}\bar{x} &= \sum_{a=0}^{4} a\ p(a \mid rk) \\ &= 0 + 0.294 + 2 \cdot 0.412 + 3 \cdot 0.118 + 4 \cdot 0.118 = 1.944\end{aligned}$$

Zusammenhang zweier Merkmale

Bedingte Lage- und Streuungsparameter
Zu den bedingten Häufigkeitsverteilungen kann man die Lage- und Streuungsparameter wie bei gewöhnlichen Häufigkeitsverteilungen bestimmen.

Beispiel 8.8
Bedingte Lage- und Streuungsparameter aus Beispiel 8.7:

- Arithmetische Mittel der Kinderzahl für Mütter mit Religionszugehörigkeit rk

$$
\begin{aligned}
\bar{x} &= \sum_{a=0}^{4} a \; p(a \mid rk) \\
&= 0 + 0.294 + 2 \cdot 0.412 + 3 \cdot 0.118 + 4 \cdot 0.118 = 1.944
\end{aligned}
$$

Zusammenhang zweier Merkmale

Beispiel 8.8

- Für Mütter mit Religionszugehörigkeit ev

$$\bar{x} = 0.190 + 2 \cdot 0.476 + 3 \cdot 0.238 = 1.856,$$

- Und für Mütter mit sonstiger Religionszugehörigkeit

$$\bar{x} = 0.08\bar{3} + 2 \cdot 0.58\bar{3} + 3 \cdot 0.1\bar{6} = 1.75.$$

- Bedingter Median für alle Religionszugehörigkeiten $x_z = 2$, da bei allen Religionszugehörigkeiten bei 2 Kindern die bedingte empirische Verteilungsfunktion erstmals den Wert 0.5 überschreitet.

Zusammenhang zweier Merkmale

Beispiel 8.8

- Für Mütter mit Religionszugehörigkeit ev

$$\bar{x} = 0.190 + 2 \cdot 0.476 + 3 \cdot 0.238 = 1.856,$$

- Und für Mütter mit sonstiger Religionszugehörigkeit

$$\bar{x} = 0.08\bar{3} + 2 \cdot 0.58\bar{3} + 3 \cdot 0.1\bar{6} = 1.75.$$

- Bedingter Median für alle Religionszugehörigkeiten $x_z = 2$, da bei allen Religionszugehörigkeiten bei 2 Kindern die bedingte empirische Verteilungsfunktion erstmals den Wert 0.5 überschreitet.

Zusammenhang zweier Merkmale

Beispiel 8.8

- Für Mütter mit Religionszugehörigkeit ev

$$\bar{x} = 0.190 + 2 \cdot 0.476 + 3 \cdot 0.238 = 1.856,$$

- Und für Mütter mit sonstiger Religionszugehörigkeit

$$\bar{x} = 0.08\bar{3} + 2 \cdot 0.58\bar{3} + 3 \cdot 0.1\bar{6} = 1.75.$$

- Bedingter Median für alle Religionszugehörigkeiten $x_z = 2$, da bei allen Religionszugehörigkeiten bei 2 Kindern die bedingte empirische Verteilungsfunktion erstmals den Wert 0.5 überschreitet.

Zusammenhang zweier Merkmale

Beispiel 8.8

- Bedingte Varianz der Kinderzahl bei Müttern mit Religion rk

$$\begin{aligned} s^2 &= \sum_{a=0}^{4} p(a \mid rk)(a-\bar{x})^2 \\ &= 0.059(-1.944)^2 + 0.294(-0.944)^2 + 0.412 \cdot 0.056^2 + \\ &\quad 0.118(1.056)^2 + 0.118(2.056)^2 \\ &= 1.117 \end{aligned}$$

- Bei Müttern mit Religion ev

$$\begin{aligned} s^2 &= 0.095(-1.856)^2 + 0.190(-0.856)^2 + 0.476(0.144)^2 + \\ &\quad 0.238(1.144)^2 \\ &= 0.788 \end{aligned}$$

Zusammenhang zweier Merkmale

Beispiel 8.8

- Bedingte Varianz der Kinderzahl bei Müttern mit Religion rk

$$
\begin{aligned}
s^2 &= \sum_{a=0}^{4} p(a \mid rk)(a - \bar{x})^2 \\
&= 0.059(-1.944)^2 + 0.294(-0.944)^2 + 0.412 \cdot 0.056^2 + \\
&\quad 0.118(1.056)^2 + 0.118(2.056)^2 \\
&= 1.117
\end{aligned}
$$

- Bei Müttern mit Religion ev

$$
\begin{aligned}
s^2 &= 0.095(-1.856)^2 + 0.190(-0.856)^2 + 0.476(0.144)^2 + \\
&\quad 0.238(1.144)^2 \\
&= 0.788
\end{aligned}
$$

Zusammenhang zweier Merkmale

Beispiel 8.8

- Bei Müttern mit sonst. Religion

$$\begin{aligned} s^2 &= 0.1\bar{6}(-1.75)^2 + 0.08\bar{3}(-0.75)^2 + 0.58\bar{3}(0.25)^2 + 0.1\bar{6}(1.25)^2 \\ &= 0.854 \end{aligned}$$

Zusammenhang zweier Merkmale

Gegenseitige Beeinflussung zweier Merkmale
Ein Einfluss von Merkmal 1 auf Merkmal 2 (und entsprechend umgekehrt) muss sich im Datenmaterial dadurch widerspiegeln, dass die bedingte Häufigkeitsverteilung von Merkmal 2 unter der Bedingung a von der Merkmalsausprägung a beeinflusst ist. Sie verändert sich dann bei Variation von a, d.h. für verschiedene Merkmalsausprägungen a und a' sind die bedingten Häufigkeitsverteilungen von Merkmal 2 unterschiedlich.

Es liegt also **kein Einfluss von Merkmal 1 auf Merkmal 2** vor, wenn für alle $a, a' \in M_1$ mit $p(a), p(a') \neq 0$ gilt:

$$p(b \mid a) = p(b \mid a') \quad \text{für alle } b \in M_2$$

Analog für den Einfluss von Merkmal 2 auf Merkmal 1.

Zusammenhang zweier Merkmale

Gegenseitige Beeinflussung zweier Merkmale
Ein Einfluss von Merkmal 1 auf Merkmal 2 (und entsprechend umgekehrt) muss sich im Datenmaterial dadurch widerspiegeln, dass die bedingte Häufigkeitsverteilung von Merkmal 2 unter der Bedingung a von der Merkmalsausprägung a beeinflusst ist. Sie verändert sich dann bei Variation von a, d.h. für verschiedene Merkmalsausprägungen a und a' sind die bedingten Häufigkeitsverteilungen von Merkmal 2 unterschiedlich.

Es liegt also **kein Einfluss von Merkmal 1 auf Merkmal 2** vor, wenn für alle $a, a' \in M_1$ mit $p(a), p(a') \neq 0$ gilt:

$$p(b \mid a) = p(b \mid a') \quad \text{für alle } b \in M_2$$

Analog für den Einfluss von Merkmal 2 auf Merkmal 1.

Zusammenhang zweier Merkmale

Gegenseitige Beeinflussung zweier Merkmale
Ein Einfluss von Merkmal 1 auf Merkmal 2 (und entsprechend umgekehrt) muss sich im Datenmaterial dadurch widerspiegeln, dass die bedingte Häufigkeitsverteilung von Merkmal 2 unter der Bedingung a von der Merkmalsausprägung a beeinflusst ist. Sie verändert sich dann bei Variation von a, d.h. für verschiedene Merkmalsausprägungen a und a' sind die bedingten Häufigkeitsverteilungen von Merkmal 2 unterschiedlich.

Es liegt also **kein Einfluss von Merkmal 1 auf Merkmal 2** vor, wenn für alle $a, a' \in M_1$ mit $p(a), p(a') \neq 0$ gilt:

$$p(b \mid a) = p(b \mid a') \quad \text{für alle } b \in M_2$$

Analog für den Einfluss von Merkmal 2 auf Merkmal 1.

Unabhängigkeit

Wegen $p(a,b) = p(b|a) \cdot p(a)$:

Gilt

$$p(b|a) = p(b|a')$$

für alle a, a' und b, so folgt:

$$\begin{aligned} p(b) &= \sum_a p(a,b) = \sum_a p(b|a) \cdot p(a) \\ &= p(b|a) \sum_a p(a) \\ &= p(b|a) \end{aligned}$$

Unabhängigkeit

Wegen $p(a, b) = p(b|a) \cdot p(a)$:

Gilt

$$p(b|a) = p(b|a')$$

für alle a, a' und b, so folgt:

$$\begin{aligned} p(b) &= \sum_a p(a,b) = \sum_a p(b|a) \cdot p(a) \\ &= p(b|a) \sum_a p(a) \\ &= p(b|a) \end{aligned}$$

Unabhängigkeit

Wegen $p(a,b) = p(b|a) \cdot p(a)$:

Gilt

$$p(b|a) = p(b|a')$$

für alle a, a' und b, so folgt:

$$\begin{aligned} p(b) &= \sum_a p(a,b) = \sum_a p(b|a) \cdot p(a) \\ &= p(b|a) \sum_a p(a) \\ &= p(b|a) \end{aligned}$$

Zusammenhang zweier Merkmale

und damit

$$p(a, b) = p(a) \cdot p(b).$$

Umgekehrt folgt aus

$$\begin{aligned} p(a, b) &= p(a) \cdot p(b), \text{ für alle a und b} \\ p(b|a) &= \frac{p(a, b)}{p(a)} = \frac{p(a) \cdot p(b)}{p(a)} = p(b), \text{ für alle a und b} \end{aligned}$$

und damit die Übereinstimmung der bedingten Häufigkeiten $p(b|a)$ für alle Bedingungen.

Zusammenhang zweier Merkmale

und damit

$$p(a,b) = p(a) \cdot p(b).$$

Umgekehrt folgt aus

$$\begin{aligned} p(a,b) &= p(a) \cdot p(b), \text{ für alle a und b} \\ p(b|a) &= \frac{p(a,b)}{p(a)} = \frac{p(a) \cdot p(b)}{p(a)} = p(b), \text{ für alle a und b} \end{aligned}$$

und damit die Übereinstimmung der bedingten Häufigkeiten $p(b|a)$ für alle Bedingungen.

Zusammenhang zweier Merkmale

Unabhängigkeit
Zwei Merkmale heißen **unabhängig** auf einer statistischen Masse, wenn für alle Merkmalsausprägungen a von Merkmal 1 und alle Merkmalsausprägungen b von Merkmal 2 gilt:

$$p(a, b) = p(a) \cdot p(b)$$

Die Merkmale heißen **abhängig**, wenn für mindestens ein a und mindestens ein b gilt:

$$p(a, b) \neq p(a) \cdot p(b)$$

Zusammenhang zweier Merkmale

Unabhängigkeit
Zwei Merkmale heißen **unabhängig** auf einer statistischen Masse, wenn für alle Merkmalsausprägungen a von Merkmal 1 und alle Merkmalsausprägungen b von Merkmal 2 gilt:

$$p(a, b) = p(a) \cdot p(b)$$

Die Merkmale heißen **abhängig**, wenn für mindestens ein a und mindestens ein b gilt:

$$p(a, b) \neq p(a) \cdot p(b)$$

Zusammenhang zweier Merkmale

Beispiel 8.9
In Beispiel 8.3 sind die Merkmale Religionszugehörigkeit der Mütter und Anzahl der im Haushalt lebenden Kinder offensichtlich nicht unabhängig. Aus der Randverteilungen wird deutlich, dass z.B. für Mütter mit Religion rk, die keine Kinder haben, gilt:

$$p(rk, 0) = 0.02 \neq p(rk) \cdot p(0) = 0.34 \cdot 0.1.$$

Bemerkung: Die Abhängigkeit der beiden Merkmale kann bereits daran erkannt werden, dass die bedingten Häufigkeitsverteilungen (vgl. Beispiel 8.7) für die verschiedenen Religionszugehörigkeiten nicht übereinstimmen.

Zusammenhang zweier Merkmale

Beispiel 8.9
In Beispiel 8.3 sind die Merkmale Religionszugehörigkeit der Mütter und Anzahl der im Haushalt lebenden Kinder offensichtlich nicht unabhängig. Aus der Randverteilungen wird deutlich, dass z.B. für Mütter mit Religion rk, die keine Kinder haben, gilt:

$$p(rk, 0) = 0.02 \neq p(rk) \cdot p(0) = 0.34 \cdot 0.1.$$

Bemerkung: Die Abhängigkeit der beiden Merkmale kann bereits daran erkannt werden, dass die bedingten Häufigkeitsverteilungen (vgl. Beispiel 8.7) für die verschiedenen Religionszugehörigkeiten nicht übereinstimmen.

Agenda

1. Einleitung
2. Absolute bzw. relative Häufigkeiten
3. Graphische Darstellung
4. Zusammenhang zweier Merkmale
5. **Überprüfung auf Unabhängigkeit**

Überprüfung auf Unabhängigkeit

1. Möglichkeit:
Berechnung der bedingten relativen Häufigkeitsverteilung von Merkmal 2 für jede Merkmalsausprägung von Merkmal 1 als Bedingung und Überprüfung auf Übereinstimmung. Wenn ja liegt Unabhängigkeit vor.

2. Möglichkeit:
Analog zu 1. mit vertauschten Merkmalen.

3. Möglichkeit:
Überprüfung der Gleichung

$$p(a, b) = p(a) \cdot p(b)$$

für alle a und b. Falls eine Kombination (a, b) gefunden ist, bei der die Gleichung verletzt ist, sind die Merkmale abhängig und die Überprüfung ist abgebrochen.

Überprüfung auf Unabhängigkeit

1. Möglichkeit:
Berechnung der bedingten relativen Häufigkeitsverteilung von Merkmal 2 für jede Merkmalsausprägung von Merkmal 1 als Bedingung und Überprüfung auf Übereinstimmung. Wenn ja liegt Unabhängigkeit vor.

2. Möglichkeit:
Analog zu 1. mit vertauschten Merkmalen.

3. Möglichkeit:
Überprüfung der Gleichung

$$p(a, b) = p(a) \cdot p(b)$$

für alle a und b. Falls eine Kombination (a, b) gefunden ist, bei der die Gleichung verletzt ist, sind die Merkmale abhängig und die Überprüfung ist abgebrochen.

Überprüfung auf Unabhängigkeit

1. Möglichkeit:
Berechnung der bedingten relativen Häufigkeitsverteilung von Merkmal 2 für jede Merkmalsausprägung von Merkmal 1 als Bedingung und Überprüfung auf Übereinstimmung. Wenn ja liegt Unabhängigkeit vor.

2. Möglichkeit:
Analog zu 1. mit vertauschten Merkmalen.

3. Möglichkeit:
Überprüfung der Gleichung

$$p(a, b) = p(a) \cdot p(b)$$

für alle a und b. Falls eine Kombination (a, b) gefunden ist, bei der die Gleichung verletzt ist, sind die Merkmale abhängig und die Überprüfung ist abgebrochen.

Überprüfung auf Unabhängigkeit

Beispiel 8.9 - unabhängige Merkmale (Kinderzahl, Religion)

relative Häufigkeiten

Kinderzahl / Religion	0	1	2	3	4	gesamt
rk	0,02	0,04	0,08	0,04	0,02	0,2
ev	0,04	0,08	0,16	0,08	0,04	0,4
sonst	0,04	0,08	0,16	0,08	0,04	0,4
gesamt	0,1	0,2	0,4	0,2	0,1	1