9 Signifikanztests Teil III

Zoltán Zomotor

Versionsstand: 12. April 2016, 11:29

Die nummerierten Felder bitte während der Vorlesung ausfüllen.

Bitte hier notieren, was beim Bearbeiten unklar geblieben ist:

Inhaltsverzeichnis

1 Kontingenztest[1]

- Gegeben: Zwei verbundene einfache Stichproben

1

- Hypothesenpaar:

H_0: Die Merkmale X und Y sind 2

[1]Bedeutung 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	$\cdots$	B_ℓ	
X	A_1	h_{11}	h_{12}	$\cdots$	$h_{1\ell}$	h_{A_1}
	A_2	h_{21}	h_{22}	$\cdots$	$h_{2\ell}$	h_{A_2}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	A_k	h_{k1}	h_{k2}	$\cdots$	$h_{k\ell}$	h_{A_k}
		h_{B_1}	h_{B_1}	$\cdots$	h_{B_ℓ}	[8]

Dabei bezeichnen [9] die Anzahl der beobachteten Paare [10] .

2.3 Zu jeder Kombination aus $i = 1, \ldots k$ und $j = 1, \ldots, \ell$ die Größe [11] berechnen.

2.4 Den Testfunktionswert v folgendermaßen berechnen: [12]

Schritt 3: Mit dem Fraktilswert $x_{1-\alpha}$ der [13] -Verteilung den

Verwerfungsbereich [14] festlegen.

Schritt 4: H_0 genau dann ablehnen, wenn [15] gilt.

- Bemerkungen:
  (i) Test ist nur anwendbar, wenn

  [16]

  Falls nicht erfüllt ⇒ [17]

  (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

    [18]

  (iv) Falls Randwahrscheinlichkeiten $p_i = P(X \in A_i), q_i = P(Y \in B_j)$ bekannt:
  - Schritt 2.2: Randhäufigkeiten entfallen
  - Schritt 2.3: Ersetze [19]
  - Schritt 3: Ersetze [20]

- 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

Fraktilswerte $x_{1-\alpha/2}$ 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 ermitlelt (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.

  a) Unterstellen Sie normalverteilte Beobachtungen und prüfen Sie, ob die Gesprächsdauer signifikant ($\alpha = 0.05$) vom Luftdruck abhängt.

  b) 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.