Statistik I für Human- und Sozialwissenschaften
11. Übung
Lösungsvorschlag

Gruppenübung

G 37 Eine Reifenfirma hat für einen neuen Winterreifen zwei Profile entwickelt, die bezüglich ihrer Griffigkeit im Schnee und ihrer Rutschfestigkeit auf Eis nahezu gleichwertig sind. Es soll nun untersucht werden, ob sie sich im Hinblick auf ihre Bremswirkung bei trockener Fahrbahn unterscheiden. Dazu werden 10 Testfahrzeuge mit Reifen der Profilsorte A bestückt und 12 weitere Testfahrzeuge mit Reifen der Profilsorte B. Nacheinander fahren diese Fahrzeuge unter gleichen Bedingungen auf eine Teststrecke und werden jeweils bei der gleichen Geschwindigkeit abgebremst. Die ermittelten Bremswege [in m] sind in folgender Tabelle abgetragen:

Profil A	57	42	52	47	50	49	53	50	47	51		
Profil B	50	44	45	56	48	52	47	50	54	46	55	43

Wir nehmen an, dass sich die Bremswege der 10 Testfahrzeuge mit Reifen der Profilsorte A durch i.i.d. Zufallsvariablen gemäß einer Verteilungsfunktion $F(x),\ x \geq 0,$ und die Bremswege der 12 Testfahrzeuge mit Reifen der Profilsorte B durch i.i.d. Zufallsvariablen gemäß einer Verteilungsfunktion $G(x) = F(x-a),\ x \geq 0,$ für ein $a > 0$, beschreiben lassen.

(a) Formulieren Sie in Formeln und in Worten Nullhypothese und Alternativhypothese für die beschriebene Testsituation.

(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.

(a) $H_0 : G(x) = F(x)$, $x \geq 0$, *d.h. kein Unterschied bzgl. der Bremswirkung*
$H_1 : G(x) = F(x-a)$, $x \geq 0$, $a > 0$, *d.h. es besteht ein Unterschied bzgl. der Bremswirkung*

(b) U-Test von Mann-Whitney

(c)

Profil	*Wert*	*Rang*	*Rangplatz-überschreitung*
A	42	1	12
B	43	2	
B	44	3	
B	45	4	
B	46	5	
B	47	7	
A	47	7	7
A	47	7	7
B	48	9	
A	49	10	6
B	50	12.5	
B	50	12.5	
A	50	12.5	4
A	50	12.5	4
A	51	15	4
B	52	16.5	
A	52	16.5	3
A	53	18	3
B	54	19	
B	55	20	
B	56	21	
A	57	22	

Die Ränge für die Werte 47, 50 und 52 wurden als Mittelwerte der ursprünglichen Rangplätze bestimmt (für 47; $1/3 \cdot (6+7+8)$*, für 50;* $1/4 \cdot (11+12+13+14)$*, für 52;* $1/2 \cdot (16+17)$*). Des weiteren gilt*

$$n_1 = 10,\ n_2 = 12,\ k = 3,\ t_1 = 3,\ t_2 = 4,\ t_3 = 2$$

und es folgt

$$\mu_u = \frac{1}{2} \cdot 10 \cdot 12 = 60$$

$$\bar{\sigma}_u^2 = \frac{1}{12} \cdot 10 \cdot 12 \cdot (10+12+1) - \frac{10 \cdot 12 \cdot \sum_{i=1}^{3} \left(t_i^3 - t_i\right)}{12 \cdot (10+12) \cdot (10+12-1)}$$

$$\text{mit } \sum_{i=1}^{3} \left(t_i^3 - t_i\right) = (3^3-3) + (4^3-4) + (2^3-2) = 90$$

$$\bar{\sigma}_u^2 \approx 228.05.$$

aus der Quantil-Tabelle der Normalverteilung lesen wir ab $z_{0.95} = 1.645$ *(wegen* $\alpha = 0.9$*). Der Wert der Teststatistik* U *ist gleich der Summe der letzten Spalte der obigen Tabelle, d.h.* $U(x_1, \ldots, x_{10}, y_1, \ldots, y_{12}) = 50$*. Und wegen*

$$\frac{|U(x_1, \ldots, x_{10}, y_1, \ldots, y_{12}) - \mu_u|}{\sqrt{\bar{\sigma}_u^2}} \approx 0.662 < 1.645$$

kann H_0 *nicht verworfen werden. D.h. man konnte keine unterschiedliche Bremswirkung nachweisen.*

G 38 Ein Automobilclub bietet auf einem Testgelände ein Fahrtraining an, bei dem die Teilnehmer eine benzinsparende Fahrweise lernen sollen. Zum Nachweis der Wirksamkeit dieses Trainings wird von 20 Teilnehmern der Benzinverbrauch einmal vor und einmal nach dem Fahrtraining gemessen, und zwar immer im selben Fahrzeug auf derselben Teststrecke. Dabei ergeben sich folgende Messwerte [in Liter]:

Teilnehmer Nr.	1	2	3	4	5	6	7	8	9	10
Benzinverbrauch vor dem Training	1,9	2,4	3,3	2,1	3,0	1,8	2,6	3,0	3,8	2,8
Benzinverbrauch nach dem Training	2,8	3,1	1,7	2,9	1,2	2,5	2,8	3,6	2,2	1,5

Teilnehmer Nr.	11	12	13	14	15	16	17	18	19	20
Benzinverbrauch vor dem Training	3,9	2,8	4,0	3,3	2,8	3,4	3,3	3,0	3,7	3,0
Benzinverbrauch nach dem Training	1,3	1,8	1,4	2,0	3,5	1,2	1,8	1,4	2,3	1,3

Es wird angenommen, dass sich die Benzinverbräuche vor dem Fahrtraining durch i.i.d. Zufallsvariablen $X_1, \ldots, X_n$ und die Benzinverbräuche nach dem Fahrtraining durch i.i.d. Zufallsvariablen $Y_1, \ldots, Y_n$ beschreiben lassen.

(a) Formulieren Sie in Formeln und in Worten Nullhypothese und Alternativhypothese für die beschriebene Testsituation.

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

(c) Führen Sie den in (b) genannten Statistischen Test jeweils einmal auf dem Signifikanzniveau $\alpha = 0,05$ und $\alpha = 0,01$ durch und interpretieren Sie die Ergebnisse.

(a) H_0 : Die Differenzen $D_i = Y_i - X_i$, $i = 1, \ldots, n$ sind symmetrisch um Null verteilt, d.h. es gibt keine Veränderung des Benzinverbrauchs
H_1 : Die Differenzen D_i, $i = 1, \ldots, n$ sind nicht symmetrisch um Null verteilt.

(b) Der Wilcoxon-Test (Vorzeichen-Test)

(c)

Teiln. Nr.	Benzinverb. vorher X_i	Benzinverb. vorher Y_i	Differenz $D_i = Y_i - X_i$	Abs.betrag $\lvert D_i\rvert = \lvert Y_i - X_i\rvert$	Rang von $\lvert D_i\rvert$	Summanden für die Testgröße
1	1.9	2.8	0.9	0.9	7	7
2	2.4	3.1	0.7	0.7	4	4
3	3.3	1.7	-1.6	1.6	14	
4	2.1	2.9	0.8	0.8	6	6
5	3.0	1.2	-1.8	1.8	17	
6	1.8	2.5	0.7	0.7	4	4
7	2.6	2.8	0.2	0.2	1	1
8	3.0	3.6	0.6	0.6	2	2
9	3.8	2.2	-1.6	1.6	14	
10	2.8	1.5	-1.3	1.3	9.5	
11	3.9	1.3	-2.6	2.6	19.5	
12	2.8	1.8	-1.0	1.0	8	
13	4.0	1.4	-2.6	2.6	19.5	
14	3.3	2.0	-1.3	1.3	9.5	
15	2.8	3.5	0.7	0.7	4	4
16	3.4	1.2	-2.2	2.2	18	
17	3.3	1.8	-1.5	1.5	12	
18	3.0	1.4	-1.6	1.6	14	
19	3.7	2.3	-1.4	1.4	11	
20	3.0	1.3	-1.7	1.7	16	

Damit ergibt sich für den Wert der Testgröße

$$U(x_1, \ldots, x_{20}, y_1, \ldots, y_{20}) = 28,$$

wobei zu beachten ist, daß nur die Rangplätze positiver Differenzen für die Berechnung von U herangezogen werden. Es sind $n = 20$, $k = 4$,

$$\begin{aligned} t_0 = 0, \quad t_1 &= 3 && \text{0.7 kommt dreimal vor} \\ t_2 &= 2 && \text{1.3 kommt zweimal vor} \\ t_3 &= 3 && \text{1.6 kommt dreimal vor} \\ t_4 &= 2 && \text{2.6 kommt zweimal vor.} \end{aligned}$$

Daraus ergibt sich

$$\sum_{i=1}^{k} \left(t_i^3 - t_i\right) = (3^3 - 3) + (2^3 - 2) + (3^3 - 3) + (2^3 - 2) = 60$$

und somit

$$\begin{aligned} \mu_u &= \frac{1}{4}(20 \cdot 21 - 0 \cdot 1) = 105 \\ \bar{\sigma}_u^2 &= \frac{1}{24}(20 \cdot 21 \cdot 41 - 0 \cdot 1 \cdot 1) - \frac{1}{48} \cdot 60 = 716.25. \end{aligned}$$

Aus der Quantiltabelle der Normalverteilung erhalten wir $z_{0.975} = 1.960$ und $z_{0.995} = 2.576$. Außerdem gilt

$$\frac{|U(x_1, \ldots, x_{20}, y_1, \ldots, y_{20}) - \mu_u|}{\sqrt{\bar{\sigma}_U^2}} = \frac{|28 - 105|}{\sqrt{716.25}} \approx 2.877.$$

Da sowohl $2.877 > 1.9960$ als auch $2.877 > 2.576$, wird die Nullhypothese auf beiden Signifikanzniveaus verworfen. D.h. wir können eine Veränderung des Benzinverbrauchs aufgrund des Fahrtrainings nachweisen.
ABER: Da es sich um einen zweiseitigen Test handelt, können wir nicht sagen, ob der Benzinverbrauch größer oder kleiner geworden ist.

G 39 Für $\theta > 0$ sei $X_1, \ldots, X_n$ eine Folge unabhängig und identisch verteilter Zufallsvariablen mit Dichte

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & x > \theta \end{cases}$$

Zeigen Sie, daß

$$T_n = 2 \cdot \bar{X}_n$$

ein erwartungstreuer Schätzer für θ ist.
Hinweis: Berechnen Sie zuerst EX.

Es gilt

$$EX = \int_0^\infty x f(x)\, dx = \int_0^\theta x \cdot \frac{1}{\theta}\, dx = \frac{x^2}{2\theta}\Big|_0^\theta = \frac{\theta}{2}.$$

Nun berechnen wir den Erwartungswert von T_n

$$E(T_n) = 2 \cdot E(\bar{X}_n) = 2 \cdot E(X) = \theta.$$

Somit ist T_n ein erwartungstreuer Schätzer von θ.

Hausübung

H 21 Zwei Gerätetypen der gleichen Preisklasse von verschiedenen Herstellern sollen hinsichtlich ihrer Zuverlässigkeit verglichen werden. Dazu wird bei 8 Geräten vom ersten Hersteller und bei 11 Geräten vom zweiten Hersteller die Zeit [in Stunden] von der Inbetriebnahme bis zur ersten Störung ermittelt:

Geräte vom 1. Hersteller	131	213	409	556	348	227	470	503			
Geräte vom 2. Hersteller	270	144	398	196	223	152	317	120	254	266	135

Wir gehen davon aus, dass sich die Betriebszeiten der Geräte des ersten Herstellers durch i.i.d. Zufallsvariablen gemäß einer Verteilungsfunktion $F(x),\ x \geq 0,$ und die Betriebszeiten der Geräte des zweiten Herstellers durch i.i.d. Zufallsvariablen gemäß einer Verteilunsfunktion $G(x) = F(x - a),\ x \geq 0,$ für ein $a \geq 0$, beschreiben lassen.

(a) Formulieren Sie in Formeln und in Worten Nullhypothese und Alternativhypothese für die beschriebene Testsituation.

(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.

(a) $H_0 : G(x) = F(x)$, $x \geq 0$, *d.h gleiche Zuverlässigkeit beider Geräte*
$H_1 : G(x) = F(x - a)$, $x \geq 0$, $a \neq 0$, *d.h. unterschiedliche Zuverlässigkeit*

(b) U*-Test von Mann-Whitney*

(c)

Hersteller	*Zeit*	*Rang*	*Rangplatz-überschreitungen*
2	120	1	
1	131	2	10
2	135	3	
2	144	4	
2	152	5	
2	196	6	
1	213	7	6
2	223	8	
1	227	9	5
2	254	10	
2	266	11	
2	270	12	
2	317	13	
1	348	14	1
2	398	15	
1	409	16	0
1	470	17	0
1	503	18	0
1	556	19	0

Somit haben wir keine Bindungen und es gilt $n_1 = 8$, $n_2 = 11$.

$$\mu_u = \frac{1}{2} \cdot 8 \cdot 11 = 44$$
$$\bar{\sigma}_u^2 = \frac{1}{12} \cdot 8 \cdot 11 \cdot (8 + 11 + 1) \approx 146.667.$$

Aus der Quantiltabelle der Normalverteilung lesen wir ab $z_{0.95} = 1.645$. *Wegen*

$$U(x_1, \ldots, x_8, y_1, \ldots, y_{11}) = 22$$

gilt

$$\frac{|U(x_1, \ldots, x_8, y_1, \ldots, y_{11}) - \mu_u|}{\sqrt{\bar{\sigma}_u^2}} \approx 1.817.$$

Da $1.8166 > 1.645$ *wird die Nullhypothese verworfen und wir schlußfolgern, daß die Gerätetypen unterschiedlich zuverlässig sind.*

H 22 Sei $x_1, \ldots, x_n$ eine Stichprobe der iid Zufallsvariablen $X_1, \ldots, X_n$mit Dichte

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \theta \cdot e^{-\theta \cdot x} & x > 0. \end{cases}$$

Bestimmen Sie den Maximum-Likelihood-Schätzer für θ.

Sei $x_1, \ldots, x_n$ eine Stichprobe der Zufallsvariablen X. Dann sieht die Likelihood-Funktion wie folgt aus:

$$L(\theta; x_1, \ldots, x_n) = \theta^n \cdot e^{-\theta x_1} \cdot \ldots \cdot e^{-\theta x_n}.$$

Als nächstes bestimmen wir die Log-Likelihood-Funktion

$$l(\theta; x_1, \ldots, x_n) = \ln(L(\theta; x_1, \ldots, x_n)) = n \cdot \ln\theta - \theta \cdot \sum_{i=1}^{n} x_i.$$

Die erste Ableitung der Log-Likelihood-Funktion nach θ lautet

$$\frac{dl(\theta; x_1, \ldots, x_n)}{d\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i.$$

Diese Ableitung setzen wir nun gleich 0 und stellen nach θ um. Es folgt

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

Um zu überprüfen, ob es sich bei $\hat{\theta}$ wirklich um einen Maximum-Likelihood-Schätzer handelt, müssen wir uns noch die zweite Ableitung anschauen:

$$\frac{d^2 l(\theta; x_1, \ldots, x_n)}{d\theta^2} = -\frac{n}{\theta^2} < 0.$$

Somit ist $\hat{\theta}$ der gesuchte Maximum-Likelihood-Schätzer.