DHBW Mannheim
Informatik | Mannheim, 07.06.2023
Mathematik II/2

Lösungen zu Übungsblatt 10

Aufgabe 1. Eine Verbraucherzentrale möchte überprüfen, ob ein bestimmtes Milchprodukt Übelkeit bei den Konsumenten auslöst. In einer Studie mit zehn Personen wird bei sieben Personen nach dem Genuss dieses Milchprodukts eine auftretende Übelkeit registriert. Überprüfen Sie zum Signifikanzniveau $\alpha = 0.05$ die statistische Nullhypothese, dass der Anteil der Personen mit Übelkeitssymptomen nach dem Genuß dieses Produkts in der Grundgesamtheit höchstens 60% beträgt. Geben Sie zunächst das zugehörige statistische Testproblem an.

Lösung:

$$H_0 : p \leq 0.6 \text{ gegen } H_1 : p > 0.6.$$

Wenn H_0 verworfen wird, ist folgende Aussage der Verbraucherzentrale zulässig: "Wir haben herausgefunden, dass das Milchprodukt mit einer Sicherheitswahrscheinlichkeit von $1-\alpha$ Übelkeit hervorruft."Bei der Wahl eines geeigneten Tests und seiner Durchführung sind folgende Aspekte zu beachten:

- Das Merkmal (Übelkeit; Ja/Nein) ist binär,
- die Hypothese ist über einen Anteil formuliert, d.h. es ist der Binomialtest zu wählen und zwar der exakte,
- die Prüfgröße ist somit die Anzahl der Personen mit Übelkeit, kurz bezeichnet mit $\sum X_i$, wobei gilt: $\sum X_i \sim B(10, 0.6)$,
- der Ablehnungsbereich ist durch "große"Werte von X_i und $\alpha = 0.05$ festgelegt. Bei der Bestimmung des kritischen Werts nutzt man aus, dass für $p > 0.5$ gilt:

  $$B(x|n,p) = P(X \leq x|n,p) = 1 - B(n-x-1|n, 1-p),$$

  d.h. man erhält hier $B(x|10, 0.6) = 1 - B(10-x-1|10, 0.4)$.

Gesucht ist nun x, so dass

$$\begin{aligned} P(X > x|0.6) &\leq 0.05 \text{ und} \\ P(X \geq x-1|0.6) &> 0.05. \end{aligned}$$

Da

$$\begin{aligned} P(X \geq x|0.6) &= 1 - P(X < x|0.6) = 1 - P(X \leq x-1|0.6) \\ &= 1 - [1 - B(10-(x-1)-1|10, 0.4)] \\ &= B(10-x|10, 0.4), \end{aligned}$$

gilt:

$$\begin{aligned} P(X \geq 10|0.6) &= B(0|10, 0.4) = 0.006 && < 0.05, \\ P(X \geq 9|0.6) &= B(1|10, 0.4) = 0.0464 && < 0.05, \\ P(X \geq 8|0.6) &= B(2|10, 0.4) = 0.1673 && > 0.05. \end{aligned}$$

Damit ist 9 der kritische Wert, woraus sich der Ablehnungsbereich $C = \{9, 10\}$ ergibt. Also kann erst bei neun oder zehn Personen mit Übelkeit in einer Stichprobe vom Umfang zehn die Nullhypothese zum Niveau $\alpha = 0.05$ verworfen werden, d.h. diese Werte sind zu „unwahrscheinlich“, wenn H_0 wahr wäre.
Da in diesem Beispiel nur sieben Personen Übelkeitssymptome aufweisen, kann H_0 nicht verworfen werden, d.h. es kann also nicht entschieden werden, dass das Milchprodukt Übelkeit auslöst.

Aufgabe 2. Ein Unternehmen organisiert Fahrten (vor allem für Familien) auf einem Ausflugsschiff, das Platz für 60 Personen bietet. Um an der Fahrt teilzunehmen, muss man im Voraus eine Reservierung vornehmen, ohne dabei schon den Fahrpreis bezahlen zu müssen. Erfahrungsgemäß erscheinen von den Personen mit Reservierung einige nicht zur Fahrt. Die Zufallsgröße X beschreibt die Anzahl der Personen mit Reservierung, die nicht zur Fahrt antreten.
Vereinfachend soll angenommen werden, dass die Zufallsgröße X binomialverteilt mit Parametern n und p ist, wobei n die Anzahl der eingegangenen Reservierungen ist und p die Wahrscheinlichkeit dafür, dass eine Person mit Reservierung nicht zur Fahrt erscheint. Es wird vermutet, dass die Wahrscheinlichkeit dafür, dass eine Person mit Reservierung die Fahrt nicht antritt, bei 10 % liegt. Um das zu überprüfen, untersucht das Unternehmen die Nullhypothese „Die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Person mit Reservierung nicht zur Fahrt erscheint, ist höchstens 10 %“, die mithilfe einer Zufallsstichprobe von 80 Reservierungen bei einer maximale Fehlerwahrscheinlichkeit von 5 % getestet werden soll.

a) Erläutern Sie, warum es sich bei der Annahme, dass X binomialverteilt ist, im Sachzusammenhang um eine Vereinfachung handelt.

   Lösung:

   Die Annahme, dass X binomialverteilt ist, setzt voraus, dass das Verhalten der einzelnen Reservierenden, ob sie die Fahrt antreten oder nicht, unabhängig voneinander ist. Da sich das Angebot aber speziell auch an Familien richtet und bei Familien davon ausgegangen werden muss, dass die gesamte Familie nicht zur Fahrt antritt, wenn ein Familienmitglied nicht kommt, ist diese Unabhängigkeitsannahme in diesem Kontext eine Vereinfachung.

Etwas präziser und mathematischer:

Bezeichnen wir mit Y_i die Zufallsvariable, die das Nichtantreten der Fahrt bei der i–ten Reservierung beschreibt, also die Zufallsvariable

$$Y_i(\omega) = \begin{cases} 1 & \text{falls } \omega = \text{die } i\text{–te Reservierung tritt nicht an} \\ 0 & \text{sonst} \end{cases}$$

so ist Y_i eine Bernoullivariable mit Parameter p und $X = Y_1 + Y_2 + \cdots + Y_n$. Damit X tatsächlich binomialverteilt mit Parametern n und p ist, müssten die Y_i stochastisch unabhängig sein, was aber in der Regel nicht der Fall sein wird, etwa wenn Reservierungen von einer Familie (oder generell einer Gruppe) kommen.

b) Beschreiben Sie das komplette Testszenario des Hypothesentests (einschließlich aller Zufallsvariablen) und bestimmen Sie die Entscheidungsregeln, die zur Ablehnung der Nullhypothese führen.

Lösung:

Wir führen einen Binomialtest durch (wir arbeiten also, wie angegeben, mit der vereinfachenden Annahme, dass X binomialverteilt ist). Dazu bezeichnen wir mit p die (tatsächliche) Wahrscheinlichkeit dafür, dass eine Reservierung nicht angetreten wird und stellen die folgenden Hypothesen auf.

Nullhypothese H_0: $p = 0.1$.

Alternativhypothese H_1: $p > 0.1$.

Da $80 \cdot 0.1 \cdot (1 - 0.1) = 7.2 < 9$, ist ein exakter Binomialtest durchzuführen.

Ist die Nullhypothese richtig, so ist $X \sim B(80, 0.1)$, und daher wird die Nullhypothese dann abgelehnt, wenn die Realisation x von X eine Ergebnis $x \geq l$ liefert, wobei l so zu wählen ist, dass

$$p(B(80,\, 0.1) \geq l - 1) > 0.05, \qquad p(B(80,\, 0.1) \geq l) \leq 0.05$$

Die Werte der $B(80, 0.1)$–Verteilung finden sich nicht in den üblichen Tabellen, sind also aus den Formeln

$$p(B(80, 0.1) = k) = \binom{80}{k} \cdot 0.1^k \cdot 0.9^{80-k}$$

sowie

$$p(B(80, 0.1) \leq k) = \sum_{i=0}^{k} p(B(80, 0.1) = i)$$

und

$$p(B(80,0.1) \geq k) = 1 - p(B(80,0.1) \leq k-1)$$

direkt zu berechnen.

Für eine Zufallsvariable $B \sim B\,(80, 0.1)$ erhalten wir

k	9	10	11	12	13	14	15
$p(B \leq k)$	0.7234	0.8266	0.8996	0.9462	0.9733	0.9877	0.9947
$p(B \geq k)$	0.4073	0.2766	0.17338	0.1004	0.0538	0.0267	0.0123

Damit gilt also

$$p(B(80,0.1) \geq 13) > 0.05, \qquad p(B(80.0.1) \geq 14) \leq 0.05$$

und daher wird die Nullhypothese abgelehnt, wenn mindestens 14 Personen aus der Stichprobe die Fahrt nicht antreten.

c) Wie ändert sich Ihr Vorgehen, wenn eine Zufallsstichprobe von 270 Reservierungen zum Test der Nullhyptothese herangezogen wird? Beschreiben Sie die Unterschiede in der Herangehensweise und die Regeln die in dieser Situation zur Ablehnung der Nullhypothese führen.

Lösung: In diesem Fall ist $270 \cdot 0.1. \cdot (1 - 0.1) = 24.3 > 9$, und daher kann ein approximativer Binomialtest durchgeführt werden. Die Hypothesen ändern sich nicht, und auch nicht die Zufallsvariablen Y_i (nur jetzt mit $i = 1, \ldots, 270$). Ferner betrachten wir

$$X = \sum_{i=1}^{270} Y_i, \qquad \overline{X} = \frac{1}{270} \cdot X$$

und als Entscheidungsvariable

$$Z = \frac{\overline{X} - 0.10}{\sqrt{0.1 \cdot 0.9}} \cdot \sqrt{270} = 10 \cdot \left(\overline{X} - 0.10\right) \cdot \sqrt{30}$$

Ist H_0 korrekt, so ist Z standardnormalverteilt, und wir lehnen H_0 (bei der gegebenen Fehlerwahrscheinlichkeit von 5 %) ab, wenn für die Realisation z von Z gilt:

$$z > 1.65$$

d) Entscheiden Sie, ob bei der Wahl der Nullhypothese eher das Interesse, dass wenige Plätze frei bleiben oder das Interesse, dass wenige Personen mit Reservierung abgewiesen werden müssen, im Vordergrund stand.

Lösung:

Bei einem Hypothesentest wird immer gegen die Nullhypothese getestet, es werden also Argumente gesammelt, die gegen die Nullhypothese sprechen, und die Nullhypothese wird erst dann abgelehnt, wenn es sehr starke statistische Evidenz gibt, die gegen die Nullhypothese sprechen.

Im Beispiel postuliert die Nullhypothese eine geringe Wahrscheinlichkeit für das Nichtantreten einer Fahrt trotz Reservierung ($p \leq 0.1$). Gegen diese Hypothese wird nur dann entschieden, wenn die Daten deutlich für eine höhere Ausfallrate sprechen (vergleiche Teil b): die Nullhypothese wird erst abgelehnt, wenn mindesten 14 von 80 Personen mit Reservierung nicht zur Fahrt antreten, was deutlich mehr ist als die eigentlich zu erwartenden 10 %), und entsprechend wird eine höhere Nichtantrittsrate nur dann in Betracht gezogen, wenn die Argumente deutlich gegen niedrige Raten sprechen. Im Zweifel wird weiter mit $p \leq 0.1$ gearbeitet, und daher steht im Zweifel das Interesse der Kunden, nicht abgewiesen zu werden, im Vordergrund.

e) Beschreiben Sie den Fehler zweiter Art bei diesem Test sowie die daraus resultierenden Konsequenzen im Sachzusammenhang.

Lösung:

Beim Fehler zweiter Art oder β–Fehler wird die Nullhypothese nicht abgelehnt, obwohl Sie falsch ist. In dieser Situation bedeutet das also, dass mit $p \leq 0.1$ gearbeitet wird, obwohl die Wahrscheinlichkeit dafür, dass eine Person mit Reservierung die Fahrt nicht antritt, tatsächlich höher ist. Die Konsequenz davon ist, dass tendenziell weniger Reservierungen angenommen werden und damit mehr Plätze frei bleiben als nötig wäre.