Wirtschaftswissenschaftliches Prüfungsamt

BACHELOR-STUDIENGANG
SCHLIESSENDE STATISTIK
SOMMERSEMESTER 2010

Aufgabenstellung und Ergebnisse

Prof. Dr. Ralph Friedmann / Dr. Martin Becker

Hinweise für die Klausurteilnehmer

- Kleben Sie bitte **sofort** Ihr Namensschild auf obige Markierung!
- Die Klausur besteht aus insgesamt 8 Aufgaben. Prüfen Sie die Vollständigkeit Ihres Exemplares nach; spätere Reklamationen können nicht berücksichtigt werden.
- Es sind insgesamt 120 Punkte (= 24 + 24 + 6 + 13 + 14 + 11 + 10 + 18) erreichbar.
- Als Hilfsmittel sind zugelassen: Taschenrechner, 2 selbsterstellte DIN-A4 Blätter bzw. 4 selbsterstellte (einseitige) DIN-A4 Seiten. Tabellen zur Normal-, t-, χ^2- und F-Verteilung sind der Klausur beigefügt.
- Bei der Korrektur werden **nur** die Lösungen auf diesen Blättern berücksichtigt.
- Bei mehreren Lösungsvorschlägen muss die gültige Lösung eindeutig gekennzeichnet sein.
- Mit Ausnahme der Multiple-Choice-Aufgaben muss der Lösungsweg klar ersichtlich sein.

Bewertungsteil — Bitte nicht beschreiben							
Aufgabe	(a)	(b)	(c)	(d)	(e)	(f)	Σ
1							
2							
3							
4							
5							
6							
7							
8							
Σ							

Aufgabe 1 (24 Punkte)

Markieren Sie jeweils mit einem Kreuz pro Aussage im betreffenden Kästchen, ob die unten stehenden Aussagen wahr oder falsch sind.
Richtige Antworten geben +3 Punkte, falsche Antworten −1 Punkt, nicht bearbeitete Aussagen 0 Punkte (Aussagen mit zwei Kreuzen zählen als nicht bearbeitet!).
Die Aufgabe wird insgesamt mit mindestens 0 Punkten bewertet!

	Aussage	wahr	falsch
1.	Sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y. Dann sind $X_1, \ldots, X_n$ stets stochastisch unabhängig.	■	□
2.	Sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y und es gelte $E(Y) = \mu$. Dann ist $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ stets eine erwartungstreue Schätzfunktion für μ.	■	□
3.	Das Signifikanzniveau eines Hypothesentests entspricht stets der minimalen Fehlerwahrscheinlichkeit 2. Art.	□	■
4.	Bei der Entscheidung eines statistischen Hypothesentests zum Signifikanzniveau α mit Hilfe des zugehörigen p-Werts p wird H_0 genau dann abgelehnt, wenn $p > \alpha$ gilt.	□	■
5.	p-Werte zu statistischen Hypothesentests existieren nur dann, wenn die Teststatistik unter H_0 zumindest approximativ standardnormalverteilt ist.	□	■
6.	Sind die Voraussetzungen der einfachen Varianzanalyse erfüllt, so unterscheiden sich die Verteilungen zu den einzelnen Faktorstufen höchstens in der Lage.	■	□
7.	Beim Chi-Quadrat-Test für die Varianz sind nur einseitige Tests möglich.	□	■
8.	Man kann die Breite der Konfidenzintervalle für μ bei normalverteilten Grundgesamtheiten mit bekannter Varianz halbieren, indem man den Stichprobenumfang verdoppelt.	□	■

Aufgabe 2 (24 Punkte)

Markieren Sie jeweils die korrekte Antwort mit einem Kreuz im betreffenden Kästchen. Es ist jeweils genau ein Kreuz korrekt.
Richtige Antworten geben +6 Punkte, falsche Antworten −1 Punkt, nicht bearbeitete Aufgabenteile 0 Punkte (Aufgabenteile mit mehr als einem Kreuz zählen als nicht bearbeitet!).
Die Aufgabe wird insgesamt mit mindestens 0 Punkten bewertet!

1. Markieren Sie die (einzig) richtige Vervollständigung des folgenden Satzes (beachten Sie, dass die Gegenwahrscheinlichkeit eines Ereignisses A als $1 - P(A)$ definiert ist):

   Die Gütefunktion liefert, ausgewertet für die wahren Verteilungsparameter der Grundgesamtheit,

   (a) die Wahrscheinlichkeit für einen Fehler 1. Art, falls die Verteilungsparameter zur Nullhypothese gehören, und die Wahrscheinlichkeit für einen Fehler 2. Art, sonst. ☐

   (b) die Wahrscheinlichkeit für einen Fehler 1. Art, falls die Verteilungsparameter zur Nullhypothese gehören, und die Gegenwahrscheinlichkeit für einen Fehler 2. Art, sonst. ■

   (c) die Gegenwahrscheinlichkeit für einen Fehler 1. Art, falls die Verteilungsparameter zur Nullhypothese gehören, und die Wahrscheinlichkeit für einen Fehler 2. Art, sonst. ☐

   (d) die Gegenwahrscheinlichkeit für einen Fehler 1. Art, falls die Verteilungsparameter zur Nullhypothese gehören, und die Gegenwahrscheinlichkeit für einen Fehler 2. Art, sonst. ☐

2. Im einfachen linearen Regressionsmodell

$$y_t = \beta_1 + \beta_2 x_t + u_t$$

   mit $\mathrm{E}(u_t) = 0$, $\mathrm{Var}(u_t) = \sigma^2$, $\mathrm{Cov}(u_t, u_s) = 0$ für alle $t, s = 1, \ldots, T; t \neq s$ ist das Bestimmtheitsmaß R^2 umso größer,

   (a) je kleiner β_1 (betragsmäßig) ist. ☐

   (b) je größer β_1 (betragsmäßig) ist. ☐

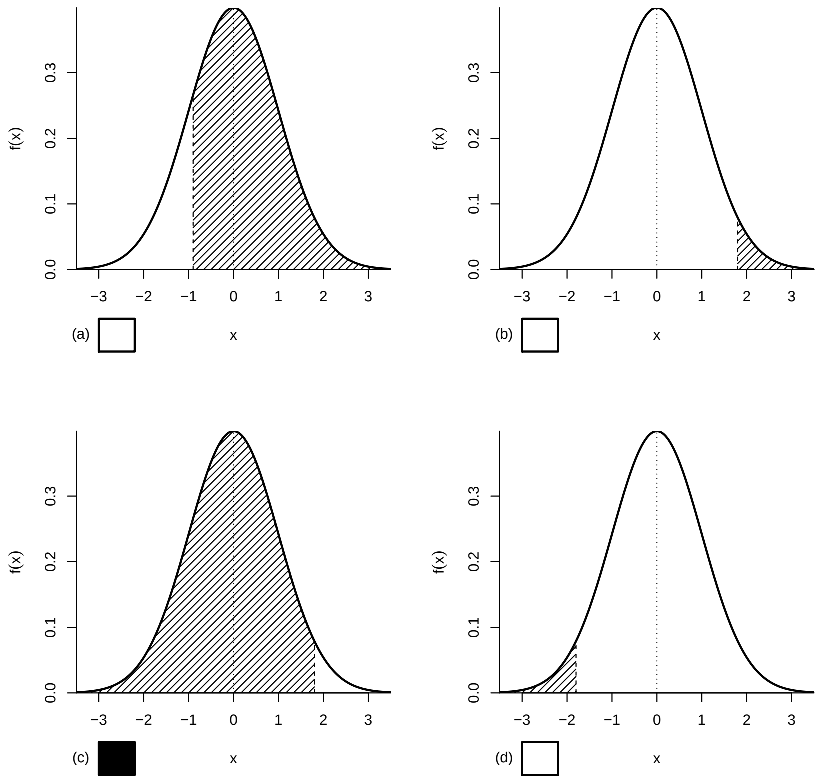
   (c) je kleiner der Korrelationskoeffizient zwischen der unabhängigen Variablen x und der abhängigen Variablen y (betragsmäßig) ist. ☐

   (d) je größer der Korrelationskoeffizient zwischen der unabhängigen Variablen x und der abhängigen Variablen y (betragsmäßig) ist. ■

3. Sei $X_1, \ldots, X_{16}$ eine einfache Stichprobe zu einer $N(\mu, \sigma_0^2)$-verteilten Zufallsvariablen Y mit unbekanntem Parameter μ und bekanntem Parameter $\sigma_0^2 = 4$. Auf der Grundlage einer Stichprobenrealisation mit realisiertem Stichprobenmittel $\overline{x} = 10.9$ soll

$$H_0 : \mu \geq 10 \qquad \text{gegen} \qquad H_1 : \mu < 10$$

getestet werden. Markieren Sie die Abbildung, welche den p-Wert des oben beschriebenen Tests korrekt als Inhalt der schraffierten Fläche unter der Dichtefunktion der Standardnormalverteilung darstellt.



4. Bei einem Chi-Quadrat-Anpassungstest soll anhand einer einfachen Stichprobe vom Umfang $n = 500$ überprüft werden, ob die Grundgesamtheit exponentialverteilt ist. Nach Wahl einer geeigneten Klassierung aus 10 Klassen wird dazu zunächst der unbekannte Parameter der Exponentialverteilung durch eine ML-Schätzung aus den klassierten Daten ermittelt. Damit ist zur Konstruktion des Ablehnungsbereiches die folgende Verteilung zu verwenden:

(a) χ^2-Verteilung mit 10 Freiheitsgraden ☐

(b) χ^2-Verteilung mit 9 Freiheitsgraden ☐

(c) χ^2-Verteilung mit 8 Freiheitsgraden ■

(d) χ^2-Verteilung mit 7 Freiheitsgraden ☐

Aufgabe 3 (3 + 3 = 6 Punkte)

Zu einer Grundgesamtheit Y mit $E(Y) = \mu$ und $\mathrm{Var}(Y) = \sigma^2 > 0$ sei zur Schätzung von μ aus einer einfachen Stichprobe $X_1, \ldots, X_n$ vom Umfang $n > 1$ (jeweils) die Schätzfunktion

$$\widetilde{\mu}_n := \frac{1}{2}(X_1 + X_n)$$

definiert, also die Schätzfunktion, die (jeweils) die erste und letzte Beobachtung in der Stichprobe mittelt.

(a) Sind die Schätzfunktionen $\widetilde{\mu}_n$ erwartungstreu für μ? Begründen Sie Ihre Antwort!

(b) Ist die Folge der Schätzfunktionen $\widetilde{\mu}_n$ für μ konsistent im quadratischen Mittel? Begründen Sie Ihre Antwort!

Ergebnisse (ohne Begründung/Rechenweg):

(a) Ja.

(b) Nein.

Aufgabe 4 (6 + 3 + 4 = 13 Punkte)

Es sei Y eine mit Parameter p, $0 < p < 1$, geometrisch verteilte Grundgesamtheit, Y habe also die Wahrscheinlichkeitsfunktion

$$p_Y(i) = \begin{cases} (1-p)^i \cdot p & \text{falls } i \in \{0, 1, 2, \ldots\} \\ 0 & \text{sonst} \end{cases} .$$

(a) Sei $X_1, \ldots, X_n$ eine einfache Stichprobe vom Umfang n zu Y und $x_1, \ldots, x_n$ die Realisation zu $X_1, \ldots, X_n$. Berechnen Sie (in Abhängigkeit von der Stichprobenrealisation $x_1, \ldots, x_n$) den ML-Schätzer für p.

(b) Bekanntlich ist die Anzahl der Misserfolge einer Bernoulli-Kette vor dem ersten Erfolg geometrisch verteilt, wobei der Parameter die Erfolgswahrscheinlichkeit des Bernoulli-Experiments ist.

Um die Wahrscheinlichkeit für das Würfeln einer „6“ bei einem bestimmten Würfel zu schätzen, wurde 10-mal solange gewürfelt, bis zum ersten Mal eine „6“ gefallen war, und die Anzahl der vorangegangenen (Fehl-)Würfe notiert. Berechnen Sie aus der so erhaltenen Stichprobenrealisation

$$(x_1, \ldots, x_{10}) = (3, 2, 14, 0, 7, 3, 6, 8, 0, 3)$$

mit Hilfe von Teil (a) den ML-Schätzer der Wahrscheinlichkeit, mit dem Würfel eine „6“ zu würfeln.

(c) Der Erwartungswert einer mit Parameter p geometrisch verteilten Zufallsvariable ist bekanntlich $\frac{1-p}{p}$. Berechnen Sie nun den zur in Teil (b) angegebenen Stichprobenrealisation gehörigen Schätzwert für p nach der Methode der Momente. In welchem Verhältnis steht der erhaltene Wert zum Ergebnis aus Teil (b)?

(Hinweis: Beachten Sie, dass Sie Teil (c) im Wesentlichen auch ohne Lösung der Teile (a) und (b) bearbeiten können!)

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\hat{p} = \frac{n}{n+\sum_{i=1}^{n} x_i} = \frac{1}{1+\overline{x}}$

(b) $\hat{p} = 0.1786$.

(c) $\hat{p}_{MM} = 0.1786$.

Aufgabe 5 (7 + 4 + 3 = 14 Punkte)

Bei der Herstellung von Desinfektionsspray weiß man aus langjähriger Erfahrung, dass die verwendete Abfüllanlage eine Varianz von $2^2 = 4[ml^2]$ für die abgefüllte Flüssigkeitsmenge hat. Nach einer Inventur hat der Hersteller den Verdacht, dass die Abfüllanlage im Mittel mehr als die auf dem Produkt ausgezeichneten $100[ml]$ in die Sprühflaschen einfüllt. Dies soll mit einem statistischen Test überprüft werden. Hierzu werden der Produktion 25 Flaschen entnommen, deren gemessene Füllmengen $x_1, \ldots, x_{25}$ als Realisation einer einfachen Stichprobe vom Umfang 25 zur annahmegemäß $N(\mu, 2^2[ml^2])$-verteilten Abfüllmenge betrachtet werden kann. Als Stichprobenmittelwert ergibt sich dabei

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 100.82[ml] \ .$$

(a) Testen Sie zum Signifikanzniveau $\alpha = 0.05$, ob der Verdacht der Herstellerfirma bestätigt werden kann. Fassen Sie das Ergebnis des Tests in einem Antwortsatz zusammen.

(b) Wie groß ist die Wahrscheinlichkeit des Fehlers 2. Art, falls $\mu = 101[ml]$ beträgt?

(c) Geben Sie auf Basis der Stichprobenrealisation ein zweiseitiges Konfidenzintervall für die mittlere Abfüllmenge zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.99$ **(!)** an.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $Z = 2.05 \in (1.645, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt! Der Test bestätigt also den Verdacht der Herstellerfirma, dass die Maschine im Mittel eine zu große Menge abfüllt.

(b) $\beta(101) = 0.1963$

(c) Realisation des Konfidenzintervalls zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.99$: $[99.79, 101.85]$

Aufgabe 6 (11 Punkte)

Ein Hotelangestellter vermutet nach langjähriger Tätigkeit, dass Häufigkeit und Art des Auftretens vergessener Gegenstände vom Geschlecht der Hotelgäste abhängig sind. Um seine Hypothese zu überprüfen, notiert er über einen Zeitraum von einem Jahr zusammen mit dem Geschlecht der Gäste in Einzelzimmern, ob und welche Gegenstände auf den Zimmern vergessen wurden. Im Einzelnen erhielt er dabei folgende Häufigkeiten für die vergessenen Gegenstände in Verbindung mit dem Geschlecht:

	keine	Kleidung	elektr. Gerät	Kleidung und elektr. Gerät
männlich	943	45	71	41
weiblich	757	105	29	9

Überprüfen Sie anhand dieses Datenmaterials zum Signifikanzniveau $\alpha = 0.05$, ob das Geschlecht der Gäste und die Art bzw. Häufigkeit des Auftretens vergessener Gegenstände abhängig sind.

Ergebnisse (ohne Begründung/Rechenweg):

$\chi^2 = 63.102 \in (7.815, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
Die Vermutung des Hotelangestellten, dass eine Abhängigkeit zwischen Geschlecht der Gäste und Art bzw. Häufigkeit des Auftretens vergessener Gegenstände besteht, wird also vom Test bestätigt.

Aufgabe 7 (8 + 2 = 10 Punkte)

Anhand der Ergebnisse der Klausur zur Veranstaltung „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“ des Sommersemesters 2010 soll mit Hilfe einer einfachen Varianzanalyse untersucht werden, ob die Verteilung der von den Studierenden im Fach „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“ erreichten Punktzahlen abhängig davon ist, in welchem Prüfungsversuch (1. Versuch – 3. Versuch) sich der Studierende befindet. Hierzu wurden folgende Daten zu den erreichten Punktzahlen erhoben:

i (Versuch)	n_i	$\overline{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$	$\sum_{j=1}^{n_i} x_{ij}^2$
1	312	59.431	1237338.75
2	41	51.841	121766.75
3	15	62.633	64820.25

Für die Durchführung der einfachen Varianzanalyse wurde hieraus bereits die Größe $SW = 152897.65$ berechnet.

(a) Überprüfen Sie mit einer einfachen Varianzanalyse (Signifikanzniveau $\alpha = 0.05$) unter der Annahme, dass die erreichten Punktzahlen Realisierungen von unabhängig $N(\mu_i, \sigma^2)$-verteilten Zufallsvariablen $X_{ij} (1 \leq i \leq 3, 1 \leq j \leq n_i)$ sind, ob die Ausprägung des Merkmals „Prüfungsversuch“ einen Einfluss auf die erreichte Punktzahl hat. Fassen Sie das Ergebnis der Varianzanalyse auch in einem Antwortsatz zusammen.

(Hinweis: Verwenden Sie beim Ablesen des Quantils der Verteilung der Teststatistik unter H_0 einen „naheliegenden“ Wert der Tabelle, da die eigentlich benötigte Verteilung nicht in der entsprechenden Tabelle enthalten ist.)

(b) Geben Sie eine Formel an, mit der die (vorgegebene) Größe SW auch aus den Angaben der obigen Tabelle berechnet werden könnte.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $F = 2.778 \notin (3.02, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!

Die einfache Varianzanalyse kommt also zum Ergebnis, dass kein signifikanter ($\alpha = 0.05$) Einfluss des Prüfungsversuchs auf die erreichte Punktzahl besteht.

(b)

$$SW = \sum_{i=1}^{3} \left(\left(\sum_{j=1}^{n_i} x_{ij}^2 \right) - n_i \cdot \overline{x}_i^2 \right)$$

Aufgabe 8 (3 + 3 + 4 + 2 + 2 + 4 = 18 Punkte)

Man gehe davon aus, dass sich die Abhängigkeit des systolischen Blutdrucks y_t vom Lebensalter x_t eines Menschen durch das einfache lineare Regressionsmodell

$$y_t = \beta_1 + \beta_2 x_t + u_t \qquad \text{mit} \quad u_t \overset{i.i.d.}{\sim} N(0, \sigma^2),\ t = 1, \ldots, T$$

erklären lässt.
In einer medizinischen Studie wurden bei $T = 62$ Personen die Merkmale Alter x_t und systolischer Blutdruck y_t erhoben und daraus für die Durchführung einer einfachen linearen Regressionsanalyse bereits die folgenden Zwischenwerte errechnet:

$$\sum_{t=1}^{62} y_t = 9193; \quad \sum_{t=1}^{62} y_t^2 = 1384977; \quad \sum_{t=1}^{62} x_t = 2882;$$

$$\sum_{t=1}^{62} x_t^2 = 148292; \quad \sum_{t=1}^{62} x_t y_t = 441517; \quad \sum_{t=1}^{62} \hat{y}_t^2 = 1377142.$$

(a) Schätzen Sie β_1 und β_2 mit Hilfe der Kleinst-Quadrate-Methode.

(b) Schätzen Sie die Varianz σ^2 mit einer erwartungstreuen Schätzfunktion und berechnen Sie $\widehat{\sigma}^2_{\widehat{\beta}_1}$ sowie $\widehat{\sigma}^2_{\widehat{\beta}_2}$.

(c) Testen Sie zum Signifikanzniveau $\alpha = 0.001$ **(!)**, ob das Alter einer Person einen signifikanten Einfluss auf den systolischen Blutdruck hat.

(d) Geben Sie ein zweiseitiges Konfidenzintervall zum Niveau $1 - \alpha = 0.95$ für β_1 an.

(e) Steigt der systolische Blutdruck eher mit zunehmendem oder eher mit abnehmendem Alter? Begründen Sie Ihre Antwort.

(f) Geben Sie ein zweiseitiges Prognoseintervall zum Niveau $1 - \alpha = 0.95$ für $\mathrm{E}(y_{63})$ an, falls $x_{63} = 33$ beträgt.

Ergebnisse (ohne Begründung/Rechenweg):

(a) $\widehat{\beta}_2 = 0.9906$, $\widehat{\beta}_1 = 102.227$

(b) $\widehat{\sigma}^2 = 130.583$, $\widehat{\sigma}^2_{\widehat{\beta}_1} = 21.802$, $\widehat{\sigma}^2_{\widehat{\beta}_2} = 0.009115$

(c) $t = 10.376 \in (-\infty, -3.46) \cup (3.46, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

Das Alter der Person hat also einen signifikanten Einfluss auf den systolischen Blutdruck.

(d) Realisation des Konfidenzintervalls für β_1 zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.95$:

$[92.889, 111.565]$

(e) Der systolische Blutdruck steigt eher mit zunehmendem Alter.

(f) Prognoseintervall für $\mathrm{E}(y_{63})$ zur Sicherheitswahrscheinlickeit $1 - \alpha = 0.95$:

$[131.037, 138.797]$