

Statistik-Klausur vom 4. Februar 2004

Die Bearbeitungszeit dieser Klausur beträgt 90 Minuten.

Aufgabe 1

Von je 1 000 Rentnerinnen bzw. Rentnern bekommen in Deutschland eine monatliche Rente:

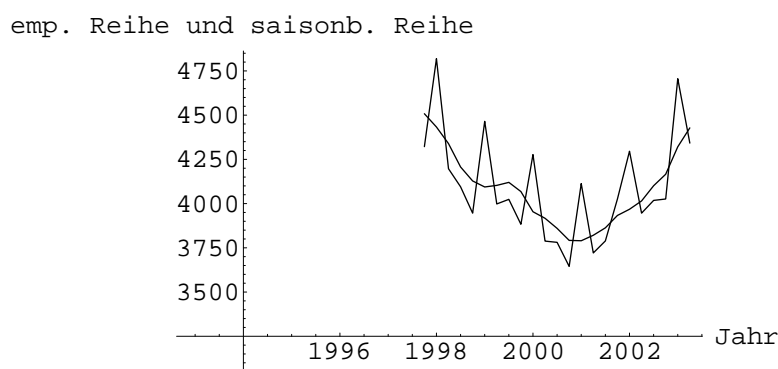
von	Frauen	Männer
unter 150 Euro	112	45
150 bis unter 300	187	44
300 bis unter 450	131	42
450 bis unter 600	162	52
600 bis unter 750	196	78
750 bis unter 900	112	113
900 bis unter 1 200	77	299
1 200 bis unter 1 500	20	226
1 500 Euro und mehr	3	101

Stand: Ende 2002
Quelle: VDR

- Vergleichen Sie das Rentenniveau von Rentnerinnen und Rentnern anhand einer statistischen Maßzahl. Welches Geschlecht bezieht die höhere Rente?
- Vergleichen Sie die Rentenunterschiede bei den Männern mit den Rentenunterschieden bei den Frauen. Bei welchem Geschlecht sind die Unterschiede stärker?

Aufgabe 2

Arbeitslose (in 1 000) in Deutschland von 1997 bis 2003 nach Quartalen



Quelle: Deutsche Bundesbank: Monatsbericht;
Saisonbereinigte Wirtschaftszahlen

- Welche Komponenten enthält die empirische Zeitreihe der Arbeitslosen in Deutschland? Belegen Sie die Existenz jeder einzelnen von Ihnen genannten Komponente durch Angabe einer charakteristischen Stelle der empirischen Zeitreihe,

aus der man diese Komponente ersehen kann! Erläutern Sie kurz die jeweiligen Stellen!

- b) Die empirische Zahl der Arbeitslosen ist vom I. zum II. Quartal im Jahr 2003 gesunken, und zwar um 364 000. Dagegen ist die saisonbereinigte Zahl der Arbeitslosen im nämlichen Zeitraum um 107 000 gestiegen. Erklären Sie, was diese gegenläufige Entwicklung der empirischen und saisonbereinigten Zahl der Arbeitslosen zu bedeuten hat!
- c) Sie sollen die längerfristige Entwicklungstendenz der Arbeitslosenzahl in Deutschland mittels der Methode der kleinsten Quadrate bestimmen.

Dabei stehen Ihnen vier verschiedene Methoden-Varianten zur Verfügung:

1. Linearer Trend, errechnet aus der Gesamtzahl der empirischen Werte:
Trendwert $\approx +4\,166 - 5,8 \cdot \text{Zeitpunkt}$
(Das IV. Quartal 1997 ist der Zeitpunkt $t = 1$; das II. Quartal 2003 ist der Zeitpunkt $t = 23$)
2. Linearer Trend, errechnet aus der Gesamtzahl der saisonbereinigten Werte:
Trendwert $\approx +4\,191 - 8,9 \cdot \text{Zeitpunkt}$
(Das IV. Quartal 1997 ist der Zeitpunkt $t = 1$; das II. Quartal 2003 ist der Zeitpunkt $t = 23$)
3. Linearer Trend, errechnet aus der Zahl der empirischen Werte für den Teilabschnitt vom I. Quartal 2001 bis zum II. Quartal 2003:
Trendwert $\approx +3\,773 + 59,2 \cdot \text{Zeitpunkt}$
(Das I. Quartal 2001 ist der Zeitpunkt $t = 1$; das II. Quartal 2003 ist der Zeitpunkt $t = 10$)
4. Linearer Trend, errechnet aus der Zahl der saisonbereinigten Werte für den Teilabschnitt vom I. Quartal 2001 bis zum II. Quartal 2003:
Trendwert $\approx +3\,664 + 68,5 \cdot \text{Zeitpunkt}$
(Das I. Quartal 2001 ist der Zeitpunkt $t = 1$; das II. Quartal 2003 ist der Zeitpunkt $t = 10$)

Sie sollen analysieren, welches der vier Verfahren einerseits aus der Sicht der statistischen Methodenlehre zulässig und zum anderen vom Verlauf der Zeitreihe der Arbeitslosen her auch sinnvoll ist. Für welches der vier Verfahren entscheiden Sie sich? Begründen Sie bitte Ihre Antwort!

Aufgabe 3

In einem Unternehmen wurde eine Studie über die Gesundheit der Mitarbeiter und Mitarbeiterinnen durchgeführt. Dabei waren auch die Fragen

- „Rauchen Sie?“
- „Trinken Sie Kaffee?“

zu beantworten. Beide Fragen konnten nur mit „Ja“ oder „Nein“ beantwortet werden.

Die Auswertung ergab, dass

- 35% der Befragten sowohl Kaffee trinken als auch rauchen.
 - 48% der Befragten rauchen.
 - 64% der Befragten Kaffee trinken.
- a) Mit welcher Wahrscheinlichkeit beantwortet eine zufällig ausgewählte Person mindestens eine der beiden Fragen mit „Ja“?
 - b) Mit welcher Wahrscheinlichkeit beantwortet eine zufällig ausgewählte Person mindestens eine der beiden Fragen mit „Nein“?
 - c) Mit welcher Wahrscheinlichkeit beantwortet eine zufällig ausgewählte Person, die raucht, die Frage nach dem Kaffee trinken mit „Ja“?
 - d) Sind die Ereignisse „Rauchen“ und „Kaffee trinken“ stochastisch unabhängig? (Begründung!)
 - e) Mit welcher Wahrscheinlichkeit beantwortet eine zufällig ausgewählte Person, die das Rauchen verneint, die Frage nach dem Kaffee trinken mit „Ja“?

Aufgabe 4

Bei der Deutschen Post AG werden die Sendungen automatisiert sortiert. Eine Kodiermaschine liest die Anschriften und ordnet die Sendungen den jeweiligen Postleitzahlen zu. Mit einer Wahrscheinlichkeit von 2% vertauscht die Kodiermaschine den Empfänger mit dem Absender, so dass die Sendung ihren Empfänger nicht erreicht, sondern zurück zum Absender geschickt wird. Ferner geschieht das Vertauschen von Empfänger und Absender der Sendungen stochastisch unabhängig voneinander.

- a) Sie haben zehn Briefe versehen mit Empfänger und Absender verschickt. Wie groß ist die Wahrscheinlichkeit, dass bei höchstens zwei Ihrer Briefe von der Kodiermaschine Empfänger und Absender vertauscht werden?
- b) Ein Unternehmen hat 1 200 Sendungen versehen mit Empfänger und Absender verschickt. Wie groß ist annähernd die Wahrscheinlichkeit, dass bei höchstens vierzehn dieser Briefe von der Kodiermaschine Empfänger und Absender vertauscht werden?

Aufgabe 5

Ein Marktforschungsinstitut möchte angeben, wie hoch im Mittel die monatlichen Ausgaben eines Haushalts für Haarshampoos sind. In einer Umfrage von 100 Haushalten betragen die durchschnittlichen monatlichen Ausgaben für Haarshampoos 15 Euro. Die Standardabweichung in der Stichprobe betrug 6 Euro.

Berechnen und interpretieren Sie ein 0,95-Konfidenzintervall für die im Mittel aufgebrauchten monatlichen Ausgaben eines Haushalts für Haarshampoos.

Lösungen:

Lösung zu Aufgabe 1:

X = Rentenhöhe (in Euro) einer Frau

Y = Rentenhöhe (in Euro) eines Mannes

von	Frauen	Männer	F_{Frauen}	$F_{\text{Männer}}$
unter 150 Euro	112	45	0,112	0,045
150 bis unter 300	187	44	0,299	0,089
300 bis unter 450	131	42	0,430	0,131
450 bis unter 600	162	52	0,592	0,183
600 bis unter 750	196	78	0,788	0,261
750 bis unter 900	112	113	0,900	0,374
900 bis unter 1 200	77	299	0,977	0,673
1 200 bis unter 1 500	20	226	0,997	0,899
1 500 Euro und mehr	3	101	1,000	1,000

$$\text{a) } x_{0,50} \approx 450 + 150 \cdot \frac{0,50 - 0,430}{0,162} = 514,82$$

$$y_{0,50} \approx 900 + 300 \cdot \frac{0,50 - 0,374}{0,299} = 1\,026,42$$

d.h. etwa 50% aller Rentnerinnen bezieht höchstens ca. 515 Euro monatliche Rente, während etwa 50% aller Rentnern mindestens ca. 1026 Euro monatliche Rente beziehen.

Das Rentenniveau ist bei den Männern etwa doppelt so hoch.

$$\text{b) } x_{0,25} \approx 150 + 150 \cdot \frac{0,25 - 0,112}{0,187} = 260,70$$

$$x_{0,75} \approx 600 + 150 \cdot \frac{0,75 - 0,592}{0,196} = 720,92$$

$$\text{Relativer Quartilsabstand: } \frac{x_{0,75} - x_{0,25}}{x_{0,50}} \approx \frac{720,92 - 260,70}{514,82} = 0,8939$$

d.h. die relative Streuung der Rentenhöhe der Frauen beträgt etwa 0,90.

$$y_{0,25} \approx 600 + 150 \cdot \frac{0,25 - 0,183}{0,078} = 728,85$$

$$y_{0,75} \approx 1\,200 + 300 \cdot \frac{0,75 - 0,673}{0,226} = 1\,302,21$$

Relativer Quartilsabstand: $\frac{y_{0,75} - y_{0,25}}{y_{0,50}} \approx \frac{1\,302,21 - 728,85}{1\,026,42} = 0,5586$

d.h. die relative Streuung der Rentenhöhe der Männer beträgt etwa 0,56.

Die Unterschiede der einzelnen Rentenhöhen sind bei den Frauen stärker.

Lösung zu Aufgabe 2:

a) 1. Komponenten der empirischen Zeitreihe:

- Trendkomponente
- Saisonkomponente
- unerklärte Restkomponente

2. Existenz der genannten Komponenten:

- Der Trend ist zunächst fallend und ab 2001 steigend (deutlich sichtbar).
- Im Winter ist bekanntlich die Anzahl der Arbeitslosen am größten. In der empirischen Reihe ist das erkennbar an den Gipfeln (saisonale Schwankungen), die bei jedem beobachteten I. Quartal zu sehen sind.
- Zwischen den insgesamt sechs Gipfeln fallen die Schwankungen unterschiedlich stark aus. Dies ist auf die unerklärte Restkomponente zurückzuführen.

b) Es liegt ein Trend zu höheren Arbeitslosenzahlen vor, was sich in der saisonbereinigten Reihe zeigt. Saisonal bedingt ist die Zahl der Arbeitslosen zwar gesunken.

c) Mit dem 4. Verfahren lässt sich der Trend am besten berechnen. Da der Trend bis 2001 fallend ist und anschließend steigend, sollte nur der letzte Teil der Zeitreihe berücksichtigt werden. Außerdem ist die saisonbereinigte Reihe zur Trendberechnung heranzuziehen, weil sie nicht so stark schwankt wie die ursprüngliche Reihe.

Lösung zu Aufgabe 3:

R = Raucher

K = Kaffee Trinker

$$0,35 = P(R \cap K)$$

$$0,48 = P(R)$$

$$0,64 = P(K)$$

Arbeitstabelle:

	R	\bar{R}	
K	0,35	0,29	0,64
\bar{K}	0,13	0,23	0,36
	0,48	0,52	1

a) $P(R \cup K) = 1 - 0,23 = 0,77$

oder

$$P(R \cup K) = P(R) + P(K) - P(R \cap K) = 0,48 + 0,64 - 0,35 = 0,77$$

d.h. die Wahrscheinlichkeit beträgt 0,77.

- b) $P(\overline{R} \cup \overline{K}) = 1 - 0,35 = 0,65$
 oder
 $P(\overline{R} \cup \overline{K}) = P(\overline{R}) + P(\overline{K}) - P(\overline{R} \cap \overline{K}) = 0,52 + 0,36 - 0,23 = 0,65$
 d.h. die Wahrscheinlichkeit beträgt 0,65.
- c) $P(K | R) = \frac{P(K \cap R)}{P(R)} = \frac{0,35}{0,48} = 0,7292$
 d.h. die Wahrscheinlichkeit beträgt 0,7292.
- d) $P(K) = 0,64 \neq 0,7292 = P(K | R)$
 d.h. K, R sind stochastisch abhängig.
- e) $P(K | \overline{R}) = \frac{P(K \cap \overline{R})}{P(\overline{R})} = \frac{0,29}{0,52} = 0,5577$
 d.h. die Wahrscheinlichkeit beträgt 0,5577.

Lösung zu Aufgabe 4:

X = Anzahl der Sendungen, bei denen von der Kodiermaschine Empfänger und Absender vertauscht werden

$X \sim \mathbf{B}(n; p = 0,02)$

- a) $X \sim \mathbf{B}(n = 10; p = 0,02)$
 $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
 $= \binom{10}{0} \cdot 0,02^0 \cdot 0,98^{10} + \binom{10}{1} \cdot 0,02^1 \cdot 0,98^9 + \binom{10}{2} \cdot 0,02^2 \cdot 0,98^8$
 $= 0,8171 + 0,1667 + 0,0153$
 $= 0,9991$
 d.h. die Wahrscheinlichkeit beträgt 0,9991.

- b) $X \sim \mathbf{B}(n = 1\,200; p = 0,02)$
 $E[X] = np = 1\,200 \cdot 0,02 = 24$
 $Var[X] = np(1 - p) = 24 \cdot 0,98 = 23,52$
 $X \approx_{ZGWS} \mathbf{N}$; da Faustregel $np \geq 10$ und $np(1 - p) \geq 10$ erfüllt ist
 $P(X \leq 14) \approx F_U \left(\frac{14 + 0,5 - 24}{\sqrt{23,52}} \right) = F_U(-1,9589) = 0,025$
 d.h. die Wahrscheinlichkeit beträgt annähernd 0,025.

Lösung zu Aufgabe 5:

μ = im Mittel aufgebrauchte Ausgaben (in €) eines Haushalts pro Monat für Haarshampoos

Stichprobe vom Umfang $n = 100$ mit $\bar{x} = 15$ und $s = 6$ (Faustregel $n \geq 30$ erfüllt)
 0,95-KI für $\mu = \left[15 \pm 1,96 \cdot \frac{6}{\sqrt{100}} \right] = [15 \pm 1,176] = [13,824; 16,176]$

d.h. $[13,8; 16,2]$ ist ein geschätzter Bereich für das Intervall, in dem die mittleren monatlichen Ausgaben für Haarshampoos eines Haushalts mit der Wahrscheinlichkeit 0,95 liegen.