Lösungsvorschläge zur Klausur

Wahrscheinlichkeitsrechnung und schließende Statistik (Wintersemester 2011/12)

Aufgabe 1 (9 Punkte)

Die Zeit, die ein Autofahrer für eine bestimmte Strecke benötigt, kann als Zufallsvariable X mit Erwartungswert μ und Standardabweichung σ angesehen werden.

I. In diesem Teil sei $\mu = 4$ (Stunden) und $\sigma = 0.2$ (Stunden).

a) Wie groß ist mindestens die Wahrscheinlichkeit, dass der Autofahrer zwischen 3.5 und 4.5 Stunden für die Strecke benötigt?

b) Wie groß ist approximativ die Wahrscheinlichkeit, dass der Autofahrer für 100 Fahrten zwischen 395 und 405 Stunden benötigt? Gehen Sie dabei davon aus, dass die einzelnen Fahrtdauern stochastisch unabhängig sind.

II. In diesem Teil sei μ unbekannt. Bei der Messung von 50 Fahrten dieser Strecke ergab sich eine mittlere Fahrtdauer von 3.8 Stunden bei einer Stichprobenstandardabweichung von 0.28. Gehen Sie davon aus, dass die 50 Fahrten eine einfache Stichprobe darstellen. Geben Sie ein Konfidenzintervall für den Erwartungswert μ zum Konfidenzniveau $1 - \alpha = 0.99$ an.

Lösung:

I. Es ist bekannt, dass $E[X] = \mu = 4$ und $V[X] = \sigma^2 = 0.2^2$.

a) Mit der Ungleichung von Tschebyscheff erhält man

$$P(3.5 \le X \le 4.5) = P(|X - \mu| \le 0.5) \ge 1 - \frac{0.2^2}{0.5^2} = 0.84\,.$$

b) Nach dem Zentralen Grenzwertsatz (Faustregel $n \ge 40$ ist erfüllt!) ist $S_{100} = \sum_{i=1}^{100} X_i$ approximativ normalverteilt,

$$S_{100} = \sum_{i=1}^{100} X_i \overset{appr.}{\sim} N(100\mu, 100\sigma^2) = N(400, 4)\,.$$

Somit ist

$$\begin{aligned} P(395 \le S_{100} \le 405) &\approx \Phi\left(\frac{405-400}{2}\right) - \Phi\left(\frac{395-400}{2}\right) = \Phi(2.5) - \Phi(-2.5) \\ &= 2\Phi(2.5) - 1 = 2 \cdot 0.9938 - 1 = 0.9876\,. \end{aligned}$$

II. Wir bestimmen das Konfidenzintervall für μ einer beliebigen Verteilung bei unbekannter Varianz. Gegeben ist $n = 50$, $\overline{X} = 3.8$ und $S_X = 0.28$. Achtung: Da hier S_X und nicht S_X^* gegeben ist, muss man entweder erst S_X^* ausrechnen oder man ersetzt in der Formel für das Konfidenzintervall $S_X^*/\sqrt{n}$ durch $S_X/\sqrt{n-1}$. Wergen $1 - \alpha = 0.99$ ist $u_{1-\alpha/2} = u_{0.995} = 2.5758$. Somit erhält man für das Konfidenzintervall

$$\overline{X} \mp u_{1-\frac{\alpha}{2}} \frac{S_X}{\sqrt{n-1}} = 3.8 \mp 2.5758 \cdot \frac{0.28}{7} = 3.8 \mp 0.1030\,.$$

Das konkrete Konfidenzintervall ist also $[3.6970, 3.9030]$.

Aufgabe 2 (9 Punkte)

I. Gegeben sei die folgende Funktion:

$$f(x) = \begin{cases} m(1+x)\,, & \text{falls } -1 \le x < 0, \\ m(1-x)\,, & \text{falls } 0 \le x \le 1, \\ 0 & \text{sonst.} \end{cases}$$

a) Skizzieren Sie die Funktion f.

b) Bestimmen Sie m so, dass f die Dichte einer Zufallsvariablen X ist.

c) Bestimmen Sie $E[X]$.

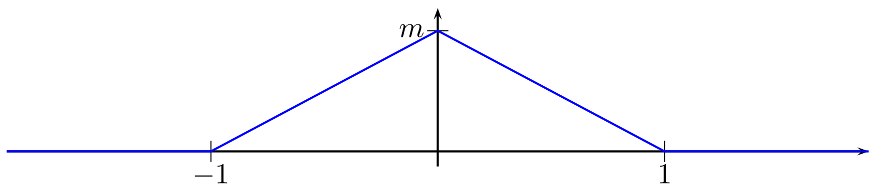
II. Gegeben sei ein medizinischer Test, welcher bezüglich einer bestimmten Krankheit entweder zu einem positiven oder einem negativen Befund führt. Es sei bekannt, dass 5% aller Personen an der betreffenden Krankheit leiden. Die Wahrscheinlichkeit, dass sich bei einer zufällig ausgewählten Person ein negativer Befund (d.h. die untersuchte Person ist vermeintlich gesund) ergibt, betrage 70%. Ergibt der Test einen positiven Befund, so ist die Wahrscheinlichkeit, dass die Person tatsächlich erkrankt ist, gleich 1/6.

a) Berechnen Sie die Wahrscheinlichkeit, dass eine gesunde Person einen positiven Befund erhält.

b) Berechnen Sie die Wahrscheinlichkeit, dass eine erkrankte Person einen positiven Befund erhält.

Lösung:

I. a)



b) Die Fläche unter der Dichte muss 1 sein. Offensichtlich ist die Fläche unter der Dichte ein Dreieck mit Grundseite der Länge 2 und der Höhe m. Die Fläche diese Dreiecks ist $\frac{1}{2} \cdot 2 \cdot m$. Da die Fläche 1 sein muss, muss also $m = 1$ gelten. Dass die Dichte dann auch überall größer oder gleich null ist, ist wegen der Skizze klar.

c) Da die Dichte symmetrisch zu null ist, gilt $E[X] = 0$ und $x_{0.5} = 0$.

II. Wir definieren die Ereignisse K: „Person ist krank" und B_P: „Befund ist positiv". Bekannt ist $P(K) = 0.05$, $P(\overline{B_P}) = 0.7$ und $P(K \mid B_P) = 1/6$. Aus der letzten Angabe erhält man $P(K \cap B_P) = \frac{1}{6} P(B_P) = \frac{1}{6} \cdot 0.3 = 0.05$. Damit erhält man die folgende Vierfeldertafel (gegebene Werte in Fettdruck):

	B_P	$\overline{B_P}$	
K	**0.05**	0	**0.05**
$\overline{K}$	0.25	0.7	0.95
	0.3	**0.7**	1

Mit der Vierfeldertafel erhält man:

a) $P(B_P \mid \overline{K}) = \dfrac{P(B_p \cap \overline{K})}{P(\overline{K})} = \dfrac{0.25}{0.95} = 0.2632,$

b) $P(B_P \mid K) = \dfrac{P(B_p \cap K)}{P(K)} = \dfrac{0.05}{0.05} = 1.$

Aufgabe 3 (9 Punkte)

I. In einer Mensa werden jeden Tag vier verschiedene Gerichte der Gruppe A (mit Fleisch), zwei verschiedene Gerichte der Gruppe B (vegetarisch) und ein Gericht der Gruppe C (vegan) angeboten. Um einen Speiseplan zusammenzustellen, stehen 20 Gerichte der Gruppe A, 10 der Gruppe B und 5 der Gruppe C zur Auswahl.

a) Wie viele verschiedene Tagespläne können erstellt werden?

b) Ein Student hat zwei Lieblingsgerichte. Eines davon gehört zur Gruppe B, das andere zur Gruppe C. Wie hoch ist die Wahrscheinlichkeit, dass in einem zufällig ausgewählten Tagesplan mindestens eines dieser Gerichte enthalten ist?

II. Um zu untersuchen, wie zufrieden die Studenten mit dem Mensaessen sind, werden 200 zufällig ausgewählte Studenten nach ihrer Zufriedenheit befragt. Von diesen geben 160 an, mit dem Essen zufrieden zu sein. Führen Sie einen geeigneten Test zum Niveau $\alpha = 0.05$ durch, um zu untersuchen, ob statistisch gesichert werden kann, dass mehr als 75% der Studenten mit dem Mensaessen zufrieden sind. Wie lautet Ihre Testentscheidung?

Lösung:

I. a) Da die Auswahl für jede Kategorie ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge geschieht, ist die Anzahl verschiedener Speisepläne

$$\binom{20}{4}\binom{10}{2}\binom{5}{1} = 4845 \cdot 45 \cdot 5 = 1\,090\,125\,.$$

b) Da die Auswahl zufällig geschieht, hat jeder Speiseplan dieselbe Wahrscheinlichkeit; es handelt sich also um ein Laplace-Experiment. Wir berechnen zunächst die Gegenwahrscheinlichkeit, d.h. die Wahrscheinlichkeit, dass keines der beiden Lieblingsgerichte im Speiseplan enthalten ist. Es gibt offensichtlich

$$\binom{20}{4}\binom{9}{2}\binom{4}{1} = 4845 \cdot 36 \cdot 4 = 697\,680$$

derartige Speisepläne. Somit ist

$$P(\text{kein Lieblingsgericht}) = \frac{697\,680}{1\,090\,125} = 0.64\,.$$

Die gesuchte Wahrscheinlichkeit ist also

$$P(\text{mindestens ein Lieblingsgericht}) = 1 - 0.64 = 0.36\,.$$

II. Da nur die Gegenhypothese statistisch gesichert werden kann, muss $H_1 : \pi > 0.75 = \pi_0$ gewählt werden. Zu testen ist also die Nullhypothese $H_0 : \pi \leq 0.75$ gegen die Alternative $H_1 : \pi > 0.75$.

Aus den Angaben in der Aufgabe erhält man $\hat{\pi} = \frac{160}{200} = 0.8$ sowie $n = 200$. Wegen $n\pi_0(1-\pi_0) = 200 \cdot 0.75 \cdot 0.25 = 37.5 > 9$ ist die Faustregel für die Anwendung des Tests erfüllt. Die Testgröße ist

$$T = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.8 - 0.75}{\sqrt{\frac{0.75 \cdot 0.25}{200}}} = 1.6330\,.$$

Da H_0 abgelehnt wird, wenn $T > u_{1-\alpha} = u_{0.95} = 1.6449$ ist, kann H_0 nicht abgelehnt werden. Es kann also nicht statistisch gesichert werden, dass mehr als 75% der Studenten mit dem Mensaessen zufrieden sind.

Aufgabe 4 (9 Punkte)

Tragen Sie die Antworten zu den folgenden Fragen in die dafür vorgesehenen Kästchen ein. Eine Begründung der Antworten ist in dieser Aufgabe nicht erforderlich.

1. Wie bezeichnet man zwei Zufallsvariablen X und Y, für die $Cov[X, Y] = 0$ ist?

   unkorreliert

2. Für die Zufallsvariablen X und Y sei A das Ereignis „$X \leq 10$ und $Y \leq 8$“ und B das Ereignis „$X \leq 9$ und $Y \leq 11$“. Dann ist

   $A \cap B =$ $\{X \leq 9, Y \leq 8\}$

3. Seien $X_1, \ldots, X_n \overset{iid}{\sim} F$ mit $E[X_i] = \mu$ und $V[X_i] = \sigma^2$. Welche Verteilung besitzt dann $\sqrt{n}(\overline{X} - \mu)$ für hinreichend großes n? Geben Sie auch an, ob die Verteilung exakt oder nur approximativ gültig ist.

   $\overset{appr.}{\sim} N(0, \sigma^2)$

4. Sei $X_1, \ldots, X_n$ eine einfache Stichprobe aus X mit $E[X] = \mu$ und $V[X] = \sigma^2$. Welche der Schätzer S und S^* sind dann erwartungstreu für σ?

   weder S noch S^*

5. Geben Sie an, ob eine Exponentialverteilung linksschief, rechtsschief oder symmetrisch ist.

   rechtsschief

6. Wie ändert sich der kritische Bereich beim t-Test, wenn das Signifikanzniveau erhöht wird? Wird der kritische Bereich größer, wird er kleiner oder ändert sich nicht?

   wird größer

7. Wie viele unterscheidbare Möglichkeiten gibt es, sechs rote, vier blaue und zwei gelbe Kugeln anzuordnen?

   13860

8. Sei $X \sim B(20, 0.8)$. Dann ist

   $P(X = 17) =$ 0.2054

9. Sei $P(A) = 0.7$, $P(B) = 0.6$ und $P(A|B) = 0.5$. Dann ist

   $P(A \cap \overline{B}) =$ 0.4

Aufgabe 5 (9 Punkte)

Die Schätzung einer linearen Regression $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + U_i$, $i = 1, \ldots, n$, mit $U_1, \ldots, U_n \overset{iid}{\sim} N(0, \sigma^2)$ wurde mit den Analyse-Funktionen in Excel durchgeführt. Dabei erhielt man den folgenden Output:

AUSGABE: ZUSAMMENFASSUNG

Regressions-Statistik	
Multipler Korrelationskoeffizient	0.6427
Bestimmtheitsmaß	0.4131
Adjustiertes Bestimmtheitsmaß	0.3696
Standardfehler	3.2995
Beobachtungen	30

ANOVA	Freiheits-grade (df)	Quadratsum-men (SS)	Mittlere Qua-dratsumme (MS)	Prüfgröße (F)	F krit
Regression	2	206.8583	103.4291	9.5007	0.0008
Residue	27	293.9362	10.8865		
Gesamt	29	500.7944			

	Koeffizienten	Standardfehler	t-Statistik	P-Wert	Untere 95%	Obere 95%
Schnittpunkt	29.4865	4.0410	7.2968	0.0000	21.1950	37.7779
X Variable 1	1.4478	0.3465	4.1782	0.0003	0.7368	2.1588
X Variable 2	–0.2068	0.6135	–0.3372	0.7386	–1.4656	1.0519

a) Wie lauten die Punktschätzungen für die Regressionskoeffizienten β_0, β_1 und β_2?

b) Wie lautet die Testentscheidung für den Test der Nullhypothese $H_0 : \beta_1 = 0$ gegen die Alternative $H_1 : \beta_1 \neq 0$ bei einem Signifikanzniveau von $\alpha = 0.05$? Führen Sie denselben Test auch für die Nullhypothese $H_0 : \beta_2 = 0$ durch.

c) Testen Sie zu einem Signifikanzniveau $\alpha = 0.05$ die Nullhypothese, dass β_1 **und** β_2 gleich null sind, gegen die Alternative, dass dies nicht der Fall ist. Welchen Wert hat die Testgröße des Tests, wie lautet der p-Wert? Wie lautet somit Ihre Testentscheidung?

d) Welchen Wert haben die Komponenten der Varianzzerlegung SST, SSR und SSE?

e) Wie lautet die Punktschätzung für die Varianz der Störterme?

f) Welcher Anteil der Varianz der Y_i wird durch die lineare Regression erklärt?

g) Wie lauten die Punktschätzungen für $\sigma_{\hat{\beta}_1}$ und $\sigma_{\hat{\beta}_2}$?

Lösung:

a) $\hat{\beta}_0 = 29.4865$, $\hat{\beta}_1 = 1.4478$, $\hat{\beta}_2 = -0.2068$

b) Beim Test auf $H_0 : \beta_1 = 0$ ist der p-Wert gleich 0.0003 und damit kleiner als $\alpha = 0.05$. Die Nullhypothese wird daher abgelehnt.

Beim Test auf $H_0 : \beta_2 = 0$ ist der p-Wert gleich 0.7386 und damit größer als $\alpha = 0.05$. Die Nullhypothese kann daher nicht abgelehnt werden.

c) Die Testgröße ist $F = 9.5007$. Da der p-Wert mit 0.0008 kleiner als $\alpha = 0.05$ ist, wird die Nullhypothese abgelehnt.

d) $SST = 500.7944$, $SSR = 206.8583$, $SSE = 293.9362$

e) $\hat{\sigma}^2 = 3.2995^2$

f) 41.31% der Varianz von Y wird durch die lineare Regression erklärt.

g) $\widehat{\sigma_{\hat{\beta}_1}} = 0.3465$, $\widehat{\sigma_{\hat{\beta}_2}} = 0.6135$