

Übungsblatt 12

Aufgabe P1 *Schraubenlänge.*

Ein Großkunde kauft Schrauben direkt vom Hersteller. Der Hersteller gibt an, dass die Länge der Schrauben produktionsbedingt schwankt und die Verteilung der Längen einer Normalverteilung mit Erwartungswert 20 mm und Varianz 0.01 mm^2 entspricht. Der Kunde hat den Verdacht, dass die gelieferten Schrauben nicht diesen Erwartungswert besitzen. Daher misst er bei 100 Schrauben die Länge exakt nach und erhält einen Mittelwert von 20.026 mm. Kann die Hypothese, dass die durchschnittliche Länge der Schrauben 20 mm beträgt, aufgrund dieser Stichprobe verworfen werden, wenn er eine Fehlerwahrscheinlichkeit von 1% zulässt und

- (ii) der angegebenen Varianz nicht glaubt, aber als Stichprobenvarianz $s^2 = 0.01 \text{ mm}^2$ bestimmt hat?

Aufgabe P2 *Losverkäufer.*

Ein Losverkäufer behauptet in seiner Lostrommel sei mindestens jedes 4. Los ein Gewinn. Eine Studentin kauft daraufhin 20 Lose, unter denen 2 Gewinne sind. Ist damit die Behauptung des Losverkäufers statistisch widerlegt, wenn wir ein Signifikanzniveau von 5% fordern?

Aufgabe H1 Tests für Anteilswerte.

- a) Bei Professor Satanis liegt die Durchfallquote der Studenten bei 20%. Bei einem neuen Professor Hilfikus, der im nächsten Semester die gleiche Vorlesung hält, wird die Durchfallquote von einigen Studenten herbeigefühlterweise deutlich niedriger eingeschätzt. Die Fachschaft bekommt zur Auswertung 640 zufällig ausgewählte Prüfungen des Professors Hilfikus zugespielt und kommt bei einer Analyse auf 152 durchgefallene Studenten. Lohnt es sich wirklich, sein Studium möglicherweise unnötig zu verzögern und bei Professor Hilfikus die Klausur zu schreiben? Benutzen Sie den approximativen Binomialtest mit einem Signifikanzniveau von 1%.
- b) Ein Großhändler hat den Verdacht, dass die ihm gelieferte Ware eine Ausschussquote hat, die größer als 3% ist. In einer Stichprobe vom Umfang 60 findet er 3 defekte Stücke. Ist sein Verdacht bei einem Signifikanzniveau von 2% gerechtfertigt, wenn der Ablehnungsbereich approximativ durch die Poisson-Approximation bestimmt wird?
Hinweis: Gehen Sie zunächst wie folgt vor: Nehmen Sie gemäß der Poissonapproximation approximativ an, dass die eigentlich binomialverteilte Teststatistik mit den entsprechenden Parametern Poissonverteilt ist. Bestimmen Sie dann für diesen Fall einen Ablehnungsbereich gemäß der Poissonverteilung.

Lösung: Insgesamt fünf Punkte.

Aufgabenteil (i) entlehnt aus Aufgabensammlung zur Statistik von Schwarze, Lösung erstellt von Alexander Klump 01.2022

(i) Die Aufgabenstellung suggeriert, dass wir anzweifeln, dass es sich lohnt. Wir testen also $H_0 : p \leq p_0 = 0,2$ gegen $H_1 : p > p_0$ (1P). Unsere Teststatistik ist wie in der Vorlesung für den approximativen Binomialtest angegeben

$$Z = \frac{Y - np_0}{\sqrt{np_0(1 - p_0)}}$$

wobei Y die Anzahl der durchgefallenen Studenten unter n Prüfungen ist (0.5P). Hier gilt $\alpha = 0.01$, $n = 640$ und $Y = 152$. (0.5P) Unser Ablehnungskriterium ist also

$$2.37171 = Z > z_{1-\alpha} = z_{0,99} = 2.3264.$$

Die Nullhypothese H_0 sollte also verworfen werden (0.5P). Es lohnt sich nicht.

Aufgabenteil (ii) aus Aufgabensammlung zur Statistik von Schwarze, Lösung erstellt von Alexander Klump 01.2022

(ii) Wir gehen zunächst wie im klassischen Binomialtest vor: Wir bezweifeln, dass die Ausschussquote p kleiner als 3% ist, also testen wir $H_0 : p \leq p_0 = 0,03$ gegen $H_1 : p > 0,03$ (0.5P). Die Teststatistik ist $X = \text{Anzahl der defekten Stücke unter den untersuchten Stücken}$. Aus der Vorlesung ergäbe sich das Ablehnungskriterium

$$X \geq l,$$

wobei l so gewählt ist, dass $\mathbb{P}_{p_0}(X \geq l) \leq \alpha$, mit $X \sim \text{Bin}(n, p_0)$. Hier ist nun $\alpha = 0,02$, $n = 60$ und die Stichprobe ergibt $X = 3$. (0.5P)

Abgesehen davon, dass $n = 60$ nicht mehr in der Binomialtabelle berücksichtigt ist, bietet es sich

an die Poissonapproximation zu verwenden, da p in unserer Nullhypothese als relativ klein angenommen wird. Entsprechend der Poissonapproximation nehmen wir also an, dass X approximativ Poissonverteilt ist mit Parameter $\lambda = np_0 = 1,8$. Dies bedeutet

$$\mathbb{P}_{p_0}(X \leq l) \approx 1 - \sum_{i=0}^{l-1} e^{-\lambda} \frac{\lambda^i}{i!}.$$

Wir suchen also das kleinste l sodass

$$1 - \sum_{i=0}^{l-1} e^{-\lambda} \frac{\lambda^i}{i!} \leq 0,02$$

oder äquivalent dazu

$$\sum_{i=0}^{l-1} e^{-\lambda} \frac{\lambda^i}{i!} \geq 0,98$$

Aus der Tabelle zur Poissonverteilung lesen wir heraus, dass $l = 6$. (1P) Da $X = 3 < 6 = l$ ist können wir H_0 nicht verwerfen. (0.5P)

Aufgabe H2 BSE.

Zur Erforschung der Übertragbarkeit der Krankheit BSE (bovine spongiforme Enzephalopathie) wurden in einem Tierversuch n biologisch gleichartigen Mäusen über einen gewissen Zeitraum täglich eine bestimmte Menge Milch von BSE-kranken Kühen verabreicht. Innerhalb dieses Zeitraums entwickelte keine dieser Mäuse irgendwelche klinischen Symptome, die auf eine BSE-Erkrankung hindeuten könnten.

Es bezeichne p die Wahrscheinlichkeit, dass eine Maus der untersuchten Art unter genau den obigen Versuchsbedingungen innerhalb des Untersuchungszeitraumes BSE-spezifische Symptome aufweist.

- a) Sei $X \sim \text{Bin}(n, p)$ mit bekanntem $n \in \mathbb{N}$ und unbekanntem $p \in [0, 1]$ und sei $\alpha \in (0, 1)$ gegeben. Ferner definieren wir für alle $k \in \{0, \dots, n\}$ die Zahl $p_o(k)$ als eindeutige Lösung der Gleichung $\sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} = \alpha$ für p . Zeigen Sie, dass $[0, p_o(X)]$ ein einseitiger Konfidenzbereich für p zum Konfidenzniveau $1 - \alpha$ ist. (Tipp: Führen Sie die Menge $M := \{k \in \{0, 1, \dots, n\} : p > p_o(k)\}$ ein.)

Hinweis: Sie dürfen verwenden, dass für alle $k \in \{0, \dots, n-1\}$ die Funktion $(0, 1) \ni p \mapsto \sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j}$ monoton fallend ist.

- b) Konstruieren Sie unter Verwendung von Teilaufgabe a) für die obige Situation ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha \in (0, 1)$ für die unbekannte Wahrscheinlichkeit p in der Form $K = [0, p_o(X)]$ (Warum ist hier die untere Grenze 0 sinnvoll?).
- c) Bestimmen Sie K in b) für $n = 275$ und $\alpha = 0.01$.
- d) Wie viele Mäuse müssen mindestens untersucht werden, damit $p_o(X) \leq 10^{-4}$ gilt, wenn man $\alpha = 0.01$ veranschlagt?

- e) Angenommen der wahre Parameter p läge tatsächlich in dem in b) berechneten Konfidenzintervall K . Wie viele BSE erkrankte Mäuse würde man dann unter 10^7 Mäusen höchstens erwarten?

Lösung: a) Es seien $M := \{k \in \{0, 1, \dots, n\} : p > p_o(k)\}$ und $k_o := \max M$ die größte Realisierung der Trefferzahl X , für die $p > p_o(k)$ gilt. Es gilt

$$\mathbb{P}_p(p \leq p_o(X)) = 1 - \mathbb{P}_p(p > p_o(X)) = 1 - \mathbb{P}_p(X \in M).$$

Weiter gilt $1 - \mathbb{P}_p(X \in M) \geq 1 - \alpha$ genau dann, wenn $\mathbb{P}_p(X \in M) \leq \alpha$. Nun gilt

$$\mathbb{P}_p(X \in M) \leq \mathbb{P}_p(X \leq k_o) \leq \mathbb{P}_{p_o(k_o)}(X \leq k_o) = \sum_{j=0}^{k_o} \binom{n}{j} p_o(k_o)^j (1 - p_o(k_o))^{n-j} = \alpha,$$

wobei wir bei der zweiten Ungleichheit den Hinweis verwendet haben. [2 Pkt]

b) Kodieren die Zufallsvariablen X_i , ob Maus i BSE-Symptome zeigte ($X_i = 1$) oder nicht ($X_i = 0$) und betrachten wir $X = \sum_{i=1}^n X_i$, so ist X , da wir annehmen können, dass die X_i unabhängig sind, binomialverteilt mit Parametern n und p , p unbekannt. Wir können also die Konstruktion aus a) verwenden. [1 Pkt] Nach a) wählen wir $p_o := p_o(0)$ (hier ist $k = 0$, da keine Maus infiziert wurde) als eindeutige Lösung der Gleichung $(1 - p_o)^n = \alpha$, was $p_o = 1 - \alpha^{1/n}$ liefert. [1 Pkt]

c) Einsetzen der Werte liefert $K = [0, 0.01661]$. [5 Pkt]

d) Lösen von $1 - 0.01^{1/n} \leq 10^{-4}$ liefert $n \geq 46050$. [5 Pkt]

e) Die unter diesen Voraussetzungen höchste Anzahl der erwarteten erkrankten Mäuse ist $10^7 \cdot 0.01661 = 166100$. [1 Pkt]

Aufgabe H3 Flaschenabfüllung.

Bei einer Flaschenabfüllmaschine einer oberbayerischen Brauerei wird regelmäßig überprüft, ob die mittlere Füllung 500 cm^3 beträgt. Anhand der Messung des Inhalts von zufällig herausgegriffenen Flaschen ist zu entscheiden, ob die Maschine neu eingestellt werden muss. Frühere Beobachtungen lassen die Annahme zu, dass die Zufallsvariablen X_1, \dots, X_n (zur Beschreibung der Messergebnisse) unabhängig und identisch $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind mit unbekanntem Erwartungswert $\mu \in \mathbb{R}$ und unbekannter Varianz $\sigma^2 \in \mathbb{R}^+$. Die Vermessung von 20 Flaschen liefert die folgenden (auf volle cm^3 gerundeten) Füllmengen

496, 500, 498, 497, 504, 493, 491, 501, 502, 502,
498, 498, 502, 497, 497, 494, 496, 500, 499, 502.

Muss die Maschine neu eingestellt werden, wenn als Fehler $\alpha = 0.05$ zugelassen wird?

Lösung: Insgesamt fünf Punkte.

Es soll die Nullhypothese überprüft werden, dass $\mu_0 = 500$ gilt. Also

$$H_0 : \mu = 500, \quad H_1 : \mu \neq 500.$$

(1P)

Da die Varianz unbekannt ist, verwenden wir den t -Test (0.5 P.). Dieser sagt uns, dass wir die Nullhypothese ablehnen können, falls

$$\left| \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} \right| > t_{(n-1; 1-\alpha/2)}$$

gilt. (0.5P) Hier ist $\alpha = 0.05$ und $n = 20$. (0.5P)
Aus dem Datensatz bestimmen wir

$$\bar{x} = \frac{1}{20} \cdot \sum_{i=1}^{20} x_i = 498.35 \quad \text{und} \quad s^2 = \frac{1}{19} \cdot \sum_{i=1}^{20} (x_i - 498.35)^2 = 11.40.$$

(1P) Damit gilt

$$\left| \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} \right| = 2.19 > 2.0930 = t_{(19;0.975)}.$$

(1P) Also muss die Maschine neu eingestellt werden, d.h. die Nullhypothese kann verworfen werden.
(0.5)

Abgabe der Hausübungen: Mittwoch, 26. Januar, 13:00 Uhr

Viel Erfolg! :)