

9 Signifikanztests Teil III

Zoltán Zomotor

Versionsstand: 18. Mai 2015, 09:33

Die nummerierten Felder bitte während der Vorlesung ausfüllen.



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Germany License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/de/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Bitte hier notieren, was beim Bearbeiten unklar geblieben ist:

Inhaltsverzeichnis

1	Kontingenztest	1
2	Korrelationstest	5

1 Kontingenztest¹

- Gegeben: Zwei verbundene einfache Stichproben

¹ _____
|

- Hypothesenpaar:

H_0 : Die Merkmale X und Y sind ² _____
|

¹Bedeutung von „Kontingenz“ (duden.de): Häufigkeit bzw. Grad der Wahrscheinlichkeit des gemeinsamen Auftretens zweier Sachverhalte, Merkmale usw.

H_1 : Die Merkmale X und Y sind | _____³

• Vorgehensweise entspricht dem | _____⁴

• basiert aber auf | _____⁵

• Vorgehen:

Schritt 1: Ein Signifikanzniveau | _____⁶ festlegen.

Schritt 2: Den Testfunktionswert folgendermaßen ermitteln:

2.1 Die x -Achse in $k \geq 2$ und die y -Achse in $\ell \geq 2$ disjunkte, aneinander angrenzende

Intervalle | _____⁷ unterteilen.

2.2 Eine Kontingenztabelle mit Randhäufigkeiten erstellen:

		Y				
		B_1	B_2	\dots	B_ℓ	
X	A_1	h_{11}	h_{12}	\dots	$h_{1\ell}$	h_{A_1}
	A_2	h_{21}	h_{22}	\dots	$h_{2\ell}$	h_{A_2}
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	A_k	h_{k1}	h_{k2}	\dots	$h_{k\ell}$	h_{A_k}
		h_{B_1}	h_{B_1}	\dots	h_{B_ℓ}	_____ ⁸

Dabei bezeichnen | _____⁹ die Anzahl der beobach-

teten Paare | _____¹⁰

2.3 Zu jeder Kombination aus $i = 1, \dots, k$ und $j = 1, \dots, \ell$ die Größe

| _____¹¹ berechnen.

2.4 Den Testfunktionswert v folgendermaßen berechnen:

| _____¹²

Schritt 3: Mit dem Fraktilswert $x_{1-\alpha}$ der χ^2 -Verteilung
 den Verwerfungsbereich festlegen.

Schritt 4: H_0 genau dann ablehnen, wenn $\chi^2 > x_{1-\alpha}$ gilt.

• Bemerkungen:

(i) Test ist nur anwendbar, wenn

1. $k = \ell = 2$

2. Falls nicht erfüllt \Rightarrow Schritt 2.1 entfällt

(ii) Falls diskrete Verteilungen:

- pro Ausprägung ein Intervall (falls (i) erfüllt)
- Schritt 2.1 entfällt

(iii) Falls $k = \ell = 2$:

- Schritt 2.3: kann entfallen
- Schritt 2.4: Berechne Testfunktionswert gemäß

3. $p_i = P(X \in A_i), q_j = P(Y \in B_j)$ bekannt:

(iv) Falls Randwahrscheinlichkeiten $p_i = P(X \in A_i), q_j = P(Y \in B_j)$ bekannt:

- Schritt 2.2: Randhäufigkeiten entfallen

4. Schritt 2.3: Ersetze p_{ij} durch $p_i \cdot q_j$

5. Schritt 3: Ersetze n durch $\sum_{i,j} n_{ij}$

- Beispiel: Eine regionale Tageszeitung lässt per Telefon ihre Leser befragen, wie gut sie sich durch die Zeitung informiert fühlen. Zur Auswahl stehen die Antworten gut, mäßig, schlecht. Außerdem wurde nach dem Alter des Urteilenden gefragt. Hier das Ergebnis:

		Bewertung		
		gut	mäßig	schlecht
Alter	unter 20	24	11	5
	20 bis 40	32	17	11
	über 40	64	22	14

Ist die Meinung der Leser über ihre Zeitung mit dem Signifikanzniveau $\alpha = 0.01$ signifikant vom Alter abhängig? $\Rightarrow H_0$: „Alter und Meinung sind voneinander unabhängig“

²¹
1. |

2. 2.1 Diskrete Verteilung, A_j, B_j gegeben

2.2

h_{ij}				
	24	11	5	²²
	32	17	11	
	64	22	14	
	²³			²⁴

2.3

\tilde{h}_{ij}				
	²⁵			

2.4

²⁶				
---------------	--	--	--	--

²⁷
3. |

²⁸
4. |

Die Meinung hängt somit ²⁹ _____ vom Alter ab.

2 Korrelationstest

- Gegeben: Zwei verbundene einfache Stichproben

³⁰ _____
 |
 mit
³¹ _____
 |

- Hypothesenpaare:

³² _____
 |

- Vorgehen:

Schritt 1: Ein Signifikanzniveau ³³ _____ festlegen.

Schritt 2: Mit Hilfe des empirischen oder Stichproben-Korrelationskoeffizienten

³⁴ _____
 |

der Beobachtungspaare $(x_i, y_i), i = 1, \dots, n$ den Testfunktionswert v berechnen:

³⁵ _____
 |

Schritt 3: Den Verwerfungsbereich festlegen:

³⁶ _____
 |

Fraktilewerte $x_{1-\alpha/2}$ bzw. $x_{1-\alpha}$ der ³⁷ _____ -Verteilung entnehmen.

Schritt 4: H_0 genau dann verwerfen, wenn ³⁸ _____ gilt

- Beispiel ([BV12, Aufgabe 3.3.11]): In einer Behörde wurde an 25 zufällig ausgewählten Arbeitstagen des Jahres 2009 jeweils die Gesamtlänge Y aller abgehenden Telefonate ermittelt (in Minuten). Um die Abhängigkeit der „Telefonierfreudigkeit“ vom Wetter zu untersuchen, wurde an diesen Tagen zusätzlich um 13 Uhr der Luftdruck X registriert (in Hektopascal). Aus den $n = 25$ Beobachtungspaaren sind dann
 - die empirischen Varianzen $s_x^2 = 94.09$ bzw. $s_y^2 = 1\,263\,376$ und
 - die empirische Kovarianz $s_{xy} = 3925$

berechnet worden.

- Unterstellen Sie normalverteilte Beobachtungen und prüfen Sie, ob die Gesprächsdauer signifikant ($\alpha = 0.05$) vom Luftdruck abhängt.
- Was bedeutet der empirische Korrelationskoeffizient in der vorliegenden Situation?

- Hypothesenpaar: ³⁹ _____

1. ⁴⁰ _____

2. ⁴¹ _____

3. ⁴² _____

4. ⁴³ _____

Es lässt sich somit ⁴⁴ _____, dass die Gesprächsdauer signifikant vom Luftdruck abhängt.

45

Bedeutung von r :**Literatur**

- [BV12] Udo Bankhofer und Jürgen Vogel. *Übungsbuch Datenanalyse und Statistik*. Springer Gabler, 2012.