

Musterlösung

1. (11 Punkte)

- a) Für welchen Parameter ist \bar{X} ein geeigneter Schätzer? \bar{X} ist ein geeigneter Schätzer für den Erwartungswert μ_X von X . **(1 Punkt)**
- b) Bestimmen Sie jeweils ein exaktes 1-seitiges 95% Vertrauensintervall (mit Formel) für die durchschnittliche Adrenalinmenge in den beiden Gruppen. **Tipp:** Beachten Sie hierbei die Hypothese der Schulleitung und wählen Sie eine geeignete Richtung für das jeweilige Vertrauensintervall. 1-seitiges Vertrauensintervall für μ in den beiden Gruppen:

$$\begin{aligned} & \left(-\infty, \bar{X} + t_{n_X-1, 1-\alpha} \frac{S_{n_X}}{\sqrt{n_X}}\right] \\ & \left[\bar{Y} - t_{n_Y-1, 1-\alpha} \frac{S_{n_Y}}{\sqrt{n_Y}}, \infty\right) \quad \text{(1 Punkt)} \end{aligned}$$

Einsetzen der beobachteten Werten:

$$VI_{\mu_X} = \left(-\infty, \bar{x} + t_{n_X-1, 1-\alpha} \frac{s_X}{\sqrt{n_X}}\right] = \left(-\infty, 2.59 + 1.708 \cdot \frac{0.45}{\sqrt{26}}\right] \approx (-\infty, 2.74] \quad \text{(0.5 Punkte)}$$

$$VI_{\mu_Y} = \left[\bar{y} - t_{n_Y-1, 1-\alpha} \frac{s_Y}{\sqrt{n_Y}}, \infty\right) = \left[3.25 - 1.753 \cdot \frac{0.52}{4}, \infty\right) \approx [3.02, \infty) \quad \text{(0.5 Punkte)}$$

(total: 2 Punkt)

- c) Wie beurteilen Sie die Vermutung der Schulleitung aufgrund der berechneten Vertrauensintervalle? Die Vertrauensintervalle überschneiden sich nicht. Es kann daher davon ausgegangen werden, dass Sport einen signifikanten Einfluss auf den Stresspegel hat. **(1 Punkt)**

Die Schulleitung ist nicht restlos überzeugt und möchte, dass Sie zur Sicherheit noch einen geeigneten Hypothesentest durchführen. Nehmen Sie dazu an, dass die Varianz für beide Gruppen gleich ist.

- d) Es handelt sich um einen ungepaarten Test. Warum? Die 2 Stichproben sind unabhängig. Die Versuchseinheiten (hier: Studenten) in beiden Gruppen (hier: Sport/kein Sport) haben nach der Randomisierung nichts mit einander zu tun. **(1 Punkt)**
- e) Geben Sie die Null- und die Alternativhypothese an und begründen Sie kurz Ihre Wahl.

$$H_0 : \mu_X = \mu_Y \quad \text{(0.5 Punkte)}$$

$$H_A : \mu_X < \mu_Y \quad \text{(0.5 Punkte)}$$

Man möchte im Sinne der Schulleitung statistisch nachweisen (mit gösstmöglicher Macht), dass Sport den Stresspegel signifikant reduziert, d.h. zu einem geringeren Adrenalinwert führt. Daher handelt es sich um eine einseitige Alternativhypothese von der Form $H_A : \mu_X < \mu_Y$. **(1 Punkt)**

- f) Führen Sie den geeigneten t -Test auf dem 5% Niveau durch: Bestimmen Sie den Wert der Teststatistik T , die Verteilung der Teststatistik T unter der Nullhypothese, den Verwerfungsbereich für T und den Testentscheid.

1. Modell: $X_i \sim \mathcal{N}(\mu_X, \sigma^2)$ iid und $Y_j \sim \mathcal{N}(\mu_Y, \sigma^2)$ iid, wobei X_i und Y_j unabhängig sind.

2. Nullhypothese: $H_0 : \mu_X = \mu_Y$

Alternative: $H_A : \mu_X < \mu_Y$

3. Teststatistik:

$$T = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}, \text{ wobei } \quad \textbf{(0.5 Punkt)}$$

$$S^2 = \frac{1}{n_X + n_Y - 2} ((n_X - 1)S_X^2 + (n_Y - 1)S_Y^2) \quad \textbf{(0.5 Punkt)}$$

$$t = \frac{2.59 - 3.25}{0.48 \cdot \sqrt{1/26 + 1/16}} \approx -4.35 \quad \textbf{(0.5 Punkte)}$$

Verteilung von T unter H_0 : $T \sim t_{n_X+n_Y-2} = t_{40}$. **(0.5 Punkte)**

4. Signifikanzniveau: $\alpha = 0.05$

5. Verwerfungsbereich:

$$K = (-\infty; -t_{n_X+n_Y-2; 1-\alpha}] \quad \textbf{(0.5 Punkte)}$$

$$= (-\infty; -t_{40; 0.95}] = (-\infty; -1.684] \quad \textbf{(0.5 Punkte)}$$

6. Testentscheid: $t \in K \Rightarrow H_0$ wird verworfen. **(1 Punkt, total: 4 Punkte)**

2. (11 Punkte)

a) Wie ist X verteilt? Geben Sie die Verteilungsfamilie und die Parameter an.

X ist binomialverteilt **(0.5 Punkte)** mit $n = 10$ und $\pi = 0.45$. **(0.5 Punkte)**

b) Wie gross ist die Wahrscheinlichkeit, dass mindestens 2 der befragten Personen abgestimmt haben?

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - \left(\binom{10}{1} 0.45^1 (1 - 0.45)^9 + \binom{10}{0} 0.45^0 (1 - 0.45)^{10} \right) = 0.9767 \quad \textbf{(1 Punkt)}$$

c) Wie gross ist die Wahrscheinlichkeit, dass die ersten 8 befragten Personen abgestimmt haben (von der 9. und 10. befragten Person weiss man nicht, ob sie abgestimmt haben)?

Die Wahrscheinlichkeit, dass die ersten 8 befragten Personen abgestimmt haben ist $0.45 \cdot 0.45 \cdot \dots \cdot 0.45 \cdot 1 \cdot 1 = 0.45^8 = 0.0017$ **(1 Punkt)**

d) Was sind die Null- und die Alternativhypothese?

$H_0 : \pi = 0.45$ **(0.5 Punkte)** und $H_A : \pi > 0.45$. **(0.5 Punkte)**

e) Führen Sie einen geeigneten exakten Test auf dem 1%-Signifikanzniveau durch. Geben Sie den Verwerfungsbereich und den Testentscheid an. 1. Modell: $X \sim \text{Bin}(10, \pi)$ ist die Anzahl der Freunde, die an der letzten Abstimmung teilgenommen haben

2. Nullhypothese: $H_0 : \pi = 0.45$ Alternativhypothese: $H_A : \pi > 0.45$

3. Die Teststatistik ist: $T = X$, Verteilung von T unter H_0 : $T \sim \text{Bin}(10, 0.45)$

4. Signifikanzniveau: $\alpha = 0.01$

5. Verwerfungsbereich:

Da die Alternativhypothese nach oben zeigt hat der Verwerfungsbereich die Form $K = [c, n]$. Gesucht ist das kleinste c , sodass $P[X \geq c] \leq 0.01$. Da

$$P(X = 10) = \binom{10}{10} 0.45^{10} 0.55^0 = 0.00034,$$

$$P(X = 9) = \binom{10}{9} 0.45^9 0.55 = 0.00416,$$

$$P(X = 8) = \binom{10}{8} 0.45^8 0.55^2 = 0.02289,$$

gilt

	$x = 10$	$x = 9$	$x = 8$
$P[X = x]$	0.00034	0.00416	0.02289
$P[X \geq x]$	0.00034	0.0045	0.02739

(1 Punkt für korrekte Wahrscheinlichkeiten)

Der Verwerfungsbereich ist also gegeben durch $K = \{9, 10\}$. **(0.5 Punkte)**

6. Testentscheid: Da $7 \notin \{9, 10\}$ wird die Nullhypothese auf dem 1%-Niveau nicht verworfen **(0.5 Punkte, mit Folgefehler)**.

- f) Führen Sie erneut einen statistischen Test - diesmal auf dem 5%-Niveau - durch, um die Hypothese von Jonas zu testen. Benutzen Sie dazu eine geeignete Approximation. Geben Sie die Teststatistik, die Verteilung der Teststatistik unter der Nullhypothese, den Verwerfungsbereich und den Testentscheid an.

1. Modell: $X \sim \text{Bin}(232, \pi)$ ist die Anzahl der Freunde, Arbeitskollegen und Arbeitskolleginnen, die an der letzten Abstimmung teilgenommen haben

2. Nullhypothese: $H_0: \pi = 0.45$ Alternativhypothese: $H_A: \pi > 0.45$

3. Die Teststatistik ist: $T = X$, Verteilung von T unter $H_0: T \sim \text{Bin}(232, 0.45)$ **(0.5 Punkte)**

4. Signifikanzniveau: $\alpha = 0.05$

5. Verwerfungsbereich:

Da die Alternativhypothese nach oben zeigt hat der Verwerfungsbereich die Form $K = [c, n]$ **(0.5 Punkte)**. Zur Berechnung von c benutzen wir die Normalapproximation:

$$\begin{aligned} c &\approx n\pi_0 + \Phi^{-1}(1 - \alpha)\sqrt{n\pi_0(1 - \pi_0)} \quad \textbf{(0.5 Punkte)} \\ &= 232 \cdot 0.45 + 1.64 \cdot \sqrt{232 \cdot 0.45 \cdot 0.55} = 116.83 \quad \textbf{(0.5 Punkte)} \end{aligned}$$

Der Verwerfungsbereich ist also gegeben durch $K = \{117, \dots, 232\}$. **(0.5 Punkte)**

6. Testentscheid: Da $126 \in K$ wird die Nullhypothese auf dem 5%-Niveau verworfen **(0.5 Punkte, mit Folgefehler)**.

(Total: 3 Punkte, sinngemäss bei anderen Varianten der Normalapproximation)

- g) Berechnen Sie mithilfe der Approximation aus f) das kleinste Signifikanzniveau α , bei welchem der Test die Nullhypothese gerade noch verwirft. Wie heisst dieser Wert? Damit der approximative Test aus der vorherigen Aufgabe die Nullhypothese gerade noch verwirft, muss

$$c = 232 \cdot 0.45 + \Phi^{-1}(1 - \alpha)\sqrt{232 \cdot 0.45 \cdot 0.55} = 104.4 + \Phi^{-1}(1 - \alpha) \cdot 7.5776 = 126$$

sein **(0.5 Punkte)**. Daraus folgt

$$\alpha = 1 - \Phi\left(\frac{126 - 104.4}{7.5776}\right) = 0.0022 \quad \textbf{(0.5 Punkte)}$$

Dieser Wert wird P-Wert genannt. **(1 Punkt)**

3. (11 Punkte)

- a) Wie gross ist die Wahrscheinlichkeit, dass im Frühling genau 15 Schneeglöckchen in Oma Gerdas Garten blühen?

(1 Punkt)

Oma Gerda hat $n_s = 16$ Zwiebeln für Schneeglöckchen vergraben. Die Wahrscheinlichkeit für das Keimen einer Zwiebel ist $p_s = 0.9$. Sei S die Zufallsvariable, die die Anzahl blühender Schneeglöckchen beschreibt. Dann gilt:

$$P[S = 15] = \binom{16}{15} (0.9)^{15} (1 - 0.9)^1 = 0.3294.$$

- b) Wie gross ist die Wahrscheinlichkeit, dass mindestens ein Krokus und mindestens eine Tulpe blühen?

(2 Punkte)

Sei K die Zufallsvariable, die die Anzahl blühender Krokusse beschreibt, und T die Zufallsvariable für die Anzahl blühender Tulpen. Dann gilt: $K \sim \text{Bin}(2, 0.85)$ und $T \sim \text{Bin}(3, 0.9)$

Gesucht ist $P[K \geq 1 \cap T \geq 1]$.

Da die Blumenzwiebeln unabhängig voneinander sind, gilt

$$P[K \geq 1 \cap T \geq 1] = P[K \geq 1]P[T \geq 1] = (1 - P[K = 0])(1 - P[T = 0]) \quad \textbf{(1 Punkt)}$$

$$P[K = 0] = \binom{2}{0} (0.85)^0 (1 - 0.85)^2 = 0.0225$$

$$P[T = 0] = \binom{3}{0} (0.9)^0 (1 - 0.9)^3 = 0.001$$

$$P[K \geq 1 \cap T \geq 1] = 0.9775 * 0.999 = 0.9765 \quad \textbf{(1 Punkt)}.$$

- c) Wie gross ist die Wahrscheinlichkeit, dass Oma Gerda im Beet, in dem sie die Zwiebeln der Schneeglöckchen und Tulpen vergraben hat, genau 14 Blumen zählt?

(2 Punkte)

Sei S die Zufallsvariable, die die Anzahl blühender Schneeglöckchen beschreibt, und T die Zufallsvariable für die Anzahl blühender Tulpen. Dann ist $P[S + T = 14]$ gesucht. S und T sind beide binomial verteilt mit $p = 0.9$. Die Summe der beiden unabhängigen binomial verteilten Zufallsvariablen ist $S + T \sim \text{Bin}(16 + 3, 0.9)$ **(1 Punkt)**

Damit ist $P[S + T = 14] = \binom{19}{14} (0.9)^{14} (1 - 0.9)^5 = 0.0266$ **(1 Punkt)**

- d) Oma Gerda möchte die Wahrscheinlichkeit für das Blühen der Osterglocken schätzen. Im Vorjahr hat sie 21 Osterglockenzwiebeln gesetzt und davon keimten 17. Was wäre ein geeigneter Schätzer? Berechnen Sie den Schätzer.

(1 Punkt)

Wir verwenden den Momentenschätzer:

$$p_O = 17/21 = 0.8095.$$

- e) Was ist die erwartete Anzahl blühender Blumen in Oma Gerdas Garten? Mit welcher Varianz? (Falls Sie Aufgabe d) nicht lösen konnten, verwenden Sie für die Osterglocken $P[\text{Keimen}] = 0.85$.)

(2 Punkte)

Die erwartete Anzahl blühender Blumen ist gegeben als:

$$E[S + K + T + O] = E[S] + E[K] + E[T] + E[O] = 16 * 0.9 + 2 * 0.85 + 3 * 0.9 + 19 * 0.81 = 34$$

(1 Punkt) mit der Varianz

$$16 * 0.1 * 0.9 + 2 * 0.85 * 0.15 + 3 * 0.9 * 0.1 + 19 * 0.19 * 0.81 = 4.9.$$

(1 Punkt)

Mit Alternativlösung:

$$E[S + K + T + O] = 35 \text{ mit Varianz } 4.3875.$$

- f) Oma Gerda hat die Blumenzwiebeln bei der Gärtnerei Zwiebeli gekauft. In der Gärtnerei werden 1000 Tulpenzwiebeln gesetzt. Berechnen Sie die Wahrscheinlichkeit, dass zwischen 880 und 900 Zwiebeln keimen? Benutzen Sie eine geeignete Normalapproximation.

(3 Punkte)

Die Anzahl blühender Tulpen D kann approximativ als $\mathcal{N}(1000 * 0.9, 1000 * 0.9 * 0.1)$ -verteilt angesehen werden **(1 Punkt)**.

Die gesuchte Wahrscheinlichkeit $P[880 \leq D \leq 900]$ kann durch Standardisieren berechnet werden.

Es gilt:

$$P[880 \leq D \leq 900] = P\left[\frac{880-900}{\sqrt{90}} \leq \frac{D-\mathbf{E}[D]}{\sqrt{\text{Var}(D)}} \leq \frac{900-900}{\sqrt{90}}\right] = P[-2.11 \leq Z \leq 0]$$

wobei $Z := \frac{D-\mathbf{E}[D]}{\sqrt{\text{Var}(D)}} \sim \mathcal{N}(0, 1)$. **(1 Punkt)**

Mit der Tabelle der kumulativen Standardnormalverteilung erhält man:

$$\begin{aligned} P[-2.11 \leq Z \leq 0] &= \phi(0) - \phi(-2.11) = \phi(0) - (1 - \phi(2.11)) \\ &= 0.5 - (1 - 0.9826) = 0.4826. \end{aligned} \quad \mathbf{(1 Punkt)}$$

4. **(8 Punkte)** In einem Fahrsimulator wird die Auswirkung von Alkoholkonsum auf die Reaktionszeit eines Autofahrers erforscht. Dazu wird bei 35 stark betrunkenen Probanden jeweils die Blutalkoholkonzentration (BAK) in Promille und die Reaktionszeit (RZ) in Sekunden gemessen. Mit den beobachteten Werten wird folgendes Modell gefittet:

$$RZ_i = \beta_0 + \beta_1 \cdot \text{BAK}_i + E_i, \quad E_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2).$$

Der (unvollständige) Regressionsoutput sieht wie folgt aus:

```
lm(formula = RZ ~ BAK)
```

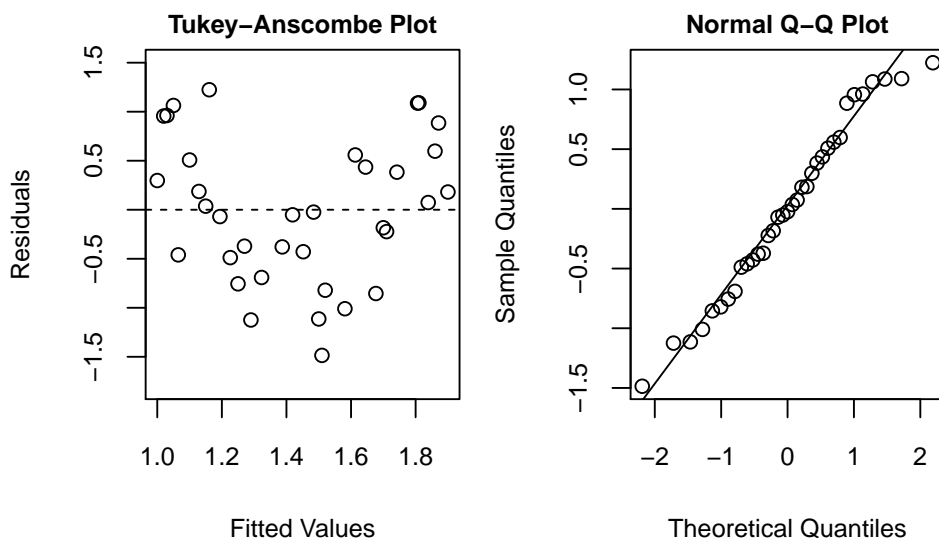
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.311	???	0.48	0.635
BAK	1.142	0.443	2.58	???

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.743 on ?? degrees of freedom
 Multiple R-squared: 0.168, Adjusted R-squared: 0.142
 F-statistic: 6.64 on 1 and ?? DF, p-value: 0.0146

- 1) Mit wievielen Freiheitsgraden wurde der residual standard error berechnet?
 - a) 33
 - b) 34
 - c) 35
 - d) 37
 - a. Mit 33 Freiheitsgraden
- 2) Gibt es in diesem Modell auf dem 1%-Niveau mindestens eine signifikante erklärende Variable?
 - a) Ja
 - b) Nein
 - c) Dies kann aus den obigen Angaben nicht entschieden werden.
 - b. Der P-Wert des globalen F-Tests ist 0.0146. Daher gibt es auf dem 1%-Niveau keine signifikante Co-Variable.
- 3) Wie gross ist der Standardfehler für $\hat{\beta}_0$?
 - a) 0.648
 - b) 0.149
 - c) 1.543
 - d) 0.443
 - a. $\frac{\hat{\beta}_0}{t\text{-value}(\hat{\beta}_0)} = \frac{0.311}{0.48} = 0.648$
- 4) Berechnen Sie das exakte zweiseitige 95%-Konfidenzintervall für β_1 .
 - a) [0.240, 2.044]
 - b) [0.392, 1.892]
 - c) [-0.069, 2.353]
 - d) [0.256, 2.028]
 - a. $[\hat{\beta}_1 - t_{33,0.975} \cdot se(\hat{\beta}_1), \hat{\beta}_1 + t_{33,0.975} \cdot se(\hat{\beta}_1)] = [1.142 - 2.035 \cdot 0.443, 1.142 + 2.035 \cdot 0.443] = [0.240, 2.044]$
- 5) Betrachten Sie die nachfolgenden Plots. Welche der folgenden Aussagen trifft zu?
 - a) Alle Modellannahmen sind erfüllt.
 - b) Die Fehlervarianz ist konstant, aber die Normalverteilungsannahme trifft nicht zu.
 - c) Sowohl konstante Fehlervarianz als auch Normalverteilungsannahme treffen nicht zu.
 - d) Es scheint in diesem Modell systematische Fehler zu geben.
 - d. Man sieht einen klaren Trend im Tukey-Anscombe Plot. Dies weist auf systematische Fehler im Modell hin.



- 6) Wie würde sich $\hat{\beta}_1$ verändern, falls zusätzlich bei einem nüchternen Probanden (Blutalkoholkonzentration = 0 Promille) eine Reaktionszeit von 1 Sekunde gemessen wurde?
- $\hat{\beta}_1$ wird grösser
 - $\hat{\beta}_1$ bleibt gleich gross
 - $\hat{\beta}_1$ wird kleiner
 - Dies kann man nicht entscheiden, ohne die Regressionskoeffizienten erneut zu berechnen.

c. Das Modell sagt in nüchternem Zustand eine Reaktionszeit von $\hat{\beta}_0 = 0.311$ Sekunden voraus. Kommt eine Beobachtung von 1 Sekunde dazu, wird die Gerade "flacher", d.h. $\hat{\beta}_1$ wird kleiner.

- 7) Was kann man über den P-Wert für die Variable BAK aussagen?

- P-Wert > 0.2
- $0.05 < \text{P-Wert} \leq 0.2$
- $0.01 < \text{P-Wert} \leq 0.05$
- P-Wert ≤ 0.01
- Man kann nichts über den P-Wert aussagen.

c. Berechnet man den Verwerfungsbereich für $\alpha = 0.2$ erhält man:

$$K = (-\infty, -t_{33,0.9}] \cup [t_{33,0.9}, \infty) = (-\infty, -1.308] \cup [1.308, \infty).$$

Der t-value von 2.53 liegt darin, also verwirft der Test auf dem 20%-Niveau. Folglich ist der P-Wert kleiner als 0.2. Analog findet man für $\alpha = 0.05$:

$$K = (-\infty, -t_{33,0.975}] \cup [t_{33,0.975}, \infty) = (-\infty, -2.035] \cup [2.035, \infty).$$

Also verwirft der t-Test und der P-Wert ist kleiner als 0.05. Für $\alpha = 0.01$ erhält man:

$$K = (-\infty, -t_{33,0.995}] \cup [t_{33,0.995}, \infty) = (-\infty, -2.733] \cup [2.733, \infty).$$

Hier verwirft der t-Test die Nullhypothese nicht, da $2.53 \notin K$. Also ist der P-Wert grösser als 0.01. Damit ist Antwort c) richtig.

- 8) Bei welchem der folgenden Regressionsmodelle handelt es sich **nicht** um ein lineares Modell?

- $\log(Y_i) = \beta_1 X_i + E_i$
- $Y_i = \log(\beta_1) X_i + E_i$
- $Y_i = \beta_1 \log(X_i) + E_i$
- $\log(Y_i) = \beta_1 \log(X_i) + E_i$
- Keines der obigen Modelle ist linear.

b. Der Term $\log(\beta_1)$ ist nicht linear, also handelt es sich hier nicht um ein lineares Regressionsmodell.

5. (10 Punkte) Die folgenden Aufgaben sind zufällig angeordnet und insbesondere nicht nach Schwierigkeitsgrad sortiert.

- 1) Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ verteilt mit $\mu = 3$ und $\sigma^2 = 5$. Wie ist $Y = 2X$ verteilt?

- $\mathcal{N}(3, 5)$
- $\mathcal{N}(2 \cdot 3, 2 \cdot 5)$
- $\mathcal{N}(2 \cdot 3, 2^2 \cdot 5)$
- $\mathcal{N}(2^2 \cdot 3, 2^2 \cdot 5)$
- dies kann mit obigen Angaben nicht entschieden werden

c). Es gilt $E[Y] = E[2X] = 2 \cdot E[X] = 2 \cdot 3$ und $\text{Var}(Y) = \text{Var}(2X) = 2^2 \text{Var}(X) = 2^2 \cdot 5$.

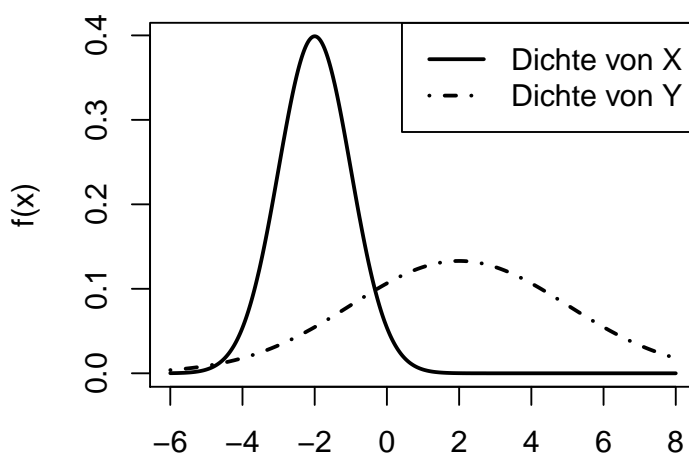
- 2) Eine Zufallsvariable X hat die Dichtefunktion

$$f_X(x) = \begin{cases} 0.1 & -1 \leq x < 0 \\ 0.3 & 0 \leq x < 3 \\ 0 & \text{sonst} \end{cases}.$$

Dann beträgt $P(X \leq 2)$

- a) 0
 b) 0.1
 c) 0.4
 d) 0.7
 e) f_X ist keine Dichtefunktion, deshalb kann die gesuchte Wahrscheinlichkeit nicht berechnet werden.
- d). Die Wahrscheinlichkeit $P(X \leq 2)$ ist gerade die Fläche unter der Dichtefunktion bis $x = 2$. Deshalb gilt: $P(X \leq 2) = 1 * 0.1 + 2 * 0.3 = 0.7$.
- 3) Sei X, Y und Z Zufallsvariablen mit $Y = 3X + 2$ und $Z = 2X - 3$. Dann gilt
- a) $\text{Corr}(X, Y) > \text{Corr}(X, Z)$
 b) $\text{Corr}(X, Y) = \text{Corr}(X, Z)$
 c) $\text{Corr}(X, Y) < \text{Corr}(X, Z)$
 d) Die Korrelationen $\text{Corr}(X, Y)$ und $\text{Corr}(X, Z)$ können aus diesen Angaben nicht bestimmt werden.
- b). Da sowohl Y als auch Z lineare Funktionen von X sind mit positiver Steigung, gilt: $\text{Corr}(X, Y) = \text{Corr}(X, Z) = 1$.
- 4) Sie testen die Wirkung eines Medikaments und teilen die 62 Probanden der Studie in 2 Gruppen auf. 22 Personen verabreichen Sie den Wirkstoff (W), den restlichen 40 Personen ein Placebo (P). Sie testen die Nullhypothese $H_0 : \mu_W = \mu_P$ (der Wirkstoff hat keinen Effekt) gegen die einseitige Alternative $H_A : \mu_W > \mu_P$ (der Wirkstoff hat einen positiven Effekt). Zur Auswertung der Ergebnisse führen Sie einen ungepaarten 2-Stichproben t-Test durch und erhalten für die Teststatistik
- $$T = \frac{\bar{W}_n - \bar{P}_m}{s_{\text{pool}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$
- den realisierten Wert $t = 2$. Welches ist das kleinste Signifikanzniveau, bei dem der Test die Nullhypothese gerade noch verwirft?
- a) 0.01
 b) 0.025
 c) 0.05
 d) 0.075
 e) 0.1
- b). Der Verwerfungsbereich des obigen Tests lautet: $K = [t_{n+m-2, 1-\alpha}, \infty) = [t_{60, 1-\alpha}, \infty)$. Damit die Nullhypothese für den Wert $t = 2$ gerade noch verworfen wird, muss $t_{60, 1-\alpha} = 2$ sein. Mit der Tabelle der t-Verteilung findet man: $t_{60, 0.975} = 2.000$. Daraus folgt $\alpha = 0.025$.
- 5) Nehmen Sie an, die Ereignisse A und B seien unabhängig, und $A \cap B$ ist disjunkt von einem dritten Ereignis C . Dann gilt
- a) $P(A \cap B \cap C) = 0$
 b) $P(A \cap B \cap C) = P(A)P(B)P(C)$
 c) $P(A \cap B \cap C) = P(A \cap B)$
 d) $P(A \cap B \cap C) = P(C)$
 e) $P(A \cap B \cap C) = 1$
- a). Da $A \cap B$ disjunkt ist von C folgt direkt: $P(A \cap B \cap C) = P((A \cap B) \cap C) = P(\emptyset) = 0$.
- 6) Ruben wohnt an einer stark befahrenen Strasse und ärgert sich über die vielen Autos. Deshalb möchte er ein statistisches Modell erstellen für die Anzahl Autos pro Minute, die an seinem Schlafzimmerfenster vorbeifahren. Welche Verteilung wäre dazu am besten geeignet?
- a) Uniformverteilung
 b) Exponentialverteilung
 c) Bernoulliverteilung
 d) Poissonverteilung
 e) Binomialverteilung
- d). Die Poissonverteilung beschreibt die Anzahl Ereignisse während eines Zeitintervalls.

- 7) Seien X und Y zwei unabhängige Zufallsvariablen mit $\text{Var}(X) = 2$, $\text{Var}(Y) = 1$. Berechnen Sie $\text{Var}(2X + 3Y - 5)$.
- 2
 - 7
 - 12
 - 17
 - 22
- d). $\text{Var}(2X + 3Y - 5) = 2^2 \text{Var}(X) + 3^2 \text{Var}(Y) = 2^2 \cdot 2 + 3^2 \cdot 1 = 17$
- 8) Im nachfolgenden Plot sind die Dichten von $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ eingezeichnet. Es gilt:
- $\mu_X > \mu_Y$, $\sigma_X^2 > \sigma_Y^2$
 - $\mu_X > \mu_Y$, $\sigma_X^2 < \sigma_Y^2$
 - $\mu_X < \mu_Y$, $\sigma_X^2 > \sigma_Y^2$
 - $\mu_X < \mu_Y$, $\sigma_X^2 < \sigma_Y^2$
- d). Aus dem Plot entnimmt man $\mu_X = -2$ und $\mu_Y = 2$, daher gilt $\mu_X < \mu_Y$. Zudem ist die Dichte von X "schmäler", d.h. $\sigma_X^2 < \sigma_Y^2$.



- 9) Der Schüler Patric musste während eines Schuljahres n Prüfungen ($n > 2$) schreiben. In der Hälfte der Prüfungen hat er die Note 6 erreicht, für die andere Hälfte der Prüfungen hat er nichts gelernt und folglich die Note 1 erhalten. Sein Lehrer hat Mitleid und lässt ihn aus den folgenden vier Strategien wählen, um seine Zeugnisnote zu berechnen. Welche Strategie sollte Patric wählen, wenn er eine möglichst gute Gesamtnote erreichen möchte (beste Note: 6, schlechteste Note: 1)?
- Strategie 1: Die Zeugnisnote ergibt sich aus dem Median der Noten während des Schuljahres.
 - Strategie 2: Die Zeugnisnote ergibt sich aus dem arithmetischen Mittel der Noten während des Schuljahres.
 - Strategie 3: Die schlechteste Note wird gestrichen. Die Zeugnisnote ergibt sich aus dem Median der übrigen Noten während des Schuljahres.
 - Strategie 4: Die schlechteste Note wird gestrichen. Die Zeugnisnote ergibt sich aus dem arithmetischen Mittel der übrigen Noten während des Schuljahres.
- c). Löscht man einmal die Note 1, ist der Median der übrigen Noten gleich 6.
- 10) Die kumulative Verteilungsfunktion der Zufallsvariablen X ist gegeben durch

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_{\min}}{x}\right)^k & x \geq x_{\min} \\ 0 & x < x_{\min} \end{cases},$$

wobei $x_{\min} > 0$ und $k > 0$ zwei Konstanten sind. Bestimmen Sie den Median m von X in Abhängigkeit von x_{\min} und k .

- a) $m = 1 - (2x_{\min})^k$
 - b) $m = 2x_{\min}$
 - c) $m = 2(x_{\min})^k$
 - d) $m = \sqrt[k]{2}x_{\min}$
 - e) $m = \sqrt[k]{2}(x_{\min})^k$
- d). Für den Median gilt: $F(m) = 0.5$. Auflösen nach m ergibt Antwort d).