- Einführung: Was ist Statistik?
1. Datenerhebung und Messung
2. Univariate deskriptive Statistik
3. **Multivariate Statistik**
4. Regression
5. Ergänzungen

Einführung

In den meisten Anwendungen und in den Beispielen werden an jeder Einheit *gleichzeitig* mehrere Merkmale $X, Y, Z, \ldots$ erhoben:

$\Rightarrow$ mehrdimensionale oder multivariate Daten

Grundgesamtheit

Werte (x_i, y_i, z_i) der
Merkmale (X, Y, Z)

Einheit i

Daten

- Daten $(x_1, y_1, z_1), \ldots, (x_i, y_i, z_i), \ldots, (x_n, y_n, z_n)$

  Im weiteren: Meistens nur *zwei* Merkmale X, Y
- Fragestellungen:

  $X \leftrightarrow Y$ Wie hängen X und Y zusammen?
  Assoziation, Korrelation

  $X \rightarrow Y$ Wie beeinflusst X das (Ziel-)Merkmal Y ?
  Regression

Diskrete und gruppierte Merkmale

- Darstellung, Präsentation von (zwei) diskreten Merkmalen X und Y mit den Ausprägungen

$$\begin{array}{ll} a_1, \dots, a_k & \text{für } X \\ b_1, \dots, b_m & \text{für } Y \end{array}$$

- Skalenniveau von X, Y beliebig; X, Y können auch gruppierte metrische Merkmale sein.
- Benutzt wird nur das Nominalskalenniveau der Merkmale.

Kontingenztabellen

Sonntagsfrage: “Welche Partei würden Sie wählen, wenn am nächsten Sonntag Bundestagswahlen wären?” werden üblicherweise in Prozenten (%) wiedergegeben. Für den Befragungszeitraum ergab sich folgende Tabelle:

	CDU/CSU	SPD	FDP	Grüne	Rest	
Männer	33	35	4	6	22	100
Frauen	40	29	6	10	15	100
insgesamt	37	32	5	8	18	100

Aus den ersten beiden Zeilen ergibt sich, dass die Parteipräferenzen für Männer und Frauen unterschiedlich sind.

Sonntagsfrage

In der angegebenen Tabelle sind die ursprünglichen Daten bereits in Prozenten für die geschlechtsspezifischen Populationen angegeben. Die Rückrechnung auf 435 Männer und 496 Frauen ergibt:

	CDU/CSU	SPD	FDP	Grüne	Rest	
Männer	144	153	17	26	95	435
Frauen	200	145	30	50	71	496
	344	298	47	76	166	931

Beispiel: Arbeitslosigkeit

Zwei Merkmale:

- X Ausbildungsniveau mit den Kategorien
  - "keine Ausbildung",
  - "Lehre",
  - "fachspezifische Ausbildung"
  - "Hochschulabschluß"
- Y Dauer der Arbeitslosigkeit mit den Kategorien
  - "Kurzzeitarbeitslosigkeit" ($\leq$ 6 Monate),
  - "mittelfristige Arbeitslosigkeit" (7–12 Monate),
  - "Langzeitarbeitslosigkeit" ($\geq$ 12 Monate)

Arbeitslosigkeit

	Kurzzeit-arbeitslosigkeit	mittelfristige Arbeitslosigkeit	Langzeit-arbeitslosigkeit	
K A	86	19	18	123
Lehre	170	43	20	233
Fachspez	40	11	5	56
Hoch	28	4	3	35
	324	77	46	447

Ausbildungsspezifische Dauer der Arbeitslosigkeit für männliche Deutsche

Allgemeine Darstellung

Kontingenztafel der absoluten Häufigkeiten:

Eine $(k \times m)$-*Kontingenztafel der absoluten Häufigkeiten* besitzt die Form

$$
\begin{array}{c|ccc|c}
 & b_1 & \dots & b_m & \\
\hline
a_1 & h_{11} & \dots & h_{1m} & h_{1\cdot} \\
a_2 & h_{21} & \dots & h_{2m} & h_{2\cdot} \\
\vdots & \vdots & & \vdots & \vdots \\
a_k & h_{k1} & \dots & h_{km} & h_{k\cdot} \\
\hline
 & h_{\cdot 1} & \dots & h_{\cdot m} & n
\end{array}
$$

Notation

$h_{ij} = h(a_i, b_j)$ die absolute Häufigkeit der Kombination (a_i, b_j),

$h_{1\cdot}, \ldots, h_{k\cdot}$ die Randhäufigkeiten von X,

$h_{\cdot 1}, \ldots, h_{\cdot m}$ die Randhäufigkeiten von Y.

Die Kontingenztabelle gibt die gemeinsame Verteilung der Merkmale X und Y in absoluten Häufigkeiten wieder.

Kontingenztafel der relativen Häufigkeiten

Die $(k \times m)$-*Kontingenztafel der relativen Häufigkeiten* hat die Form

$$
\begin{array}{c|ccc|c}
 & b_1 & \dots & b_m & \\
\hline
a_1 & f_{11} & \dots & f_{1m} & f_{1\cdot} \\
\vdots & \vdots & & \vdots & \vdots \\
a_k & f_{k1} & \dots & f_{km} & f_{k\cdot} \\
\hline
 & f_{\cdot 1} & \dots & f_{\cdot m} & 1
\end{array}
$$

Noation

$$f_{ij} = h_{ij}/n \qquad \text{die relative Häufigkeit der Kombination } (a_i, b_j),$$

$$f_{i\cdot} = \sum_{j=1}^{m} f_{ij} = h_{i\cdot}/n, \quad i = 1, \ldots, k, \text{ die relativen Randhäufigkeiten zu } X,$$

$$f_{\cdot j} = \sum_{i=1}^{k} f_{ij} = h_{\cdot j}/n, \quad j = 1, \ldots, m, \text{ die relativen Randhäufigkeiten zu } Y$$

Die Kontingenztabelle gibt die gemeinsame Verteilung von X und Y wieder.

Bedingte Häufigkeiten

Zusammenhang zwischen X und Y aus *gemeinsamen* Häufigkeiten h_{ij} bzw. f_{ij} schwer ersichtlich.
Deshalb: Blick auf *bedingte* Häufigkeiten $\Rightarrow$ Verteilung des einen Merkmals für einen festgehaltenen Wert des zweiten Merkmals

Beispiel: Sonntagsfrage Prozentzahlen für Parteipräferenz in den

Schichten (Subpopulationen) „weibliche Wähler" und „männliche Wähler" $\stackrel{\wedge}{=}$ bedingte relative Häufigkeiten für Parteipräferenzen gegeben das Geschlecht.

Bedingte relative Häufigkeitsverteilung

Die *bedingte Häufigkeitsverteilung von Y unter der Bedingung* $X = a_i$, kurz $Y|X = a_i$, ist bestimmt durch

$$f_Y(b_1|a_i) = \frac{h_{i1}}{h_{i\cdot}}, \ldots, f_Y(b_m|a_i) = \frac{h_{im}}{h_{i\cdot}}.$$

Die *bedingte Häufigkeitsverteilung von X unter der Bedingung*

$Y = b_j$, kurz $X|Y = b_j$, ist bestimmt durch

$$f_X(a_1|b_j) = \frac{h_{1j}}{h_{\cdot j}}, \ldots, f_X(a_k|b_j) = \frac{h_{kj}}{h_{\cdot j}}.$$

Bemerkung

Wegen

$$\frac{h_{i1}}{h_{i\cdot}} = \frac{h_{i1}/n}{h_{i\cdot}/n} = \frac{f_{i1}}{f_{i\cdot}}$$

gilt auch

$$f_Y(b_1|a_i) = \frac{f_{i1}}{f_{i\cdot}}, \ldots, f_Y(b_m|a_i) = \frac{f_{im}}{f_{i\cdot}}$$

$$f_X(a_1|b_j) = \frac{f_{1j}}{f_{\cdot j}}, \ldots, f_X(a_k|b_j) = \frac{f_{kj}}{f_{\cdot j}}.$$

Merksatz:
Bedingte Häufigkeitsverteilungen werden durch Division der h_{ij} bzw. f_{ij} durch die entsprechende Zeilen- bzw. Spaltensumme gebildet.

Beispiel: Sonntagsfrage

- Zeile $X = a_1 =$ Männer

  Bedingte Häufigkeiten für Männer ($X = a_1$):

  1.Zeile / Randhäufigkeit für Männer

  $\frac{h(a_1,b_1)}{h(a_1)} = f(b_1|a_1), \ldots, \frac{h(a_j,b_j)}{h(a_1)} = f(b_j|a_1), \ldots$

  $\frac{144}{435} \approx 33\%$, $\frac{153}{435} \approx 35\%$ usw.

- Zeile $X = a_2 =$ Frauen analog,

  z.B. $\frac{200}{496} \approx 40\%$ usw.

Bedingte und gemeinsame Häufigkeiten

Man kann auch umgekehrt aus bedingten Häufigkeiten und Randhäufigkeiten die gemeinsamen Häufigkeiten ausrechnen. Bei der Sonntagsfrage ist die Tabelle der bedingten Häufigkeiten gegeben und dazu die Randhäufigkeiten

$$h(a_1) = 435 \text{ Männer}, \quad h(a_2) = 496 \text{ Frauen}; \quad n = 931.$$

$$\Rightarrow \quad \begin{array}{l} h(a_1) \cdot f(b_1|a_1) = h(a_1, b_1) \\ \\ 435 \cdot 33\% \approx 144 \quad \text{usw.} \end{array}$$

So wurde die Tabelle der gemeinsamen Häufigkeiten h_{ij} rekonstruiert.

Beispiel: Arbeitslosigkeit

$$f(\quad |a_i), \quad X = a_i, \ \ i = 1, \ldots, 4 \ \text{Ausbildungsniveau}$$

z.B. $\frac{86}{123} = 0.699, \ \frac{19}{123} = 0.154, \ldots$

$\frac{170}{233} = 0.730, \ldots$

usw.

Für festgehaltenes Ausbildungsniveau $(X = a_i)$ erhält man die relative Verteilung über die Dauer der Arbeitslosigkeit durch die folgende Tabelle.

Bedingte Verteilung

	Kurzzeit-arbeitslosigkeit	mittelfristige Arbeitslosigkeit	Langzeit-arbeitslosigkeit	
Keine Ausb.	0.699	0.154	0.147	1
Lehre	0.730	0.184	0.086	1
Fachspez. Aus.	0.714	0.197	0.089	1
Hochschula.	0.800	0.114	0.086	1

- Bedingen auf das Ausbildungsniveau:
  $\Rightarrow$ Verteilung der Dauer der Arbeitslosigkeit für die Subpopulationen "Keine Ausbildung", "Lehre", usw.
- Verteilungen lassen sich nun miteinander vergleichen

$\Rightarrow$Nun ersichtlich: Relative Häufigkeit für Kurzzeitarbeitslosigkeit ist in der Subpopulation "Hochschulabschluß" mit 0.8 am größten.

Zusammenhangsanalyse in Kontingenztabellen

Bisher: Tabellarische / graphische Präsentation
Jetzt: Maßzahlen für Stärke des Zusammenhangs zwischen X und Y.

Chancen und relative Chancen

- Zunächst 2×2 - Kontingenztafel

$$\begin{array}{cc|cc|c} & & \multicolumn{2}{c}{\text{Y}} & \\ & & 1 & 2 & \\ \cline{3-4} X & 1 & h_{11} & h_{12} & h_{1\cdot} \\ & 2 & h_{21} & h_{22} & h_{2\cdot} \\ \cline{3-4} & & h_{\cdot 1} & h_{\cdot 2} & n \end{array}$$

Chancen („Odds“)

- Wir betrachten die Merkmale X und Y zunächst asymmetrisch: Die Ausprägungen von X definieren (hier 2) Subpopulationen, Y ist das interessierende binäre Merkmal in diesen Subpopulationen
- Unter einer Chance (“odds”) versteht man nun das Verhältnis zwischen dem Auftreten von $Y = 1$ und $Y = 2$ in einer Subpopulation $X = a_i$.

Odds Ratio

- Die (empirische) bedingte Chance für festes $X = a_i$ ist bestimmt durch
$$\gamma(1,2|X = a_i) = \frac{h_{i1}}{h_{i2}}.$$
- Ein sehr einfaches Zusammenhangsmaß stellen die empirischen relativen Chancen (Odds Ratio) dar, die gegeben sind durch
$$\gamma(1,2|X = 1, X = 2) = \frac{\gamma(1,2|X = 1)}{\gamma(1,2|X = 2)} = \frac{h_{11}/h_{12}}{h_{21}/h_{22}} = \frac{h_{11}h_{22}}{h_{21}h_{12}},$$
d.h. $\gamma(1,2|X = 1, X = 2)$ ist das Verhältnis zwischen den Chancen der 1. Population ($X = 1$, 1. Zeile) zu den Chancen der 2. Population ($X = 2$, 2. Zeile).

Beispiel: Dauer der Arbeitslosigkeit

Beschränkt man sich jeweils nur auf zwei Kategorien der Merkmale Ausbildungsniveau und Dauer der Arbeitslosigkeit, erhält man beispielsweise die Tabelle

	Kurzzeit-arbeitslosigkeit	Mittel- und langfristige Arbeitslosigkeit
Fachspezifische Ausbildung	40	16
Hochschulabschluß	28	7

Daraus ergibt sich für Personen mit fachspezifischer Ausbildung die "Chance", kurzzeitig arbeitslos zu sein, im Verhältnis dazu, mittel –oder längerfristig arbeitslos zu sein, durch

$$\gamma(1,2|\text{fachspezifisch}) = \frac{40}{16} = 2.5.$$

Für Arbeitslose mit Hochschulabschluß erhält man

$$\gamma(1,2|\text{Hochschulabschluß}) = \frac{28}{7} = 4.$$

Für fachspezifische Ausbildung stehen die "Chancen" somit 5 : 2, für Arbeitslose mit Hochschulabschluß 4 : 1.

Man erhält für fachspezifische Ausbildung und Hochschulabschluß die relativen Chancen (Odds Ratio)

$$\gamma(1,2|\text{fachsp. Ausbildung, Hochschule}) = \frac{2.5}{4} = 0.625 = \frac{40 \cdot 7}{16 \cdot 28}$$

Interpretation „Odds Ratio“

- Wegen der spezifischen Form $\gamma(1,2|X=1,X=2)=(h_{11}h_{22})/(h_{21}h_{12})$ werden die relativen Chancen auch als Kreuzproduktverhältnis bezeichnet. Es gilt

$\gamma = 1$	Chancen in beiden Populationen gleich
$\gamma > 1$	Chancen in Population $X = 1$ besser als in Population $X = 2$
$\gamma < 1$	Chancen in Population $X = 1$ schlechter als in Population $X = 2$.

- Die relativen Chancen geben somit an, welche der Populationen die besseren Chancen besitzen und um wieviel besser diese Chancen sind.

Symmetrie

- Für die Kontingenztafel

h_{11}	h_{12}
h_{21}	h_{22}

  ist das *Kreuzproduktverhältnis* (*relative Chance* oder *Odds Ratio*) bestimmt durch

$$\gamma = \frac{h_{11}/h_{12}}{h_{21}/h_{22}} = \frac{h_{11}h_{22}}{h_{21}h_{12}}.$$

- Die asymmetrische Betrachtung der Merkmale X und Y wird aufgehoben

Fall - Kontroll - Studien

Beispiel: Morbus Alzheimer und Genetik

	ApoE3	ApoE4	Summe
Kontrolle	2258	803	3061
Fall	593	620	1213
	2851	1423	4274

$$OR = \frac{593/620}{2258/803} = 0.34$$

$\Rightarrow$ Chance für ApoE3 bei Fällen um den Faktor 3 niedriger als bei Kontrollen

$\overset{?}{\Rightarrow}$ ApoE4 Risiko-Faktor für Morbus Alzheimer

Fall - Kontroll - Studien

Zentrale Argumentation:

Odds Ratio ist symmetrisches Maß
d.h. Chancenverhältnis für Auftreten von ApoE4 bei Kontrolle zu Auftreten von ApoE4 bei Fällen

Person ist krank bei ApoE3
zu
Person ist krank bei ApoE4

$\Rightarrow$ Interpretation als **Risikofaktor** zulässig

Verallgemeinerung auf $k \times m$ Kontingenztafeln, Anmerkungen

- Verallgemeinerung des Verfahres auf mehr als zwei Ausprägungen mindestens eines Merkmals: Man beschränkt sich auf jeweils zwei Zeilen $X = a_i$ und $X = a_j$ und zwei Spalten $Y = b_r$ und $Y = b_s$ und die zugehörigen vier Zellen einer $(k \times m)$-Kontingenztafel.
- Verwendung einer Referenzkategorie
- Statt Odds Ratio wird oft der logarithmierte Odds Ratio verwendet

Anwendung: Apolipoprotein E und Morbus Alzheimer

Etablierter Zusammenhang zwischen Apolipoprotein Eε4 und Morbus Alzheimer

Daten aus Metaanalyse

ApoE genotype	$\varepsilon2\varepsilon2$	$\varepsilon2\varepsilon3$	$\varepsilon2\varepsilon4$	$\varepsilon3\varepsilon3$	$\varepsilon3\varepsilon4$	$\varepsilon4\varepsilon4$
Clinical controls	27	425	81	2258	803	71
Clinical Alzheimer	7	74	41	593	620	207
PM controls	3	75	18	358	120	8
PM Alzheimer	1	20	17	249	373	97

Anwendung: Apolipoprotein E und Morbus Alzheimer

OR im Vergleich zu $\varepsilon 3 \varepsilon 3$ (Referenz)

	$\varepsilon 2 \varepsilon 2$	$\varepsilon 2 \varepsilon 3$	$\varepsilon 2 \varepsilon 4$	$\varepsilon 3 \varepsilon 3$	$\varepsilon 3 \varepsilon 4$	$\varepsilon 4 \varepsilon 4$
OR (klinisch)	1	0.7	2.94	1	2.94	11.1
OR (post mortem)	0.5	0.4	1.4	1	4.5	17.4

Kontingenz- und χ^2-Koeffizient

Ausgangspunkt: Wie sollten gemeinsame Häufigkeiten $\tilde{h}_{ij}$ bzw. $\tilde{f}_{ij}$ verteilt sein, damit - bei vorgegebenen Randverteilungen - die Merkmale X und Y als „empirisch unabhängig“ angesehen werden können?

$$\begin{array}{c|ccc|c} & b_1 & \dots & b_m & \\ \hline a_1 & & & & h_{1\cdot} \\ \vdots & & ? & & \vdots \\ a_k & & & & h_{k\cdot} \\ \hline & h_{\cdot 1} & \dots & h_{\cdot m} & n \end{array}$$

Empirische Unabhängigkeit

Idee: X und Y „empirisch unabhängig“

$\Leftrightarrow$ Bedingte relative Häufigkeiten

$$f_Y(b_1|a_i), \ldots, f_Y(b_m|a_i), \quad i = 1, \ldots, k$$

sind in jeder Schicht $X = a_i$ identisch, d.h. unabhängig von a_i.
Formal:

$$\begin{aligned}
f_Y(b_1|a_1) &= f(b_1), \ldots, f_Y(b_m|a_1) = f_Y(b_m) \\
f_Y(b_1|a_2) &= f(b_1), \ldots, f_Y(b_m|a_2) = f_Y(b_m) \\
\vdots &= \vdots \\
f_Y(b_1|a_k) &= f(b_1), \ldots, f_Y(b_m|a_k) = f_Y(b_m)
\end{aligned}$$

Kunstbeispiel:

	b_1	b_2	b_3	
a_1	10	20	30	60
a_2	20	40	60	120
	30	60	90	180

$$
\begin{aligned}
f_Y(b_1|a_1) &= f_Y(b_1|a_2) = f_Y(b_1) = \frac{1}{6} \\
f_Y(b_2|a_1) &= f_Y(b_2|a_2) = f_Y(b_2) = \frac{1}{3} \\
f_Y(b_3|a_1) &= f_Y(b_3|a_2) = f_Y(b_3) = \frac{1}{2}
\end{aligned}
$$

Bemerkung: Lokale Odds Ratios sind alle 1

Wie sehen die "erwarteten" (absoluten und relativen) Häufigkeiten $\tilde{h}_{ij}$ und $\tilde{f}_{ij}$ also aus?

$$f_Y(b_1|a_i) = f(b_1), \ldots, f_Y(b_m|a_i) = f_Y(b_m),\ i = 1, \ldots, k$$

$$\Leftrightarrow \frac{\tilde{h}_{ij}}{h_{i\cdot}} = \frac{h_{\cdot j}}{n}$$

$$\Leftrightarrow \tilde{h}_{ij} = \frac{h_{i\cdot} h_{\cdot j}}{n}$$

$$\Leftrightarrow \tilde{f}_{ij} = f_{i\cdot} f_{\cdot j}$$

„Unabhängigkeitstabelle"

Idee: Vergleiche für jede Zelle (i, j) $\tilde{h}_{ij}$ mit tatsächlich beobachteten h_{ij}

$\Rightarrow \chi^2$-**Koeffizient**

Der χ^2-*Koeffizient* ist bestimmt durch

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(h_{ij} - \frac{h_{i\cdot} h_{\cdot j}}{n}\right)^2}{\frac{h_{i\cdot} h_{\cdot j}}{n}} = n \sum_{i} \sum_{j} \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}}$$

Eigenschaften des χ^2-Koeffizienten:

- $\chi^2 \in [0, \infty)$
- $\chi^2 = 0 \Leftrightarrow X$ und Y „empirisch unabhängig“
- χ^2 groß $\Leftrightarrow$ starker Zusammenhang
- χ^2 klein $\Leftrightarrow$ schwacher Zusammenhang
- Nachteil: χ^2 hängt vom Stichprobenumfang n und von der Dimension der Tafel ab.

Kontingenzkoeffizient und korrigierter Kontingenzkoeffizient

Weitere Normierung $\Rightarrow$ Kontingenzkoeffizient

Der Kontingenzkoeffizient ist bestimmt durch

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

und besitzt den Wertebereich $K \in \left[0, \sqrt{\frac{M-1}{M}}\right]$, wobei $M = \min\{k, m\}$.

Der korrigierte Kontingenzkoeffizient ergibt sich durch

$$K^* = K / \sqrt{\frac{M-1}{M}}$$

mit dem Wertebereich $K^* \in [0, 1]$.

Eigenschaften des Kontingenzkoeffizienten

- Es wird nur die *Stärke* des Zusammenhangs gemessen, nicht die Richtung wie beim Odds Ratio.
- Vorsicht ist geboten bei einem Vergleich von Kontingenztafeln mit stark unterschiedlichen Stichprobenumfängen, da χ^2 mit wachsendem Stichprobenumfang wächst, beispielsweise führte eine Verzehnfachung von h_{ij} und $\tilde{h}_{ij}$ zu zehnfachem χ^2.
- Sämtliche Maße benutzen nur das Nominalskalenniveau von X und Y.

Beispiel: Sonntagsfrage

Für die Kontingenztafel aus Geschlecht und Parteipräferenz für das Beispiel der Sonntagsfrage erhält man die in der folgenden Tabelle wiedergegebenen zu erwartenden Häufigkeiten $\tilde{h}_{ij}$.

	CDU/CSU	SPD	FDP	Grüne	Rest	
Männer	160.73 (144)	139.24 (153)	21.96 (17)	35.51 (26)	77.56 (95)	435
Frauen	183.27 (200)	158.76 (145)	25.04 (30)	40.49 (50)	88.44 (71)	496
	344	298	47	76	166	

Zu erwartende Häufigkeiten $\tilde{h}_{ij}$ und tatsächliche Häufigkeiten h_{ij} (in Klammern)

Interpretation:

- Wenn Geschlecht und Parteipräferenz keinen Zusammenhang aufweisen, wären 160.73 die CDU/CSU präferierende Männer zu erwarten.
- Tatsächlich wurden aber nur 144 beobachtet.

$$\Rightarrow \quad \chi^2\text{-Wert von } 20.065,$$

$$K = 0.145,$$

$$K^* = 0.205$$

Spezialfall: (2×2)-Tafel

Für den Spezialfall einer (2×2)-Tafel

$$
\begin{array}{|cc|c}
\hline
a & b & a+b \\
c & d & c+d \\
\hline
\multicolumn{1}{c}{a+c} & \multicolumn{1}{c}{b+d} &
\end{array}
$$

erhält man χ^2 aus

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}.$$

Beispiel: Arbeitslosigkeit

Aus der Kontingenztafel

	Mittelfristige Arbeitslosigkeit	Langfristige Arbeitslosigkeit	
Keine Ausbildung	19	18	37
Lehre	43	20	63
	62	38	100

erhält man also unmittelbar

$$\chi^2 = \frac{100(19 \cdot 20 - 18 \cdot 43)^2}{37 \cdot 63 \cdot 62 \cdot 38} = 2.826$$

und $K = 0.165$, $K^* = 0.234$.

Mehrdimensionale Kontingenztabellen

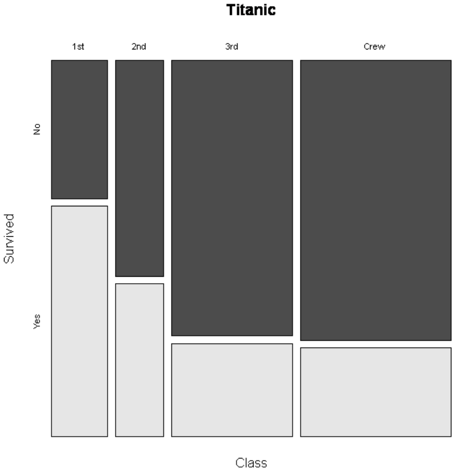
Beispiel: Überleben beim Titanic-Untergang

- Mehrere diskrete Merkmale: Geschlecht, Klasse, Kind/ Erwachsene, Überleben (Ja/Nein)
- Darstellung durch geeignete bedingte und marginale Verteilungen
- Berechnung von Odds-Ratio zweier Merkmale bedingt auf ein drittes Merkmal
- Graphische Darstellung durch Mosaik-Plot

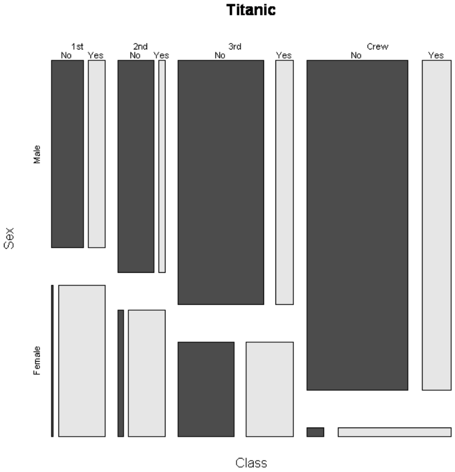
Mosaik-Plot

- Flächentreue Darstellung von Häufigkeiten
- Aufteilung schrittweise
- Zuerst Einflussgröße, zum Schluss nach Zielgröße aufteilen
- Gut geeignet für mehrkategoriale ordinale Daten
- Auch für höhere Dimensionen geeignet

Beispiel: Überlebende bei Titanic



Beispiel: Überlebende bei Titanic



Maße zur Übereinstimmung

Problemstellung

- Stimmen zwei oder mehrere Beobachter in ihrer Einschätzung überein? (im engl.: *rater agreement* oder *interrater agreement*)
- Beispiel: Zwei Professoren beurteilen die Referate oder Seminararbeiten von Studenten. Stimmen Sie in ihrer Bewertung (durch Noten) überein?
- Beispiel: Stimmen zwei oder mehrere Ärzte in ihrer Diagnose überein?
- Wahre Diagnose ("gold standard") unbekannt

Medizinisches Beispiel

Eine Klinik macht computertomografische Aufnahmen (CT–Bilder) oder Röntgenaufnahmen von sogenannten Kalkschultern, also Schultern, die Kalkablagerungen aufweisen. Mehrere Ärzte sollen anhand von 56 Patienten beurteilen, um welchen Typ es sich bei diesen Ablagerungen handelt (sog. Gärtner–Skala, ordinal):

- Typ 1: Aufbauphase (kann Monate oder Jahre dauern). Der Patient hat chronische Beschwerden, Schmerztherapie und Krankengymnastik bringen keine Linderung, eine Operation muss erwogen werden.
- Typ 2: Beginnende Auflösungsphase.
- Typ 3: Auflösung des Kalkdepots (kann Wochen oder Monate dauern). Behandlung meist konservativ durch Krankengymnastik und Schmerztherapie.

Medizinisches Beispiel (Fortsetzung)

- Die Beobachtungen lassen sich für zwei Ärzte in einer quadratischen Kontingenztafel veranschaulichen
- Die folgenden Kontingenztafeln geben die Ergebnisse für jeweils zwei verschiedene Ärzte bei den gleichen 56 Patienten wieder

Medizinisches Beispiel (Fortsetzung)

Tabelle: 3×3–Tafel für die Einschätzung von Arzt A und Arzt B

	Arzt A			
Arzt B	1	2	3	$\sum$
1	8	20	3	31
2	2	15	1	18
3	0	5	2	7
$\sum$	10	40	6	56

Medizinisches Beispiel (Fortsetzung)

Tabelle: 3×3–Tafel für die Einschätzung von Arzt C und Arzt D

	Arzt C			
Arzt D	1	2	3	$\sum$
1	27	6	3	36
2	2	6	3	11
3	0	4	5	9
$\sum$	29	16	11	56

Medizinisches Beispiel (Fortsetzung)

- Vollständige Übereinstimmung in der Einschätzung des Patienten liegt vor, wenn beide Ärzte den gleichen Typ (1, 2 oder 3) zuordnen
- *Bemerkung:* Gleiche Einstufung bedeutet nicht unbedingt, dass diese auch *richtig* im Sinne einer validen Einstufung ist. Beide können sich irren!

Üblichen Maßzahlen?

- Üblich bei Kontingenztafeln: χ^2, Kontingenzkoeffizient, etc.
- Man geht von vornherein davon aus, dass ein Zusammenhang bezüglich der Bewertung der Beobachter besteht
- Beobachter führen Bewertung *unabhängig voneinander durch*, d.h. kein Beobachter kennt die Bewertung des oder der anderen Beobachter, die Beurteilungen hängen aber zusammen, da sie jeweils am gleichen Subjekt (hier: Patient) durchgeführt werden

Medizinisches Beispiel (Fortsetzung)

- Für die Tabelle der Ärzte A und B erhält man $(8 + 15 + 2) = 25$ vollständige Übereinstimmungen
- Für die Tabelle der Ärzte C und D erhält man $(27 + 6 + 5) = 38$ vollständige Übereinstimmungen
- Kann man daraus sofort schließen, dass Übereinstimmung von C und D größer ist als von A und B? Im Prinzip ja, allerdings müssen wir beachten, dass ein gewisser *Teil der Übereinstimmung zufällig sein kann*

Maßzahl: Kappa–Koeffizient

- Kappa–Koeffizient nach Cohen (1960) dient zur Messung der Übereinstimmung in quadratischen $I \times I$–Kontingenztafeln
- Betrachtet wird primär die Hauptdiagonale der Kontingenztafel (vollständige Übereinstimmung)

Allgemeine Darstellung

Tabelle: Schema einer $I \times I$-Kontingenztafel. Die fettgedruckten Häufigkeiten liegen auf der Diagonalen und werden zur Berechnung von Kappa verwendet.

		Beobachter 1					
		1		i		I	$\sum$
	1	$\mathbf{n_{11}}$	$\cdots$	n_{1i}	$\cdots$	n_{1I}	n_{1+}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
Beobachter 2	i	n_{i1}	$\cdots$	$\mathbf{n_{ii}}$	$\cdots$	n_{iI}	n_{i+}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	I	n_{I1}	$\cdots$	n_{Ii}	$\cdots$	$\mathbf{n_{II}}$	n_{I+}
	$\sum$	n_{+1}	$\cdots$	n_{+i}	$\cdots$	n_{+I}	n

Allgemeine Darstellung

- Kappa–Koeffizient berücksichtigt darüber hinaus die *zufällige Übereinstimmung* die man auch bekommen würde, wenn die Einschätzungen der Beobachter keinen Zusammenhang aufweisen würden
- Wir berechnen daher zwei Größen

Berechnung von Kappa

- Relativer Anteil der Übereinstimmung beider Beobachter:

$$f_o = \sum_{i=1}^{I} f_{ii} = \sum_{i=1}^{I} \frac{n_{ii}}{n} = \frac{\sum_{i=1}^{I} n_{ii}}{n} \tag{3.1}$$

- Zufällige Übereinstimmung, wenn kein Zusammenhang besteht: dies ist äquivalent zur Bestimmung der sogenannten erwarteten relativen Häufigkeiten unter Unabhängigkeit (wie bei χ^2):

$$f_e = \sum_{i=1}^{I} f_{i+} f_{+i} = \sum_{i=1}^{I} \frac{n_{i+}}{n} \frac{n_{+i}}{n} = \sum_{i=1}^{I} \frac{n_{i+} n_{+i}}{n^2} = \frac{\sum_{i=1}^{I} n_{i+} n_{+i}}{n^2} \tag{3.2}$$

Berechnung von Kappa (Fortsetzung)

Der Kappa–Koeffizient ist definiert durch

$$\kappa = \frac{f_o - f_e}{1 - f_e} \ , \qquad (3.3)$$

Interpretation von Kappa

- Der Zähler ist die Differenz aus der beobachteten Übereinstimmung und der unter Zufälligkeit zu erwartenden Übereinstimmung. Dies ist damit ein Maß für die über die Zufälligkeit hinausgehende Übereinstimmung der Beobachter (*engl.: chance corrected agreement*)
- Die Eins im Nenner stellt die maximal mögliche relative Häufigkeit für Übereinstimmung dar, nämlich wenn alle Beobachtungen auf der Diagonalen der Kontingenztafel liegen und sämtliche Nebendiagonalen nur Nullen enthalten

Interpretation von Kappa (Fortsetzung)

- Der Kappa–Koeffizient ist damit ebenfalls Eins, wenn alle Beobachtungen auf der Diagonalen der Kontingenztafel liegen und sämtliche Nebendiagonalen nur Nullen enthalten. Er kann auch negativ werden, wenn zum Beispiel keine Übereinstimmung da ist (im Extremfall: Nullen auf der Diagonalen)
- Der Kappa–Koeffizient ist Null, wenn für alle $i = 1, \ldots, I$ gilt: $f_{ii} = f_{i+}f_{+i}$, das heißt, wenn exakte Unabhängigkeit in der beobachteten Tafel vorliegt

Medizinisches Beispiel (Fortsetzung), Arzt A und B

$$f_o = \frac{8+15+2}{56} = \frac{25}{56}$$

und

$$\begin{aligned} f_{11} &= \frac{31 \cdot 10}{56^2} \\ f_{22} &= \frac{18 \cdot 40}{56^2} \\ f_{33} &= \frac{7 \cdot 6}{56^2} \\ f_e &= \frac{31 \cdot 10 + 18 \cdot 40 + 7 \cdot 6}{56^2} \\ &= \frac{1072}{56^2} \end{aligned}$$

Damit erhalten wir

$$\kappa = \frac{\frac{25}{56} - \frac{1072}{56^2}}{1 - \frac{1072}{56^2}} = 0.159$$

Medizinisches Beispiel (Ergebnis)

- Entsprechend: Arzt D und C: $\kappa = 0.445$
- Die Einschätzung von A und B ist „schwach übereinstimmend“
- Die Einschätzung von C und D ist „mäßig übereinstimmend“

Veranschaulichung

Zwei mögliche Ergebnisse für die Einschätzung zweier Beobachter für 20 Objekte bezüglich eines dichotomen Merkmals:

Tabelle: Problematik von Kappa

	1	0
1	10	1
0	0	9

	1	0
1	18	1
0	0	1

In beiden Fällen erhält man eine Übereinstimmung in 19 von 20 Objekten, oder einen Wert von $f_o = 0.95$. Allerdings ist Kappa für die linke Tafel 0.9 und für die rechte Tafel nur 0.64

Erklärung

- Die Randverteilungen der beiden Beobachter unterscheiden sich dabei jeweils nur gering, d.h. sie scheinen gut kalibriert zu sein (daran liegts also nicht)
- Offenbar ist die Prävalenz (d.h. der Grundanteil in der untersuchten Population/Stichprobe) in der rechten Tafel für die Ausprägung „1“ wesentlicher größer ist als für die Ausprägung „0“. Damit ist aber auch die zufällige Übereinstimmung wahrscheinlicher! Genau dieser Effekt wird bei Kappa berücksichtigt und herausgerechnet
- Befürworter von Kappa sehen diesen Effekt als wünschenswert an!

Kritik an Kappa

Zwei Aspekte werden vermischt:

- Die Beobachter können einen Bias aufweisen, das heisst, die Nicht–Übereinstimmung beruht darauf, dass zum Beispiel Lehrer 1 generell bessere Noten vergibt als Lehrer 2. Man sagt dann auch, dass die Beobachter nicht *kalibriert* sind
- Die Beobachter schätzen die Subjekte verschieden ein. Die Nicht–Übereinstimmung beruht darauf, dass Beobachter 1 zum Beispiel Subjekt 1 höher einstuft als Subjekt 2, Beobachter 2 dagegen Subjekt 2 höher als Subjekt 1. Beispiel: Lehrer 1 gibt Schüler 1 eine bessere Note als Schüler 2, Lehrer 2 dagegen gibt Schüler 2 eine bessere Note als Schüler 1.

Aspekt 2 ist der eigentlich uns interessierende Aspekt, während Aspekt 1 (Bias) nach Möglichkeit durch Kalibrierung vermieden werden sollte

Veranschaulichung der Kritik an Kappa

Tabelle: Problematik von Kappa

9	3		5	7
5	3		1	7

In der linken Tafel ergibt sich ein Kappa von 0.13, in der rechten Tafel ein Kappa von 0.26, obwohl wieder in beiden Fällen 12 von 20 Objekten ($f_o = 0.6$) übereinstimmend eingestuft wurden

Veranschaulichung der Kritik an Kappa

Erklärung:

- Die Randverteilungen in der rechten Tafel divergieren wesentlich stärker als in der linken Tafel (d.h. hier kann ein Kalibrierungsproblem vorliegen)
- Rechte Tafel:
  - Beobachter 1: $((5+7)/20, (1+7)/20) = (0.6, 0.4)$
  - Beobachter 2: $((5+1)/20, (7+7)/20) = (0.3, 0.7)$
- Linke Tafel:
  - Beobachter 1: $(12/20, 8/20) = (0.6, 0.4)$
  - Beobachter 2: $(14/20, 6/20) = (0.7, 0.3)$

Fazit: Beobachter müssen unbedingt kalibriert werden!

Erweiterungen von Kappa

- Ein weiterer Nachteil von Kappa ist, dass nur die Diagonale berücksichtigt wird
- Wenn die Bewertungsskala sehr viele verschiedene Merkmalsausprägungen besitzt, so liegen aufeinanderfolgende Ausprägungen oft nicht so weit auseinander
- Wenn zwei Beobachter sich nur gering in der Bewertung unterscheiden, so sollte eine Maßzahl dies auch berücksichtigen können
- Eine solche Maßzahl ist das *gewichtete Kappa*

Gewichtetes Kappa

- Das gewichtete Kappa wurde von Cohen (1968) vorgeschlagen
- Formal gehen alle Zellen der Kontingenztafel in die Berechnung ein
- Die Zellen auf der Hauptdiagonalen erhalten das höchste Gewicht (in der Regel Gewicht Eins), während die anderen Zellen ein geringeres Gewicht erhalten
- Die Idee ist, die Zellen umso geringer zu gewichten, je schlechter die Übereinstimmung der beiden Beobachter ist

Definition: Gewichtetes Kappa

Das gewichtete Kappa ist definiert als

$$\kappa_w = \frac{f_o^* - f_e^*}{1 - f_e^*} \; , \qquad (3.4)$$

mit

$$\begin{aligned} f_o^* &= \sum_{i=1}^{I} \sum_{j=1}^{I} w_{ij} f_{ij} \\ f_e^* &= \sum_{i=1}^{I} \sum_{j=1}^{I} w_{ij} f_{i\cdot} f_{\cdot j} \end{aligned}$$

Dabei wird f_o^* wie beim ungewichteten Kappa als relativer Anteil der Übereinstimmung beider Beobachter aufgefasst, während f_e^* die zufällige Übereinstimmung darstellt, wenn kein Zusammenhang bestehen würde

Wahl der Gewichte w_{ij}

Zwei populäre Vorschläge sind

$$w_{ij} = 1 - \frac{(i-j)^2}{(I-1)^2} \tag{3.5}$$

und

$$w_{ij}^* = 1 - \frac{|i-j|}{I-1} \ . \tag{3.6}$$

Gewichte w_{ij} und w_{ij}^* im 3×3-Fall

Tabelle: Gewichte w_{ij} einer 3×3–Tafel

1.0	0.75	0.0
0.75	1.0	0.75
0.0	0.75	1.0

Tabelle: Gewichte w_{ij}^* einer 3×3–Tafel

1.0	0.5	0.0
0.5	1.0	0.5
0.0	0.5	1.0

Die Zellen der größten Nichtübereinstimmung (Zelle $(1,3)$ und $(3,1)$ bei einer 3×3–Tafel) werden in beiden Fällen mit 0 gewichtet, die Zellen auf der Diagonalen mit Gewicht 1

Gewichte w_{ij} und w_{ij}^* im 4×4-Fall

Tabelle: Gewichte w_{ij} einer 4×4–Tafel

1.0	0.89	0.56	0.0
0.89	1.0	0.89	0.56
0.56	0.89	1.0	0.89
0.0	0.56	0.89	1.0

Tabelle: Gewichte w_{ij}^* einer 4×4–Tafel

1.0	0.67	0.33	0.0
0.67	1.0	0.67	0.33
0.33	0.67	1.0	0.67
0.0	0.33	0.67	1.0

Medizinisches Beispiel (Fortsetzung), Arzt A und B

Verwendung der Gewichte gemäß Formel (3.5):

$$
\begin{aligned}
nf_o \quad = \quad & 8 \cdot 1.0 + 20 \cdot 0.75 + 3 \cdot 0.0 \\
& + 2 \cdot 0.75 + 15 \cdot 1.0 + 1 \cdot 0.75 \\
& + 0 \cdot 0.0 + 5 \cdot 0.75 + 2 \cdot 1.0
\end{aligned}
$$

und damit

$$f_o = \frac{46}{56} = 0.8214 \ ,$$

sowie

$$
\begin{aligned}
n^2 f_e \quad = \quad & 31 \cdot 10 \cdot 1.0 + 31 \cdot 40 \cdot 0.75 + 31 \cdot 6 \cdot 0.0 \\
& + 18 \cdot 10 \cdot 0.75 + 18 \cdot 40 \cdot 1.0 + 18 \cdot 6 \cdot 0.75 \\
& + 7 \cdot 10 \cdot 0.0 + 7 \cdot 40 \cdot 0.75 + 7 \cdot 6 \cdot 1.0
\end{aligned}
$$

Medizinisches Beispiel (Fortsetzung), Arzt A und B

Also

$$f_e = \frac{2428}{56^2} = 0.7742$$

Damit erhalten wir

$$\kappa_w = \frac{0.8214 - 0.7742}{1 - 0.7742} = 0.209$$

Medizinisches Beispiel (Fortsetzung), Arzt A und B

Verwendung der Gewichte gemäß Formel (3.6):

$$\begin{aligned} nf_o &= 8 \cdot 1.0 + 20 \cdot 0.5 + 3 \cdot 0.0 \\ &\quad + 2 \cdot 0.5 + 15 \cdot 1.0 + 1 \cdot 0.5 \\ &\quad + 0 \cdot 0.0 + 5 \cdot 0.5 + 2 \cdot 1.0 \end{aligned}$$

und damit

$$f_o = \frac{39}{56} = 0.6964 \ ,$$

sowie

$$\begin{aligned} n^2 f_e &= 31 \cdot 10 \cdot 1.0 + 31 \cdot 40 \cdot 0.5 + 31 \cdot 6 \cdot 0.0 \\ &\quad + 18 \cdot 10 \cdot 0.5 + 18 \cdot 40 \cdot 1.0 + 18 \cdot 6 \cdot 0.5 \\ &\quad + 7 \cdot 10 \cdot 0.0 + 7 \cdot 40 \cdot 0.5 + 7 \cdot 6 \cdot 1.0 \end{aligned}$$

Medizinisches Beispiel (Fortsetzung), Arzt A und B

Also

$$f_e = \frac{1976}{56^2} = 0.6301$$

Damit erhalten wir

$$\kappa_{w^*} = \frac{0.6964 - 0.6301}{1 - 0.6301} = 0.179$$

Für dieses Beispiel erhalten wir also

$$\kappa = 0.159 < \kappa_{w^*} = 0.179 < \kappa_w = 0.209$$

Medizinisches Beispiel (Fortsetzung), Arzt C und D

Wir erhalten unter Verwendung der Gewichte aus (3.5)

$$\kappa_w = 0.601$$

und unter Verwendung der Gewichte aus (3.6)

$$\kappa_{w^*} = 0.525$$

Auch hier erhalten wir

$$\kappa = 0.445 < \kappa_{w^*} = 0.525 < \kappa_w = 0.601$$

Cohens Kappa: Zusammenfassung

- Nützliches Maß zur Übereinstimmung bei binären und metrischen Merkmalen
- Im ordinalen Fall sollte gewichtetes Kappa verwendet werden
- Zusätzlich sollte Kalibrierung geprüft werden