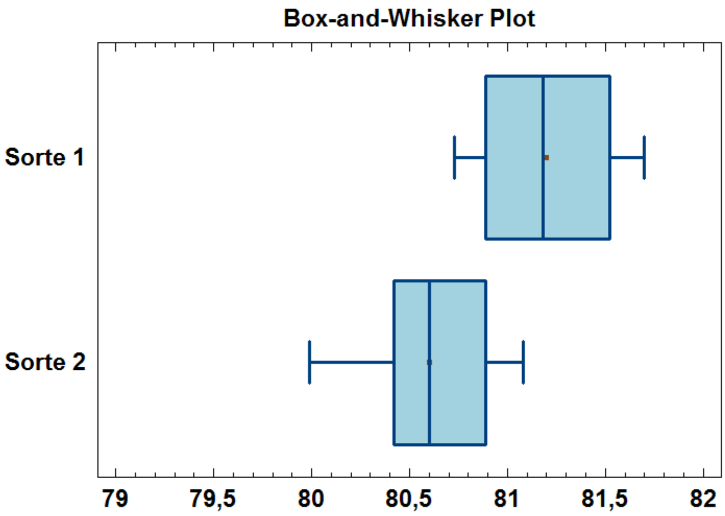
3. Lösung weitere Übungsaufgaben Statistik II WiSe 2019/2020

1. Aufgabe: Es ist bekannt, dass Mineralwasser mit einem relativ hohen Magnesiumgehalt empfehlenswert ist. Für zwei verschiedene Mineralwassersorten wurden unabhängig voneinander die folgenden Magnesiumgehalte in $\frac{mg}{l}$ erhoben.

Sorte 1	80,89	81,70	80,73	81,39	81,52	81,05	80,98	81,31	80,83	81,60
Sorte 2	80,56	79,99	81,03	80,64	80,21	80,66	80,52	81,08	80,42	80,89



Die Magnesiumgehalte sind normalverteilt mit der gleichen Varianz bei beiden Sorten.

Für das Wasser der Sorte 1 ist der mittlere Magnesiumgehalt 81,2 und die empirische Varianz $s_1^2 = 0{,}1206$. Bei der zweiten Sorte ist die empirische Varianz $s_2^2 = 0{,}1188$.

Testen Sie zum Niveau $\alpha = 0{,}01$, ob der erwartete Magnesiumgehalt bei Sorte 1 signifikant größer ist als bei Sorte 2. Begründen Sie die Wahl des verwendeten Tests kurz!

Lösung:

X_1 - zufälliger Mangnesiumgehalt bei Sorte 1.
$X_{1i} \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $i = 1,..,10$, d.h. $n_1 = 10$ und aus den Daten erhält man $\overline{x}_1 = 81{,}2$ und $s_1^2 = 0{,}1206$.

X_2 - zufälliger Mangnesiumgehalt bei Sorte 2.
$X_{2i} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $i = 1,..,10$, d.h. $n_2 = 10$ und aus den Daten erhält man $\overline{x}_2 = 80{,}6$ und $s_2^2 = 0{,}1188$.

Man wählt hierbei den doppelten t-Test, da zwei unabhängige, normalverteilte Stichproben vorliegen, beide die gleiche Varianz besitzen und auf (Lage) Erwartungswerte getestet wird.

1. $H_0 : \mu_1 \leq \mu_2$ gegen $H_A : \mu_1 > \mu_2$
2. $\alpha = 0{,}01$
3. $T = \frac{\overline{X}_1 - \overline{X}_2}{S_g} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$
4. $K = \{t \mid t \geq t_{n_1+n_2-2;1-\alpha}\}$ $\qquad t_{18;\,0{,}99} = 2{,}55$
5. $s_g^2 = \frac{1}{n_1+n_2-2}\left((n_1-1)s_1^2 + (n_2-2)s_2^2\right) = \frac{1}{18}(9 \cdot 0{,}1206 + 9 \cdot 0{,}1188) = 0{,}1197$

   $t = \frac{81{,}2-80{,}6}{\sqrt{0{,}1197}} \cdot \sqrt{\frac{100}{20}} = 3{,}88$
6. $t = 3{,}88 > 2{,}55 \implies t \in K \implies H_0$ wird abgelehnt.

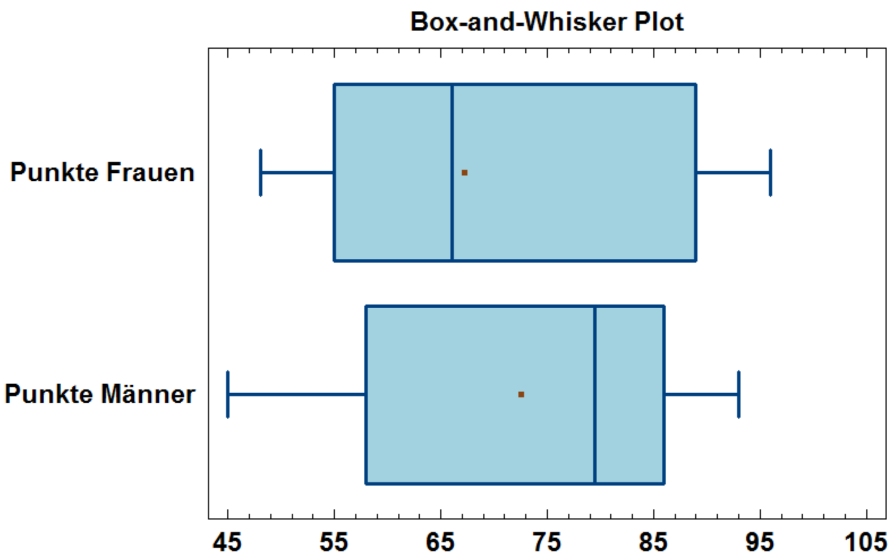
   Der erwartete Magnesiumgehalt von Sorte 1 ist signifikant größer als der von Sorte 2.

---

2. Aufgabe: Für die Vergabe von Praktikumsplätzen führt eine Firma einen Test durch. Bei diesem Test sind maximal 100 Punkte erreichbar. Der Praktikumsleiter vermutet, dass Frauen bei diesem Test besser abschneiden als Männer.

Testen Sie diese Vermutung anhand der vorliegenden Daten (Punktezahl) zum Signifikanzniveau von 5%!

Frauen	48	49	55	55	59	66	67	67	89	89	96
Männer	45	49	58	63	78	81	83	86	89	93	



Hinweis: Die erreichte Punktezahl ist **nicht** normalverteilt.

Lösung:

X_1 - zufällige Punktezahl der Frauen ($n_1 = 11,\ \mu_1 = \mathbf{E}X_1$)
X_2 - zufällige Punktezahl der Männer ($n_2 = 10,\ \mu_2 = \mathbf{E}X_2$)

Da die erreichte Punktezahl nicht normalverteilt ist wird als Test für einen Lagevergleich der Wilcoxon-Rangsummentest verwendet.

1. $H_0 : \mu_1 \leq \mu_2$ gegen $H_A : \mu_1 > \mu_2$
2. $\alpha = 0{,}05$
3. $n_1 = 11 \geq 4,\ n_2 = 10 \geq 4$ und $n_1 + n_2 = 21 \geq 20$, damit kann der approximative Test verwendet werden.
$$T = \frac{R_1 - \frac{1}{2} \cdot n_1 \cdot (n_1 + n_2 + 1)}{\sqrt{\frac{1}{12} \cdot n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}}$$
4. $K = \{t \mid t \geq z_{1-\alpha}\}$ $z_{0{,}95} = 1{,}6449$
5. Bestimmung der Ränge:

Frauen	48	49	55	55	59	66	67	67	89	89	96
Rang	2	3,5	5,5	5,5	8	10	11,5	11,5	18	18	21
Männer	45	49	58	63	78	81	83	86	89	93	
Rang	1	3,5	7	9	13	14	15	16	18	20	

$r_1 = 2 + 3{,}5 + \ldots + 18 + 21 = 114{,}5$ und damit
$$t = \frac{114,5 - \frac{1}{2} \cdot 11 \cdot (11 + 10 + 1)}{\sqrt{\frac{1}{12} \cdot 11 \cdot 10 \cdot (11 + 10 + 1)}} = -0{,}457$$
6. $t = -0{,}457 \not\geq 1{,}6449 \implies t \notin K \implies H_0$ wird angenommen.

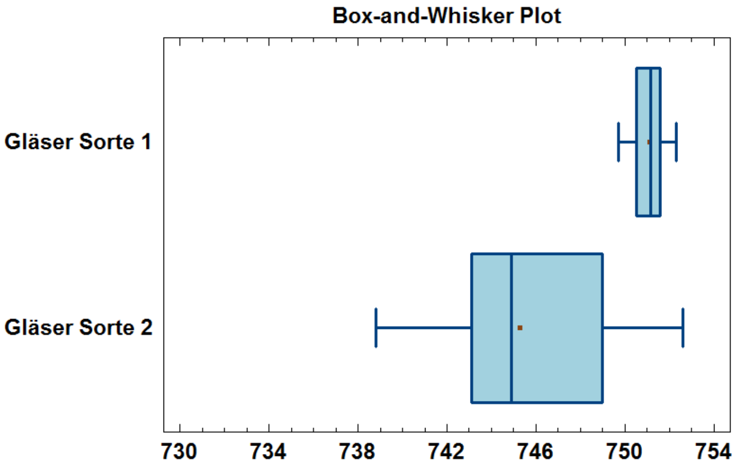
Die erreichte Punktezahl der Frauen ist nicht signifikant größer als die der Männer.

Beim exakten Wilcoxon-Rangsummentest wird auch H_0 angenommen. Im Unterschied zum approximativen Test ist dort:

3. $T = R_1$
4. $K = \{t \mid t \geq w_{n_1,n_2;1-\alpha}\}$ $n_1 = 11, n_2 = 10$
$$\begin{aligned} w_{11,10;0{,}95} &= 11 \cdot (11 + 10 + 1) - w_{11,10;0{,}05} \\ &= 242 - 97 = 145 \end{aligned}$$
5. $t = r_1 = 114{,}5$
6. $t = 114{,}5 \not\geq 145 \implies t \notin K \implies H_0$ wird angenommen. ...

3. Aufgabe: Eine neue Sorte von Reagenzgläsern soll bezüglich ihrer Schmelztemperatur mit einer gebräuchlichen Sorte verglichen werden. Aus der Tagesproduktion wurden zufällig und unabhängig voneinander jeweils 10 Reagenzgläser entnommen und deren Schmelztemperaturen in °C wie folgt bestimmt.

	i	1	2	3	4	5	6	7	8	9	10
Sorte 1	X_{1i}	751,1	751,2	750,9	751,3	752,3	750,3	751,6	752,1	750,5	749,7
Sorte 2	X_{2i}	750,1	743,2	747,2	739,2	745,3	743,1	749,0	752,6	744,5	738,8



Sorte 1 sind die neuen Reagenzgläser und Sorte 2 die herkömmlichen Reagenzgläser.

Die empirischen Standardabweichungen sind $s_1 = 0{,}80$ bei Sorte 1 und $s_2 = 4{,}51$ bei Sorte 2. Der Mittelwert ist bei Sorte 1 $\overline{x}_1 = 751{,}1$ und bei Sorte 2 $\overline{x}_2 = 745{,}3$. Die Schmelztemperaturen beider Sorten sind jeweils normalverteilt.

a) Testen Sie zum Niveau $\alpha = 0{,}05$, ob die Varianzen der Schmelztemperatur der beiden Sorten gleich sind, oder sich signifikant unterscheiden.

b) Testen Sie zum Niveau $\alpha = 0{,}05$, ob die erwartete Schmelztemperatur bei Sorte 1 signifikant größer ist als bei Sorte 2.

Lösung:

X_1 - zufällige Schmeltztemperatur bei Sorte 1.
$X_{1i} \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad i = 1, .., 10$, d.h. $n_1 = 10$ und aus den Daten erhält man $\overline{x}_1 = 751{,}1$ und $s_1 = 0{,}80$.

X_2 - zufällige Schmeltztemperatur bei Sorte 2.
$X_{2i} \sim \mathcal{N}(\mu_2, \sigma_2^2), \quad i = 1, .., 10$, d.h. $n_2 = 10$ und aus den Daten erhält man $\overline{x}_2 = 745{,}3$ und $s_2 = 4{,}51$.

Es liegen zwei unabhängige normalverteilte Stichproben vor.

a) Test auf Gleichheit der Varianzen:

1. $H_0 : \sigma_1^2 = \sigma_2^2$ gegen $H_A : \sigma_1^2 \neq \sigma_2^2$
2. $\alpha = 0{,}05$
3. $$T = \frac{S_1^2}{S_2^2}$$
4. $K = \left\{ t \mid t \leq F_{n_1-1, n_2-1; \frac{\alpha}{2}} \text{ oder } t \geq F_{n_1-1, n_2-1; 1-\frac{\alpha}{2}} \right\}$
   $F_{9,9;\,0{,}025} = \frac{1}{F_{9,9;\,0{,}975}} = \frac{1}{4{,}03} = 0{,}248$ und $F_{9,9;\,0{,}975} = 4{,}03$
5. $$t = \frac{0{,}8^2}{4{,}51^2} = 0{,}031$$
6. $t = 0{,}031 < 0{,}248 \implies t \in K \implies H_0$ wird abgelehnt.

   Die Varianzen der Schmelztemperatur der beiden Sorten unterscheiden sich signifikant.

b) Man wählt hierbei den Welch-Test, da zwei unabhängige, normalverteilte Stichproben vorliegen, die Varianzen sich signifikant voneinander unterscheiden (vgl. a)) und auf (Lage) Erwartungswerte getestet wird.

1. $H_0 : \mu_1 \leq \mu_2$ gegen $H_A : \mu_1 > \mu_2$
2. $\alpha = 0{,}05$
3. $T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
4. $K = \{ t \mid t \geq t_{m;1-\alpha} \}$ $t_{9;\,0{,}95} = 1{,}83$

   $$m = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2 - 1}} \right]_{int} = \left[\frac{\left(\frac{0{,}8^2}{10} + \frac{4{,}51^2}{10} \right)^2}{\frac{\left(\frac{0{,}8^2}{10} \right)^2}{9} + \frac{\left(\frac{4{,}51^2}{10} \right)^2}{9}} \right]_{int} = [9{,}566]_{int} = 9$$
5. $$t = \frac{751{,}1 - 745{,}3}{\sqrt{\frac{0{,}8^2}{10} + \frac{4{,}51^2}{10}}} = 4,00$$
6. $t = 4{,}00 > 1{,}83 \implies t \in K \implies H_0$ wird abgelehnt.

   Die erwartete Schmelztemperatur von Sorte 1 ist signifikant größer als die von Sorte 2.

---

4. **Aufgabe:**

In der ersten Novemberwoche des Jahres 2016 wurden bei den Wetterstationen Nossen und Dresden-Strehlen, laut wetter.com, die folgenden Tageshöchsttemperaturen (in °C) gemessen.

Tag	1.Nov.	2.Nov.	3.Nov.	4.Nov.	5.Nov.	6.Nov.	7.Nov.
Nossen	12,4	7,7	5,4	7,2	9,2	6,2	3,4
Dresden-Strehlen	15,3	10,2	7,3	9	7,5	8	5,4

Testen Sie zum Niveau $\alpha = 0,05$, ob die Tageshöchsttemperaturen der beiden Orte sich signifikant unterscheiden.

Lösung:

X_i - zufällige Tageshöchsttemperaturen in Nossen
Y_i - zufällige Tageshöchsttemperaturen in Dresden-Strehlen

Da die Tageshöchsttemperaturen an gleichen Tagen bestimmt wurden handelt es sich um eine verbundene Stichprobe (X_i, Y_i), $i = 1, .., 7$.

Durch die Bildung der Differenz der beiden Stichproben erhält man eine Stichprobe $D_i = X_i - Y_i$. Wir verwenden im Folgenden den Wilcoxon-Vorzeichentest, da Normalverteilung der Tageshöchsttemperaturen nicht gegeben ist.

M ist der Median von D.

1. $H_0:\ M = 0$ gegen $H_A:\ M \neq 0$
2. $\alpha = 0{,}05$
3. $T = \sum_{i=1}^{n} R_i^+ \cdot Z_i$
4. $K = \left\{ t \mid t \leq w^+_{n,\frac{\alpha}{2}} \text{ oder } t \geq w^+_{n,1-\frac{\alpha}{2}} \right\}$,

$n = 7$, $w^+_{7;0{,}025} = 2$ und damit $w^+_{7;0{,}975} = \frac{7\cdot 8}{2} - w^+_{7;0{,}025} = 28 - 2 = 26$

$$\Longrightarrow K = \{t \mid t \leq 2 \text{ oder } t \geq 26\}$$

5.

i	Nossen x_i	Dresden-Strehlen y_i	$d_i = x_i - y_i$	$\lvert d_i\rvert$	Rang $\lvert d_i\rvert$	Rang $\lvert d_i\rvert$ bei $d_i > 0$
1	12,7	15,3	-2,9	2,9	7	
2	7,7	10,2	-2,5	2,5	6	
3	5,4	7,3	-1,9	1,9	4	
4	7,2	9	-1,8	1,8	2,5	
5	9,2	7,5	1,7	1,7	1	1
6	6,2	8	-1,8	1,8	2,5	
7	3,4	5,4	-2	2	5	
						t=1

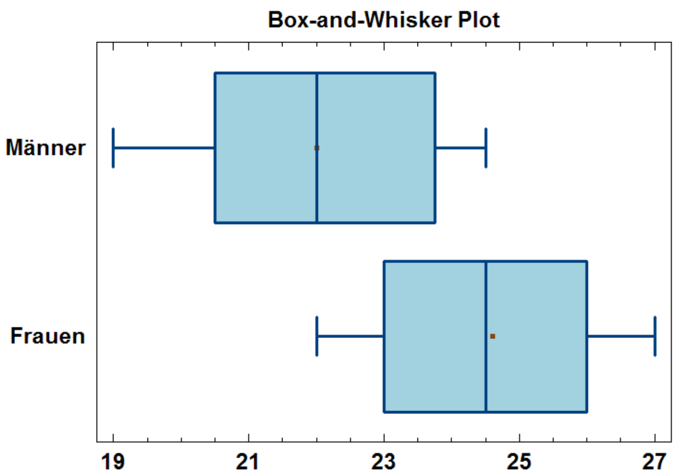
6. $t = 1 < 2 \quad \Longrightarrow \quad t \in K \quad \Longrightarrow H_0$ wird abgelehnt. D.h. die Tageshöchsttemperaturen an den beiden Orten Nossen und Dresden-Strehlen unterscheiden sich signifikant.

Bemerkungen:

- Da die Symmetrie von D_i nicht überprüft wurde und damit nicht gesichert ist, sollte man besser nur den (einfachen) Vorzeichentest verwenden.
- Achtung: Die Differenzen D_i, $i = 1, .., 7$ der Tagestemperaturen der beiden Orte sind nicht unabhängig, da es sich um aufeinanderfolgende Tage handelt. Es wäre besser gewesen, die 7 Tage in einen Jahr zufällig auszuwählen.

5. Aufgabe: Der Abteilungsleiter bemüht sich um möglichst optimale Arbeitsbedingungen für seine Mitarbeiter. So wurde bei einer Befragung unter anderem die Wohlfühltemperatur am Arbeitsplatz erfragt. Der Abteilungsleiter vemutet, dass es geschlechtsspezifische Unterschiede gibt. Es liegen folgende Daten (in °C) vor.

Männer	19	20	21	21,5	22,5	23,5	24	24,5		
Frauen	22	23	23	24	24	25	25,5	26	26,5	27



Die Wohlfühltemperatur ist nicht normalverteilt. Testen Sie zum Signifikanzniveau $\alpha = 0{,}05$, ob die (erwartete) Wohlfühltemperatur der Frauen signifikant größer als die der Männer ist.

Lösung:

X_1 - zufällige Wohlfühltemperatur der Männer ($n_1 = 8$, $\mu_1 = \mathbf{E}X_1$)
X_2 - zufällige Wohlfühltemperatur der Frauen ($n_2 = 10$, $\mu_2 = \mathbf{E}X_2$)

Da die Wohlfühltemperatur nicht normalverteilt ist wird als Test für einen Lagevergleich der Wilcoxon-Rangsummentest verwendet.

1. $H_0: \mu_1 \geq \mu_2$ gegen $H_A: \mu_1 < \mu_2$
2. $\alpha = 0{,}05$

3. $T = R_1$ (Rangsumme der ersten Stichprobe)
4. $K = \{t \mid t \leq w_{n_1,n_2,\alpha}\}$, $n_1 = 8$, $n_2 = 10$, $w_{8,10,0.05} = 56 \Longrightarrow K = \{t \mid t \leq 56\}$
5.

x_1	19	20	21	21,5	22,5	23,5	24	24,5		
x_2	22	23	23	24	24	25	25,5	26	26,5	27
Rang(x_1)	1	2	3	4	6	9	11	13		
Rang(x_2)	5	7,5	7,5	11	11	14	15	16	17	18

$$t = 1 + 2 + 3 + 4 + 6 + 9 + 11 + 13 = 49$$

6. $t = 49 < 56 \quad \Longrightarrow \quad t \in K \quad \Longrightarrow H_0$ wird abgelehnt. D.h. die erwartete Wohlfühltemperatur der Frauen ist signifikant größer als die der Männer.

<u>Hinweise</u> zur Bestimmung der Ränge bei den Bindungen:

Die 23 kommt zwei mal vor und die Ränge 7 und 8 müssen vergeben werden:

$$\frac{7+8}{2} = 7{,}5.$$

Beide erhalten den Rang 7,5.

Die 24 kommt drei mal vor und die Ränge 10, 11 und 12 müssen vergeben werden:

$$\frac{10+11+12}{3} = 11.$$

Alle drei erhalten den Rang 11.