
Bivariate Verfahren

WS 2023-24

DI Emil Marinov | DI David Bechtold

Übersicht

1.	Zusammenhang von metrischen Merkmalen <ul style="list-style-type: none">- Streudiagramm- Kennzahlen des Zusammenhangs- Regression	3 – 9
2.	Zusammenhang von ordinalen Merkmalen <ul style="list-style-type: none">- Korrelationskoeffizient von Spearman	10 – 13
3.	Zusammenhang von nominale Merkmalen <ul style="list-style-type: none">- Zweidimensionale Verteilungen- Kennzahlen des Zusammenhangs	14 – 23

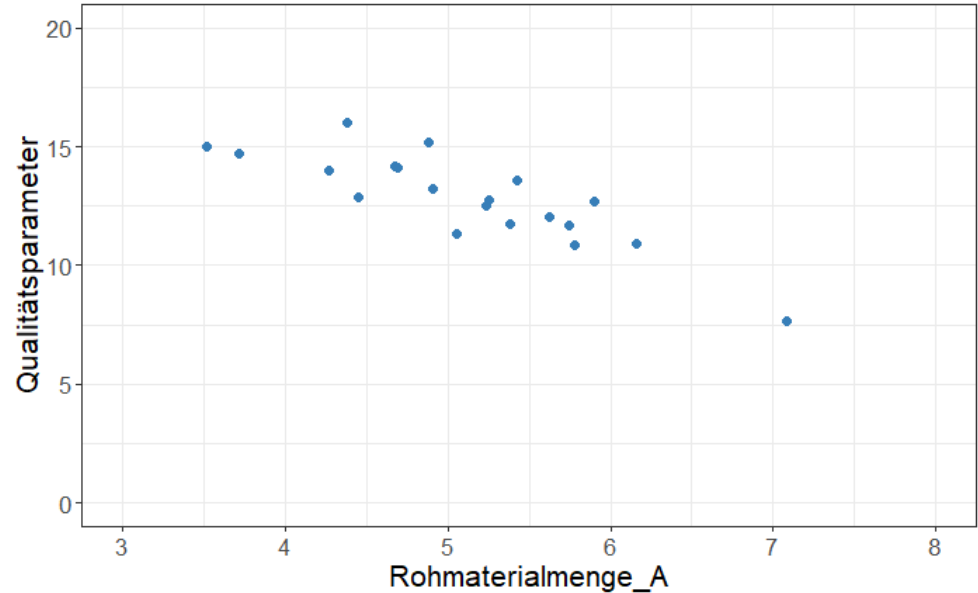
Zusammenhang von metrischen Merkmale

Streudiagramm

Darstellung des Zusammenhangs zwischen 2 metrischen Merkmalen im (x, y)- Koordinatensystem

Beispiel:

Untersuchung, ob die Menge an Rohmaterial A einen Einfluss auf einen bestimmten Qualitätsparameter hat.



Kennzahlen des Zusammenhangs

Kovarianz

- Maß für die Streuung um den Mittelpunkt (\bar{x}, \bar{y})

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Korrelationskoeffizient von Bravais-Pearson

- Maß für die Stärke des linearen Zusammenhangs von metrischen Merkmalen

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Beispiel:

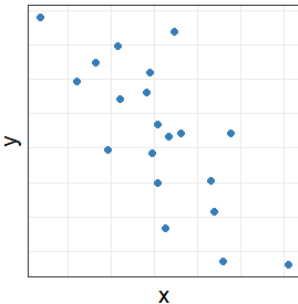
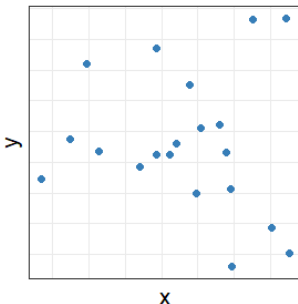
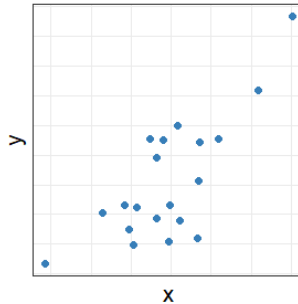
Zusammenhang Menge an Rohmaterial A und Qualität:

$$\sigma_{XY} = -1.36$$

$$r = -0.85$$

Kennzahlen des Zusammenhangs

Interpretation des Korrelationskoeffizienten

$-1 < r < 0$	$r = 0$	$0 < r < 1$
negative Korrelation, Wertepaare liegen um eine Gerade mit negativer Steigung	keine Korrelation, kein linearer Zusammenhang	positive Korrelation, Wertepaare liegen im eine Gerade mit positiver Steigung
		

Regression

- metrische Merkmale X, Y
Werte von X: x_1, x_2, \dots, x_n
Werte von Y: y_1, y_2, \dots, y_n
- Annäherung des Zusammenhangs durch eine Funktion

$$\hat{y} = f(x)$$

- **Least-Squares-Ansatz:**

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \rightarrow \text{Min.}$$

Regression

Beispiel:

Zusammenhang Menge an Rohmaterial A und Qualität:

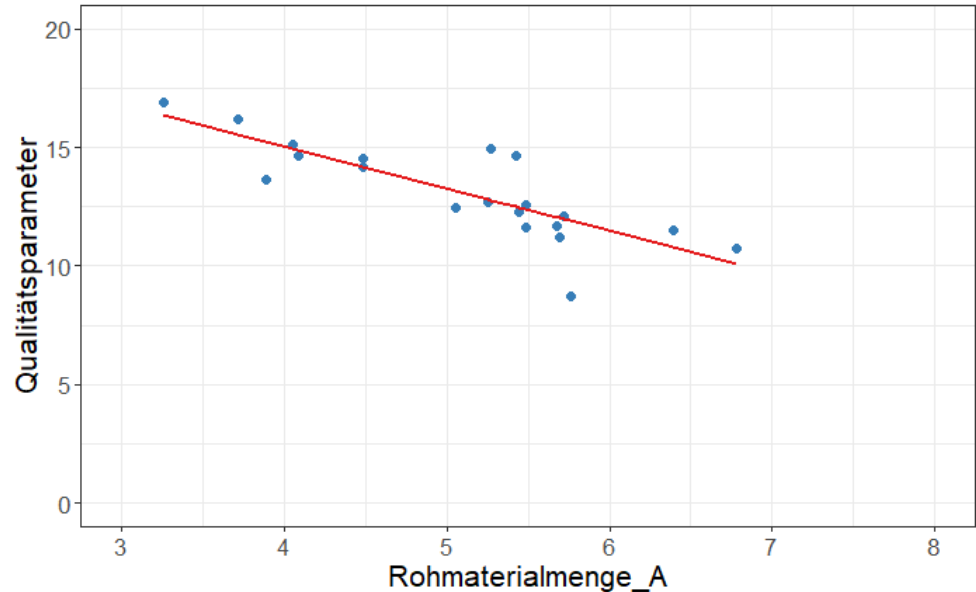
x ... Rohmaterialmenge_A

y ... Qualitätsparameter

Näherung durch lineare Regressionsgerade:

$$\hat{y} = f(x) = kx + d$$

$$\hat{y} = -1.781x + 22.154$$



Kontrollfragen

Thema

Zusammenhang von metrischen Merkmalen

Fragen

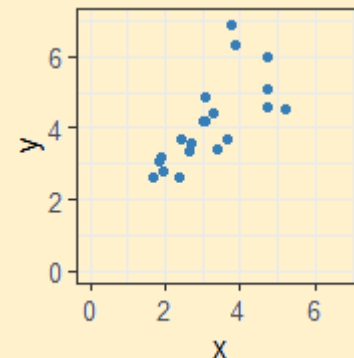
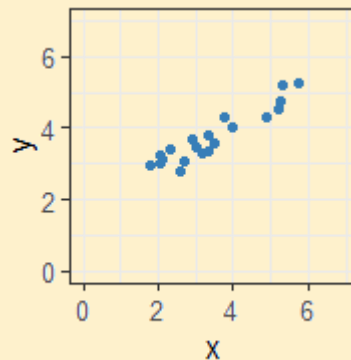
Ordnen Sie die Werte der Korrelationskoeffizienten und Regressionsgeraden zu den Grafiken zu.

$$r = 0.7$$

$$r = 0.9$$

$$y = 0.6x + 1.8$$

$$y = 0.8x + 1.6$$



Zusammenhang von ordinalen Merkmalen

Kennzahlen des Zusammenhangs

Korrelationskoeffizient von Spearman

- für zumindest ordinal skalierte Merkmale X, Y
- misst die Stärke des monotonen Zusammenhangs
- Vorgehensweise bei Berechnung:
 - > Rangordnung der Merkmalsträger von beiden Merkmalen erstellen und Rangziffer zuordnen
 - > bei gleichen Merkmalswerten als Rang das arithmetische Mittel von benachbarten Plätzen verwenden
 - > Korrelationskoeffizient von den Rangziffernpaaren berechnen

Kennzahlen des Zusammenhangs

Beispiel:

Merkmalsträger	Merkmal X	Merkmal Y	Rang X	Rang Y
A	1	7	1	4
B	2	5	2.5	3
C	2	3	2.5	1
D	4	4	4	2

$$r_{SP} = -0.63$$

Kontrollfragen

Thema

Zusammenhang von ordinalen Merkmalen

Fragen

Für welcher der folgenden Merkmalspaare kann der Korrelationskoeffizient von Spearman sinnvoll bestimmt werden?

- Anzahl PS – Anzahl Zylinder eines Autos
- Interesse an Statistik – Alter
- Beruf – Einstiegsgehalt
- Klausurnoten – Geschlecht

Zusammenhang von nominalen Merkmalen

Zweidimensionale Verteilungen

Darstellung der gemeinsamen Verteilung von zwei Merkmalen X und Y

Werte von X: a_1, a_2, \dots, a_k

Werte von Y: b_1, b_2, \dots, b_k

- relative Häufigkeiten $h(a_i, b_j)$ werden in einer Tabelle dargestellt (**Kontingenztafel**)
- Spalten bzw. Zeilensummen sind eindimensionale Verteilungen (**Randverteilungen**)

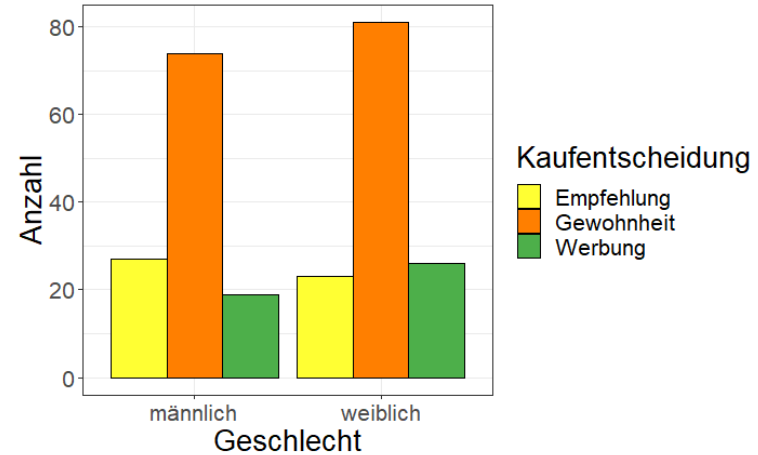
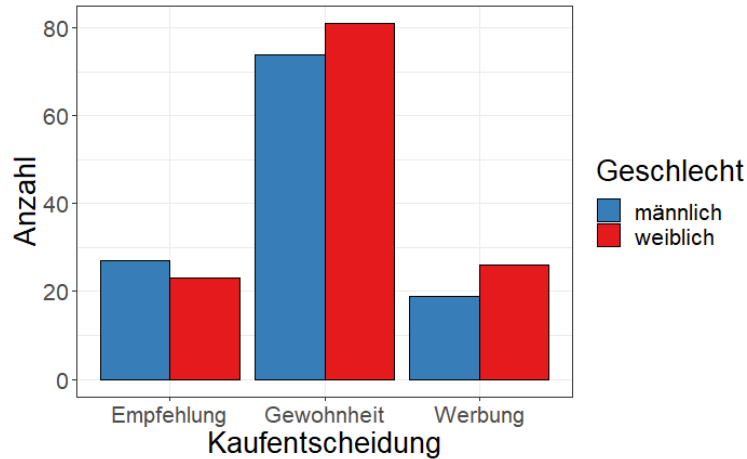
Beispiel:

Kaufentscheidung für ein Produkt

	männlich	weiblich	Summe
Gewohnheit	74	81	155
Werbung	19	26	45
Empfehlung	27	23	50
Summe	120	130	250

Zweidimensionale Verteilungen

Grafische Darstellung von zweidimensionalen Verteilungen



Zweidimensionale Verteilungen

Bedingte Häufigkeiten

bedingte Häufigkeit des Merkmalswerts b_j von Y unter der Bedingung $X = a_i$:

$$f(b_j|a_i) = \frac{h(a_i, b_j)}{h(a_i)}$$

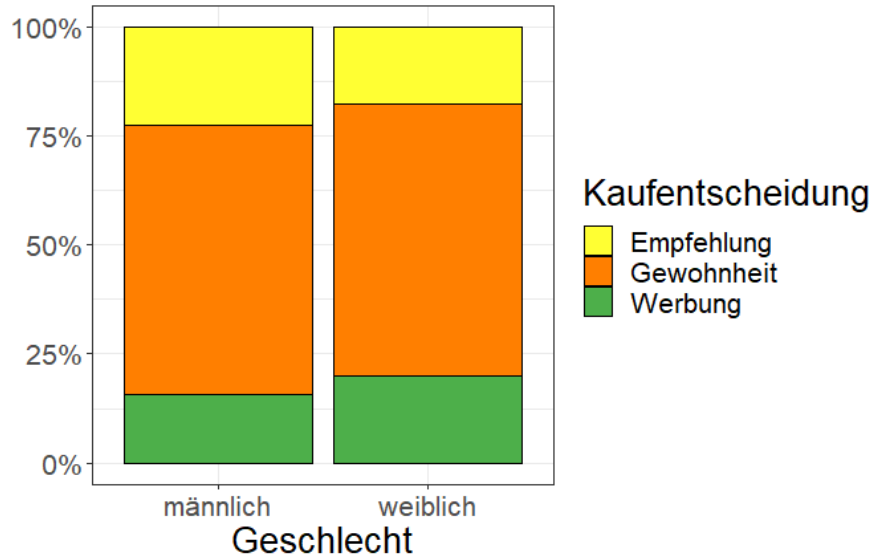
Beispiel:

	männlich	weiblich	Summe
Gewohnheit	74	81	155
Werbung	19	26	45
Empfehlung	27	23	50
Summe	120	130	250

$$f(Werbung|männlich) = \frac{19}{120} = 15.8\%$$

Zweidimensionale Verteilungen

Grafische Darstellung von bedingten Verteilungen



Kennzahlen des Zusammenhangs

Unabhängigkeiten von Merkmalen

Bedingung von Merkmal X hat keinen Einfluss auf Merkmal Y

→ X und Y sind unabhängig

$$f(b_j|a_i) = \frac{h(a_i, b_j)}{h(a_i)} = \frac{h(b_j)}{n}$$

Erwartete Häufigkeit h^e bei Unabhängigkeit:

$$h^e(a_i, b_j) = \frac{h(a_i)h(b_j)}{n}$$

Kennzahlen des Zusammenhangs

Chi-Quadrat-Koeffizient

- misst die Abweichung der beobachteten Häufigkeiten von den erwarteten Häufigkeiten bei Unabhängigkeit
- beschreibt nur die Stärke und nicht die Richtung des Zusammenhangs
- Berechnung:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{\left(h(a_i, b_j) - h^e(a_i, b_j)\right)^2}{h^e(a_i, b_j)}$$

- Wertebereich: $\chi^2 \in [0, \infty[$ (ist abhängig vom Stichprobenumfang)

Kennzahlen des Zusammenhangs

Cramer'sches Assoziationsmaß V

- normiertes Maß für die Abhängigkeit
- misst nur die Stärke und nicht die Richtung des Zusammenhangs
- Berechnung:

$$V = \sqrt{\frac{\chi^2}{n(M-1)}} \text{ mit } M = \min\{k, l\}^e$$

- Wertebereich: $V \in [0, 1]$

Kennzahlen des Zusammenhangs

Beispiel:

beobachtete Häufigkeiten

	männlich	weiblich	Summe
Gewohnheit	74	81	155
Werbung	19	26	45
Empfehlung	27	23	50
Summe	120	130	250

erwartete Häufigkeiten bei Unabhängigkeit

	männlich	weiblich	Summe
Gewohnheit	74.4	80.6	155
Werbung	21.6	23.4	45
Empfehlung	24.0	26.0	50
Summe	120	130	250

Chi-Quadrat Koeffizient: $\chi^2 = 1.327$,
Cramer'sches V: $M = \min(2, 3) = 2, V = 0.073$

Kontrollfragen

Thema

Zusammenhang von nominale Merkmalen

Fragen

Die folgende Tabelle gibt das Ergebnis einer Ernährungsumfrage wieder. Berechnen Sie daraus

- die erwartete Häufigkeiten bei Unabhängigkeit
- den Chi-Quadrat-Koeffizient
- das Cramer'sche Assoziationsmaß V

Geschlecht	Gesunde Ernährung	
	ja	nein
weiblich	144	16
männlich	16	24