ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik

Musterlösung 10

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

1. Gemüsehändler Hase behauptet seine Karotten seien im Durchschnitt mindestens $30cm$ lang.
2. 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.
3. Konservenfabrikant Hamster wirbt für seine extrazarten jungen Erbsen mit der Garantie, die durchschnittliche Dicke der jungen Erbsen betrage höchstens $3mm$.
4. Er behauptet auch, dass der Anteil von holzigen Spargeln in seinen Konserven unter $0,3\%$ liege.
5. Ausserdem lobt er die ausgewogene Mischung seiner "Erbsen mit Karotten", die genau 40% zu 60% Gewichtsanteil betrage.

a) Gib jeweils an, ob ein linksseitiger, ein rechtsseitiger oder ein zweiseitiger Test nötig ist um den Wahrheitsgehalt[1] der Aussagen zu testen. Auf welchen Parameter bezüglich welcher Zufallsvariable wird getestet? Stelle jeweils Nullhypothese und Gegenhypothese (Alternativhypothese) auf.

(Der Test selber muss (und kann mangels fehlender Angaben) nicht durchgeführt werden.)

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)
1. Sei X die Zufallsvariable welche die Länge der Karotten (in cm) modelliert. Getested wird auf den Parameter $\mu = \mathbb{E}(X)$, die durchschnittliche Länge der Karotten. Die Nullhypothese ist die Behauptung des Gemüsehändlers. Wir wollen testen ob sie sich widerlegen lässt. Daraus ergibt sich ein linksseitiger Test mit $H_0 : \mu \geq 30cm \quad H_A : \mu < 30cm$.
2. Sei X die Zufallsvariable welche das Gewicht der Kartoffeln (in g) modelliert. Getested wird auf den Parameter $\mu = \mathbb{E}(X)$. Wie oben, ist die Nullhypothese die Behauptung des Gemüsehändlers. Wir wollen testen ob sie sich widerlegen lässt. Daraus ergibt sich ein zweiseitiger Test mit den Hypothesen $H_0 : \mu = 50g \quad H_A : \mu \neq 50g$.
3. Nun sei X die Zufallsvariable, welche die Dicker der Erbsen (in mm) modelliert. Hier nehmen wir einen rechtsseitigen Test mit den Hypothesen $H_0 : \mu \leq 3mm \quad H_A : \mu > 3mm$.

[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.

4. 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 $H_0 : \mu \leq \mu_0 \quad H_A : \mu > \mu_0$.
5. Zweiseitiger Test mit $\mu_0 = 0.4$ sowie den Hypothesen $H_0 : \mu = \mu_0 \quad H_A : \mu \neq \mu_0$.

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

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

Bemerkung 1: Wir betrachten hier Tests mit einer Nullhypothese der Form $\mu \leq \mu_0$, während wir uns in der Vorlesung auf Tests mit einer Nullhypothese der Form $\mu = \mu_0$ beschränkt haben. Für einen Test mit Nullhypothese $\mu \leq \mu_0$ würde dann die Bedingung 'W'keit für Fehler 1. Art $\leq \alpha$' wie folgt lauten: Der Verwerfungsbereich K muss so gewählt werden, dass für *alle* $\mu \leq \mu_0$ erfüllt ist, dass $P_\mu(X \in K) \leq \alpha$. In unserem Fall muss dies nur für $\mu = \mu_0$ erfüllt sein, weil die W'keit für $\mu = \mu_0$ am grössten ist. In allen konkreten Tests, die wir in dieser Vorlesung und den Übungen betrachten, kann man die Nullhypothese auf die Form $\mu = \mu_0$ vereinfachen.

Bemerkung 2: 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. Lisa hat eine Maschine gebaut, die das Abwerfen einer Münze filmt und anhand einer komplizierten Videoanalyse das Ergebnis ("Kopf" oder "Zahl") vorhersagt.
Sie glaubt aber, dass dies reiner Zufall ist und die Vorhersagen der Maschine im Schnitt nur in der Hälfte der Fälle zutreffen. Ihre Freundin Laura hält sie für eine Pessimistin und möchte in einem statistischen Test nachweisen, dass die Maschine nicht einfach nur zufällige Vorhersagen macht.
Für einen Testlauf wirft sie 20 mal eine Münze und lässt die Maschine den Ausgang vorhersagen. In 13 der 20 Würfe sagt die Maschine das richtige Ergebnis voraus.

Sei X die Anzahl der von der Maschine korrekt geratenen Ergebnisse (aus $n = 20$ Versuchen). Zur Modellierung von X verwenden wir eine Binomialverteilung mit Parametern $n = 20$ und Erfolgswahrscheinlichkeit p, d.h. $X \sim \text{Bin}(n, p)$, $n = 20$.

a) Gib die Null- und Alternativhypothese für den statistischen Test an. Begründe deine Wahl.

b) Gib die Verteilung unter H_0 an. Wird die Nullhypothese auf dem 1%-Signifikanzniveau verworfen?

c) Wie wahrscheinlich ist es, dass die Nullhypothese verworfen wird, falls die Vorhersagewahrscheinlichkeit in Tat und Wahrheit $p = 0.75$ ist?

Lösung:

a) Die Null- und Alternativhypothesen lauten

$$H_0 : p = 0.5 \quad \text{und} \quad H_A : p \neq 0.5\,.$$

Die Nullhypothese besagt, dass die Maschine zufällig Kopf oder Zahl voraussagt. Laura möchte Lisas Behauptung, dass die Maschine im Schnitt in der Hälfte der Fälle zutreffende Vorhersagen macht, widerlegen. Dazu verwendet sie eine zweinseitige Alternativhypothese der Form $p \neq 0.5$. (Bemerkung: Wir verwenden hier eine zweiseitige Alternativhypothese, da die Maschine auch eine im Durchschnitt schlechte Vorhersage mit p¡ 0.5 machen koennte - und das Ergebnis damit nicht mehr reiner Zufall wäre.)

b) Unter der Nullhypothese ist die Teststatistik $\mathrm{Bin}(20, 0.5)$ verteilt. Um den Verwerfungsbereich zu bestimmen sucht Laura das grösste c_1 und das kleinste c_2, sodass $P[X \leq c_1] \leq 0.005$ und $P[X \geq c] \leq 0.005$. Dazu berechnet sie die folgenden Wahrscheinlichkeiten:

$$\begin{aligned}
P[X \geq 20] &= \binom{20}{20} 0.5^{20} \approx 9.54 \cdot 10^{-7} < 0.005 \\
P[X \geq 19] &= \binom{20}{19} 0.5^{20} + P[X \geq 20] \approx 2.00 \cdot 10^{-5} < 0.005 \\
P[X \geq 18] &= \binom{20}{18} 0.5^{20} + P[X \geq 19] \approx 2.01 \cdot 10^{-4} < 0.005 \\
P[X \geq 17] &= \binom{20}{17} 0.5^{20} + P[X \geq 18] \approx 0.00129 < 0.005 \\
P[X \geq 16] &= \binom{20}{16} 0.5^{20} + P[X \geq 17] \approx 0.0059 > 0.005 \\
P[X \leq 0] &= \binom{20}{0} 0.5^{20} \approx 9.53 \cdot 10^{-7} < 0.005 \\
P[X \leq 1] &= \binom{20}{1} 0.5^{20} + P[X \leq 0] \approx 2.00 \cdot 10^{-5} < 0.005 \\
P[X \leq 2] &= \binom{20}{2} 0.5^{20} + P[X \leq 1] \approx 2.01 \cdot 10^{-4} < 0.005 \\
P[X \leq 3] &= \binom{20}{3} 0.5^{20} + P[X \leq 2] \approx 0.00129 < 0.005 \\
P[X \leq 4] &= \binom{20}{4} 0.5^{20} + P[X \leq 3] \approx 0.0059 > 0.005.
\end{aligned}$$

Daher ist $c_1 = 3$, $c_2 = 17$ und der Verwerfungsbereich für den Test gegeben durch: $K = \{0, 1, 2, 3\} \cup \{17, ..., 20\}$. Da $13 \notin K$ kann die Nullhypothese nicht verworfen werden.

c) Gesucht ist:

$$\begin{aligned}
P_{p=0.75}[X \in K] &= P_{p=0.75}[X \leq 3] + P_{p=0.75}[X \geq 17] \\
&= \sum_{k=0}^{3} \binom{20}{k} 0.75^k\, 0.25^{20-k} + \sum_{k=17}^{20} \binom{20}{k} 0.75^k\, 0.25^{20-k} \\
&\approx 0.2252.
\end{aligned}$$

Bei der berechneten Wahrscheinlichkeit handelt es sich gerade um die *Macht* des Tests.

3. Die Anreicherung einer Legierung mit einem Metall soll den Volumenausdehnungskoeffizienten, der bei der Standardlegierung (ohne Anreicherung) 1.0085 beträgt, reduzieren. Um diese Hypothese nachzuprüfen, wurde der Koeffizient an 12 Proben der neuen Legierung bei gleicher Temperaturänderung gemessen, mit folgenden Ergebnissen:

1.00781	1.00646	1.00801	1.00833	1.00738	1.00687
1.00783	1.00936	1.00564	1.00543	1.00794	1.01060,

die zugehörigen empirischen Werte sind $\overline{x} = 1.00764$ und $s = 0.00146$, außerdem nehmen wir an dass diese Daten normalverteilt sind.

a) Konstruiere das 99%-Vertrauensintervall für den Parameter μ.

b) Lässt sich auf dem Niveau von 5% tatsächlich nachweisen, dass der Volumenausdehnungskoeffizient der neuen Legierung kleiner ist? Formuliere dazu geeignete Hypothesen, gib an ob der Test ein- oder zweiseitig ist, und führe den Test durch.

c) Berechne den p-Wert.

Wir idealisieren nun die Situation und nehmen an, dass $\sigma = 0.0014$ bekannt ist.

d) Konstruiere das 99%-Vertrauensintervall für den Parameter μ. Ist es grösser oder kleiner als das Vertrauensintervall von a)? Warum?

e) Führe den analogen Test zu b) durch. Wie lautet nun die Teststatistik und wie entscheidet dieser Test?

f) Berechne die Wahrscheinlichkeiten eines Fehlers 2. Art für den in d) bestimmten Test, wenn der wahre Wert $\mu = 1.008$ bzw. $\mu = 1.007$ ist.

g) Wie groß muss der Stichprobenumfang n gewählt werden, damit das Vertrauensintervall zum 5% Niveau schmäler als 0.0002 ist? Welcher empirisch gemessene Mittelwert würde dann schon zum Verwerfen der Nullhypothese führen?

Lösung:

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ der gemessene Volumenausdehnungskoeffizient und seien wie immer die tatsächlich gemessenen Werte $x_1, \ldots, x_{12}$ unabhängige Realisierungen davon. Der Parameter μ soll getestet werden, da die Standardabweichung σ unbekannt ist, muss sie aus den Daten geschätzt werden, was uns zum t-Test führt.

a) Mit $t_{n-1;1-\frac{\alpha}{2}} = t_{11;0.995} = 3.106$ ergibt sich die Realisierung des 99%-Vertrauensintervalls zu

$$I_{99\%} = \left[\overline{x} - t_{11,0.995}\frac{s}{\sqrt{12}} \ , \ \overline{x} + t_{11,0.995}\frac{s}{\sqrt{12}}\right] = [1.00633, 1.00895].$$

b) Wir wollen wissen ob das μ der neuen Legierung kleiner ist als 1.0085, also jenes der Standardlegierung. Dazu führen wir einen linksseitigen Test zum Niveau 5% mit Null- und Alternativhypothese

$$H_0 : \mu = \mu_0 := 1.0085 \qquad\qquad H_A : \mu < \mu_0$$

durch, der Verwerfungsbereich ist somit gegeben durch $K = (-\infty, t_{11,0.05}] = (-\infty, -1.796]$. Die Realisierung der Teststatistik ist

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{12}} = \frac{1.00764 - 1.0085}{0.00146/\sqrt{12}} = -2.04.$$

Da $-2.04 \in K$ wird die Nullhypothese verworfen.

c) Wir suchen die Wahrscheinlichkeit unter der Nullhypothese, dass die (t_{n-1}-verteilte) Teststatistik in den Bereich $(-\infty, -2.04]$ fällt; das heisst, wir suchen die Wahrscheinlichkeit p, für die $t_{n-1,p} = -2.04$ gilt. Dank Symmetrie der t-Verteilung gilt $t_{n-1,1-p} = 2.04$. Durch Rückwärtsauslesen aus der Tabelle der t-Quantile sieht man lediglich, dass der Wert zwischen 2.5% und 5% liegen muss. Mit R (oder analog mit Matlab) erhält man mit

```
> pt(-2.04,11)
[1] 0.03305204
```

also einen Wert von $\alpha = 3.3\%$.

Wenn $\sigma = 0.0014$ bekannt ist, verwenden wir jetzt den z-Test, der einzige Unterschied ist, dass die Quantile der t-Verteilung durch jene der Standardnormalverteilung, und die geschätzte Streuung durch die tatsächlich Streuung ersetzt wird.

d) Mit $z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.57$ ergibt sich die Realisierung des 99%-Vertrauensintervalls zu

$$I_{99\%} = \left[\overline{x} - z_{0.995}\frac{\sigma}{\sqrt{12}}\ ,\ \overline{x} + z_{0.995}\frac{\sigma}{\sqrt{12}}\right] = [1.00660, 1.00868].$$

Somit ist das Vertrauensintervall kleiner als in Teilaufgabe a). Dieser Fakt trägt der im Vergleich kleineren Ungewissheit Rechnung, die sich durch das Bekanntwerden von σ ereignet hat.

e) $K = (-\infty, z_{0.05}] = (-\infty, -1.645]$.
Die Teststatistik lautet nun $Z = \frac{\overline{X}_{12} - \mu_0}{\sigma/\sqrt{12}}$ und ihre Realisierung ist gegeben durch

$$z = \frac{\overline{x}_{12} - \mu_0}{\sigma/\sqrt{12}} = \frac{1.00764 - 1.0085}{0.0014/\sqrt{12}} = -2.13 \in K,$$

also wird H_0 auch hier verworfen.

f) Der Verwerfungsbereich ist $K = (-\infty, -1.645]$. Wir nehmen zuerst an $\mu = 1.008$. Wegen $Z \sim \mathcal{N}\left(\frac{\mu-\mu_0}{\sigma/\sqrt{n}}, 1\right)$ ist $Z \overset{\mu=1.008}{\sim} \mathcal{N}(-1.2372, 1)$ und daher $Z' := Z + 1.2372 \overset{\mu=1.008}{\sim} \mathcal{N}(0,1)$ ist standardnormalverteilt. Somit ergibt sich der Fehler 2.Art zu

$$\begin{aligned}
\mathbb{P}_{\mu=1.008}(Z \notin K) &= \mathbb{P}_{\mu=1.008}(Z > -1,645) \\
&= 1 - \mathbb{P}_{\mu=1.008}(Z \leq -1,645) \\
&= 1 - \mathbb{P}_{\mu=1.008}(Z' \leq -1,645 + 1.2372) \\
&= 1 - \Phi(-0.41) = 0.6591 \approx 66\%,
\end{aligned}$$

was ziemlich hoch ist, die selbe Rechung für $\mu = 1.007$ ergibt schon einen wesentlich kleineren Wert von ca. 2%.

g) Die Breite[2] des Vertrauensintervalls ist $2\frac{\sigma}{\sqrt{n}}z_{0.975}$, somit ergibt sich die Bedingung $\frac{\sigma}{\sqrt{n}}z_{0.975} < 0.0001$ und weiters $\sqrt{n} > \frac{\sigma z_{0.975}}{0.0001} = \frac{0.0014 \cdot 1.96}{0.0001} = 27.44$ bzw. $n > 27.44^2 = 752.95$, d.h. n sollte mindestens 753 sein.

Die Frage ist nun wie klein $\overline{x}_{753}$ sein muss, damit $z = \frac{\overline{x}_{753} - 1.0085}{0.0014/\sqrt{753}} \leq -1.645$ gilt. Diese Bedingung ist äquivalent zu $\overline{x}_{753} \leq -1.645\frac{0.0014}{\sqrt{753}} + 1.0085 = 1.008416$. Bei einem Stichprobenumfang von $n = 753$ würde also bereits ein empirisch gemessener Mittelwert von 1.008416 zum Verwerfen der Nullhypothese führen!

4. Eine Klimaanlage schafft es, die Raumtemperatur bis auf eine Standardabweichung von einem halben Grad Celsius konstant zu halten. Die angestrebte Raumtemperatur beträgt 20.00 Grad Celsius. An zehn aufeinanderfolgenden Tagen wurden die folgenden Temperaturen gemessen:

20.71	19.76	20.56	21.39	21.00	19.67	20.92	20.31	20.39	20.72

Aus diesen Daten ergibt sich $\overline{x}_{10} = 20.543$.

a) Nehme an, dass die gemessenen Temperaturen unabhängig voneinander und identisch normalverteilt sind, und führe damit einen geeigneten Test auf dem 5%-Niveau durch, um zu beurteilen, ob die Klimaanlage wirklich auf den Sollwert von 20.00 Grad geeicht ist. Führe diesen Test durch mittels Berechnung des Verwerfungsbereiches.

b) Führe den Test nochmals durch, jetzt mittels Berechnung des p-Werts statt Berechnung des Verwerfungsbereichs.

Lösung:

a) Test durchführen:

Modellannahme:	X_i: $i-$te Temperaturmessung. X_i i.i.d. $\mathcal{N}(\mu, \sigma^2)$ mit $\sigma = 0.5$.
Test:	z-Test zum Niveau $\alpha = 5\%$
Nullhypothese H_0:	$\mu = \mu_0 := 20$
Alternative H_A:	$\mu \neq \mu_0$ (zweiseitig)
Teststatistik:	$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$
Verwerfungsbereich:	Aus der Normalverteilungstabelle: $K = (-\infty, -1.96] \cup [1, 96, \infty)$.
Wert der Teststatistik:	$z = \frac{\overline{x}_{10} - \mu_0}{\sigma/\sqrt{10}} = \frac{20.543 - 20}{0.5/\sqrt{10}} = 3.43 \in K$
Testentscheid:	$3.43 \in K$, also wird die Nullhypothese (ziemlich klar) verworfen. Die Überschreitung des Grenzwertes ist auf dem 5%-Niveau signifikant.

b) Der p-Wert für die gegebene Realisierung der Teststatistik errechnet sich zu $\mathbb{P}(|Z| \geq 3.43) = 2\mathbb{P}(Z \geq 3.43) = 2(1 - \Phi(3.43)) = 2(1 - 0.999698) = 0.0006$. Wir könnten also einen 2-seitigen Test zum Niveau $\alpha = 0.06\%$ durchführen und die Nullhypothese würde immer noch verworfen.

[2] Beachte, dass diese beim z-Test nicht von der jeweiligen Stichprobe $x_1, \ldots, x_{12}$ abhängt, sondern nur vom Stichprobenumfang (hier $n = 12$) und vom gewählten Niveau.