Lösung zu Aufgabe 7: t-Test für Erwartungswert

Statistische Tests dienen dazu ***Hypothesen*** *abzusichern oder begründet zu verwerfen. Hypothesen entstehen aus experimentellen Beobachtungen oder formalen Überlegungen, die einer Prüfung unterzogen werden müssen. Im Allgemeinen kann man sie in experimentellen Wissenschaften weder beweisen, noch widerlegen. Aber mit Hilfe statistischer Test können sie abgesichert oder begründet verworfen werden.*

Betrachten wir eine Messgröße X *als Zufallsvariable, so kann die Hypothese zum Beispiel eine Aussage über den (unbekannten) Erwartungswert oder die (unbekannte) Standardabweichung oder die zugrunde liegende Verteilungsfunktion sein. Die zu untersuchende Hypothese wird als* ***Nullhypothese*** H_0 *bezeichnet. Zur Nullhypothese existiert stets eine* ***Alternativhypothese*** H_1.

Es wird eine Testgröße $T = T(x_1, ..., x_n, ...)$ *definiert, die so konstruiert wird, dass T einer bestimmten Verteilungsfunktion entspricht (zum Beispiel:* χ^2*-Verteilung oder Studentsche t-Verteilung). Die Messwerte werden in die Gleichung für T eingesetzt und die Testgröße wird berechnet. Der Wert der Testgröße wird mit einem Schwellenwert verglichen, der sich aus der Verteilungsfunktion von T ergibt. Aus diesem Vergleich folgt die Entscheidung, ob* H_0 *verworfen oder beibehalten wird.*

Anschaulich entspricht diese Vorgehensweise der Berechnung der Wahrscheinlichkeit p, dass unter Annahme der Gültigkeit der Hypothese H_0 *das tatsächlich beobachtete Ergebnis auftritt. Diese Wahrscheinlichkeit heißt Signifikanzniveau* α *oder p-Wert. Wenn der p-Wert klein ist, also das Eintreten des tatsächlich beobachteten Ereignisses unter der Annahme der Gültigkeit von* H_0 *klein ist, wird die Hypothese* H_0 *verworfen.*

Bei der Entscheidung muss stets mit einer Fehlentscheidung gerechnet werden. Folgende Situationen sind möglich:

	Tatsächlich: H_0 richtig	***Tatsächlich: H_0 falsch***
Nichtablehnung von H_0	*richtige Entscheidung mit Wahrscheinlichkeit* $1-\alpha$	***Fehlentscheidung 2. Art*** *mit Wahrscheinlichkeit* β
Ablehnung von H_0	***Fehlentscheidung 1. Art*** *mit Wahrscheinlichkeit* α	*richtige Entscheidung mit Wahrscheinlichkeit* $1-\beta$

α *: Irrtumswahrscheinlichkeit oder Signifikanzniveau, typische Werte sind 5%; 1%; 0,1%*

$1-\beta$ *: Güte des Tests*

Beide Fehler sind in der Praxis unvermeidbar und lassen sich im Allgemeinen nicht unabhängig voneinander beeinflussen. Verkleinern von α *hat im Allgemeinen eine Vergrößerung von* β *zur Folge. Die Güte des Tests kann durch Vergrößern des Stichprobenumfangs erhöht werden.*

a) Werden mit einer statistischen Sicherheit von 95% mindestens 1000 ml ausgeschenkt?

Bei der vorliegenden Aufgabe soll überprüft werden, ob die von Wirt Alois ausgeschenkte Biermenge je Maßkrug signifikant geringer ist, als die geforderte Füllmenge von 1000 ml. Es gilt also, den mittels einer Stichprobe experimentell bestimmten Schätzwert für den Erwartungswert der vorliegenden Verteilung mit dem vorgegebenen Referenzwert von 1000 ml zu vergleichen. Diese Fragestellung kann mit einem t-Test für den Erwartungswert beantwortet werden:

Der t-Test für den Erwartungswert testet anhand einer Stichprobe, ob der Mittelwert dieser Stichprobe von einem vorgegebenen Testwert in der Grundgesamtheit signifikant verschieden ist. Dabei sei X normalverteilt und μ und σ seien unbekannt. Der Erwartungswert wird als μ_0 vermutet, woraus die Hypothese $H_0 : \mu_x = \mu_0$ folgt. Zur Berechnung der Testgröße werden aus den Messwerten $x_1, ..., x_n$ Mittelwert und Streuung berechnet.

Die Formel zur Berechnung der t-verteilten Testgröße lautet im Falle des t-Tests für den Erwartungswert wie folgt:

$$t_0 = \frac{\overline{x} - \mu_0}{S / \sqrt{n}}$$

Hierin ist $\overline{x}$ der aus den experimentell erhobenen Messwerten berechnete arithmetische Mittelwert als bester Schätzwert für den Erwartungswert der Verteilung, μ_0 ist der vorgegebene Vergleichswert, S ist die empirische Streuung der Messwerte und n der Stichprobenumfang.

Wir berechnen zunächst den Mittelwert $\overline{x}$ und die Streuung S der vorliegenden Messwerte:

$$\overline{x} = 990 \text{ ml}$$

$$S \approx 16{,}623 \text{ ml}$$

Ferner wissen wir, dass der Stichprobenumfang n = 20 beträgt.

Die Testgröße t_0 ergibt sich daher zu:

$$t_0 = \frac{990 \text{ ml} - 1000 \text{ ml}}{\frac{16{,}623 \text{ ml}}{\sqrt{20}}} \approx -2{,}689$$

Den Wert dieser Testgröße müssen wir nun mit einem kritischen Wert vergleichen. Die Bestimmung dieses kritischen Wertes sowie die Testregel hängen von der interessierenden Alternativhypothese H_1 ab. Prinzipiell können wir zwischen den folgenden drei Varianten unterscheiden:

1) $H_0 : \mu_x = \mu_0$ gegen $H_1 : \mu_x < \mu_0$ (einseitige Hypothese)

   Ist $t_0 < -t_{n-1;1-\alpha}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

2) $H_0 : \mu_x = \mu_0$ gegen $H_1 : \mu_x > \mu_0$ (einseitige Hypothese)

Ist $t_0 > t_{n-1;1-\alpha}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

3) $H_0 : \mu_x = \mu_0$ gegen $H_1 : \mu_x \neq \mu_0$ (zweiseitige Hypothese)

Ist $|t_0| > t_{n-1;1-\alpha/2}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

Wie zu erkennen, ist die Nullhypothese H_0 in allen drei Fällen identisch. Der Unterschied liegt in der Alternativhypothese H_1. Bei den Fällen 1 und 2 wird jeweils eine einseitige Gegenhypothese überprüft, das heißt, es interessiert nur der Fall, dass der zu testende Wert μ_x kleiner als der Vergleichswert μ_0 (Fall 1) bzw. größer als der Vergleichswert μ_0 (Fall 2) ist. Im Fall 3 wird hingegen eine zweiseitige Alternativhypothese getestet, bei der sowohl Abweichungen nach unten als auch nach oben von Interesse sind und die Alternativhypothese daher $H_1 : \mu_x \neq \mu_0$ lautet.

Im vorliegenden Fall besteht das Interesse des Mitarbeiters des Ordnungsamtes darin, zu bestätigen oder auszuschließen, dass Wirt Alois **weniger** als die geforderten 1000 ml je Maßkrug ausschenkt, da dies eine Ordnungswidrigkeit darstellen würde. Der umgekehrte Fall, dass der Wirt mehr als die geforderten 1000 ml ausschenkt ist für den Prüfer hingegen nicht von Interesse, da dies rechtlich zulässig wäre.

Vergleichen wir das so formulierte Interesse mit den drei oben unterschiedenen Fällen, so stellen wir fest, dass für uns der Fall 1 der relevante ist, in dem die Gegenhypothese $H_1 : \mu_x < \mu_0$ lautet. Die in diesem Fall anzuwendende Testregel lautet:

Ist $t_0 < -t_{n-1;1-\alpha}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

Der kritische Wert, mit dem unsere oben berechnete Testgröße t_0 verglichen werden muss, lautet demnach $-t_{n-1;1-\alpha}$. Den Zahlenwert von $t_{n-1;1-\alpha}$ bestimmen wir wie gewohnt mit Kenntnis des Stichprobenumfangs n und des Signifikanzniveaus α aus der Tabelle der Student'schen t-Verteilung. Mit n = 20 und α = 0,05 erhalten wir:

$$t_{n-1;1-\alpha} = t_{19;0,95} = 1{,}729$$

Die auszuwertende Testbedingung lautet damit:

$$-2{,}689 < -1{,}729$$

Diese Bedingung ist offensichtlich erfüllt. Die Testregel besagt nun weiter, dass für den Fall, dass obige Bedingung erfüllt ist, die Nullhypothese H_0 auf dem Signifikanzniveau α abzulehnen ist. Wie schließen also:

Die Nullhypothese $H_0 : \mu_x = \mu_0$ wird auf dem Signifikanzniveau α = 0,05 abgelehnt!

Inhaltlich besagte unsere Nullhypothese, dass Wirt Alois mit einer statistischen Sicherheit von $P = 95\%$ im Mittel mindestens die geforderten 1000 ml je Maßkrug ausschenkt. Die Ablehnung der Nullhypothese bedeutet daher im vorliegenden Fall:

Es kann nicht mit einer statistischen Sicherheit von P = 95% davon ausgegangen werden, dass Wirt Alois im Mittel 1000 ml oder mehr je Maß ausschenkt!

b) Mittlere Fehlmenge je Krug bei „ertragsoptimiertem" Zapfen?

Um ein Bußgeld wegen zu geringer Füllmenge zu vermeiden, darf der Test der Nullhypothese $H_0 : \mu_x = \mu_0$ mit der Alternativhypothese $H_1 : \mu_x < \mu_0$ nicht zu einer Ablehnung der Nullhypothese führen. Eine Ablehnung der Nullhypothese würde für den Fall erfolgen, dass gilt:

$$t_0 < -t_{n-1;1-\alpha}$$

Unsere Forderung, diesen Fall zu vermeiden, können wir daher wie folgt formulieren:

$$t_0 \overset{!}{\geq} -t_{n-1;1-\alpha}$$

Mit den in der Aufgabenstellung gegebenen Parametern, $n = 20$ und $\alpha = 0{,}025$, ergibt sich für den kritischen Wert $-t_{n-1;1-\alpha}$:

$$-t_{n-1;1-\alpha} = -t_{19;0{,}975} = -2{,}093$$

Für die Testgröße t_0 fordern wir daher:

$$t_0 \overset{!}{\geq} -2{,}093$$

Die Formel zur Berechnung der Testgröße lautet wie bereits oben eingeführt:

$$t_0 = \frac{\overline{x} - \mu_0}{S/\sqrt{n}}$$

Im vorliegenden Fall sind folgende Zahlenwerte durch die Aufgabenstellung vorgegeben:

$$\mu_0 = 1000 \text{ ml}$$

$$S = 20 \text{ ml}$$

$$n = 20$$

Anmerkung: Bei der Annahme, dass die Standardabweichung ***mindestens*** *20 ml beträgt, handelt es sich um eine Abschätzung des für den Wirt ungünstigsten Falls. Denn je größer die Standardabweichung ist, desto schwieriger wird es, den Nachweis zu führen, dass eine Abweichung vom Erwartungswert tatsächlich statistisch signifikant ist. Bei einer kleineren Standardabweichung steigt hingegen die Signifikanz einer Abweichung vom Erwartungswert. Für* *den Grenzfall einer Standardabweichung von Null wäre jede Abweichung automatisch signifikant.*

Für die Testgröße t_0 gilt folglich:

$$t_0 = \frac{\overline{x} - 1000\ \text{ml}}{\frac{20\ \text{ml}}{\sqrt{20}}}$$

Setzen wir dies in unsere obige Forderung für die Testgröße t_0 ein, erhalten wir:

$$\frac{\overline{x} - 1000\ \text{ml}}{\frac{20\ \text{ml}}{\sqrt{20}}} \overset{!}{\geq} -2{,}093$$

Diesen Ausdruck können wir nach der gesuchten mittleren Füllmenge $\overline{x}$ auflösen und erhalten:

$$\overline{x} \overset{!}{\geq} -2{,}093 \cdot \frac{20\ \text{ml}}{\sqrt{20}} + 1000\ \text{ml} \approx 990{,}64\ \text{ml}$$

Die mittlere Füllmenge je Maß, bei der gerade noch ein Bußgeld vermieden wird, beträgt also ungefähr $\overline{x} = 990{,}64\ \text{ml}$. Da in der Aufgabenstellung nach der mittleren Fehlmenge $\Delta_{\overline{x}}$ je Maß gefragt ist, berechnen wir weiter:

$$\Delta_{\overline{x}} = 1000\ \text{ml} - 990{,}64\ \text{ml} = 9{,}36\ \text{ml}$$

Die mittlere Fehlmenge je Maßkrug, bei der ein Bußgeld gerade noch vermieden wird, beträgt etwa 9,36 ml.