

3 Zweidimensionale Häufigkeitsverteilung

3.1 Kontingenztafel

Um zu untersuchen, ob die Wirkung eines bestimmten Giftes vom Geschlecht abhängt, wurde 20 Mäusen das Gift verabreicht – mit folgendem Ergebnis:

Tier Nr.	Wirkung		Geschlecht	Tier Nr.	Wirkung		Geschlecht
	vorhanden	fehlt			vorhanden	fehlt	
1	x		m	11		x	w
2		x	m	12		x	m
3	x		w	13	x		m
4		x	m	14	x		w
5	x		w	15	x		m
6	x		w	16	x		w
7		x	m	17	x		w
8	x		m	18		x	w
9	x		w	19	x		m
10	x		w	20	x		w

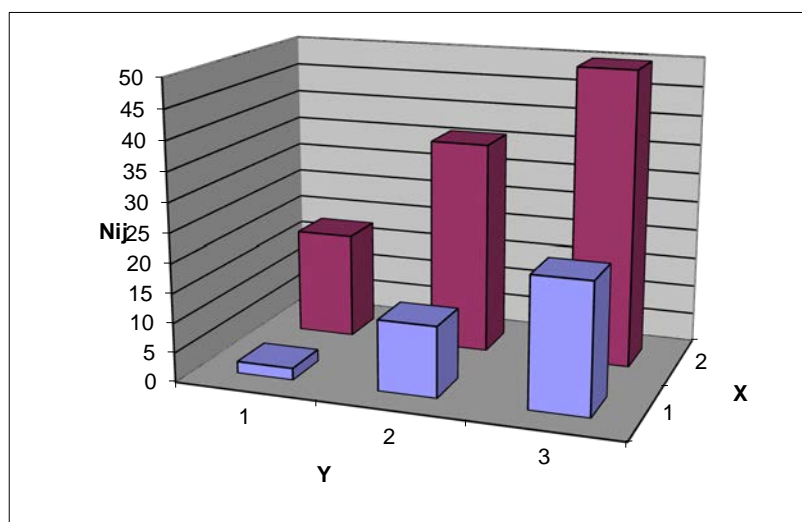
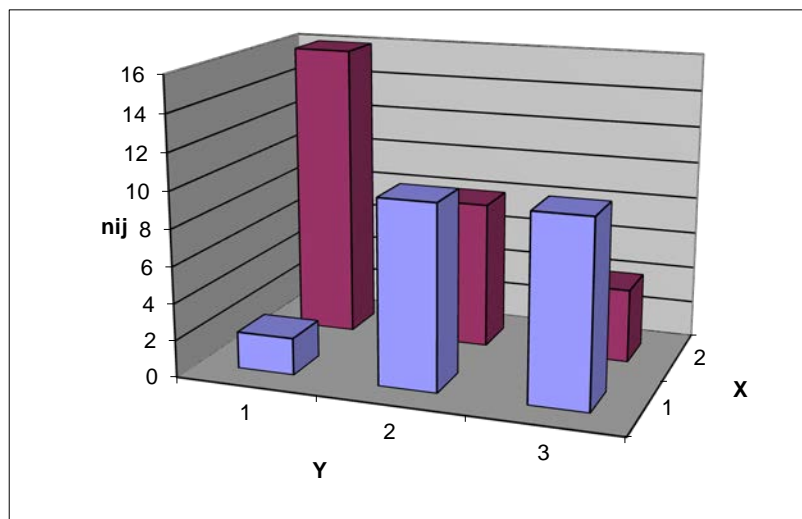
- Legen Sie für die beiden Merkmale eine Kontingenztafel an.
- Bei wie viel Prozent der männlichen (weiblichen) Tiere wirkt das Gift?
- Wie viel Prozent der Tiere, bei denen das Gift wirkt, sind männlich?

3.2 Darstellung bei zweidimensionalen Verteilungen

Für ein bestimmtes Lebensmitteleinzelhandels-Kettenunternehmen wurden $n = 50$

Mitgliedsunternehmen nach den Merkmalen Umsatzkategorie und Anzahl der Ladenlokale pro Unternehmen untersucht.

Merkmal Y Anzahl Läden	y_1 1 Laden	y_2 2L	y_3 3L	Σ
Merkmal X Kategorie				
x_1 Kategorie K_I	$n_{11} = 2$	$n_{12} = 10$	$n_{13} = 10$	
x_2 K_{II}	$n_{21} = 16$	$n_{22} = 8$	$n_{23} = 4$	
Σ				



Aufsummierte Häufigkeiten

Merkmal Y Anzahl Läden		y_1 1 Laden	y_2 2L	y_3 3L
Merkmal X Kategorie				
x_1 Kategorie	K_I	$N_{11} = 2$	$N_{12} = 12$	$N_{13} = 22$
x_2	K_{II}	$N_{21} = 18$	$N_{22} = 36$	$N_{23} = 50$

3.3 Analyse empirischer Zusammenhänge

Angenommen im Rahmen einer Studie wurden 100 erwachsene Personen im Alter zwischen 18 und 29 Jahren zu deren Raucherstatus befragt. Demnach waren unter den 100 befragten Personen insgesamt 40 Frauen und 60 Männer. Unter den Frauen rauchten 4 Frauen regelmäßig, 8 gelegentlich und 28 Frauen überhaupt nicht. Bei den Männern waren entsprechend 12 Raucher, 12 Gelegenheitsraucher und 36 Nichtraucher.

→ Tragen Sie die Werte in die Kontingenztabelle ein.

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich				
Männlich				
Summe				

Zeilen von $i = 1$ bis m .

Spalten von $j = 1$ bis l .

→ Stellen Sie eine Formel für n auf:

$n =$

→ Stellen Sie eine Formel für die Zeilensumme $n_{i\bullet}$ und die Spaltensumme $n_{\bullet j}$ auf.

→ Bestimmen Sie die gemeinsame relative Verteilung von Geschlecht und Raucherstatus.

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich				
Männlich				
Summe				

→ Bestimmen Sie die Verteilung von Raucherstatus bedingt auf Geschlecht:

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich				
Männlich				

→ Bestimmen Sie die Verteilung von Geschlecht bedingt auf Raucherstatus:

	Raucher	Gelegenheitsraucher	Nichtraucher
Weiblich			
Männlich			
Summe			

→ Empirische Abhängigkeit – Sind die Merkmale Raucherstatus und Geschlecht voneinander abhängig?

→ Welche „Richtung“ der Abhängigkeit würden Sie wählen?

- Ist der Raucherstatus abhängig vom Geschlecht?
- Ist das Geschlecht abhängig vom Raucherstatus?

Vergleichen Sie mit dieser Verteilung:

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich	10	10	20	40
Männlich	15	15	30	60
Summe	25	25	50	100

3.4 Chi-Quadrat-Koeffizient

Gegeben sei eine $m \times l$ Kontingenztafel der absoluten Häufigkeiten n_{ij} mit positiven Randhäufigkeiten.

Dann ist der Chi-Quadrat-Koeffizient definiert als:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^l \frac{\left(n_{ij} - \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \right)^2}{\frac{n_{i\bullet} \cdot n_{\bullet j}}{n}}$$

→ Berechnen Sie den Chi-Quadrat-Koeffizient für folgende Verteilungen und interpretieren Sie:

Beispiel 1

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich	4	8	28	
Männlich	12	12	36	
Summe				

Beispiel 2

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich	10	10	20	40
Männlich	15	15	30	60
Summe	25	25	50	100

Beispiel 3

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich	0	0	40	
Männlich	30	30	0	
Summe				

3.5 Parameter zweidimensionaler Verteilungen

In einer Studie zu Rauchergewohnheiten wurden von einem Marktforschungsinstitut $n=1000$ Testpersonen (Rauchern) zwei verschiedene Sorten von Zigaretten vorgelegt, von denen eine einen Nikotingehalt von 0,8 mg/Zigarette und die andere 8,5 mg/Zigarette aufwies. Jede der beiden Sorten wurde in den drei unterschiedlichen Längen 8 cm, 10 cm und 12 cm angeboten, wobei auch bei unterschiedlicher Länge der Nikotingehalt bei den oben angegebenen Werten konstant gehalten wurde. Die Testpersonen, die über die unterschiedlichen Eigenschaften genau informiert waren, wurde aufgefordert, sich spontan eine Zigarette zu nehmen, die ihrem Rauchbedürfnis am meisten entsprach. Aus den Wünschen der Testpersonen ergab sich die folgende zweidimensionale Häufigkeitsfunktion:

Merkmal Y: Zigarettenlänge (in cm)		y_1 8	y_2 10	y_3 12	Σ
Merkmal X: Nikotingehalt (in mg)					
x_1	0,8	150	250	300	
x_2	8,5	150	50	100	
Σ					

Welche Nikotinmenge (in mg) wurde im Durchschnitt von jeder Testperson aufgenommen? – Berechnen Sie \bar{x} .

Formel:

Was ist die Durchschnittslänge der gerauchten Zigaretten? – Berechnen Sie \bar{y} .

Formel:

Berechnen Sie Varianz und Standardabweichung des Zeilenmerkmals X.

Formel:

Berechnen Sie Varianz und Standardabweichung des Spaltenmerkmals Y.

Formel:

Datum: _____

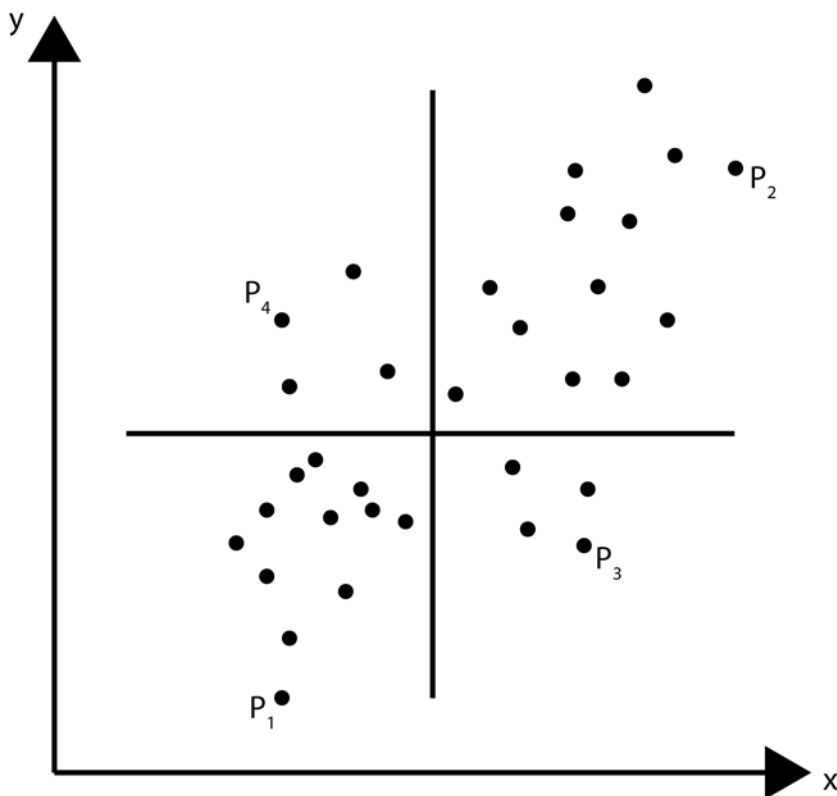
3.6 Kovarianz

Die Formeln für die Kovarianz lauten wie folgt:

$$\text{COV}(X, Y) = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^l (x_i - \bar{x})(y_j - \bar{y}) \cdot n_{ij} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^l x_i y_j \cdot n_{ij} - \bar{x} \cdot \bar{y}$$

$$\text{COV}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}$$

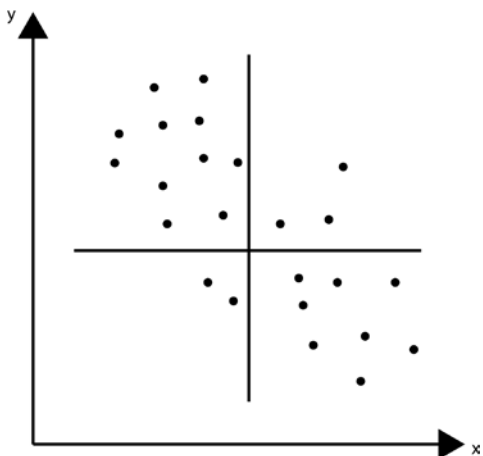
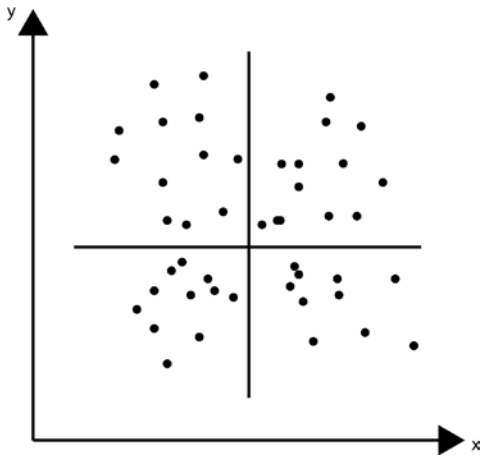
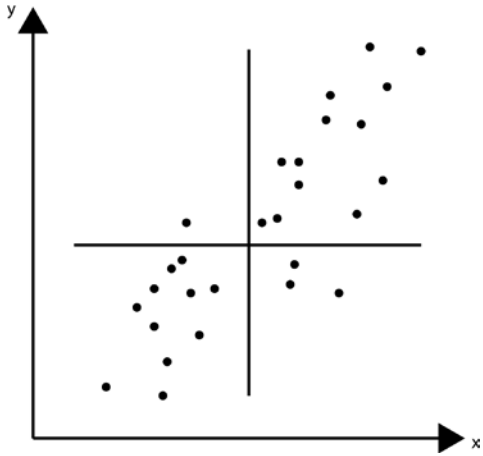
Versuchen Sie die Formel zu verstehen, indem Sie sich aus dem Schaubild jeweils für die eingezeichneten Punkte das Ergebnis des Produktes $(x_i - \bar{x})(y_i - \bar{y})$ abschätzen.



Datum: _____

Betrachten Sie nun für die drei Verteilungen die Summe der Produkte $(x_i - \bar{x})(y_i - \bar{y})$.

Welche Werte nimmt die Kovarianz jeweils an?



Datum: _____

3.6.1 Übung zur Kovarianz

Berechnen Sie für folgende Verteilungen jeweils die Kovarianz:

a)

X \ Y	$y_1 = 2$	$y_2 = 3$	$y_3 = 4$	
$x_1 = 2$	1	1	1	3
$x_2 = 3$	1	2	1	4
$x_3 = 4$	1	1	1	3
	3	4	3	10

b)

X \ Y	$y_1 = 2$	$y_2 = 3$	$y_3 = 4$	
$x_1 = 2$	1	1	0	2
$x_2 = 3$	2	3	1	6
$x_3 = 4$	0	0	2	2
	3	4	3	10

c)

X \ Y	$y_1 = 2$	$y_2 = 3$	$y_3 = 4$	
$x_1 = 2$	0	0	3	3
$x_2 = 3$	1	2	1	4
$x_3 = 4$	3	0	0	3
	4	2	4	10

Berechnungstabellen: (Prod. = Produkt $(x_i - \bar{x}) \cdot (y_j - \bar{y}) \cdot n_{ij}$)

a)

$(x_i - \bar{x})$	$(y_j - \bar{y})$	n_{ij}	Prod.

b)

$(x_i - \bar{x})$	$(y_j - \bar{y})$	n_{ij}	Prod.

c)

$(x_i - \bar{x})$	$(y_j - \bar{y})$	n_{ij}	Prod.

3.7 Übungen

3.7.1 Übung: Kontingenztabelle

X: soziale Stellung des Vaters (Arbeiter, Angestellter, Beamter, Selbständiger)

Y: Schultyp des Kindes (Hauptschule, Realschule, Gymnasium)

X	x ₁ : Arbeiter	x ₂ Angest.	x ₃ Beamter	x ₄ Selbstän.	Σ
Y					
y ₁ HS	6	2	1	1	10
y ₂ RS	4	2	2	2	10
y ₃ Gymnas.	0	1	2	2	5
Σ	10	5	5	5	25

Eine Abhängigkeit oder Unabhängigkeit kann festgestellt werden durch das Ermitteln der bedingten Verteilungen der relativen Häufigkeiten für X und Y.

Für X: Berechnen Sie die relativen Häufigkeiten bei variablem x und festem y (für alle y).

X	x ₁ : Arbeiter	x ₂ Angest.	x ₃ Beamter	x ₄ Selbstän.
Y				
y ₁ HS				
y ₂ RS				
y ₃ Gymnas.				

Für Y: Berechnen Sie die relativen Häufigkeiten bei variablem y und festem x (für alle x).

X	x ₁ : Arbeiter	x ₂ Angest.	x ₃ Beamter	x ₄ Selbstän.
Y				
y ₁ HS				
y ₂ RS				
y ₃ Gymnas.				

3.7.2 Übung: Bedingte Verteilungen

450 Personen wurden nach Alter und Einkommen befragt. Das Befragungsergebnis ist in der folgenden Tabelle festgehalten.

Alter Einkommen	20 bis unter 30	30 bis unter 40	40 bis unter 50	50 bis unter 60	60 bis unter 70	
0 bis unter 2000	10	20	10	10	10	
2000 bis unter 2500	20	10	20	30	10	
2500 bis unter 3000	30	60	60	30	20	
3000 und mehr	10	30	20	20	20	

Bestimmen Sie für jede Altersklasse die bedingten Verteilungen des Einkommens, d.h. die relativen Häufigkeiten bei variablem Einkommen und fester Altersklasse werden errechnet. Geben Sie die relativen Häufigkeiten der bedingten Verteilungen in der folgenden Tabelle an.

Alter Einkommen	20 bis unter 30	30 bis unter 40	40 bis unter 50	50 bis unter 60	60 bis unter 70
0 bis unter 2000					
2000 bis unter 2500					
2500 bis unter 3000					
3000 und mehr					

3.7.3 Übung zur Kovarianz

Ein fiktives Beispiel: Bruttogehalt X und Bildungsjahre Y

Berechnen Sie die arithmetischen Mittel, die Standardabweichung und die Kovarianz

x_i	y_i				
2000	9				
5000	16				
4000	16				
1500	9				
2500	10				