

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

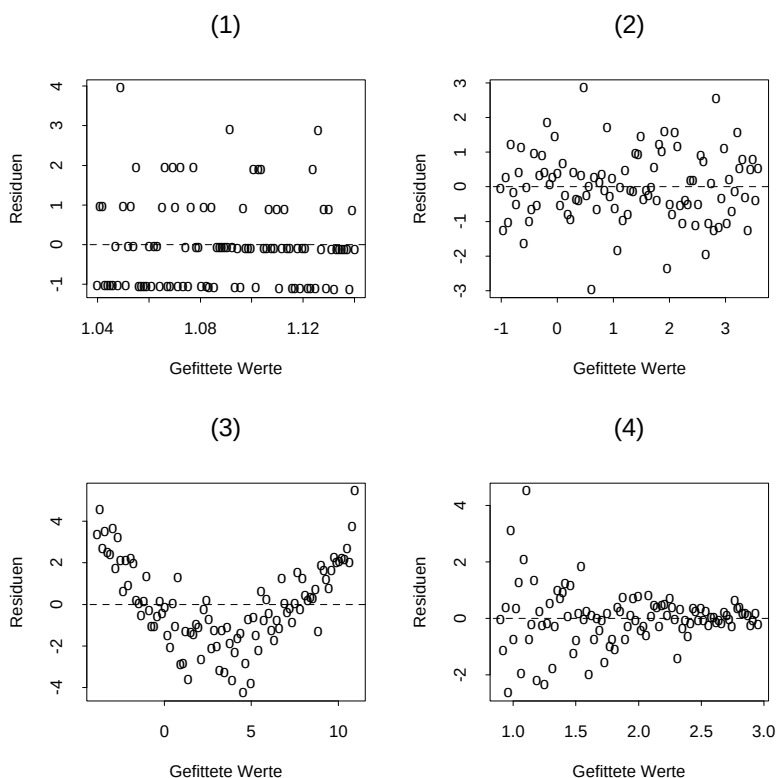
Aufgabe 94

Zur Überprüfung der Güte eines Schuleignungstests X zur Vorhersage der Schulreife ermittelt ein Schulpsychologe an $n = 500$ Vorschulkindern die Werte $\bar{x} = 40$ und $\tilde{s}_X = 5$. Nach Ablauf des ersten Schuljahres werden mit einem Schulleistungstest Y die tatsächlichen Leistungen der Kinder gemessen und man erhielt die folgenden Kennwerte $\bar{y} = 30$, $\tilde{s}_Y = 4$. Die Kovarianz in der Stichprobe zwischen Schuleignungstest und dem Schulleistungstest betrug $\tilde{s}_{XY} = 10$.

- Berechnen Sie einen Korrelationskoeffizienten zwischen Schuleignungs- und Schulleistungstest.
- Prüfen Sie die Hypothese, der Korrelationskoeffizient sei größer als 0.6 ($\alpha = 0.05$).
- Ermitteln Sie die Stichproben-Regressionsgerade zur Vorhersage der tatsächlichen Leistung nach dem ersten Schuljahr aufgrund des Schuleignungstests.
- Überprüfen Sie, ob der Schuleignungstest einen signifikanten Beitrag zur Vorhersage der tatsächlichen Leistung nach dem ersten Schuljahr besitzt ($\alpha = 0.05$, $\hat{\sigma} = 1.75$).
- Ein Kind erreicht im Eignungstest einen Wert von $x_0 = 48$. Mit welcher schulischen Leistung ist bei ihm zu rechnen?
- Man ermittle ein 95%-Konfidenzintervall zur Vorhersage der schulischen Leistung des Kindes.

Aufgabe 95

Ein exploratives Mittel zur Überprüfung der Modellannahmen einer linearen Regression sind die sogenannten Residualplots in denen die geschätzten Werte $\hat{y}_i$ und die geschätzten Residuen $\hat{\epsilon}_i$ gegeneinander abgetragen werden. Nachfolgend sind für verschiedene Datensätze Residualplots abgebildet. Geben Sie für jede Grafik an, ob und wenn ja welche Modellannahme verletzt sein könnte.



Aufgabe 96* (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!)

Aufgabe 97

Zur Schätzung des Werbe-Effekts in einem Getränke-Unternehmen wird das folgende lineare Modell aufgestellt:

$$y_t = \beta_0 + x_t \beta_1 + \epsilon_t,$$

mit y_t : Umsatz aus Getränkeverkauf (in Millionen Euro/Jahr) und x_t : Werbekosten (in Millionen Euro/Jahr). Aus einer statistischen Analyse für den Zeitraum 1983-2001 sind folgende Zwischenergebnisse bekannt:

$$X'X = \begin{bmatrix} 19 & 13.8 \\ 13.8 & 13.56 \end{bmatrix}, \quad \sum_{t=1983}^{2001} y_t = 30, \quad \sum_{t=1983}^{2001} x_t y_t = 26.4, \quad \sum_{t=1983}^{2001} y_t^2 = 61.$$

Das Unternehmen stellt Ihnen die neuesten Daten für das Jahr 2002 zur Verfügung, die Sie bei der nachfolgenden Analyse berücksichtigen sollen:

$$x_{2002} = 1.2, \quad y_{2002} = 3.$$

- (a) (i) Welche Modellannahmen treffen Sie über den Störterm und die Matrix X bei der Anwendung der KQ-Methode?
- (ii) Welche Annahme muss zusätzlich zu den genannten Annahmen bezüglich der Störgröße ϵ getroffen werden, damit Hypothesentests über die Parameter durchgeführt werden können?
- (b) Benutzen Sie die Beobachtungen bis einschließlich 2002, um den Koeffizientenvektor $\beta' = (\beta_0, \beta_1)$ mit der KQ-Methode zu schätzen.
- (c) Berechnen Sie die geschätzte Fehlervarianz $\hat{\sigma}^2$.
- (d) Berechnen Sie den Standardfehler für $\hat{\beta}_1$.
- (e) Welchen Aussagewert hat das Bestimmtheitsmaß? Würden Sie das Bestimmtheitsmaß oder das korrigierte Bestimmtheitsmaß bei einem Vergleich von alternativen Modellen benutzen? Begründen Sie Ihre Entscheidung.
- (f) Für das Jahr 2003 sind Werbekosten in einer Höhe von insgesamt 1.25 Millionen Euro vorgesehen. Mit welchem Umsatz kann das Unternehmen rechnen? Geben Sie die Varianz des Prognosewerts an.