Wahrscheinlichkeit und Statistik

Musterlösung Serie 14

1. Die Zeitschrift "Gemüsetest" testet den Wahrheitsgehalt der folgenden Werbeaussagen:

   i. Gemüsehändler Hase behauptet seine Karotten seien im Durchschnitt mindestens $30cm$ lang.

   ii. Gleichzeitig preist er seine Kartoffeln der Sorte "Pellworm" als ideale Pellkartoffeln an. Sie seien mit einem durchschnittlichen Gewicht von $50g$ weder zu dick noch zu dünn.

   iii. Konservenfabrikant Hamster wirbt für seine extrazarten jungen Erbsen mit der Garantie, die durchschnittliche Dicke der jungen Erbsen betrage höchstens $3mm$.

   iv. Er behauptet auch, dass der Anteil von holzigen Spargeln in seinen Konserven unter 0.3% liege.

   v. Ausserdem lobt er die ausgewogene Mischung seiner "Erbsen mit Karotten", die genau 40% Gewichtsanteil betrage.

   a) Geben Sie jeweils an, ob ein linksseitiger, ein rechtsseitiger oder ein zweiseitiger Test nötig ist, um den Wahrheitsgehalt[1] der Aussagen zu testen, und stellen jeweils Nullhypothese und Alternativhypothese auf.

   b) Die Karotten von Hase sind tatsächlich durchschnittlich 30 cm lang. In ihrem Test kommt die Zeitschrift jedoch zu dem Ergebnis, dass die Werbeaussage falsch sei. Was ist passiert? (Fehler 1. Art oder 2. Art?)

   c) Der tatsächliche Anteil an holzigen Spargeln liegt bei 2%, trotzdem akzeptiert die Zeitschrift nach ihrem Test die Aussage von Fabrikant Hamster. Warum? (Fehler 1.Art oder 2.Art?)

Lösung:

a) i. Sei X die Zufallsvariable, welche die Länge der Karotten modelliert. Getestet wird auf den Parameter $\mu = \mathbb{E}[X]$, die durchschnittliche Länge der Karotten. Die Nullhypothese ist die Behauptung des Gemüsehändlers, d.h. $H_0\colon \mu \geq \mu_0$ mit $\mu_0 = 30cm$. Wir wollen testen, ob sie sich widerlegen lässt. Daraus ergibt sich ein linksseitiger Test mit den Hypothesen

$$H_0\colon \mu \geq \mu_0 \quad \text{und} \quad H_A\colon \mu < \mu_0.$$

   ii. Zweiseitiger Test mit $\mu_0 = 50g$ sowie den Hypothesen $H_0\colon \mu = \mu_0$ und $H_A\colon \mu \neq \mu_0$.

   iii. Rechtsseitiger Test mit $\mu_0 = 3mm$ sowie den Hypothesen $H_0\colon \mu \leq \mu_0$ und $H_A\colon \mu > \mu_0$.

   iv. Hier modelliert die Zufallsvariable X mit Wertebereich $W_X = [0,1]$ den Anteil an holzigen Spargel. Getestet wird wieder auf den durchschnittlichen Wert davon, also auf $\mu = \mathbb{E}[X]$. Aus der Behauptung des Gemüsehändlers ergibt sich ein rechtsseitiger Test mit $\mu_0 = 0.003$ sowie den Hypothesen $H_0\colon \mu \leq \mu_0$ und $H_A\colon \mu > \mu_0$.

   v. Zweiseitiger Test mit $\mu_0 = 0.4$ sowie den Hypothesen $H_0\colon \mu = \mu_0$ und $H_A\colon \mu \neq \mu_0$.

b) Die Nullhypothese wurde verworfen obwohl sie stimmt, d.h. Fehler 1. Art.

[1] Genauer geht es darum die Aussagen zu widerlegen, gelingt dies nicht bleibt uns wohl nichts anders übrig als die Aussage zu akzeptieren, womit noch lange nicht belegt ist, dass sie wahr sind.

c) Die Nullhypothese wurde nicht verworfen obwohl sie falsch ist, d.h. Fehler 2. Art.

Bemerkung: Da wir ehrliche Gemüsehändler nicht ungerechtfertigt in Misskredit bringen wollen, müssen wir die Wahrscheinlichkeit für einen Fehler 1. Art gering halten, d.h. wenn wir seine Behauptung anzweifeln, dann müssen wir einen guten Grund haben dafür, und dieser ist die geringe Wahrscheinlichkeit für eine bestimmte Gesamtheit von Testausgängen falls die Behauptung stimmt (Fehler 1. Art). Da es im Allgemeinen nicht gelingt beide Fehler klein zu halten, bedeutet ein Akzeptieren der Nullhypothese nicht, dass die Behauptung stimmt, sie konnte lediglich nicht widerlegt werden. Der Gemüsehändler ist also auf der sicheren Seite, sagt er die Wahrheit, dann hat er kaum was zu befürchten (Fehler 1. Art), und wenn er lügt, dann hat er in bestimmten Fällen sogar ein gute Chance (Fehler 2. Art) nicht entlarvt zu werden.

2. Die Firma *Runners* hat ein Sportgetränk für Marathonläufer entwickelt. Mit einem statistischen Test will sie nun untersuchen, ob das Getränk einen positiven Einfluss auf die Leistung hat (gemessen an der benötigten Zeit für den Lauf). Ein ausgewählter Läufer rennt 8-mal in einem Monat eine 15-km Probestrecke: 4-mal nach Einnahme des Getränkes (Laufzeiten $x_1, \ldots, x_n$) und 4-mal nach Einnahme eines Placebo-Getränks, das gleich schmeckt und aussieht (Laufzeiten $y_1, \ldots, y_n$). Die Reihenfolge der Verabreichung ist zufällig und dem Läufer nicht bekannt. Übertragungseffekte sind ausgeschlossen, da der Läufer maximal einmal an einem Tag die Strecke rennt.

   a) Handelt es sich um eine gepaarte oder um eine ungepaarte Stichprobe?

   b) Wie muss der zugehörige Test durchgeführt werden: einseitig oder zweiseitig?

   c) Welche der folgenden Aussagen ist die korrekte Alternativhypothese?

      i. Das Getränk hat eine Wirkung.
      ii. Das Getränk hat keine Wirkung.
      iii. Das Getränk bewirkt eine bessere Leistung.
      iv. Das Getränk bewirkt eine schlechtere Leistung.

   d) Geben Sie an, ob die folgenden Aussagen richtig oder falsch sind.

      i. Die Nullhypothese darf auf dem 1%-Niveau beibehalten werden, wenn der P-Wert des Tests 0.034 beträgt.
      ii. Der Verwerfungsbereich wird grösser, wenn das Niveau verkleinert wird.
      iii. Ein Fehler 1. Art kann nur dann eintreten, wenn die Teststatistik im Verwerfungsbereich liegt.
      iv. Die Voraussetzungen der Wilcoxon-Tests sind strenger, als diejenigen vom t-Test.
      v. Ein zweiseitiges 99%-Vertrauensintervall umfasst immer das entsprechende zweiseitige 95%-Vertrauensintervall.
      vi. Angenommen, die Hypothese $\mu = 0$ wird zweiseitig getestet mit dem t-Test. Die Wahrscheinlichkeit, die Nullhypothese beizubehalten, obwohl sie falsch ist, ist grösser unter der Alternativhypothese $\mu = 0.15$ als unter $\mu = 0.20$.

Lösung:

   a) Die Stichprobe ist ungepaart.

   b) Da getestet werden soll, ob das Getränk eine *positive* Wirkung hat, wird einseitig getestet.

c) Die Alternativhypothese, die von den vier Möglichkeiten am besten passt ist in dem Fall die Nummer iii., d.h. *Das Getränk bewirkt eine bessere Leistung.*
Bemerkung: Die Nullhypothese lautet, dass die Leistung nach Einnahme des Getränkes genausogut oder schlechter ist.

d)
   i. Richtig: Der P-Wert ist das kleinste Niveau, auf dem die Nullhypothese verworfen wird, und $0.01 < 0.034$.
   ii. Falsch: Das Niveau gibt die Wahrscheinlichkeit an, die Nullhypothese zu verwerfen, obwohl sie richtig ist. Bei kleinerem Niveau wird daher weniger oft verworfen.
   iii. Richtig: Fehler 1. Art heisst, die Nullhypothese zu verwerfen, obwohl sie richtig ist.
   iv. Falsch: Beim t-Test ist das Niveau nur exakt richtig unter Normalverteilung, beim Wilcoxon-Test hingegen für beliebige stetige Verteilungen.
   v. Richtig: Folgt aus dem Dualitätssatz. Der Annahmebereich bei $\alpha = 1\%$ ist grösser als bei $\alpha = 5\%$.
   vi. Richtig: Je kleiner $|\mu|$, desto näher liegt die Alternative bei der Nullhypothese, und desto eher entscheidet man sich für die Nullhypothese statt für die Alternative.

3. Angenommen, die Anzahl der Defekte in einer 1200-Fuss-Rolle eines Magnetaufzeichnung Bands hat eine Poisson-Verteilung, für die der Wert des Erwartungswerts θ unbekannt ist, und die a-priori-Verteilung von θ ist die Gamma-Verteilung mit den Parametern $\alpha = 3$ und $\beta = 1$. Wenn fünf Rollen dieses Bandes zufällig ausgewählt und untersucht werden, sind die Anzahl der auf den Rollen gefundenen Defekter $2, 2, 6, 0$ und 3.

   a) Wie lautet die a-posteriori-Dichte von θ?
   b) Wie lautet der Bayes-Schätzer von θ, wenn der mittlere quadratische Fehler verwendet wird?

Lösung:

a) Wir erinnern uns, dass X der Poisson-Verteilung mit Erwartungswert θ folgt, also

$$\mathbb{P}(X = x) = e^{-\theta}\frac{\theta^x}{x!},$$

und θ hat die Dichtefunktion

$$\mathrm{w}(\theta) = \frac{\theta^{3-1}e^{-\theta}}{\Gamma(3)} = \frac{1}{2}\theta^2 e^{-\theta}.$$

Die bedingte Verteilung von θ erhält man:

$$\begin{aligned}
\mathrm{w}(\theta|2,2,6,0,3) &= \frac{\frac{1}{2}\theta^2 e^{-\theta}\cdot\frac{1}{2!2!6!0!3!}e^{-5\theta}\theta^{2+2+6+0+3}}{\int_0^\infty \theta^2 e^{-\theta}/2\cdot e^{-5\theta}\theta^{2+2+6+0+3}/(2!2!6!0!3!)d\theta} \\
&= \frac{e^{-6\theta}\theta^{15}}{\int_0^\infty e^{-6\theta}\theta^{15}d\theta} \\
&= \frac{\Gamma(16)}{6^{16}}e^{-6\theta}\theta^{15}
\end{aligned}$$

Wir bemerken, dass die a-posteriori-Verteilung von θ die Gamma-Verteilung mit den Parametern $\alpha = 16$ und $\beta = 6$ ist.

b) Der mittlere quadratische Fehler lautet

$$L(\theta, a) = (\theta - a)^2.$$

Der Bayes-Schätzer von θ ist der a, der minimiert

$$\mathbb{E}[L(\theta, a)|\text{Beobachtung}] = \int L(\theta, a)\text{w}(\theta|\text{Beobachtung})d\theta,$$

wobei $\text{w}(\theta|\text{Beobachtung})$ ist die a-posteriori-Verteilung von θ gegeben die Beobachtung. Sei Y eine Gamma-verteilte Zufallsvariable mit Parametern α und β. Dann gilt es, dass

$$\mathbb{E}[Y] = \frac{\alpha}{\beta}, \quad \mathbb{E}[Y^2] = \frac{\alpha(\alpha+1)}{\beta^2}.$$

Also folgt

$$\begin{aligned}
\mathbb{E}[(\theta - a)^2|2,2,6,0,3] &= \mathbb{E}[\theta^2|2,2,6,0,3] - 2a\mathbb{E}[\theta|2,2,6,0,3] + a^2 \\
&= \frac{17 \times 16}{6^2} - 2a\frac{16}{6} + a^2 \\
&= \left(a - \frac{8}{3}\right)^2 + \frac{4}{9}.
\end{aligned}$$

Der Bayes-Schätzer für die Beobachtung $2, 2, 6, 0, 3$ ist somit $8/3$.

4. Die Australier Mr. Smith und Dr. Thurston streiten sich über das Durchschnittsgewicht von *Strausseneiern*. Beide sind damit einverstanden, das Gewicht approximativ als normalverteilt aufzufassen. Mr. Smith behauptet aber, das mittlere Gewicht sei 1100g, während Dr. Thurston darauf besteht, dass die Eier schwerer seien, und zwar im Schnitt 1200g. Um ihren Streit beilegen zu können, reisen die beiden nach Afrika, um in der Savanne Strausseneier zu suchen. Weil diese aber meistens gut versteckt sind, finden sie nur acht, und zwar mit folgenden Gewichten (in g):

$$1090, 1150, 1170, 1080, 1210, 1230, 1180, 1130.$$

Sie wollen sich auf *einen* Test einigen, der ihren Streit über das mittlere Gewicht von Strausseneiern entscheiden soll. Deren Gewicht (in g) kann approximativ als normalverteilte Zufallsgrösse mit unbekanntem Erwartungswert μ und bekannter Streuung $\sigma = 55$ auffassen.

a) Testen Sie die Hypothese $\mu_0 = 1100$ gegen die Alternative $\mu_1 = 1200$ zum Niveau 2.5% mit einem mächtigsten Test.

b) Wie viele Beobachtungen (d.h. Strausseneier) muss man mindestens haben, damit die Macht des Tests grösser als 97.5% wird?

c) Bestimmen Sie für das n aus b) die Machtfunktion $\beta(\mu)$ des Tests für alle

$$\mu \in \{1000, 1050, 1100, 1150, 1200, 1250, 1300\},$$

und zeichnen Sie ihren Graphen.

d) Bestimmen Sie das realisierte 97.5%- Vertrauensintervall für μ.

e) Kommentieren Sie die Ergebnisse.

Lösung:

a) Sei $X = (X_1, \ldots, X_n)$ die n-dimensionale Zufallsvariable, welche die Beobachtungen modelliert. Wir nehmen an, dass die X_i unter $\mathbb{P}_0$ resp. $\mathbb{P}_1$ i.i.d. normalverteilt mit Varianz $\sigma^2 = 55^2$ und Erwartungswert μ_0 resp. μ_1 sind. Wir bezeichnen mit f_0 resp. f_1 die Dichte von X bezüglich dem n-dimensionalen Lebesgue-Mass unter $\mathbb{P}_0$ resp. $\mathbb{P}_1$. Nach dem Satz von Neyman-Pearson suchen wir nach einem Test φ und einer Konstanten c, die vom Niveau $\alpha = 0.025$ abhängt, so dass für $x \in \mathbb{R}^n$, $\varphi(x) = 1$, wenn $f_1(x)/f_0(x) \geq c$ und $\varphi(x) = 0$, sonst. Es gilt:

$$\frac{f_1(x)}{f_0(x)} = \exp\Big(\frac{\mu_1 - \mu_0}{\sigma^2}\sum_{i=1}^n x_i\Big) \exp\Big(- \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2}\Big).$$

Der Verwerfungsbereich zum 2.5%-Niveau ist folglich von der Form:

$$K_{2.5\%} = \{x \in \mathbb{R}^n; \sum_{i=1}^n x_i \geq c_\alpha\}.$$

Zur Bestimmung von c_α fordern wir:

$$\alpha \geq \mathbb{P}_0(X \in K_{2.5\%}) = \mathbb{P}_0\left(\underbrace{\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu_0}{\sigma/\sqrt{n}}}_{\mathcal{N}(0,1)} \geq \frac{\frac{1}{n}c_\alpha - \mu_0}{\sigma/\sqrt{n}}\right).$$

Daher muss gelten:

$$\frac{\frac{1}{n}c_\alpha - \mu_0}{\sigma/\sqrt{n}} \geq q_{1-\alpha} = 1.96.$$

Mit den Daten x_i gilt:

$$\frac{\frac{1}{n}\sum_{i=1}^n x_i - \mu_0}{\sigma/\sqrt{n}} = 2.89,$$

also verwerfen wir die Nullhypothese zum 2.5%-Niveau.

b) Wir berechnen nun die Macht unter $\mathbb{P}_1$:

$$\begin{aligned}\mathbb{P}_1(X \in K_{2.5\%}) &= \mathbb{P}_1\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu_0}{\sigma/\sqrt{n}} \geq q_{1-\alpha}\right) \\ &= \mathbb{P}_1\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu_1}{\sigma/\sqrt{n}} \geq q_{1-\alpha} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(q_{1-\alpha} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right).\end{aligned}$$

Damit die Macht unter $\mathbb{P}_1$ grösser als 0.975 wird, muss gelten:

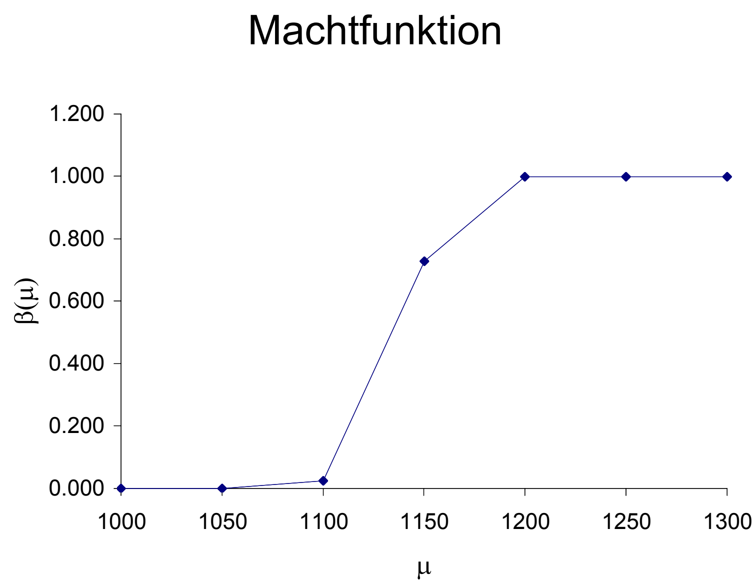
$$\Phi\Big(q_{1-\alpha} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\Big) \leq 0.025,$$

also $q_{1-\alpha} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} \leq q_{0.025} = -1.96$. Es folgt:

$$\sqrt{n} \geq \sigma\frac{-q_{0.025} + q_{1-\alpha}}{\mu_1 - \mu_0},$$

also $n \geq 5$.

c)



μ	1000	1050	1100	1150	1200	1250	1300
$\beta(\mu)$	0	0	0.025	0.528	0.982	1	1

d) Um das 97.5%-Vertrauensintervall zu bestimmen, bemerken wir, dass:

$$\mathbb{P}_\mu\left(-q_{0.0125} \leq \frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}} \leq q_{0.9875}\right) \geq 0.975.$$

Also ist das realisierte Vertrauensintervall:

$$[1156.25 - \frac{\sigma}{\sqrt{n}} q_{0.9875}, 1156.25 + \frac{\sigma}{\sqrt{n}} q_{0.9875}] = [1112.69, 1199.8].$$

Die Hypothesen μ_0 und μ_1 liegen beide knapp ausserhalb des Vertrauensintervalls und werden deshalb beide verworfen.

e) Die Daten erlauben es nicht auf μ_0 oder μ_1 zu schliessen. Der wahre Erwartungswert liegt zwischen μ_0 und μ_1.