

14.3.) Chi-Quadrat - Unabhängigkeitstest

05.06.19

nominal, ordinal, mehrfach - Geuplest
t-Test

Dr. S. 13.3.

H_0 : X, Y stochastisch unabhängig

H_1 : X, Y sind nicht stochastisch unabhängig;
d.h. X, Y sind abhängig voneinander

Ablehnung von $H_0 \Leftrightarrow p\text{-Wert} \leq 0,05$

$$\chi^2 = X^2 \rightarrow \text{Chi-Quadrat}$$

χ^2 - Quadrat

Der p-Wert wird mit der Chi-Quadrat verteilung berechnet.

Anmerkung:

Binomialverteilung $B(n, p)$

Normalverteilung $N(\mu, \sigma)$

Chi-Quadrat hängt ab von der Anzahl der Freiheitsgrade
= degree of freedom = df.

Was ist df?

Beispiele:

e1

			Σ
	5	15	20
	5	25	30
Σ	10	40	50 = n

df = 1

↳ ein Wert kann frei gewählt werden

b)

	5	10	5	20	df=2
	5	5	20	30	
Σ	10	15	25	n=50	

df = (Anzahl Zeilen minus eins) mal
(Anzahl Spalten minus eins)

Bsp:

X = Ausbildungsdauer (in Jahren)

Y = Jobzufriedenheit

X	Y					Σ
	sehr unzf.	unzf.	teils/teils	zuf.	sehr zuf.	
≤ 12	213	282	319	316	295	1.425
	$\cdot x^1$ 275,6	$\cdot x^2$ 296,7	311,2	289,0	252,5	
13-15	348	326	315	315	280	1.634
	316,0	340,2	356,9	331,4	289,5	
≥ 16	406	433	408	383	311	1.941
	375,4	404,1	423,9	383,6	343,9	
Σ	967	1.041	1.092	1.014	886	5.000

H₀: Ausbildungsdauer und Jobzufriedenheit sind stochastisch unabhängig

H₁: ... hängen voneinander ab

Welche Häufigkeiten wären unter der Nullhypothese in jeder Zelle zu erwarten?

= erwartete Häufigkeiten

z.B. 213 = beobachtete Häufigkeiten

05.06.19

$$(P(A \cap B) = P(A) \cdot P(B)) \quad *1 \quad \frac{1425 \cdot 967}{5.000} = 275,6$$

$$*2 \quad \frac{1425 \cdot 1041}{5.000} = 296,7$$

Idee: Ist der Unterschied zwischen den beobachteten und den erwarteten Häufigkeiten „zu groß“, dann wird H_0 abgelehnt.

empirischer Wert der Teststatistik =

$$\chi_{\text{emp.}}^2 = \frac{(213 - 275,6)^2}{275,6}$$

$$+ \frac{(282 - 296,7)^2}{296,7} + \dots + \frac{(311 - 343,9)^2}{343,9}$$

$$= 38,54962$$

$$df = (3-1) \cdot (5-1) = 2 \cdot 4 = 8$$

Ablehnung von $H_0 \Leftrightarrow$

$\chi_{\text{emp.}}^2$ > oberer 5%-Punkt der χ^2 -Quadratverteilung mit $df=8$

$$\chi_{\text{emp.}}^2 = 38,54962 > 15,507$$

d.h. Ablehnung von H_0 , d.h. die Ausbildungssteuer und die Jobzufriedenheit hängen voneinander ab.

Kurzer Lösungsweg:

$$\chi^2_{\text{emp.}} \geq \frac{(213 - 275,6)^2}{275,6} + \frac{(295 - 252,5)^2}{252,5}$$

$$\geq 14 + 7 = 21 \geq 15,507$$

d.h. Ablehnung von H_0 .

Damit der p-Wert mit der χ^2 -Verteilung berechnet werden darf, muss die folgende Faustregel erfüllt sein:

1.) Höchstens 20% aller Zellen dürfen eine erwartete Häufigkeit kleiner als fünf haben.

Hier: keine, also $0 \leq 20\%$ erfüllt

2.) Die minimale erwartete Häufigkeit muss mind. eins betragen

Hier: min. erwartete Häufigkeit = 252,5 > 1
erfüllt.

Bsp:

Sind Stress-Level und tägliche Pendelzeit

stoch. - unabhängig?

χ^2 -unabh.-test zum Niveau 0,05

Stichprobe:

05.06.19

Pendelzeit	Stress-Level			Σ
	hoch	mittel	niedrig	
< 15 min	10 12,8	6 6,3	19 15,9	35
15-45 min	15 18,4	9 9,5	29 24,1	53
> 45 min	20 12,8	7 6,3	8 15,9	35
Σ	45	22	56	123

Feuertregel:

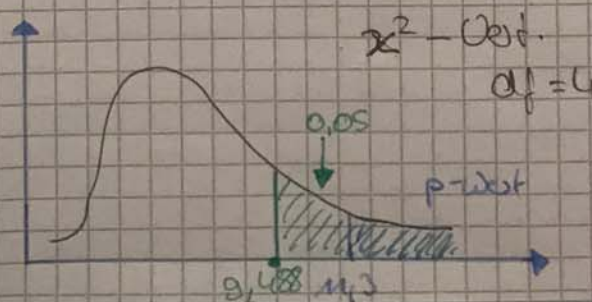
1) keine Zelle hat eine erw. H. kleiner 5,
also nicht wären 20%

2) min. erw. H. = 6,3 ≥ 1 erfüllt

$$df = 4$$

$$\chi^2_{\text{emp.}} = \frac{(10 - 12,8)^2}{12,8} + \dots + \frac{(8 - 15,9)^2}{15,9} = 11,30464 \geq 9,488$$

d.h. Ablehnung von H_0 , d.h. Stress-Level hängt ab von der täglichen Pendelzeit, d.h. Stress-Level und Pendelzeit hängen voneinander ab.



Eigentlich: Ablehnung von H_0 (\Leftrightarrow) p-Wert $\leq 0,05$
Dies ist aber dasselbe:
 $\chi^2_{\text{emp.}} \geq 9,488$

Am 05.06.2019 um 20:34 schrieb Prof. Dr. Jutta Arrenberg <jutta.arrenberg@th-koblenz.de>:

Liebe Frau Rose,

leider ist die Mitschrift nicht vollständig bei mir angekommen. Könnten Sie mir bitte Ihre Mitschrift erneut zuschicken?

Vielen Dank und liebe Grüße,

Jutta Arrenberg
Professorin für Mathematik
Schmalenbach Institut