4. Zusammenhangsmaße für diskrete Merkmale

Yichen Han

16. November 2023

Inhalt

Ein Multivariates Experiment I

Die heutige Wiederholung wird durch ein interaktives multivariates Experiment ersetzt.

Ihr werdet euer Verständnis für die in der Vorlesung behandelten Konzepte vertiefen. Erlebt den Prozess einer deskriptiven Datenanalyse hautnah – von der Datenerhebung bis zur Auswertung.

Die Daten für das Experiment stammen direkt von euch und euren Interaktionen.

Dieses Experiment ist nur als Lehrmittel konzipiert. Es soll keine umfassende wissenschaftliche Forschungsarbeit repräsentieren.

Ein Multivariates Experiment II: Überblick

Gaming-Verhalten und Entscheidungsfindung

Ziel des Experiments

- Untersuchung des Zusammenhangs zwischen Spielpräferenzen und Entscheidungsfindungsprozessen.
- Identifikation von Mustern im Spielverhalten und deren Korrelation mit realen Entscheidungsstilen.

Teilnahme

- Daten werden durch eine Online-Umfrage anonym erhoben.
- Keine Vorkenntnisse über Gaming oder Psychologie erforderlich.
- Möglichkeit, Einblicke in die eigene Entscheidungsfindung zu gewinnen.

Ein Multivariates Experiment III

Abbildung: Füllen Sie diese Umfrage aus

Methoden

- χ^2-Koeffizient, Kontingenzkoeffizient, Odds Ratio
- Kontingenztafel, Mosaikplots

Anmerkung

- f und h gut unterscheiden...
- Stärke des Zusammenhangs sinnvoll vergleichen?
  $\Rightarrow$ Immer korrigierter Kontingenzkoeffizient:
  Ein Vergleich zwischen $[0, 1]$ ist viel machbarer als in
  $[0, n \cdot (\min\{m, k\} - 1)]$!
- Einschränkung der Odds Ratio: nur für binäre Ausprägungen anwendbar.
- Interpretation der Zusammenhangsmaße besonders beachten.

Inhalt

Achtung

Achtung: ausführliche Rechenschritte der komplizierten Zusammenhangsmaße werden gespart!

	w	m	$h_{i\cdot}$
nein	10	14	24
ja	12	8	20
$h_{\cdot j}$	22	22	44

44 Teilnehmer: 27.27%: weiblich UND machen ein Auslandssemester, 63.64% (männlich): kein Auslandssemester machen, insgesamt 45.45% wollen ein Auslandssemester machen.

a Erstellen Sie die zugehörige Kontingenztabelle der absoluten Häufigkeiten.

Merkmale: Geschlecht (G) und Bereitschaft für Auslandssemester (A)
Ausprägungen: $\{w, m\}$ und $\{ja, nein\}$.

$$f(ja \cap m) = f(ja) - f(ja \cap w) = 0{,}1818$$

$$f(nein) = 1 - f(ja) = 0.5455$$

$$f(m) = \frac{f(ja \cap m)}{f(ja \mid m)} = 0{,}5 \Rightarrow f(w) = 0{,}5$$

$$f(nein \cap w) = f(w) - f(ja \cap w) = 0{,}2273$$

44 Teilnehmer: 27.27%: weiblich UND machen ein Auslandssemester, 63.64% (männlich): kein Auslandssemester machen, insgesamt 45.45% wollen ein Auslandssemester machen.

a Erstellen Sie die zugehörige Kontingenztabelle der absoluten Häufigkeiten.

Merkmale: Geschlecht (G) und Bereitschaft für Auslandssemester (A)
Ausprägungen: $\{w, m\}$ und $\{ja, nein\}$.

	G		
A	w	m	$h_{i\bullet}$
ja	12	8	20
nein	10	14	24
$h_{\bullet j}$	22	22	44

Aufgabe 1.2

b Zsmh. Geschlecht $\sim$ Auslandsaufenthalt? Verwenden Sie hierfür eine geeignete Maßzahl mit dem Wertebereich $[0, 1]$ und interpretieren Sie das Ergebnis.

b Zsmh. Geschlecht $\sim$ Auslandsaufenthalt? Verwenden Sie hierfür eine geeignete Maßzahl mit dem Wertebereich $[0,1]$ und interpretieren Sie das Ergebnis.

$$K^* = \frac{\sqrt{\frac{\chi^2}{\chi^2+n}}}{\sqrt{\frac{M-1}{M}}} \in [0,1], \quad M = \min\{k,l\} = 2$$

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

b Zsmh. Geschlecht $\sim$ Auslandsaufenthalt? Verwenden Sie hierfür eine geeignete Maßzahl mit dem Wertebereich $[0,1]$ und interpretieren Sie das Ergebnis.

$$K^* = \frac{\sqrt{\frac{\chi^2}{\chi^2+n}}}{\sqrt{\frac{M-1}{M}}} \in [0,1], \quad M = \min\{k, l\} = 2$$

chisq.test()

$$\begin{aligned} \chi^2 &= \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \\ &= \frac{44(10\times 8 - 14\times 12)^2}{(10+14)(10+12)(14+8)(12+8)} \approx 1.467 \\ K* &= \frac{\sqrt{1.467/(1.467+44)}}{\sqrt{1/2}} \approx 0.254 \Rightarrow \text{schwacher Zsmh.} \end{aligned}$$

Aufgabe 1.3

c Vergleichen Sie die Chancen für einen geplanten Auslandsaufenthalt zwischen den Geschlechtern mit einer geeigneten Maßzahl.

c Vergleichen Sie die Chancen für einen geplanten Auslandsaufenthalt zwischen den Geschlechtern mit einer geeigneten Maßzahl.

$$\gamma(ja, nein|w, m) = \frac{h_{11} h_{22}}{h_{21} h_{12}} = \frac{12 \cdot 14}{10 \cdot 8} = 2.1$$

Interpretation: Die Chance (Odds) weiblicher Befragten auf einen Auslandsaufenthalt ist etwas mehr als doppelt so hoch wie die Chance der männlichen Befragten.

d Wie groß ist der Anteil unter den weiblichen Befragten, ein Auslandssemester absolvieren zu wollen?

c Vergleichen Sie die Chancen für einen geplanten Auslandsaufenthalt zwischen den Geschlechtern mit einer geeigneten Maßzahl.

$$\gamma(ja, nein | w, m) = \frac{h_{11} h_{22}}{h_{21} h_{12}} = \frac{12 \cdot 14}{10 \cdot 8} = 2.1$$

Interpretation: Die Chance (Odds) weiblicher Befragten auf einen Auslandsaufenthalt ist etwas mehr als doppelt so hoch wie die Chance der männlichen Befragten.

d Wie groß ist der Anteil unter den weiblichen Befragten, ein Auslandssemester absolvieren zu wollen?

$$f(ja | w) = \frac{f(ja \cap w)}{f(w)} = \frac{0.2727}{0.5} \approx 0.545$$

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

(3) $f_{13} = f_{1\cdot} \cdot f_{\cdot 3} = 0.3 \times 0.6$

$= 0.18 \Rightarrow f_{12} = 0{,}1$

(4) $f_{\cdot 1} = \frac{0.3}{0.6} = 0.5$

	b_1	b_2	b_3	$\sum$
a_1	0.3	f_{12} 0.12	f_{13} 0.18	$f_{1\bullet}$ 0.6
a_2	f_{21} 0.2	f_{22}	0.12	0.4
$\sum$	$f_{\bullet 1}$ 0.5	$f_{\bullet 2}$ 0.2	$f_{\bullet 3}$ 0.3	$f_{\bullet\bullet}$ 1

Hinweis: $f_{ab} = f_{a\cdot} f_{\cdot b}$

(1) $f_{12} + f_{13} = 0.3$

$f_{21} + f_{22} = 0{,}28$

(2) $f_{\cdot 3} = \frac{f_{23}}{f_{2\cdot}} = 0.3$

(5) $f_{\cdot 2} = \frac{f_{12}}{f_{1\cdot}} = 0.2$

$\Rightarrow f_{21} = 0.08$

Aufgabe 2

	b_1	b_2	b_3	$\sum$
a_1	0.3	f_{12}	f_{13}	$f_{1\bullet}$
a_2	f_{21}	f_{22}	0.12	0.4
$\sum$	$f_{\bullet 1}$	$f_{\bullet 2}$	$f_{\bullet 3}$	$f_{\bullet\bullet}$

Hinweis: $f_{ab} = f_{a\cdot} f_{\cdot b}$

	b_1	b_2	b_3	$\sum$
a_1	0.3	**0.12**	**0.18**	**0.6**
a_2	**0.2**	**0.08**	0.12	0.4
$\sum$	**0.5**	**0.2**	**0.3**	**1**

Y

		Anzahl der reparierten Autos								
		3	5	6	8	10	11	12	15	
	2	3	2	0	0	0	0	0	0	5
Anzahl der	3	1	2	2	0	0	0	0	0	
Beschäftigten	5	1	0	4	4	1	0	0	0	
X	8	0	1	4	5	3	5	2	0	
	10	0	0	0	1	1	0	3	5	

a Bestimmen Sie die Randhäufigkeiten von X und Y. Geben Sie außerdem die bedingten relativen Häufigkeiten von Y unter der Bedingung $X = 8$ an.

		Anzahl der reparierten Autos								
		3	5	6	8	10	11	12	15	$h_{i\bullet}$
Anzahl der Beschäftigten	2	3	2	0	0	0	0	0	0	5
	3	1	2	2	0	0	0	0	0	5
	5	1	0	4	4	1	0	0	0	10
	8	0	1	4	5	3	5	2	0	20
	10	0	0	0	1	1	0	3	5	10
	$h_{\bullet j}$	5	5	10	10	5	5	5	5	50

a Bestimmen Sie die Randhäufigkeiten von X und Y. Geben Sie außerdem die bedingten relativen Häufigkeiten von Y unter der Bedingung $X = 8$ an.

b_j	3	5	6	8	10	11	12	15
$f_Y(b_j \mid X = 8)$	0	$\frac{1}{20} = 0.05$	0.2	0.25	0.15	0.25	0.1	0

		Anzahl der reparierten Autos								
		3	5	6	8	10	11	12	15	$h_{i\bullet}$
	2	3	2	0	0	0	0	0	0	5
Anzahl der	3	1	2	2	0	0	0	0	0	5
Beschäftigten	5	1	0	4	4	1	0	0	0	10
	8	0	1	4	5	3	5	2	0	20
	10	0	0	0	1	1	0	3	5	10
	$h_{\bullet j}$	5	5	10	10	5	5	5	5	50

b Geben Sie unter den Betrieben mit 8 Beschäftigten den Anteil derjenigen an, die 10 Autos repariert haben.

$$f_Y(Y = 10 | X = 8) = \frac{3}{20} = 15\%$$

		Anzahl der reparierten Autos								
		3	5	6	8	10	11	12	15	$h_{i\bullet}$
	2	3	2	0	0	0	0	0	0	5
Anzahl der	3	1	2	2	0	0	0	0	0	5
Beschäftigten	5	1	0	4	4	1	0	0	0	10
	8	0	1	4	5	3	5	2	0	20
	10	0	0	0	1	1	0	3	5	10
	$h_{\bullet j}$	5	5	10	10	5	5	5	5	50

c Geben Sie den Anteil der Betriebe an, die genau 8 Beschäftigte haben und höchstens 10 Autos repariert haben.

$$f(Y \leq 10, X = 8) = \frac{0 + 1 + 4 + 5 + 3}{50} = 26\%$$

		Anzahl der reparierten Autos								
		3	5	6	8	10	11	12	15	$h_{i\bullet}$
Anzahl der Beschäftigten	2	3	2	0	0	0	0	0	0	5
	3	1	2	2	0	0	0	0	0	5
	5	1	0	4	4	1	0	0	0	10
	8	0	1	4	5	3	5	2	0	20
	10	0	0	0	1	1	0	3	5	10
	$h_{\bullet}$	5	5	10	10	5	5	5	5	50

d Geben Sie den Anteil der Betriebe an, die höchstens zehn Autos repariert haben.

$$f(Y \leq 10) = 1 - f(Y > 10) = 1 - \frac{5+2+3+5}{50} = 70\%$$

Aufgabe 3.4

e Berechnen Sie die bedingten arithmetischen Mittel von Y unter der Bedingung $X = a_i$ für alle $i = 1, \ldots, 5$.

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

- e Berechnen Sie die bedingten arithmetischen Mittel von Y unter der Bedingung $X = a_i$ für alle $i = 1, \ldots, 5$.

$f_Y(b_j|a_i)$

$= f(Y = b_j | X = a_i)$

$$\bar{y}_{X=a_i} = \frac{1}{\sum_{j=1}^{8} w_j} \sum_{j=1}^{8} b_j \cdot w_j$$

$$= \frac{1}{\underbrace{\cancel{\sum_{j=1}^{8} f_Y(b_j|a_i)}}_{=1}} \sum_{j=1}^{8} b_j \cdot f_Y(b_j|a_i) = \sum_{j=1}^{8} b_j \cdot \frac{h_{ij}}{h_{i\cdot}}$$

$Y = 3 = b_1$

$X = 2$

$3 \cdot \frac{3}{5}$

a_i	2	3	5	8	10
$\bar{y}_{X=a_i}$	3.8	5	6.9	8.9	12.9

Aufgabe 3.5

f Sind X und Y empirisch unabhängig? Begründen Sie Ihre Entscheidung.

f Sind X und Y empirisch unabhängig? Begründen Sie Ihre Entscheidung.

z.Z. $\forall i, j,\; h_{ij} = \tilde{h}_{ij} = \frac{h_{i\cdot} h_{\cdot j}}{n} \cancel{\Leftrightarrow f_{ij} = f_{i\cdot} f_{\cdot j}}.$

f Sind X und Y empirisch unabhängig? Begründen Sie Ihre Entscheidung.

z.Z. $\forall i, j,\ h_{ij} = \tilde{h}_{ij} = \frac{h_{i.}h_{.j}}{n} \Leftrightarrow f_{ij} = f_{i.}f_{.j}$.

Gegenbeispiel: $\tilde{h}_{11} = \frac{5 \times 5}{50} = 0.5 \neq 3 = h_{11}$.

Daraus folgt, Merkmale X und Y sind nicht empirisch unabhängig.