6 KLASSISCHE SCHLIESSENDE STATISTIK

6.25 Punktschätzungen

1. $X_1, \ldots, X_n$ sei eine Stichprobe aus einer Verteilung mit Mittelwert μ und Varianz σ^2; $\overline{X}_n$ sei der Stichprobenmittelwert und S_n^2 die Stichprobenvarianz:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

   Zeigen Sie:

   (a) $\overline{X}_n$ ist ein unverzerrter Schätzer für μ. (Varianz des Schätzers? *Ist der Schätzer konsistent für μ ?)

   (b) S_n^2 ist ein unverzerrter Schätzer für σ^2. (*Hinweis:* Vgl. **B-5-12**.)

2. $X_1, \ldots, X_n$ sei eine Stichprobe von $X \sim U_{(0,\theta)}$ $(\theta > 0)$.

   (a) Zeigen Sie, daß $T_1 = 2\overline{X}_n$ ein unverzerrter Schätzer für θ ist. (Varianz des Schätzers? *Ist der Schätzer konsistent für θ ?)

   (b) Ist $T_2 = \max\{X_i,\, i = 1, \ldots, n\}$ ein unverzerrter Schätzer für θ ?

   (c) Wie lautet die effiziente lineare Schätzfunktion für θ ?

3. Eine Münze wird n Mal geworfen und fällt k $(\leq n)$ Mal auf „Kopf". Bestimmen Sie den plausiblen Schätzwert für die Wahrscheinlichkeit θ, daß die Münze auf Kopf fällt. (Überzeugen Sie sich davon, daß dieser Schätzwert die Plausibilitätsfunktion maximiert.) Ist der plausible Schätzer unverzerrt?

4. Bei 50 Autoblechen wurden die folgenden Anzahlen von Lakierungsfehlern registriert:

Fehlerzahl	0	1	2	3	4
Häufigkeit	20	18	9	2	1

   Wenn man davon ausgeht, daß es sich um eine Stichprobe aus einer P_μ–Verteilung handelt, bestimmen Sie den plausiblen Schätzwert von μ. (Überzeugen Sie sich davon, daß dieser Schätzwert die Plausibilitätsfunktion maximiert.) Ist der plausible Schätzer unverzerrt?

5. Die folgenden Daten sind Zeiten (Betriebsstunden) zwischen aufeinander folgenden Ausfällen eines Systems:

   67 102 47 85 34 9 4 21 229 262 43 152 1 14 79

   Ermitteln Sie unter der Annahme, daß es sich um eine Stichprobe aus einer Ex_τ–Verteilung handelt, den plausiblen Schätzwert von (a) τ und von (b) $\lambda = 1/\tau$. (Überzeugen Sie sich in beiden Fällen davon, daß der Schätzwert die Plausibilitätsfunktion maximiert.) Sind die Schätzer unverzerrt?

6.26 Konfidenzbereiche

6. Bei zehn Batterien findet man die folgenden Kapazitäten [Amperestunden]:

   140 136 150 144 148 152 138 141 143 151

   Wenn man davon ausgeht, daß es sich um eine Stichprobe aus einer $N(\mu, \sigma^2)$–Verteilung handelt, ermitteln Sie (a) einen Schätzwert, (b) ein 95%– und (c) ein 99%–Konfidenzintervall für die mittlere Kapazität μ dieses Batterietyps.

7. Ermitteln Sie auf Basis der Daten des vorigen Beispiels (a) einen Schätzwert für σ^2, (b) ein 95%–Konfidenzintervall für σ^2 und (c) ein 95%–Konfidenzintervall für σ. Gehen Sie dabei wieder von der Annahme normalverteilter Beobachtungen aus.

8. Eine zufällige Stichprobe von 300 Kreditkarteninhabern einer Bank ergab einen mittleren Schuldenstand von 1220 EUR bei einer Stichprobenstreuung von 840 EUR. Ermitteln Sie ein 95%–Konfidenzintervall für den mittleren Schuldenstand aller Karteninhaber dieser Bank. (*Hinweis:* Berufen Sie sich auf den zentralen Grenzverteilungssatz.)

9. $X_1, \ldots, X_n$ sei eine Stichprobe aus einer Alternativverteilung A_p. Die folgende sG ist eine approximative Pivotgröße (*Begründung?):

$$Z_n = \frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)/n}} \dot{\sim} N(0,1)$$

   (a) Entwickeln Sie auf Basis von Z_n ein approximatives $(1-\alpha)$–Konfidenzintervall für den Parameter p.

   (b) Bei einer Befragung (dafür/dagegen) zu einem Projekt waren 168 von 300 Personen dafür. Gibt es eine Mehrheit für das Projekt?

6.27 Statistische Hypothesen und Tests

10. Prüfen Sie mittels Wahrscheinlichkeitsnetz, ob die Daten von **B-6** aus einer Normalverteilung stammen. (*Hinweis:* Nehmen Sie dazu ein vorgefertigtes Papier – z.B. von der UE–Homepage; die R–Funktionen `qqnorm`, `qqline`; die R–Funktion `net.normal` von der UE–Homepage, ...)

11. [R–Aufgabe] Prüfen Sie mittels W–Netz, ob die Variablen (a) `GR` (= Größe) und (b) `GW` (= Gewicht) aus dem Datensatz `meddat.dat` (UE–Homepage) approximativ einer Normalverteilung folgen. Betrachten Sie jeweils beide Geschlechter getrennt voneinander. (*Hinweis:* Vgl. den Hinweis zum vorigen Beispiel.)
    [`net.normal.r`]

12. Ein Lieferant behauptet, daß höchstens 1% seiner Produkte fehlerhaft ist. Zur Überprüfung ziehen Sie eine Stichprobe von 100 Stück und beschließen nicht zu kaufen, wenn die Stichprobe mindestens 3 fehlerhafte Einheiten enthält.

    (a) Formulieren Sie die Null– und Alternativhypothese.

    (b) Wie groß ist bei diesem Test die Wahrscheinlichkeit eines Fehlers 1. Art?

    (c) Wie groß ist die Wahrscheinlichkeit eines Fehlers 2. Art, wenn der Ausschußanteil tatsächlich 5% beträgt?

6.28 Tests für Normalverteilungen

13. Einem Los von elektrischen Widerständen, deren Größe normalverteilt ist, wurden 20 Widerstände entnommen. Als Mittel ergab sich $\overline{x} = 111.0\ \Omega$ und als Standardabweichung $s = 3.4\ \Omega$. Sind die Stichprobenergebnisse mit der Vorgabe vereinbar, daß der Mittelwert der Widerstände $\mu = 110.0\ \Omega$ betragen soll? ($\alpha = 5\%$)

14. [Fortsetzung des vorhergehenden Beispiels] Sind die Stichprobenergebnisse mit der Vorgabe vereinbar, daß die Standardabweichung der Widerstände $\sigma = 3.0\ \Omega$ betragen soll? ($\alpha = 5\%$)

15. Die Zeit für die Ausführung eines (standardisierten) Proggramms wurde auf zwei verschiedenen Computersystemen gemessen. Die Stichprobengrößen waren $n_X = 10$ bzw. $n_Y = 20$; für die Stichprobenmittel und die Stichprobenvarianzen ergab sich:

$$\overline{x} = 104\ [\mathrm{s}], \quad \overline{y} = 114\ [\mathrm{s}]$$
$$s_X^2 = 290\ [\mathrm{s}^2], \quad s_Y^2 = 510\ [\mathrm{s}^2]$$

    (a) Unter der Annahme normalverteilter Beobachtungen (mit gleicher Varianz) teste man, ob die mittleren Ausführungszeiten auf beiden Systemen gleich sind, d.h. man teste $\mu_X = \mu_Y$ gegen $\mu_X \neq \mu_Y$. Die Fehlerwahrscheinlichkeit 1. Art soll dabei gleich 10% sein. Was ändert sich, wenn die letztere Wahrscheinlichkeit nur 5% betragen soll?

    *(b) Ermitteln Sie ein 90%–Konfidenzintervall für die Differenz $\Delta = \mu_Y - \mu_X$ der Mittelwerte. Gibt es einen Zusammenhang mit (a)? (*Hinweis:* Eine passende Pivotgröße wurde in der Vorlesung angegeben.)

16. [Fortsetzung des vorhergehenden Beispiels] Testen Sie, ob die beiden Varianzen als gleich angesehen werden können, d.h. testen Sie $\sigma_X^2 = \sigma_Y^2$ gegen $\sigma_X^2 \neq \sigma_Y^2$ ($\alpha = 10\%$).

6.29 Der Chiquadrat–Anpassungstest

17. Die folgende Tabelle enthält die Häufigkeiten der Ziffern $0, 1, \ldots, 9$ unter den ersten 2000 Stellen von π:

Ziffer	0	1	2	3	4	5	6	7	8	9
Häufigkeit	181	213	207	189	195	205	200	197	202	211

Prüfen Sie mittels Chiquadrat–Anpassungstest (mit $\alpha = 5\%$), ob die Ziffern einer diskreten Gleichverteilung $D_{\{0,1,\ldots,9\}}$ folgen.

18. Die folgenden (der Größe nach geordneten) 25 Zahlen stammen von einem Zufallszahlengenerator:

0.1254	0.1584	0.1995	0.2472	0.2704	0.3153	0.4163	0.4936	0.4973	0.5150
0.5295	0.5333	0.5712	0.6169	0.6377	0.6545	0.6863	0.8039	0.8347	0.8495
0.8804	0.9699	0.9746	0.9767	0.9854					

Prüfen Sie mittels Chiquadrat–Anpassungstest (mit $\alpha = 5\%$), ob die Zahlen als Beobachtungen einer nach $U_{(0,1)}$ verteilten stochastischen Größe X angesehen werden können. Achten Sie auf die Einhaltung der Faustregel $nw_i \geq 5$.

19. Prüfen Sie mittels Chiquadrat–Anpassungstest (mit $\alpha = 5\%$), ob die folgenden Beobachtungen aus einer Poissonverteilung P_μ stammen. Achten Sie dabei auf die Einhaltung der Faustregel $n\widehat{w}_i \geq 5$. (*Hinweis:* Ermitteln Sie zuerst den plausiblen Schätzwert von μ.)

k	0	1	2	3	4	5	6	7	8
Häufigkeit	16	23	25	20	9	4	0	2	1

20. Die Generierung von 100 nach der Methode von **B-5-5** bestimmter Zufallszahlen erbrachte zusammengefaßt das folgende Ergebnis:

Klasse	$[-3, -2]$	$(-2, -1]$	$(-1, 0]$	$(0, 1]$	$(1, 2]$	$(2, 3]$	$(3, 4]$
Häufigkeit	2	16	36	31	12	2	1

Prüfen Sie mittels Chiquadrat–Anpassungstest (mit $\alpha = 5\%$), ob die Zahlen als Beobachtungen einer Standardnormalverteilung angesehen werden können. Achten Sie dabei auf die Einhaltung der Faustregel $nw_i \geq 5$.

6.30 Regressionsrechnung

21. Betrachten Sie die folgenden Beobachtungspaare:

y	20	40	50	70	80
x	50	51	52	53	54

(a) Zeichnen Sie ein Streudiagramm.

(b) Bestimmen Sie die Parameter einer Ausgleichsfunktion der Form $y = \alpha + \beta x$.

(c) Zeichnen Sie die geschätzte Funktion in das Streudiagramm ein.

22. Verbunden mit einem Computerjob sind zwei sGn: CPU–Zeit (Y) und Zahl der I/O–Operationen (X). Bei 10 Jobs ergeben sich die folgenden Werte:

	Zeit(s) y	I/O-Op. x
1	40	398
2	37	390
3	43	410
4	48	502
5	62	590
6	32	305
7	20	220
8	27	252
9	41	398
10	39	382

(a) Zeichnen Sie ein Streudiagramm. Ist ein linearer Zusammenhang sinnvoll?

(b) Angenommen, man möchte die CPU–Zeit auf Basis der I/O–Operationen prognostizieren und verwendet einen linearen Ansatz der Form:

$$Y = a + bx + U, \quad \mathbb{E}(U) = 0,\ \mathrm{Var}(U) = \sigma^2$$

Ermitteln Sie Schätzwerte für a und b sowie einen Schätzwert für σ^2. Zeichnen Sie die geschätzte Regressionsgerade in das Streudiagramm ein.

(c) Angenommen, man möchte die Zahl der I/O–Operationen auf Basis der CPU–Zeit prognostizieren und verwendet einen linearen Ansatz der Form:

$$X = c + dy + V, \quad \mathbb{E}(V) = 0,\ \mathrm{Var}(V) = \tau^2$$

Ermitteln Sie Schätzwerte für c und d sowie einen Schätzwert für τ^2. Zeichnen Sie die geschätzte Regressionsgerade in das Streudiagramm ein.

23. [R–Aufgabe] Lösen Sie die beiden vorhergehenden Aufgaben mit Hilfe von R. (*Hinweis:* Nehmen Sie zur Bestimmung der Parameter die Funktion `lm`.)