Übungsaufgaben Kapitel 15
Lösungen

Prof. Dr. Torben Kuhlenkasper

Aufgabe 1

In einer empirischen Studie zum Rauchverhalten wurden 10 Raucher befragt, wieviele Zigaretten sie durchschnittlich pro Tag rauchen. Es wurden folgende Angaben gemacht:

```
26  34  5  20  50  44  18  39  29  19
```

(a) Überprüfen Sie anhand des Vorzeichen-Tests zum Niveau $\alpha = 0.1$, ob der Median der Anzahl der gerauchten Zigaretten größer ist als 25.

- Dem Vorzeichen-Test liegt folgendes statistische Testproblem zugrunde:

$$H_0\text{: M} \leq 25$$
$$H_1\text{: M} > 25$$

- Da $M_0 = 25$, nehmen wir als Testgröße S die Anzahl aller Beobachtungen mit einem Wert z.B. größer als 25.
- Diese ist unter H_0 binomialverteilt mit Parametern $n = 10$ und $p = 0.5$.
- Zu viele Beobachtungen größer als 25 sprechen für H_1 und somit gegen H_0.
- Wir lehnen dann H_0 ab.
- Hier aus der Stichprobe ergibt sich $S = 6$.
- gesucht ist also $P(S \geq 6|H_0 = 1 - P(S \leq 5|H_0)$.
- Der Tabelle in der Tabellensammlung auf S. 8 entnimmt man in der Spalte mit $n = 10$ und der Zeile $x \leq 5$ den Wert 0.6230

$\Longrightarrow$ Somit ergibt sich hier für den p-Wert: 1-0.6230 = 0.377 und H_0 wird mit $\alpha = 0.1$ nicht abgelehnt!

- Es kann also nicht davon ausgegangen werden, daß der Median der Anzahl der gerauchten Zigaretten größer als 25 ist.

(b) Nehmen Sie nun an, daß für die durchschnittliche Anzahl der gerauchten Zigaretten pro Tag eine Normalverteilung unterstellt werden kann. Führen Sie zum Niveau $\alpha = 0.1$ einen geeigneten Test durch. Vergleichen Sie Ihr Ergebnis mit dem aus (a).

- Der t-Test kann unter der zusätzlichen Annahme durchgeführt werden, daß die durchschnittliche Anzahl gerauchter Zigaretten X pro Tag normalverteilt ist, d.h. $X \sim N(\mu, \sigma^2)$, wobei σ^2 unbekannt ist und aus der Stichprobe mit s^2 geschätzt werden muss.
- Daher wird der t-Test verwendet.
- Die Annahme der Normalverteilung ist allerdings etwas problematisch, da es sich bei X um eine diskrete Zufallsvariable handelt.
- Nun wird das statistische Testproblem über den Erwartungswert formuliert als:

$$H_0\colon\ \mu \leq 25$$
$$H_1\colon\ \mu > 25$$

- Die Prüfgröße (Teststatistik) ist gegeben als:

$$t = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s}$$

- Mit $\overline{x} = 28.4$ und $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 186.04$ ergibt sich für die Teststatistik hier:

$$t = \frac{\sqrt{10}(28.04 - 25)}{13.64} = 0.789$$

- Hier wird H_0 abgelehnt, falls die Prüfgröße *zu groß* ist, d.h. wenn $t > t_{9;0.9} = 1.383$.
- Auch hier wird H_0 nicht abgelehnt.

Aufgabe 2

Eine Versicherungsfirma hatte folgende Schadenshöhen (in Mio. EUR) in einem Monat:

`0 1 1 2 4 4 5 6 10 12 14 19 22 36 51 75`

(a) Überprüfen Sie **ohne** Annahme der Normalverteilung, ob von mehr als 5 Mio. EUR Schadenshöhe ausgegangen werden kann. ($\alpha = 0.05$)

Ohne Annahme der Normalverteilung testen wir folgende Hypothesen mit dem Vorzeichentest:

$$H_0\text{: M} \leq 5$$
$$H_1\text{: M}>5$$

- Da $M_0 = 5$, verwenden wir als Teststatistik S: Anzahl an Beobachtungen, die größer als 5 sind.

VORSICHT: 5 ist ein beobachteter Wert im Datensatz. Diese Beobachtung wird aus dem Datensatz für den Vorzeichentest gelöscht und wir nutzen den sog. bedingten Vorzeichentest. Es gilt also hier nur noch $n = 15$.

- Unter H_0 ist die Teststatistik binomialverteilt mit den Parametern $n = 15$ und $p = 0.5$.
- Sehr viele Beobachtungen, die größer als 5 sind würden hier für H_1 und somit gegen H_0 sprechen.
- Bei sehr vielen Beobachtungen größer 5 würden wir H_0 dann ablehnen.
- In unserer Stichprobe sind $S = 9$ Beobachtungen größer als 5.
- Wir suchen also $P(S \geq 9|H_0 = 1 - P(S \leq 8|H_0)$.
- Wir können die tabellierten Werte der Binomialverteilung mit $n = 15$ und der Zeile $x \leq 5$ verwenden. Dort lesen wir 0.6964 ab.

$\Longrightarrow$ Wir erhalten somit für die Überschreitungswahrscheinlichkeit: 1-0.6230 = 0.3036 und H_0 und H_0 wird somit für $\alpha = 0.05$ nicht abgelehnt!

- Wir können also nicht davon ausgehen, dass der Median der Schadenshöhen signifikant größer als 5 Million EUR pro Jahr ist.

(b) Überprüfen Sie **mit** Annahme der Normalverteilung, ob von mehr als 5 Mio. EUR Schadenshöhe ausgegangen werden kann. ($\alpha = 0.05$, Rechenhilfen: $\overline{x} = 16.375$, $s^2 = 441.05$)

- Wir können hier den t-test mit der Annahme verwenden, dass die Schadenshöhen X normalverteilt sind.

Somit gilt $X \sim N(\mu, \sigma^2)$, mit unbekanntem σ^2. Dieser Wert muss aus der Stichprobe mit s^2 geschätzt werden.

- Hier verwenden wir natürlich alle 16 Beobachtungen.

- Wir wollen also für den Erwartungswert μ testen:

$$H_0\colon\ \mu \leq 5$$
$$H_1\colon\ \mu > 5$$

- Wir verwenden die Teststatistik

$$t = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s}$$

- Mit $\overline{x} = 16.375$ und $s^2 = 441.05$ erhalten wir für den Wert der Teststatistik:

$$t = \frac{\sqrt{16}(16.375 - 5)}{\sqrt{441.05}} = 2.1665$$

- Wir lehnen H_0 ab, wenn hier der Wert der Teststatistik *zu groß* ist. D.h., wenn die Teststatistik größer als der kritische Wert $t_{15;0.95} = 1.7531$ aus der Tabelle ist.

- Hier lehnen wir H_0 ab und gehen davon aus, dass die Schadenshöhen signifikant größer sind als 5 Million EUR pro Monat.

- Der t-Test liefert hier ein anderes Ergebnis als der Vorzeichentest für eigentlich die gleiche Fragestellung. Das liegt an einigen sehr großen Werten in der Stichprobe. Diese großen Werte werden vom t-Test z.B. im Mittelwert mit verwendet. Der Vorzeichentest zählt aber nur, ob eine Beobachtung größer ist als 5 und nicht, wie viel sie größer ist als 5.

Aufgabe 3

Bei seinen weltberühmten Experimenten zur Vererbungslehre erhielt Mendel bei einem seiner Kreuzungsversuche von Erbsenpflanzen folgende Werte:

- 315 runde gelbe Erbsen,
- 108 runde grüne Erbsen,
- 101 kantige gelbe Erbsen,
- 32 kantige grüne Erbsen.

Sprechen diese Beobachtungen für oder gegen die Theorie, daß das Verhältnis der 4 Sorten 9:3:3:1 sein müsste ($\alpha = 0.05$)?

Sei X der Ausgang des Kreuzungsexperiments mit

$$X = \begin{cases} 1, & \text{falls rund und gelb} \\ 2, & \text{falls rund und grün} \\ 3, & \text{falls kantig und gelb} \\ 4, & \text{falls kantig und grün} \end{cases}$$

Die hypothetischen Wahrscheinlichkeiten sollen im Verhältnis $9:3:3:1$ stehen, d.h.

$$p_1 = \frac{9}{16}, \quad p_2 = \frac{3}{16}, \quad p_3 = \frac{3}{16}, \quad p_4 = \frac{1}{16}$$

Zu testen ist

$$H_0 : P(X = i) = p_i \ \text{ für } i = 1, 2, 3, 4$$

gegen

$$H_1 : P(X = i) \neq p_i \ \text{ für mindestens ein } i = 1, 2, 3, 4.$$

Als Teststatistik verwenden wir:

$$X^2 = \sum_{i=1}^{4} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

Dabei bezeichnen n_i die absoluten beobachteten Häufigkeiten und $n = 556$ den Stichprobenumfang. Es ergibt sich:

n_i	$n \cdot p_i$	$n_i - n \cdot p_i$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
315	312.75	2.25	0.0162
108	104.25	3.75	0.1349
101	104.25	-3.25	0.1013
32	34.75	-2.75	0.2176

Es ergibt sich somit:

$$X^2 = 0.0162 + 0.1349 + 0.1013 + 0.2176 = 0.47$$

Unter H_0 ist X^2 approximativ χ^2-verteilt mit hier 3 Freiheitsgraden, d.h. H_0 wird abgelehnt, falls $X^2 > \chi^2_{3,0.95} = 7.815$.

Hier wird H_0 somit nicht abgelehnt und wir können von dem Verhältnis 9:3:3:1 ausgehen.

Aufgabe 4

Ein Statistiker sieht sich neun Folgen der Show WER WIRD MILLIONÄR an und notiert sich am Ende der Sendung den realisierten Gesamtgewinn des Tages. Es ergaben sich folgende Werte (in Tausend EUR):

`34 17 96 33 189 282 33 66 64`

Der Statistiker will überprüfen, ob der erwartete Gewinn eines Tages mehr als 100000 EUR beträgt. Führen Sie einen geeigneten Test zum Niveau $\alpha = 0.05$ durch. Gehen Sie von Normalverteilung aus.

Einstichprobenproblem, t-Test auf Grund der angenommenen Normalverteilung

1. $H_0 : \mu \leq 100$ gegen $H_1 : \mu > 100$

2. t-Test mit der Teststatistik $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s}$

3. t ist unter H_0 t-verteilt mit $n-1$ Freiheitsgraden.

   Tabellenwert für den einseitigen Test: $t_{8,0.95} = 1.8595$

4. Wir lehnen H_0 ab, wenn $t > t_{8,0.95}$

5. Es gilt $n = 9$ und $\bar{x} = 90.44$ und wir erhalten $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}) = \ldots = 7864.28$ und somit $s = 88.68$

   Wir erhalten $t = \frac{\sqrt{9}\cdot(90.44-100)}{88.68} = -0.32$

6. Da $-0.32 \not> 1.8595$, lehnen wir H_0 nicht ab.

7. Wir können also nicht mehr als 100 000 EUR pro Tag bei "Wer wird Millionär" erwarten.

Aufgabe 5

Sie möchten nachweisen, dass der Anteil an unzufriedenen Kunden beim Unternehmen SPEEDTEL geringer als 10% ist. Dazu fragen Sie acht Personen, von denen zwei unzufrieden sind. Bestimmen Sie die Überschreitungswahrscheinlichkeit.

Dieser Test ist mit dem Vorzeichentest vergleichbar. Es handelt sich um den allgemeinen Binomialtest.

1. $H_0 : p \geq 0.1$ gegen $H_1 : p < 0.1$

2. Binomialtest mit $S =$ Anzahl der unzufriedenen Kunden.

3. S ist unter H_0 Binomial-verteilt mit $n = 8$ und $\mathbf{p = 0.1}$

4. Wir lehne H_0 ab, falls der p-Wert (Überschreitungswahrscheinlichkeit) kleiner oder gleich $\alpha = 0.05$ ist.

5. Wir haben $S = 2$ unzufriedene Kunden in der Stichprobe vom Umfang $n = 8$.

   Wir berechnen $P(S = s|H_0) = \binom{8}{s} \cdot 0.1^s \cdot 0.9^{8-s}$

   Um die Überschreitungswahrscheinlichkeit auszurechnen nutzen wir den gefundenen Wert $S = 2$ und Werte die noch mehr für H_1 sprechen würden. Das sind hier die Werte $S = 1$ und $S = 0$.

   $$P(S = 2|H_0) + P(S = 1|H_0) + P(S = 0|H_0) = 0.43 + 0.38 + 0.15 = 0.96$$

6. Die Überschreitungswahrscheinlichkeit ist nicht kleiner als 0.05. Somit lehnen wir H_0 hier nicht ab.

7. Wir können somit nicht annehmen, dass die Wahrscheinlichkeit für einen unzufriedenen Kunden kleiner als 0.1 ist.