Lösungshinweise zur Einf. in die Math. Statistik

für Winf, LaB etc. im SS 2004, 7. Übung

G19 Die Daten $(x_1, y_1), \dots, (x_{12}, y_{12})$ des nach Verfahren A bzw. B festgestellten Alkoholspiegel werden als Realisierung der unabhängigen, zweidimensionalen Zufallsvariablen $(X_1, Y_1), \dots, (X_{12}, Y_{12})$ aufgefasst.

Dann sind die Zufallsvariablen

$$D_i = \begin{cases} 1 & \text{falls } X_i \geq Y_i \\ 0 & \text{sonst} \end{cases} \qquad i = 1, \dots, 12$$

ebenfalls unabhängig.

Geeigneter Test: Vorzeichen-Test

Der Zweistichprobentest von Wilcoxon-Mann-Whitney kann ebenso wie der Run-Test von Wald und Wolfowitz hier NICHT angewendet werden, da für festes i X_i und Y_i nicht als unabhängig angenommen werden können.

Nullhypothese H_0: $P(D_i = 1) = P(D_i = 0) = \frac{1}{2} \quad \forall i$

Testgröße: $V((X_1, Y_1), \dots, (X_{12}, Y_{12})) = \sum_{i=1}^{12} D_i$

Realisierung der Testgröße:

$$V((x_1, y_1), \dots, (x_{12}, y_{12})) = 1+1+0+1+0+0 +0+1+0+0+1+0 = \underline{5}$$

Kritischer Bereich: Unter H_0 gilt $V \sim B(12, 0.5)$

gesucht: max. $k \in \mathbb{N}$ mit $2P(V<k) \leq 0.05 = \alpha$

$$2P(V<k) = 2 \sum_{i=0}^{k-1} \binom{12}{i} \left(\frac{1}{2}\right)^{12} = \left(\frac{1}{2}\right)^{11} \sum_{i=0}^{k-1} \binom{12}{i}$$

k	1	2	3	4
$2P(V<k)$	$\left(\frac{1}{2}\right)^{11}$	0.0063	0.0386	0.1460

$\Rightarrow k = 3$

Daher ist der kritische Bereich

$$K = \{V < 3 \text{ oder } V > 12-3\}$$

Entscheidung: Wegen $3 \leq 5 \leq 9$ wird H_0 nicht abgelehnt.

G20 a) Zweistichprobentest von Wilcoxon-Mann-Whitney

Annahmen s. Aufgabenstellung

Nullhypothese H_0: $F = G \qquad H_1: F < G$ oder $F > G$

Testgröße: Anzahl der Inversionen „y vor x"

$$U(X_1, \dots, X_{10}, Y_1, \dots, Y_{10}) = \sum_{i=1}^{10} \sum_{j=1}^{10} Z_{ij}$$

mit

$$Z_{ij} = \begin{cases} 1 & \text{falls } X_i > Y_j \\ 0 & \text{sonst} \end{cases}$$

Realisierung der Testgröße:

Nach Sortieren erhält man

xxxy x x y x x y x y x yyyy x yy

$U(x_1, \dots, x_{10}, y_1, \dots, y_{10}) =$

$$7\cdot 1 + 5\cdot 1 + 3\cdot 1 + 2\cdot 1 + 1\cdot 4 = \underline{21}$$

für das y ganz links in der geordneten Messreihe

für die Anzahl der x, die rechts von y stehen

<u>Kritischer Bereich:</u>

U ist näherungsweise $N\left(\frac{10\cdot 10}{2}, \frac{1}{12}\, 10\cdot 10\cdot 21\right)$ - verteilt

Formel (120) liefert mit $u_{0.975} = 1.96$ den Wert

$$k = 50 - u_{0.975}\sqrt{175} \approx \underline{24.0716}$$

Wir lehnen H_0 daher ab, falls $U < 24.0716$ oder $U > 10\cdot 10 - 24.0716$ gilt.

<u>Entscheidung:</u>

Wegen $21 < 24.0716$ wird <u>H_0 abgelehnt.</u>

b) <u>Run-Test von Wald und Wolfowitz</u>

Annahmen s. Aufgabenstellung

<u>Nullhypothese H_0:</u> $F = G$ $\quad H_1$: $F \neq G$

<u>Testgröße:</u> Anzahl $R(X_1, \dots, X_{10}, Y_1, \dots, Y_{10})$ von Runs in der sortierten Folge der Messwerte

<u>Realisierung der Testgröße:</u> $R(x_1, \dots, x_{10}, y_1, \dots, y_{10}) = 12$

(vgl. xxx y xx y xx y x y x yyyy x yy)

1 2 3 4 5 6 7 8 9 10 11 12

<u>Kritischer Bereich:</u>

R ist näherungsweise $N\left(1 + \frac{2\cdot 10\cdot 10}{20}, \frac{2\cdot 10\cdot 10(2\cdot 100 - 10 - 10)}{(10+10)^2(10+10-1)}\right)$

verteilt, also $N(11, \frac{90}{19})$ - verteilt.

Formel (123) liefert mit $u_{0.95} = 1.645$ den Wert

$$k = 11 - u_{0.95}\sqrt{90/19} \approx \underline{7.4198}$$

Wir lehnen H_0 ab, falls $R < 7.4198$ gilt.

<u>Entscheidung:</u>

Wegen $12 \geq 7.4198$ wird <u>H_0 nicht abgelehnt.</u>

<u>G21</u> <u>Kruskal-Wallis-Test</u>

<u>Voraussetzungen:</u> $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}, Z_1, \dots, Z_{n_3}$ unabhängig und jeweils identisch verteilt wie X, bzw. Y bzw. Z

<u>Nullhypothese H_0:</u> $F_X = F_Y = F_Z$

<u>Testgröße</u> (Formel 129):

$$K(x_{11}, \dots, x_{kn_k}) = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left(\bar{r}_i - \frac{n+1}{2}\right)^2$$

<u>Realisierung der Testgröße:</u>

$n = 12$

i	n_i	r_{i1}	...		r_{in_i}	$\bar{r}_i$
1	4	1	3	4	5	13/4
2	4	2	6	8	10	26/4
$(k=)$ 3	4	7	9	11	12	39/4

$$K(\dots) = \frac{12}{12\cdot 13}\, 4 \sum_{i=1}^{3} \left(\bar{r}_i - \frac{13}{2}\right)^2 = \underline{6.5}$$

Kritischer Bereich:

Die Testgröße ist näherungsweise χ^2_{k-1}- verteilt.

Wir lehnen H_0 daher ab, falls gilt:

$$K(\ldots) > \chi^2_{k-1;\,1-\alpha} = \chi^2_{2;0.9} = 4.605$$

Entscheidung:

Wegen $K(\ldots) = 6.5 > 4.605$ wird H_0 verworfen.

G22 Geeigneter Test: Einfache Varianzanalyse

Annahmen: s. Aufgabentext

Nullhypothese H_0: $\mu_1 = \mu_2 = \mu_3$

Testgröße:

$$T = \frac{\frac{1}{k-1}\,SST}{\frac{1}{n-k}\,SSE} \qquad k=3,\ n=28$$

Realisierung der Testgröße:

Gesamtmittel: $\bar{x} = \frac{1}{n}\sum_{i=1}^{3} n_i \bar{x}_i$

$$= \frac{1}{28}(11\cdot 7 + 7\cdot 7.1 + 10\cdot 6.9)$$

$$\approx \underline{6.9893}$$

$$\frac{1}{2}\,sst = \frac{1}{2}\sum_{i=1}^{3} n_i(\bar{x}_i - \bar{x})^2$$

$$= \frac{1}{2}\left[11\,(7-6.9893)^2 + 7\,(7.1-6.9893)^2 + 10\,(6.9-6.9893)^2\right]$$

$$\approx \underline{0.0834}$$

$$\frac{1}{25}\,sse = \frac{1}{25}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 \overset{\text{Hinweis}}{=} \frac{1}{25}\sum_{i=1}^{3}(n_i - 1)\,s_i^2$$

$$= \frac{1}{25}\left[10\cdot 0.03 + 6\cdot 0.02 + 9\cdot 0.033\right]$$

$$\approx \underline{0.0287}$$

$$T(\ldots) = \frac{\frac{1}{2}\,sst}{\frac{1}{25}\,sse} \approx \underline{2.9059}$$

Kritischer Bereich:

H_0 wird abgelehnt, falls $T > F_{k-1,n-k;\,1-\alpha} = F_{2,25;0.95}$

Entscheidung:

Wegen $2.9059 < F_{2,25;0.95} = 3.3852$ wird H_0 nicht verworfen.