

zugehörige Seiten in Fahrmeir et al. (2007): Kap. 11.4- 11.5

Aufgabe 88

Bei einer Erhebung hinsichtlich der Parteipräferenz bei Männern und Frauen ergaben sich die folgenden Ergebnisse:

	CDU/CSU	SPD	FDP	Grüne	Rest	Gesamt
Männer	144	153	17	26	95	435
Frauen	200	145	30	50	71	496
Gesamt	344	298	47	76	166	931

Untersuchen Sie zum Niveau $\alpha = 0.05$, ob die voneinander abweichenden Häufigkeiten für Männer und Frauen rein zufällige Schwankungen darstellen, oder ob zwischen Geschlecht und Parteipräferenz ein Zusammenhang besteht.

Aufgabe 89

Um zu überprüfen, ob ein Arzneimittel gegen Grippe wirkt, wurden von 326 Kranken 163 mit dem Arzneimittel behandelt (Gruppe A), die andere Hälfte (Gruppe B) erhielt Tabletten ohne Wirkstoffe. Bei Gruppe A zeigte sich bei 105 Patienten eine baldige Besserung, bei Gruppe B bei 88 Patienten. Man prüfe die Hypothese, dass Tabletten mit und ohne Wirkstoffe zu gleichen Resultaten führen ($\alpha = 0.05$).

Aufgabe 90

Eine Vertriebsgesellschaft vermutet, dass es hinsichtlich der Nachfrage nach bestimmten Wochenzeitschriften regionale Unterschiede gibt. Um diese Vermutung statistisch zu überprüfen, zieht sie aus der Menge der Käufer von Wochenzeitschriften zwei voneinander unabhängige einfache Zufallsstichproben, und zwar eine ($n_1 = 510$) aus der norddeutschen und eine ($n_2 = 450$) aus der süddeutschen Region.

Die Käufer in beiden Stichproben werden gefragt, welche Wochenzeitschrift sie bevorzugen (keine Mehrfachnennungen!). Das Ergebnis dieser Befragung zeigt folgende Tabelle:

Wochenzeitschrift	Norddeutschland	Süddeutschland
A	195	145
B	140	200
C	175	105
Summe	510	450

- Formulieren Sie für die Vermutung der Betriebsgesellschaft „Es gibt regionale Unterschiede“ Null- und Alternativhypothese.
- Begründen Sie, welcher Test - unter Berücksichtigung der erhobenen Daten - geeignet ist, um die Vermutung der Vertriebsgesellschaft zu überprüfen.
- Führen Sie den Test auf einem Signifikanzniveau von $\alpha = 0.01$ durch und interpretieren Sie Ihr Ergebnis.
- Kann man aus dem Ergebnis von Teilaufgabe (c) schließen, dass der Verkauf der Wochenzeitschriften unabhängig von der Region ist, in der sie verkauft werden? Begründen Sie Ihre Entscheidung.

12 Regression

zugehörige Seiten in Fahrmeir et al. (2007): Kap. 12.1

Aufgabe 91

Leiten Sie für das einfache Regressionsmodell

$$y_i = \beta_0 + x_{i1}\beta_1 + \epsilon_i, \quad i = 1, \dots, n$$

die Normalgleichungen und KQ-Schätzformeln in Summen- und Matrixschreibweise her.

Aufgabe 92

Zwischen der Wachstumsrate der Bruttonominallöhne y_i (in %) und der Wachstumsrate der durchschnittlichen nominalen Arbeitsproduktivität x_{i1} (in %) wird ein linearer Zusammenhang vermutet. Um diesen Zusammenhang zu schätzen, wird das folgende lineare Regressionsmodell aufgestellt:

$$y_i = \beta_0 + x_{i1}\beta_1 + \epsilon_i, \quad i = 1, \dots, n. \quad (1)$$

In fünf aufeinanderfolgenden Jahren wurden die folgenden Wachstumsraten beobachtet:

i	1	2	3	4	5
y_i	2.5	3.0	2.0	2.5	3.0
x_{i1}	2.0	2.0	1.5	2.5	3.0

- (a) Welche Annahmen treffen Sie über Erwartungswert, Varianz und Kovarianz des Störterms ϵ_i in Gleichung (1)? Was müssen Sie für den Regressor x_{i1} annehmen?
- (b) Stellen Sie mit Hilfe des Modells in (1) die verfügbare Stichprobe in Matrixschreibweise $y = X\beta + \epsilon$ dar.
- (c) Berechnen Sie $(X'X)^{-1}$, $X'y$, $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\epsilon} = y - X\hat{\beta}$, $\hat{\epsilon}'\hat{\epsilon}$, $\hat{\sigma}^2 = \frac{1}{3}\hat{\epsilon}'\hat{\epsilon}$.
- (d) Warum ist $\hat{\sigma}^2$ ein geeigneter Schätzer für die Varianz des Störterms ϵ_i in Gleichung (1)?

Aufgabe 93* (16 Punkte)

Es wird davon ausgegangen, dass Alkohol selbst nicht krebserregend ist, aber die durch ihn bewirkte Schädigung der Zellen insbesondere in Mund-, Rachen- und Kehlkopfbereich gibt Anlass zur Vermutung, dass ein erhöhter Alkoholkonsum zu einer wachsenden Anzahl von Krebserkrankungen in diesem Bereich führt. Die folgende Tabelle zeigt das Ergebnis eines internationalen Vergleichs, bei dem der Alkoholkonsum X (in Litern) und die Anzahl Y der Todesfälle pro 100 000 Einwohner erhoben wurde. (Quelle: E.R.Koch u.a.: *Die Gesundheit der Nation*, Köln 1986)

x_i	5	6	8	9	10	11	12	13	13	14	17
y_i	3.5	3	7	4	7	10	8.5	6	13	11.5	26.5

- (a) Stellen Sie die Meßwerte mit Hilfe eines Scatterplots graphisch dar.
- (b) Stellen Sie ein geeignetes Modell zur Analyse des Zusammenhanges zwischen X und Y auf.
- (c) Wie lautet das Optimalitätskriterium zur Berechnung der optimalen Ausgleichsgerade?
- (d) Bestimmen Sie die optimale Ausgleichsgerade und zeichnen Sie sie in den Scatterplot ein.
- (e) Zeichnen und interpretieren Sie den zugehörigen Residuenplot.
- (f) Berechnen und interpretieren Sie das Bestimmtheitsmaß.
- (g) Geben Sie ein 95 Prozent-Konfidenzintervall für einen Alkoholkonsum von 7 Litern an.
- (h) Berechnen Sie die optimale Ausgleichsgerade für den um die Beobachtung (17, 26.5) reduzierten Datensatz. Vergleichen Sie die Ergebnisse.

(Hinweis: Runden Sie Ihre Ergebnisse auf die zweite Nachkommastelle!)