

Chi – Quadrat – Test mit Software

– Lösungshinweis –

Sie sind der Personalchef einer Firma und suchen nach neuen Mitarbeitenden. Da momentan nur Teilzeitstellen frei sind, fragen Sie sich, ob Sie in der Anzeige eine Altersspanne festlegen sollten oder ob Sie es offenlassen sollten.



Sind die beiden Eigenschaften der Erwerbstätigkeit (Voll- und Teilzeit) und Alter der Beschäftigten in Bayern voneinander abhängig?

Hinweis: Im Rahmen des Vertiefungskurses gibt es zwei unterschiedliche Möglichkeiten:

- *entweder es wird nur die obere, offene Fragestellung den Lernenden gegeben und die Teilaufgaben unten dienen als Hilfestellungen, die bei Bedarf gegeben werden können*
- *oder es erfolgt eine schrittweise, vorgegebene Bearbeitung anhand der unteren Teilaufgaben*

Aufgabe 1: Vermuten Sie einen Zusammenhang zwischen dem Alter und dem Grad der Erwerbstätigkeit in Bayern. Begründen Sie ohne weitere Recherchen oder Daten hinzuziehen.

Lösungshinweise:

Eigene Ideen.

Um eine Entscheidung treffen zu können, möchten Sie Daten benutzen, die diese beiden Eigenschaften in den letzten Jahren erhoben haben. Das statistische Bundesamt erhebt verschiedene Daten in der deutschen Bevölkerung, z. B. hinsichtlich der Wirtschaft oder Gesellschaft. Diese Daten findet man in vielen Datensätzen auf Destatis.

Aufgabe 2: Suchen Sie nach einem geeigneten Datensatz bei Destatis:

https://www.destatis.de/DE/Home/_inhalt.html



Lösungshinweise:

Man könnte als Beispiel im Rahmen dieser Aufgabe den folgenden Datensatz betrachten, der Daten zur Erwerbstätigkeit (Voll- und Teilzeit) in Deutschland in Abhängigkeit des Bundeslandes, Geschlechts und Alters beinhaltet:

<https://www.genesis.destatis.de/datenbank/online/statistic/12211/table/12211-1011/search/s/MTlyMTEgdW5kIGVyd2VyYnN0JUMzJUE0dGlnZQ==>

Zusatzinformation: Die Datensätze, die man auf Destatis findet, beinhalten die beobachteten Werte, die man direkt in die Kontingenztafeln einsetzen kann. Weiterhin ist es auch möglich, diese vorab in die Rohdaten zu überführen und die Lernenden anschließend mit den Rohdaten arbeiten zu lassen. In diesem Fall müssen die beobachteten Häufigkeiten zunächst selbst bestimmt werden.

Aufgabe 3: Erstellen Sie die Kontingenztafel für die beobachteten Häufigkeiten der Erwerbstätigkeit in Abhängigkeit des Alters in Bayern. Werfen Sie anschließend einen Blick auf die Kontingenztafel (ohne weitere Berechnungen durchzuführen). Inwieweit stützt diese Ihre Hypothesen?

Lösungshinweise:

Die Kontingenztafel für die beobachteten Häufigkeiten der Erwerbstätigkeit in Abhängigkeit des Alters sieht folgendermaßen aus:

insgesamt, beobachtet	Vollzeit	Teilzeit	
15 bis unter 25 Jahre	547	181	728
25 bis unter 35 Jahre	1206	291	1497
35 bis unter 45 Jahre	986	465	1451
45 bis unter 55 Jahre	945	459	1404
55 bis unter 65 Jahre	872	484	1356
65 bis unter 75 Jahre	35	115	150
	4591	1995	6586

Aufgabe 4: Führen Sie eine Chi-Quadrat-Test mit einer Software Ihrer Wahl durch. Inwieweit stützt er Ihre Hypothesen?

Lösungshinweise:

Die Nullhypothese für den Chi-Quadrat Test lautet:

Die beiden Eigenschaften der Erwerbstätigkeit (Voll- und Teilzeit) und Alter der Beschäftigten in Bayern sind voneinander unabhängig.

Berechnung der erwarteten Häufigkeiten:

insgesamt, erwartet	Vollzeit	Teilzeit	
15 bis unter 25 Jahre	507,47768	220,52232	728
25 bis unter 35 Jahre	1043,53583	453,464166	1497
35 bis unter 45 Jahre	1011,46994	439,530064	1451
45 bis unter 55 Jahre	978,706954	425,293046	1404
55 bis unter 65 Jahre	945,246887	410,753113	1356
65 bis unter 75 Jahre	104,562709	45,4372912	150
	4591	1995	6586

Berechnung der Freiheitsgrade: $df = (2 - 1) \cdot (6 - 1) = 5$

Durchführung des Chi-Quadrat-Tests mit jamovi:

Zunächst kopiert man den Rohdatensatz in die Tabellenkalkulation unter „Daten“:

A	Alter	Erwerbstä...
1	15-25	Vollzeit
2	15-25	Vollzeit
3	15-25	Vollzeit
4	15-25	Vollzeit
5	15-25	Vollzeit
6	15-25	Vollzeit
7	15-25	Vollzeit
8	15-25	Vollzeit
9	15-25	Vollzeit
10	15-25	Vollzeit
11	15-25	Vollzeit
12	15-25	Vollzeit
13	15-25	Vollzeit
14	15-25	Vollzeit
15	15-25	Vollzeit
16	15-25	Vollzeit
17	15-25	Vollzeit
18	15-25	Vollzeit
19	15-25	Vollzeit
20	15-25	Vollzeit
21	15-25	Vollzeit
22	15-25	Vollzeit
23	15-25	Vollzeit
24	15-25	Vollzeit
25	15-25	Vollzeit
26	15-25	Vollzeit
27	15-25	Vollzeit
28	15-25	Vollzeit
29	15-25	Vollzeit
30	15-25	Vollzeit
31	15-25	Vollzeit
32	15-25	Vollzeit
33	15-25	Vollzeit
34	15-25	Vollzeit
35	15-25	Vollzeit
36	15-25	Vollzeit
37	15-25	Vollzeit
38	15-25	Vollzeit
39	15-25	Vollzeit
40	15-25	Vollzeit
41	15-25	Vollzeit
42	15-25	Vollzeit

Danach wählt man unter „Analysen“ die Option „Häufigkeiten“ und dort den „ χ^2 -Unabhängigkeitstest“ aus. Dann wählt man die Spalte mit dem Alter für die Zeilen und die Spalte mit der Erwerbstätigkeit für die Spalten aus. Wählt man bei „Zellen“ zusätzlich „Erwartete Anzahl“, so erhält man folgendes:

Kreuztabellen

Zeilen: A, B, C
Spalten:
Anzahl (optional):
Schichten:

Anzahl

☒ Beobachtet (Anzahl)
☒ Erwartete Anzahl

Prozentanteile

☐ Zeile
☐ Spalte
☐ Insgesamt

Ergebnisse

Proportionstest (N Ergebnisse)

Anteile - ...

Stufe	Anzahl	Anteil
15-25	547	181
25-35	1206	291
35-45	986	465
45-55	945	459
55-65	872	484
65-75	35	115
Insgesamt	4591	1995

Kreuztabellen

B		C		Insgesamt
		Vollzeit	Teilzeit	
15-25	Beobachtet	547	181	728
	Erwartet	507	220.5	728
25-35	Beobachtet	1206	291	1497
	Erwartet	1044	453.5	1497
35-45	Beobachtet	986	465	1451
	Erwartet	1011	439.5	1451
45-55	Beobachtet	945	459	1404
	Erwartet	979	425.3	1404
55-65	Beobachtet	872	484	1356
	Erwartet	945	410.8	1356
65-75	Beobachtet	35	115	150
	Erwartet	105	45.4	150
Insgesamt	Beobachtet	4591	1995	6586
	Erwartet	4591	1995.0	6586

Chi-Quadrat-Tests

	Wert	df	p
χ^2	271	5	< .001
N	6586		

Es ergibt sich ein empirischer Chi-Quadrat-Wert von $\chi^2_{emp} = 271$.

Für ein Signifikanzniveau von $\alpha = 5\%$ und einen Freiheitsgrad von $df = 5$ folgt ein kritischer Chi-Quadrat-Wert von $\chi^2_{krit} = 11,07$.

Also ist $\chi^2_{emp} = 271 > 11,07 = \chi^2_{krit}$ und somit der Unterschied signifikant. Es kann also von einem Zusammenhang zwischen dem Alter und der Erwerbstätigkeit in Bayern ausgegangen werden.



Warum sind die Testwerte signifikant? Begründen Sie dies mit Hilfe des Datensatzes, indem Sie einen rechnerischen Zusammenhang zwischen erwarteten und beobachteten Werten herstellen.

Lösungshinweise:

Die Testwerte sind nicht signifikant, wenn die beobachteten Werte nur wenig von den erwarteten Werten abweichen, das Verhältnis der erwarteten Werte also ähnlich zu dem der beobachteten Werte ist.

Aufgrund des folgenden Zusammenhangs können das Verhältnis der beobachteten Werte mit dem Verhältnis der Spaltensummen der beobachteten Werte verglichen werden:

$$\frac{w_e}{y_e} = \frac{\frac{H_b(M1_1) \cdot H_b(M2_1)}{n}}{\frac{H_b(M1_1) \cdot H_b(M2_2)}{n}} = \frac{H_b(M1_1) \cdot H_b(M2_1)}{H_b(M1_1) \cdot H_b(M2_2)} = \frac{H_b(M2_1)}{H_b(M2_2)}$$

Liegt ein ähnliches Verhältnis vor, gibt es keinen signifikanten Unterschied.
Man geht insgesamt davon aus, dass sich die Verhältnisse innerhalb der einzelnen Gruppen bei Unabhängigkeit analog zu den Verhältnissen der Gesamtgruppe verhalten.

Da in diesem Datensatz große Unterschiede in den Verhältnissen vorliegen, ist der Unterschied signifikant und man kann von einer Abhängigkeit ausgehen.

DIDAKTIK DER MATHEMATIK

Universität Würzburg

DIDAKTIK DER MATHEMATIK

Universität Würzburg