6. Lösung weitere Übungsaufgaben Statistik II WiSe 2018/2019

1. Aufgabe: In einer Studie soll der Erfolg von 3 verschiedenen Düngemitteln verglichen werden. Diese wurden auf 22 Erdbeerfeldern verwendet. Dabei wurden folgende Erträge in Kilogramm erzielt. (Für die einzelnen Düngemittel sind die Erträge schon der Größe nach geordnet.)

Düngemittel 1	44	45	55	61	63	67	78		
Düngemittel 2	62	68	75	79	80	81	86	90	92
Düngemittel 3	72	85	89	100	101	121			

Die Erträge sind **nicht** normalverteilt. Testen Sie zum Niveau $\alpha = 0{,}01$, ob die erwarteten Erträge der 3 Düngemittel gleich sind oder sich signifikant voneinander unterscheiden.

Lösung:

X_{ij} - zufällige Ertrag bei Düngemittel i
$\mathbf{E}(X_{ij}) = \mu_i \qquad i = 1, \dots, 3$ $(p = 3)$ und $j = 1, \dots, n_i$
$n_1 = 7$, $n_2 = 9$ und $n_3 = 6$.

Da die Erträge nicht normalverteilt sind verwenden wird den Kruskal-Walis-Test.

1.) $H_0 : \mu_1 = \mu_2 = \mu_3$ gegen $H_A : \mu_i \neq \mu_j$ für mindestens ein Paar i, j

2.) $\alpha = 0{,}05$

3.)

$$T = \frac{1}{B}\left[\frac{12}{n(n+1)}\sum_{i=1}^{p}\frac{1}{n_i}r_{i.}^2 - 3(n+1)\right]$$

4.) $n_i > 5$ für $i = 1, \dots, 3$ und $n_2 = 9 > 8 \Longrightarrow$ approximativ möglich

$$K = \{T \mid T > \chi^2_{p-1,1-\alpha}\} = \{T \mid T > \chi^2_{2,0.99} = 9,21\}$$

5.) Bestimmung der Ränge:

Rang(D_1)	1	2	3	4	6	7	11			$r_{1.} = 34$
Rang(D_2)	5	8	10	12	13	14	16	18	19	$r_{2.} = 115$
Rang(D_3)	9	15	17	20	21	22				$r_{3.} = 104$

Bindungskorrektur: Es liegen keine Bindungen vor also ist $B = 1$.

$$\begin{aligned} t &= \frac{1}{B}\left[\frac{12}{n(n+1)}\sum_{i=1}^{p}\frac{1}{n_i}r_{i.}^2 - 3(n+1)\right] \\ &= \frac{1}{1}\left[\frac{12}{22\cdot 23}\left(\frac{1}{7}34^2 + \frac{1}{9}115^2 + \frac{1}{6}104^2\right) - 3\cdot 23\right] \\ &= \underline{12,52} \end{aligned}$$

6.) $t = 12,52 > 9,21 \quad \Longrightarrow \quad H_0$ wird abgelehnt, d.h. die erwarteten Erträge unterscheiden sich sich signifikant bei den unterschiedlichen Düngemitteln.

---

2. Aufgabe:

Ein Gesundheitsmagazin möchte untersuchen, ob sich der Kaloriengehalt von Fetakäse, hergestellt aus Kuh-, Schafs- bzw. Ziegenmilch, unterscheidet.

Dazu wurde bei verschiedenen Produkten aus unterschiedlichen Supermärkten der Kaloriengehalt pro 100g Fetakäse ermittelt.

Die Ergebnisse für 21 verschiedene Fetakäse liegen in der folgenden Tabelle vor (Messwerte in kcal/100g). Dabei sind die Werte je Milchsorte schon sortiert.

Fetakäse aus Schafsmilch	220	227	231	246	265	268			
Fetakäse aus Kuhmilch	237	242	243	245	251	266	270	288	298
Fetakäse aus Ziegenmilch	145	207	212	218	230	235			

Die Kaloriengehalte sind **nicht** normalverteilt. Testen Sie zum Niveau $\alpha = 0{,}01$, ob die erwarteten Kaloriengehalte der Fetakäse der drei Milchsorten gleich sind oder sich signifikant voneinander unterscheiden.

Lösung:

In dieser Aufgabe wird das Merkmal *Kaloriengehalt von Fetakäse* untersucht und man möchte wissen, ob es vom Einflussfaktor *Art der Milch* abhängt. Dieser Einflussfaktor tritt in $p = 3$ Stufen auf (Schaf-, Kuh-, Ziegenmilch).

Insgesamt liegen $N = 21$ Beobachtungen vor, davon jeweils $n_1 = 6$, $n_2 = 9$ und $n_3 = 6$ zu den Faktorstufen Schaf-, Kuh-, Ziegenmilch. Zudem gelten die Bezeichnungen

X_{ij} ... Kaloriengehalt der j-ten Beobachtung in der i-ten Faktorstufe
$i = 1, ..., n_j$ und $j = 1, 2, 3$
μ_i ... Erwartungswert in in der i-ten Faktorstufe ($E[X_{ij}] = \mu_i$)

Da das Merkmal keiner Normalverteilungsvorraussetung genügt, wird hier nicht die Varianzanalyse sondern der Kruskal-Wallis Test durchgeführt. Die Durchführung des Test folgt dem gewohnten Schema:

1) Das Hypothesenpaar lautet

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad v.s. \quad H_A : \mu_i \neq \mu_j \text{ für ein Paar } (i, j)$$

2) Das Konfidenzniveau ist $\alpha = 0, 01$.

3) Die Teststatistik lautet

$$T = \frac{1}{B}\left[\frac{12}{N(N+1)}\left(\sum_{i=1}^{p}\frac{1}{n_i}r_{i\bullet}^2\right) - 3(N+1)\right] \tag{1}$$

Der Variable B entspricht einem Korrekturglied, für den Fall, dass Bindung in den Daten vorliegt. Da hier keine Bindung (also keine Werte mehrfach vorkommen) vorliegt, gilt

$$B = 1$$

4) $n_i > 5$ für $i = 1, \ldots, 3$ und $n_2 = 9 > 8 \Longrightarrow$ approximativ möglich.

   Der Kritische Bereich lautet

$$K = \{t \,|\, t > \chi^2_{p-1;\,1-\alpha}\}$$

   Aus der Tabelle für die Quantile der Chi-quadrat-Verteilung erhält man $\chi^2_{p-1;\,1-\alpha} = \chi^2_{2;\,0{,}99} = 9{,}21$ und damit $K = \{t|t > 9{,}21\}$.

5) Die gemittelten Rangwerte $r_{i\bullet}$ berechnen sich wie folgt: Zunächst bestimmt man die Ränge der vorliegenden Daten:

$r_{i,j}$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$
Schafsmilch, $i=1$	5	6	8	14	16	18			
Kuhmilch, $i=2$	10	11	12	13	15	17	19	20	21
Ziegenmilch, $i=1$	1	2	3	4	7	9			

   Die Werte $r_{i\bullet}$ erhält man dann durch Summieren über die einzelnen Faktorstufen:

$$\begin{aligned} r_{1\bullet} &= 5+6+8+14+16+18 = 67 \\ r_{2\bullet} &= 10+11+12+13+15+17+19+20+21 = 138 \\ r_{3\bullet} &= 1+2+3+4+7+9 = 26 \end{aligned}$$

   Einsetzen in die Gleichung (1) ergibt

$$\begin{aligned} t &= \frac{12}{21(21+1)}\left(\frac{1}{6}\cdot 67^2 + \frac{1}{9}\cdot 138^2 + \frac{1}{6}\cdot 26^2\right) - 3\cdot(21+1) \\ &\approx 11{,}32 \end{aligned}$$

6) Es gilt $t = 11{,}32 > 9{,}21 \Longrightarrow t \in K \Longrightarrow H_0$ wird abgelehnt.

D.h. die Erwartungswerte bezüglich der Faktorstufen unterscheiden sich signifikant. Der Einflussfaktor *Milchsorte* hat also einen signifikanten Einfluss auf den durchschnittlichen Kaloriengehalt.