

Musterlösung zu Serie 10

1. a) **Gepaarte Stichprobe:** Zu jeder Blutplättchenmenge vor dem Rauchen gehört die Blutplättchenmenge der selben Person nach dem Rauchen.

Einseitiger Test: Wir wollen nicht wissen, ob sich die Blutplättchenmenge *verändert* hat, sondern ob sie sich *erhöht* hat.

H_0 : Rauchen hat keinen Einfluss auf die Anhäufung der Blutplättchen. ($\mu_R = \mu_{NR}$)

H_A : Durch Rauchen erhöht sich die Anhäufung der Blutplättchen. ($\mu_R > \mu_{NR}$)

- b) **Gepaarte Stichprobe:** Zu jeder Höhe eines selbstbefruchteten Setzlings gehört die Höhe des fremdbefruchteten "Partners".

Einseitiger Test: Wir wollen nicht wissen, ob sich die Höhen *unterscheiden*, sondern ob die fremdbefruchteten Setzlinge *grösser* werden als die selbstbefruchteten.

H_0 : Die Höhen unterscheiden sich nicht. ($\mu_f = \mu_s$)

H_A : Fremdbefruchtete Setzlinge werden grösser als selbstbefruchtete. ($\mu_f > \mu_s$)

- c) **Ungepaarte Stichprobe:** Ungleiche Anzahl in den Gruppen. Zu einem Blutdruck aus der Versuchsgruppe gehört nicht ein spezifischer aus der Kontrollgruppe.

Zweiseitiger Test: Wir wollen nur wissen, ob das Kalzium einen Einfluss hat auf den Blutdruck, *egal* ob nach oben oder unten.

H_0 : Kalzium hat keinen Einfluss auf den Blutdruck. ($\mu_{\text{Kalz}} = \mu_{\text{Kontr}}$)

H_A : Kalzium hat einen Einfluss auf den Blutdruck. ($\mu_{\text{Kalz}} \neq \mu_{\text{Kontr}}$)

- d) **Ungepaarte Stichprobe:** Die Anzahlen in den beiden Gruppen brauchen nicht gleich zu sein. Zur Eisenmessung einer "Fe²⁺-Maus" gehört nicht eine bestimmte Messung einer "Fe³⁺-Maus".

Zweiseitiger Test: Wir wollen nur wissen, ob die Mäuse die verschiedenen Eisenformen *unterschiedlich* gut aufnehmen.

H_0 : Die Eisenaufnahme ist von der Form unabhängig. ($\mu_2 = \mu_3$)

H_A : Die Eisenaufnahme ist von der Form abhängig. ($\mu_2 \neq \mu_3$)

2. a) Es handelt sich um **gepaarte** Stichproben. Am gleichen Ort wird mit beiden Geräten gemessen. Deshalb führen wir einen Ein-Stichproben t-Test basierend auf der Differenz der Werte durch.

- b) • **Modell:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, σ wird durch $\hat{\sigma}$ geschätzt.

• **Nullhypothese:** $H_0 : \mu = \mu_0 = 0$.

Alternative: $H_A : \mu < \mu_0$.

• **Teststatistik:**

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}}$$

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} = \frac{\sqrt{9}(-5.78)}{6.2} = -2.8$$

Verteilung der Teststatistik unter H_0 : $T \sim t_{n-1}$.

• **Signifikanzniveau:** $\alpha = 5\%$.

• **Verwerfungsbereich:**

$$K = (-\infty, -t_{9-1,0.95}] = (-\infty, -1.86]$$

• **Testentscheid:** Der Wert t der Teststatistik liegt im Verwerfungsbereich, d.h. eine neue Eichung der Geräte ist angezeigt.

- c) Z wäre binomialverteilt mit Parametern $n = 9$ und $p = \frac{1}{2}$. Darauf aufbauend kann man auch einen Test durchführen (man spricht vom sogenannten Vorzeichentest). Der Vorteil ist, dass man keine Normalverteilung mehr annehmen muss.

3. a) Es handelt sich um **ungepaarte** Stichproben, da zu den einzelnen Männchen nicht jeweils ein bestimmtes Weibchen gehört. Die Anzahlen in den beiden Stichproben brauchen auch gar nicht gleich gross zu sein.

b) **Zweiseitiger t-Test:**

Seien X_i : i -ter Wert der Kieferlänge der männlichen Tiere, ($i = 1, \dots, n = 10$) und Y_j : j -ter Wert der Kieferlänge der weiblichen Tiere, ($j = 1, \dots, m = 10$)

1. Modell: $X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu_X, \sigma^2)$, $Y_1, \dots, Y_m \text{ i.i.d. } \sim \mathcal{N}(\mu_Y, \sigma^2)$
2. Nullhypothese: $H_0 : \mu_X = \mu_Y$
 Alternativhypothese: $H_A : \mu_X \neq \mu_Y$. Man will herausfinden, ob sich die Kieferlängen unterscheiden, oder nicht. Dabei spielt es keine Rolle, ob die von den männlichen oder die von den weiblichen Goldschakalen länger sind. Deshalb handelt es sich um eine zweiseitige Alternative.
3. Teststatistik:

$$T = \frac{(\bar{X}_n - \bar{Y}_m)}{S_{pool} \sqrt{1/n + 1/m}},$$

wobei

$$S_{pool}^2 = \frac{1}{n + m - 2} \left((n - 1)\hat{\sigma}_X^2 + (m - 1)\hat{\sigma}_Y^2 \right) = 9.49.$$

Unter H_0 gilt: $T \sim t_{n+m-2} = t_{18}$.

Wert der Teststatistik:

$$t = \frac{113.4 - 108.6}{\sqrt{9.49} \sqrt{1/10 + 1/10}} = 3.48.$$

4. Signifikanzniveau: $\alpha = 0.05$
5. Verwerfungsbereich:

$$\begin{aligned} K &= (-\infty, -t_{n+m-2, 1-\alpha/2}] \cup [t_{n+m-2, 1-\alpha/2}, \infty) = (-\infty, -t_{18, 0.975}] \cup [t_{18, 0.975}, \infty) \\ &= (-\infty, -2.101] \cup [2.101, \infty). \end{aligned}$$

6. Testentscheid: Da der Wert der Teststatistik ($t = 3.48$) im Verwerfungsbereich K liegt, wird die Nullhypothese auf dem 5%-Niveau verworfen. Die Kieferlängen unterscheiden sich signifikant.

- c) Der **R-Output** für den t-Test sieht folgendermassen aus:

```
> t.test(jackals[, "M"], jackals[, "W"])
Welch Two Sample t-test

data: jackals[, "M"] and jackals[, "W"]
t = 3.4843, df = 14.894, p-value = 0.00336
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.861895 7.738105
sample estimates:
mean of x mean of y
 113.4      108.6
```

Der p-Wert ist $0.0034 < 0.05$, also wird die Nullhypothese verworfen.

- d) Der **R-Output** für den Wilcoxon-Test sieht folgendermassen aus:

```
> wilcox.test(jackals[, "M"], jackals[, "W"])
Wilcoxon rank sum test with continuity correction

data: jackals[, "M"] and jackals[, "W"]
W = 87.5, p-value = 0.004845
alternative hypothesis: true location shift is not equal to 0
```

Der p-Wert ist $0.0048 < 0.05$, also wird auch bei diesem Test die Nullhypothese verworfen.

- e) Das Resultat des Wilcoxon-Tests ist vertrauenswürdiger, da er im Gegensatz zum t-Test nicht annimmt, dass die Daten normalverteilt sind und wir diese Voraussetzung in keiner Weise überprüft haben. Allerdings ist die stark unterschiedliche Standardabweichung in den zwei Gruppen problematisch für beide Tests.