Statistik I für Human- und Sozialwissenschaften

10. Übung

Lösungsvorschlag

Gruppenübung

G 33 In einer Kaffeerösterei füllt eine Maschine gemahlenen Kaffee in 500g–Packungen ab. Der Kundendienst dieser Maschine wurde bei der routinemäßigen Wartung damit beauftragt, für die Varianz der Füllmenge den Sollwert $\sigma_0^2 = 14\ [g^2]$ einzustellen. Nach der Wartung kommen bei den Mitarbeitern Zweifel auf, ob die Neueinstellung erfolgreich war. Sie vermuten eine höhere Varianz der Füllmenge, als sie 8 Packungen Kaffee nach der Abfüllung gewogen haben und folgende Gewichte feststellten:

492,3 499,0 509,1 493,8 510,4 508,9 491,6 494,5

Wir nehmen an, dass sich die Abfüllmengen durch i.i.d. $N(\mu, \sigma^2)$–verteilte Zufallsvariablen beschreiben lassen mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2.

(a) Stellen Sie Nullhypothese und Alternativhypothese auf, um die Vermutung der Mitarbeiter mit Hilfe eines Statistischen Tests zu überprüfen.

(b) Benennen Sie einen geeigneten Statistischen Test zur Überprüfung der aufgestellten Hypothesen.

(c) Führen Sie den in (b) genannten Statistischen Test auf dem Signifikanzniveau $\alpha = 0,001$ durch und interpretieren Sie das Ergebnis.

Nullhypothese: $H_0 : \sigma^2 \leq 14$, *Alternative* $H_1 : \sigma^2 > 14$

Geeignetes Testverfahren: χ^2-*Streuungstest*
Begründung: Der Erwartungswert μ ist unbekannt, und man sucht ein Testverfahren zum Vergleich von σ^2 mit dem vorgegebenen Zahlenwert ($\sigma_0^2 = 14$).

Testgröße:

$$T(X_1, \ldots, X_n) = \frac{n-1}{\sigma_0^2} \cdot \bar{S}^2$$

Realisierung der Testgröße:

$$\bar{x} = 499.95 \qquad \bar{s}^2 = \frac{469.7}{7}$$

$$T(x_1, \ldots, x_8) = \frac{7}{14} \cdot \frac{469.7}{7} \approx 33.55\ .$$

Kritischer Bereich: *Wir lehnen H_0 ab, falls gilt:*
$T(x_1, \ldots, x_n) > \chi^2_{n-1;1-\alpha}$

Entscheidung: *Wegen $T(x_1, \ldots, x_8) = 33.55 > 24.322 = \chi^2_{n-1;1-\alpha}$ wird die Nullhypothese* **abgelehnt**.

D.h. die Zweifel der Mitarbeiter sind berechtigt: Die Neueinstellung der Maschine war nicht erfolgreich, die Varianz der Füllmenge nach der Wartung ist größer als der beauftragte Sollwert.

G 34 Der verantwortliche Manager eines Unternehmens möchte untersuchen, ob klassische Hintergrundmusik eine Erhöhung der mittleren Tagesproduktion von Fließbandarbeitern bewirkt

oder nicht. Dazu werden die Tages–Produktionsmengen von 7 zufällig asgewählten Arbeitern einmal ohne klassische Hintergrundmusik und einmal mit gemessen.
Es wird angenommen, dass sich die Produktionsmenge der Arbeiter ohne klassische Hintergrundmusik durch i.i.d. $N(\mu_1, \sigma_1^2)$–verteilte Zufallsvariablen $X_1, \ldots, X_n$ und mit klassischer Hintergrundmusik durch $N(\mu_2, \sigma_2^2)$–verteilte Zufallsvariablen $Y_1, \ldots, Y_n$ beschreiben lässt. Ferner wird angenommen, dass sich für jeden einzelnen Arbeiter die Produktionsmengendifferenz $D_i = Y_i - X_i$, $i = 1, \ldots, n$, ebenfalls durch normalverteilte Zufallsvariablen beschreiben lässt. Von den ausgewählten Arbeitern liegen uns folgende Tages–Produktionsmengen vor (in kg):

Arbeiter Nr.	1	2	3	4	5	6	7
Prod.menge ohne kl. H.M.	29,0	21.2	18,6	22,9	16,5	25.1	23,7
Prod.menge mit kl. H.M.	23,7	25,5	19,3	20,4	19,8	24,3	26,2

(a) Stellen Sie Nullhypothese und Alternativhypothese für die beschriebene Testsituation auf.

(b) Benennen Sie einen geeigneten Statistischen Test zur Überprüfung der aufgestellten Hypothesen.

(c) Führen Sie den in (b) genannten Statistischen Test auf dem Signifikanzniveau $\alpha = 0,1$ durch und interpretieren Sie das Ergebnis.

Nullhypothese: $H_0 : \mu_2 \leq \mu_1$, *Alternative* $H_1 : \mu_2 > \mu_1$

Geeignetes Testverfahren: *t-Test für zwei abhängige Stichproben*
Begründung: zwei abhängige Messreihen, Testen auf Unterschied in Erwartungswerten mit ubekannten Varianzen.

Testgröße:

$$T(X_1, \ldots, X_n, Y_1, \ldots Y_n) = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{\bar{D}^2}{n}}}$$

$$\bar{D}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \bar{Y} - X_i + \bar{X} \right)^2$$

Realisierung der Testgröße:

$$\bar{x} = \frac{157}{7} \approx 22.43$$

$$\bar{y} = \frac{159.2}{7} \approx 22.74$$

$$\begin{aligned}\bar{d}^2 = \frac{1}{6} \Big(& (0.96 - 6.57)^2 + (2.76 - (-1.23))^2 \\ & + (-3.44 - (-3.83))^2 + (-2.34 - 0.47)^2 \\ & + (-2.94 - (5.93))^2 + (1.56 - 2.67)^2 \\ & + (3.46 - 1.27)^2 \Big) \approx 10.9393\end{aligned}$$

$$T(x_1, \ldots, y_7) = \frac{22.74 - 22.43}{\sqrt{\frac{10.9393}{7}}} \approx 0.25 \,.$$

Kritischer Bereich: *Wir lehnen H_0 ab, falls gilt:*
$T(x_1, \ldots, x_n, y_1, \ldots, y_n) > t_{n-1;1-\alpha}$

Entscheidung: *Wegen* $T(x_1, \ldots, x_7, y_1, \ldots, y_7) \approx 0.25 < 1.44 = t_{6;0.9}$ *wird die Nullhypothese* **nicht abgelehnt**.

Das Testergebnis liefert keinerlei Bestätigung dafür, dass klassische Hintergrundmusik die mittlere Tagesproduktionsmenge von Fließbandarbeitern erhöht.

G 35 Eine bestimmte Weizensorte wird auf 9 vergleichbaren, gleich großen Versuchsflächen angebaut. Aus Erfahrung weiß man, dass die Erträge der einzelnen Versuchsflächen als eine Stichprobe unabhängiger, identisch $N(\mu, 3.24)$-verteilter Zufallsvariablen angesehen werden können. Es ergibt sich ein arithmetisches Mittel von 105.0 [dz].

a) Überprüfen Sie mit einem geeigneten Testverfahren die Nullhypothese $H_0 : \mu = 106.0$ auf dem Signifikanzniveau $\alpha = 0.1$.

b) Welche Entscheidung würde sich auf dem Niveau $\alpha = 0.05$ ergeben?

c) Überprüfen Sie mit einem geeigneten Testverfahren die Nullhypothese $H_0 : \mu \geq 106.0$ auf dem Niveau $\alpha = 0.01$.

d) Überprüfen Sie mit einem geeigneten Testverfahren die Nullhypothese $H_0 : \mu \leq 106.0$ auf dem Niveau $\alpha = 0.01$.

Geeignetes Testverfahren: *Gauß-Test*
Begründung: Die Varianz σ^2 wird als bekannt vorausgesetzt, und man sucht ein Testverfahren zum Vergleich von μ mit dem vorgegebenen Zahlenwert ($\mu_0 = 106.0$).

a)

Nullhypothese: $H_0 : \mu = 106.0$, *Alternative* $H_1 : \mu \neq 106.0$

Testgröße:

$$T(X_1, \ldots, X_n) = \frac{\bar{X}_n - \mu_0}{\sqrt{\sigma^2/n}}$$

Realisierung der Testgröße:

$$T(x_1, \ldots, x_9) = \frac{\bar{x}_9 - \mu_0}{\sqrt{\sigma^2/9}} = \frac{105.0 - 106.0}{\sqrt{3.24/9}} = -\frac{5}{3} \approx -1.667 \ .$$

Kritischer Bereich: *Wir lehnen H_0 ab, falls gilt:*
$T(x_1, \ldots, x_n) \leq z_{\alpha/2}$ *oder* $T(x_1, \ldots, x_n) \geq z_{1-\alpha/2}$

Entscheidung: *Wegen $T(x_1, \ldots, x_9) = -\frac{5}{3} \leq -1.645 = z_{0.05}$ wird die Nullhypothese* **abgelehnt**.

b) Da sich gegenüber a) lediglich das Niveau α ändert, bleiben Nullhyothese, Testgröße, deren Realisierung und der kritische Bereich identisch. Dagegen ändert sich die Entscheidung, weil nun mit anderen Quantilen verglichen wird. Wegen $z_{0.025} = -1.96 < T(x_1, \ldots, x_9) = -\frac{5}{3} < 1.96 = z_{0.975}$ wird die Nullhypothese **nicht abgelehnt**.

c)

Nullhypothese: $H_0 : \mu \geq 106.0$, *Alternative* $H_1 : \mu < 106.0$

Testgröße:
$$T(X_1, \ldots, X_n) = \frac{\bar{X}_n - \mu_0}{\sqrt{\sigma^2/n}}$$

Realisierung der Testgröße:
$$T(x_1, \ldots, x_9) = \frac{\bar{x}_9 - \mu_0}{\sqrt{\sigma^2/9}} = \frac{105.0 - 106.0}{\sqrt{3.24/9}} = -\frac{5}{3} \approx -1.667\ .$$

Kritischer Bereich: *Wir lehnen* H_0 *ab, falls gilt:* $T(x_1, \ldots, x_n) \leq z_\alpha$

Entscheidung: *Wegen* $T(x_1, \ldots, x_9) = -\frac{5}{3} > -2.326 = z_{0.01}$ *wird die Nullhypothese* **nicht abgelehnt**.

d)

Nullhypothese: $H_0 : \mu \leq 106.0$, *Alternative* $H_1 : \mu > 106.0$

Testgröße:
$$T(X_1, \ldots, X_n) = \frac{\bar{X}_n - \mu_0}{\sqrt{\sigma^2/n}}$$

Realisierung der Testgröße:
$$T(x_1, \ldots, x_9) = \frac{\bar{x}_9 - \mu_0}{\sqrt{\sigma^2/9}} = \frac{105.0 - 106.0}{\sqrt{3.24/9}} = -\frac{5}{3} \approx -1.667\ .$$

Kritischer Bereich: *Wir lehnen* H_0 *ab, falls gilt:* $T(x_1, \ldots, x_n) \geq z_{1-\alpha}$

Entscheidung: *Wegen* $T(x_1, \ldots, x_9) = -\frac{5}{3} < 2.326 = z_{0.99}$ *wird die Nullhypothese* **nicht abgelehnt**.

G 36 Gegeben sei folgende geordnete Messreihe: 1; 2; 2; 2; 2; 3; 3; 3; 3; 4.

(a) Bestimmen und Skizzieren Sie die zugehörige empirische Verteilungsfunktion F der Messreihe.

(b) Bestimmen Sie das arithmetische Mittel, die Varianz und den Median der Messreihe.

Es sei X eine diskrete Zufallsvariable deren Verteilung durch die empirische Verteilungsfunktion F der Messreihe gegeben ist.

(c) Bestimmen und Skizzieren Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen X.

(d) Bestimmen Sie den Erwartungswert, die Varianz und den Median der Zufallsvariablen X.

(a)

$$F_X(x) := \begin{cases} 0 & \text{für } x < 1 \\ \frac{1}{10} & \text{für } 1 \leq x < 2 \\ \frac{1}{2} & \text{für } 2 \leq x < 3 \\ \frac{9}{10} & \text{für } 3 \leq x < 4 \\ 1 & \text{für } 4 \leq x. \end{cases}$$

(b) $\bar{x} = 2.5$, $s^2 = \frac{13}{20}$, $Md = \frac{2+3}{2} = 2.5$

(c)

$$f_X(x) := \begin{cases} \frac{1}{10} & \text{für } x \in \{1;4\} \\ \frac{4}{10} & \text{für } x \in \{2;3\} \\ 0 & \text{sonst.} \end{cases}$$

(d)

$$EX = \sum_{i=1}^{4} i f_X(i) = \frac{1}{10} + 2\frac{4}{10} + 3\frac{4}{10} + 4\frac{1}{10} = 2.5$$

$$Var(X) = \sum_{i=1}^{4} i^2 f_X(i) - (EX)^2 = \frac{1}{10} + 4\frac{4}{10} + 9\frac{4}{10} + 16\frac{1}{10} - (2.5)^2 = \frac{13}{20}$$

Der Median einer Zufallsvariablen entspricht dem $\frac{1}{2}$-Quantil. Somit lesen wir an der Verteilungsfunktion der Zufallsvariablen X den Median der Zufallsvariablen X ab: $Md = 2$.

Hausübung

H 19 In einer Molkerei gibt es zwei Maschinen, die Milch in Milchtüten abfüllen. Die Füllmengen von 21 Milchtüten der ersten Maschine bzw. von 9 Milchtüten der zweiten Maschine wurden gemessen. Dabei erhielt man Messwerte $x_1, \ldots, x_{21}$, $y_1, \ldots, y_9$ (in ml) mit den empirischen Mittelwerten $\overline{x} = 501$ bzw. $\overline{y} = 503$ und den empirischen Varianzen $\bar{s}_x^2 = 3.24$ bzw. $\bar{s}_y^2 = 3.61$. Unter der Annahme, dass die angegebenen Messwerte eine Realisierung unabhängiger Zufallsvariablen $X_1, \ldots, X_{21}, Y_1, \ldots, Y_9$ sind, wobei $X_1, \ldots, X_{21}$ identisch $N(\mu_1, \sigma_1^2)$ - und $Y_1, \ldots, Y_9$ identisch $N(\mu_2, \sigma_2^2)$ - verteilt sind, testen Sie

a) unter der Annahme $\sigma_1^2 = \sigma_2^2$ durch Anwendung eines geeigneten Testverfahrens zum Niveau $\alpha = 0.05$ die Hypothese $\mu_1 \geq \mu_2$ gegen die Alternative $\mu_1 < \mu_2$.

b) durch Anwendung eines geeigneten Testverfahrens zum Niveau $\alpha = 0.1$, ob aufgrund des angegebenen Datenmaterials die unter a) gemachte Annahme $\sigma_1^2 = \sigma_2^2$ gegen $\sigma_1^2 \neq \sigma_2^2$ zu verwerfen ist.

Eventuell benötigte Quantile: $F_{20;\,8;\,0.05} = \frac{1}{F_{8,20;0.95}} \approx 0.4086$, $F_{20,8;0.95} = 3.1502$.

a) *Geeignetes Testverfahren:* *t-Test für zwei unabhängige Stichproben*
Begründung: zwei unabhängige Messreihen, Testen auf Unterschied in Erwartungswerten bei gleicher Varianz

Nullhypothese: $H_0 : \mu_1 \geq \mu_2$, *Alternative* $H_1 : \mu_1 < \mu_2$

Testgröße:

$$T(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = \sqrt{\frac{m \cdot n \cdot (m+n-2)}{m+n}} \cdot \frac{\overline{Y}_n - \overline{X}_m}{\sqrt{(m-1) \cdot \bar{S}_X^2 + (n-1) \cdot \bar{S}_Y^2}}$$

Realisierung der Testgröße:

$$T(x_1, \ldots, x_{21}, y_1, \ldots, y_9) = \sqrt{\frac{21 \cdot 9 \cdot 28}{30}} \cdot \frac{503 - 501}{\sqrt{20 \cdot 3.24 + 8 \cdot 3.61}} \approx 2.7444$$

Kritischer Bereich: *Wir lehnen H_0 ab, falls gilt:*
$T(x_1, \ldots, x_m, y_1, \ldots, y_n) > t_{m+n-2;1-\alpha}$

Entscheidung: *Wegen $T(\ldots) \approx 2.7444 > 1.70 = t_{28;0.95}$ wird die Nullhypothese* ***abgelehnt***.

b) <u>*Geeignetes Testverfahren:*</u> *F-Test*
Begründung: zwei unabhängige Messreihen, Erwartungswerte unbekannt, Testen auf Gleichheit der Varianzen

<u>*Nullhypothese:*</u> $H_0 : \sigma_1^2 = \sigma_2^2$, *Alternative* $H_1 : \sigma_1^2 \neq \sigma_2^2$

<u>*Testgröße:*</u> $T(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = \dfrac{S_X^2}{\widetilde{S}_Y^2}$

<u>*Realisierung der Testgröße:*</u> $T(x_1, \ldots, x_{21}, y_1, \ldots, y_9) = \dfrac{3.24}{3.61} \approx 0.8975$

<u>*Kritischer Bereich:*</u> *Wir lehnen H_0 ab, falls gilt:*
$T(x_1, \ldots, x_m, y_1, \ldots, y_n) < F_{m-1,n-1;\alpha/2}$ oder
$T(x_1, \ldots, x_m, y_1, \ldots, y_n) > F_{m-1,n-1;1-\alpha/2}$

<u>*Entscheidung:*</u> *Wegen* $F_{20;\,8;\,0.05} \approx 0.4086 \leq T(\ldots) \approx 0.8975 \leq 3.1502 = F_{20,8;0.95}$ *wird die Nullhypothese* **nicht abgelehnt**.

H 20 Für eine Stichprobe $x_1, \ldots, x_{20}$ einer $N(\mu; \sigma^2)$-verteilten Zufallsvariable X seien folgende Kenngrößen gegeben:

$$\bar{x}_{20} = 5.18; \qquad \bar{s}_{20}^2 = 0.0054.$$

(a) Geben Sie ein 99% Konfidenzintervall für den Parameter μ der Zufallsvariablen X an.

(b) Geben Sie ein 99% Konfidenzintervall für die Varianz σ^2 der Zufallsvariablen X an.

(a) Da die Varianz nicht bekannt ist, ergibt sich ein 99%-Konfidenzintervall gemäß

$$\left[\bar{X} - \sqrt{\frac{\bar{S}_{20}^2}{20}} \cdot t_{19;0.995}; \bar{X} + \sqrt{\frac{\bar{S}_{20}^2}{20}} \cdot t_{19;0.995}\right]$$

Das Quantil der t-Verteilung ist $t_{19;0.995} = 2.86$ *und demnach ergibt sich für das Konfidenzintervall*

$$\left[5.18 - \sqrt{\frac{0.0054}{20}} \cdot 2.86; 5.18 - \sqrt{\frac{0.0054}{20}} \cdot 2.86\right] = [5.123; 5.227]\,.$$

(b) Bei unbekannten Erwartungswert läßt sich das Konfidenzintervall für die Varianz wie folgt bestimmen

$$\left[\frac{(n-1)\,\bar{S}_{20}^2}{\chi^2_{19;0.995}}; \frac{(n-1)\,\bar{S}_{20}^2}{\chi^2_{19;0.005}}\right]$$

Aus den Quantiltabellen erhalten wir $\chi^2_{19;0.995} = 38.58$ *und* $\chi^2_{19;0.005} = 6.84$. *Also lautet das Konfidenzintervall*

$$\left[\frac{19 \cdot 0.0054}{38.58}; \frac{19 \cdot 0.0054}{6.84}\right] = [0.0027; 0.015]\,.$$