Schriftliche Prüfung
(120 Minuten)

Bemerkungen:

- Erlaubte Hilfsmittel: 10 hand- oder maschinengeschriebene A4 Seiten (=5 Blätter). Taschenrechner ohne Kommunikationsmöglichkeit.
- Mobiltelefone sind auszuschalten!
- Die Prüfung besteht aus insgesamt **20 Aufgaben**.
- Markieren Sie Ihre Antworten auf dem beiliegenden **Antwortblatt**.
- Jede Aufgabe besteht aus mehreren Aussagen. Pro Aufgabe können keine, eine oder mehrere Aussagen richtig sein.
- Für jede Aussage gibt es 1 Punkt, wenn sie korrekt markiert wird. 1 Punkt wird abgezogen, wenn eine Aussage falsch markiert wird. Wenn eine Aussage nicht markiert wird, gibt es keinen Punkt, es wird aber auch kein Punkt abgezogen. Die Punkte werden über die gesamte Prüfung summiert.
- Alle Rechnungsergebnisse sind auf 2 Nachkommastellen gerundet.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet!

Viel Erfolg!

Gruppe A

Binomialverteilung und -test

1. Angenommen, die Zufallsvariablen X_i für $i \in \{1, 2, 3, 4\}$ sind unabhängig und identisch verteilt mit Verteilung $P(X_i = 0) = 0.3$ und $P(X_i = 1) = 0.7$. Dann gilt ...

a) Für alle $i \in \{1, 2, 3, 4\}$ hat X_i die Verteilung $Binomial(1, 0.3)$.

b) $E(X_1) = 0.7$.

c) $E(X_1 + X_2 + X_3 + X_4) = 2.8$.

d) Angenommen, $X \sim Bernoulli(\pi)$ und $Y \sim Binomial(n, \pi)$. Dann gilt allgemein, dass $n \operatorname{Var}(X) = \operatorname{Var}(Y)$.

2. Michael tritt einen Ball gegen eine Torwand. Wir nehmen an, dass seine Schüsse unabhängig voneinander sind. Die Zufallsvariable X_i beschreibt das Ereignis, dass Michael im $i$ten Versuch trifft ($X_i = 1$) oder verfehlt ($X_i = 0$). Im Schnitt trifft Michael in 10% der Fälle. Insgesamt schiesst er 100 mal auf die Torwand. Sei die Anzahl Treffer Y definiert als $Y = X_1 + X_2 + X_3 + ... + X_{100}$. Dann gilt ...

a) Die Anzahl der fehlgeschlagenen Versuche berechnet man als $100 - Y$.

b) Y ist binomialverteilt mit $n = 100$ und $\pi = 0.1$.

c) Es gilt $P(Y < 50) > P(Y \geq 50)$.

d) Von den 100 Versuchen erwarten wir, dass Michael 10 mal trifft.

3. Bei einem Binomialtest ($n = 8,\ \alpha = 0.05$) mit $H_0 : \pi = 0.5$ und $H_A : \pi > 0.5$ und Testatistik X, ist der beobachtete Wert der Teststatistik $x = 6$. Beurteilen Sie folgende Aussagen.

a) Der Verwerfungsbereich ist $K = \{7, 8\}$.

b) Der P-Wert der Daten ist 0.145.

c) Der P-Wert ist gleich $P(X \in \{6, 7, 8\})$.

d) Je grösser der Fehler 1.Art, desto grösser ist die Macht.

4. Beurteilen Sie folgende Aussagen.

a) Die Faustregeln für die Normalapproximation einer Binomialverteilung sagen aus, dass bei einem Wert von $\pi = 0.5$ mindestens $n = 11$ Stichproben vorhanden sein müssen.

b) Wenn X_1 und X_2 bernoulliverteilt mit gleichem Parameter $\pi = 0.4$ sind, dann ist $X_1 + X_2$ binomialverteilt mit $n = 2$ und $\pi = 0.4$.

c) Wenn Y_1 und Y_2 i.i.d. $Binomial(n, \pi)$ verteilt sind, dann gilt $Y_1 + Y_2 \sim Binomial(2n, \pi)$ verteilt.

d) Auf einem Parkplatz stehen 14 rote Autos und 7 blaue Autos. Jemand bestimmt zufällig 5 Autos, welche abgeschleppt werden sollen. Die Anzahl der blauen Autos unter den 5 abgeschleppten Autos kann mit einer Binomialverteilung genau modelliert werden.

5. Bei einem Binomialtest ($n = 100, \alpha = 0.05$) mit $H_0 : \pi = 0.5$ und $H_A : \pi \neq 0.5$ wurde der Verwerfungsbereich $K = \{0, ..., 39\} \cup \{61, ..., 100\}$ konstruiert. Sei X die Teststatistik. Der beobachtete Wert der Teststatistik ist $x = 71$. Beurteilen Sie folgende Aussagen.

a) Unter H_0 ist die Wahrscheinlichkeit $P(X = 71) = 0.05$.

b) $P_{H_0}(X \in K) \leq 0.05$.

c) Angenommen $Y \sim \mathit{Binomial}(n = 100, \pi = 0.6)$. Die Macht für die konkrete Alternative ist gegeben durch $P(Y \in K)$.

d) Der P-Wert in diesem Fall ist ungefähr $2P_{H_0}(X \geq 71)$.

Gruppe A

t-Test

6. In einer Studie wird ein neues Medikament zur Senkung des Augeninnendrucks getestet. Es wurden 50 Probanden über 4 Wochen hinweg mit dem Medikament behandelt. Dafür wurde bei jedem Proband das Medikament in das linke Auge getropft. Das rechte Auge wurde jeweils unbehandelt gelassen. Der Augeninnendruck des rechten bzw. des linken Auges pro Proband nach der Behandlung ist in der folgenden (unvollständigen) Tabelle aufgeführt:

Rechtes Auge x_i	15	21	17	...	29
Linkes Auge y_i	14	19	24	...	16

Wir nehmen an, dass die Messungen der jeweiligen Probanden unabhängig voneinander sind, sowie einer Normalverteilung folgen mit Erwartungswert μ_x (rechtes Auge), beziehungsweise μ_y (linkes Auge) und gleicher Standardabweichung σ.

Beurteilen Sie folgende Aussagen.

a) Es handelt sich um ungepaarte Stichproben.

b) Bei einem gepaarten t-Test können die Stichprobengrössen in beiden Gruppen ungleich sein.

c) Wir können bei gepaarten Stichproben sowohl einen gepaarten t-Test wie auch einen ungepaarten t-Test verwenden.

d) Der z-Test setzt voraus, dass die wahre Standardabweichung bekannt ist.

7. Aus den Daten der Augenstudie (vorherige Aufgabe) wurden folgende Kennzahlen ermittelt: $\bar{x} = 20.2$, $\bar{y} = 17.6$ und die geschätzte Varianz von $X - Y$ ist $\widehat{Var} = 4.5^2$. Wir führen nun einen gepaarten t-Test durch. Wir legen den Test folgendermassen an: $H_0 : \mu_x = \mu_y,\ H_A : \mu_x > \mu_y$.

a) Die Macht des Tests für eine einseitige Alternativhypothese ist immer grösser als für $H_A : \mu_x \neq \mu_y$.

b) Der beobachtete Wert der Teststatistik liegt zwischen 3.5 und 3.55.

c) Unter H_0 folgt die Teststatistik einer t_{49}-Verteilung.

d) Angenommen, die Teststatistik wäre 2.64 und der Verwerfungsbereich des Tests wäre $[1.69, \infty)$ für das Signifikanzniveau $\alpha = 5\%$, dann wird die Nullhypothese auf dem 5%-Signifikanzniveau verworfen.

8. Beurteilen Sie folgende Aussagen.

a) Um den Wilcoxon-Test anwenden zu können, müssen die Datenpunkte i.i.d. sein und von einer symmetrischen Verteilung stammen.

b) Angenommen wir würden die Differenzen $\{x_1 - y_1, ..., x_{50} - y_{50}\}$ aus der Augenstudie betrachten. Nun führen wir auf diesem Datensatz einen Vorzeichen-Test durch, um eine Senkung des Augeninnendrucks durch das Medikament zu überprüfen. Die Verteilung der entsprechenden Teststatistik unter der Nullhypothese ist nun $Binomial(49, 0.5)$.

c) Wir führen einen ungepaarten t-Test auf den Daten der Augen-Studie durch mit $H_0 : \mu_x = \mu_y$ und $H_A : \mu_x \neq \mu_y$. Angenommen $\bar{x} = 20.2$, $\bar{y} = 17.6$, $\hat{\sigma}_x = 3.9$ und $\hat{\sigma}_y = 4.7$. Die Teststatistik des ungepaarten t-Tests liegt zwischen 3.0 und 3.05.

d) Angenommen ein t-Test liefert einen P-Wert, der es uns erlaubt, die Nullhypothese auf dem 5%-Signifikanzniveau zu verwerfen. Dann können wir die Nullhypothese immer auch auf dem 1%-Signifikanzniveau verwerfen.

9. Beurteilen Sie folgende Aussagen.

a) Beim Vorzeichen-Test muss die wahre Standardabweichung der Zufallsvariablen bekannt sein.

b) Die Nullhypothese des Vorzeichen-Tests lautet $H_0 : \mu = \mu_0$, wobei μ der Mittelwert und μ_0 ein vorgegebener, bekannter Wert ist.

c) Bei einem Einstichproben t-Test ($H_0 : \mu = 0,\ H_A : \mu > 0$, 30 Beobachtungen insgesamt) ist der P-Wert 2.5%. Daher muss der beobachtete Wert der Teststatistik 2.06 gewesen sein.

d) Wir führen einen Einstichproben t-Test ($H_0 : \mu = 0,\ H_A : \mu \neq 0$, 24 Beobachtungen insgesamt) durch. Angenommen wir haben $\bar{x} = 15.4$ und $\hat{\sigma}_x = 1.8$. Ein exaktes zweiseitiges 95%-Vertrauensintervall für μ ist $[14.64, 16.16]$.

Lineare Regression

10. Lukas, der Sprengexperte, hat bei verschiedenen Sprengungen in Silber- und Goldminen die verwendetete Menge TNT und das abgesprengte Volumen gemessen. Er möchte das abgesprengte Volumen (Variable `volumen`), gemessen in Kubikzentimetern cm^3, in Abhängigkeit der verwendeten Menge TNT (Variable `tnt`), gemessen in Gramm g, beschreiben. Hierfür verwendet er ein lineares Modell:

$$\texttt{volumen}_i = \beta_0 + \beta_1 \cdot \texttt{tnt}_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2).$$

Der (unvollständige) Regressionsoutput sieht wie folgt aus:

```
Residuals:
   Min     1Q Median     3Q    Max
 -43.9  -12.8    1.0   12.1   47.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.6384        ???    0.77     0.44
tnt           1.2423     0.0159     ???   <2e-16 ***
---
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.9 on 62 degrees of freedom
Multiple R-squared:  0.99,       Adjusted R-squared:  0.99
F-statistic: 6.12e+03 on 1 and 62 DF,  p-value: <2e-16
```

Beurteilen Sie folgende Aussagen.

a) Die lineare Regression wurde basierend auf $n = 60$ Datenpunkten berechnet.

b) Der geschätzte Standardfehler des Parameters $\hat{\beta}_0$ ist 4.73.

c) $H_0 : \beta_1 = 0$ wird auf dem 1% Niveau verworfen.

d) Ein zweiseitiges 99%-Vertrauensintervall für β_1 ist gegeben durch $[1.20, 1.28]$. (Das entsprechende Quantil der t-Verteilung ist 2.67)

11. Mit Hilfe des geschätzten Modells aus der Sprengstudie (vorherige Aufgabe) wollen wir nun Vorhersagen machen. Beurteilen Sie folgende Aussagen.

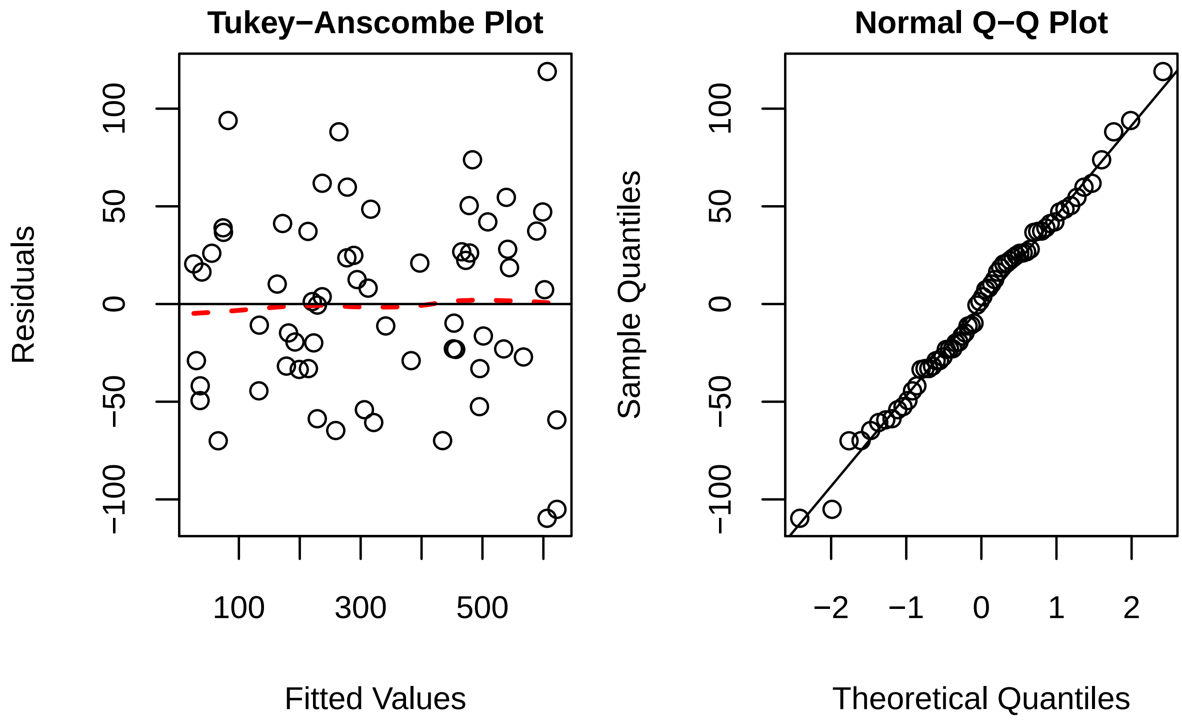
a) Wenn Lukas die Menge an eingesetztem TNT um $10\,g$ reduziert, dann reduziert sich das erwartete Sprengvolumen um $0.124\,cm^3$.

b) Angenommen, Lukas hat mit dem Modell ein erwartetes Sprengvolumen von $35\,cm^3$ vorhergesagt. Dafür muss Lukas $33\,g$ TNT einsetzen.

c) Falls Lukas $0.30\,kg$ TNT verwendet, kann er im Mittel ein Volumen von $4.01\,cm^3$ sprengen.

d) Das 95%-Vorhersageintervall für das Sprengvolumen beschreibt den Bereich, in welchem sich das erwartete Sprengvolumen mit 95% Wahrscheinlichkeit befindet.

12. Betrachten Sie die nachfolgenden Plots, welche die Residuen aus dem Regressionsmodell der Sprengstudie zeigen.



Beurteilen Sie folgende Aussagen.

a) Der Mittelwert der Residuen ist in etwa 0.

b) Die Normalitätsannahme ist erfüllt.

c) Es gibt zwei starke Ausreisser, daher sind die Parameterschätzungen der linearen Regression verfälscht.

d) Angenommen, wir verwenden den Logarithmus des Sprengvolumen als Zielvariable ($\log(\texttt{volumen}_i) = \beta_0 + \beta_1 \texttt{tnt}_i + \varepsilon_i$). Dann ist dieses Regressionsmodell kein lineares Modell.

Gruppe A

Gemischte Fragen

13. Beurteilen Sie folgende Aussagen.

- **a)** Seien $E(X) = 1$, $E(Y) = 2$ und $E(Z) = 3$, wobei X, Y und Z unabhängig sind. Dann gilt $E[X + Y] = E[Z - X + Y - 1]$.
- **b)** Sei $\text{Var}(X) = 1$ und $\text{Var}(Y) = 2$, wobei X und Y unabhängig sind. Dann gilt $\text{Var}[2X] = \text{Var}[Y]$.
- **c)** Das arithmetische Mittel der Daten $\{0, -1, 2, 3, 4\}$ ist 2.
- **d)** Die (empirisch) standardisierten Daten von $\{1, 2, 3\}$ sind $\{-1, 0, 1\}$.

14. Beurteilen Sie folgende Aussagen.

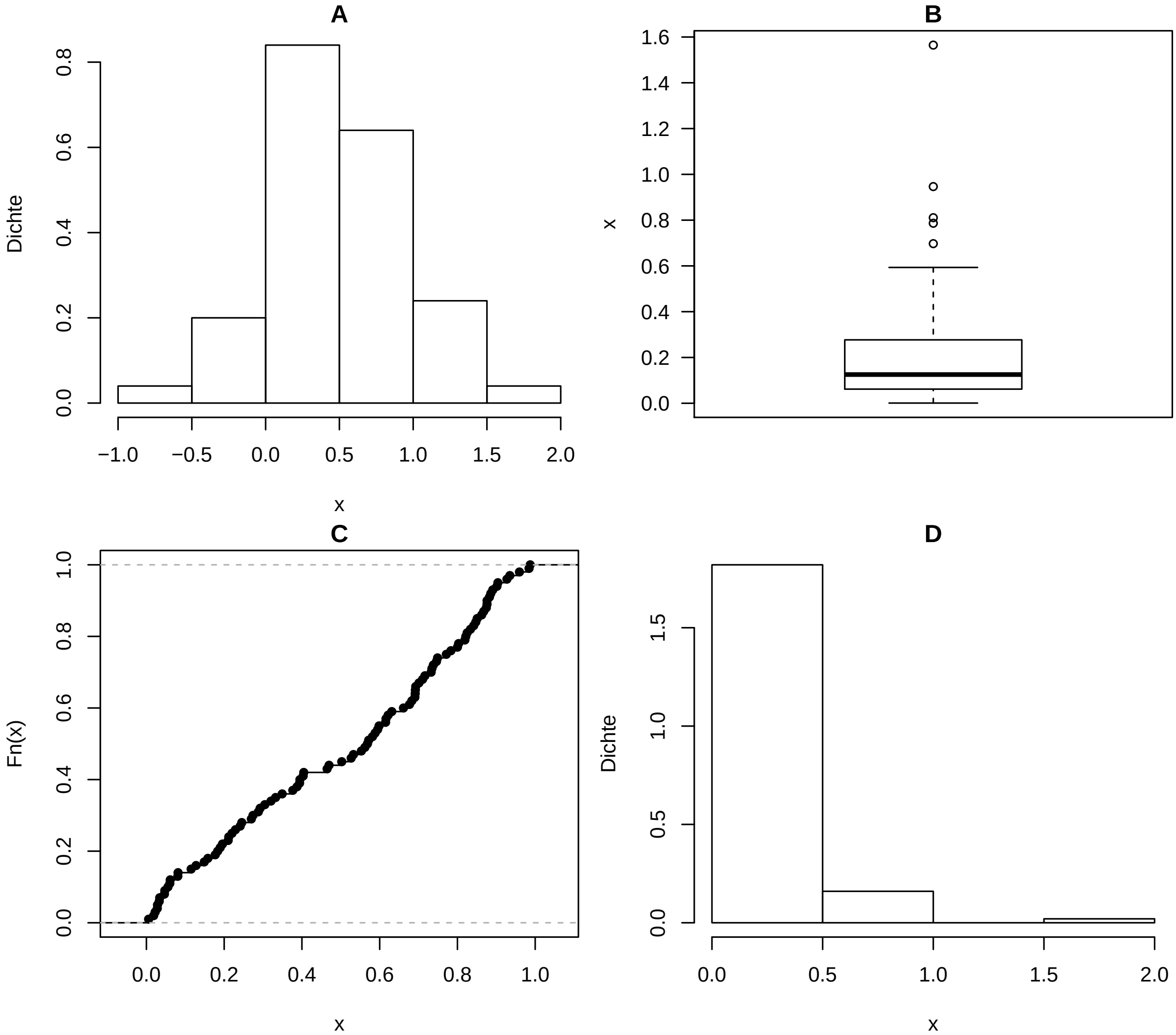
- **a)** Falls $P(A) = 0.1$ und $P(B) = 0.9$. Dann gilt immer, dass $P(A \cup B) = 1$.
- **b)** Es gilt $odds(E^c) = \frac{1}{odds(E)}$ für ein Ereignis E und sein Komplement E^c.
- **c)** Falls $P(E) < P(F)$, dann gilt $odds(E) > odds(F)$.
- **d)** Falls $A \cap B = \varnothing$ und $P(A) = P(B) = 1/8$. Dann ist $odds(A \cup B) = 1/3$.

15. Beurteilen Sie folgende Aussagen.

- **a)** Falls $P(A|B) = 0.1$ und $P(B) = 0.2$, dann muss $P(A \cap B) = 0.02$ sein.
- **b)** Falls $P(A \cup B) = 0.3$ und $P(A) = P(B) = 0.2$, dann ist $P(A|B) = 0.5$
- **c)** Falls $P(A) = P(B) = 0.1$ mit A und B unabhängig, dann gilt $P(A|B) = 0$.
- **d)** Falls $P(A) \neq 0$ und $P(B) \neq 0$, wobei A und B disjunkt sind, dann können A und B **nicht** auch unabhängig sein.

Gruppe A

16. Betrachten Sie die nachfolgenden Plots.



Beurteilen Sie folgende Aussagen.

a) Das Histogramm in Plot **A** und die empirische Verteilungsfunktion in Plot **C** beschreiben symmetrische Verteilungen.

b) Das Histogramm in Plot **A** und die empirische Verteilungsfunktion in Plot **C** stammen **nicht** von denselben Daten.

c) Der Boxplot in Plot **B** und das Histogram in Plot **D** können **nicht** von denselben Daten stammen.

d) Im Boxplot in Plot **B** sind 25% der Daten grösser als 0.6.

Gruppe A

17. Es seien $X_1, ..., X_n \sim F$, wobei F eine Verteilung ist mit Erwartungswert $E[X_i] = \mu$ und endlicher Varianz σ_F^2. Beurteilen Sie folgende Aussagen.

a) Falls $X_1 = \ldots = X_n$ gilt, kann der zentralen Grenzwertsatz anwendet werden und die Varianz von $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ geht gegen 0.

b) Falls Unabhängigkeit gilt ($X_1, ..., X_n \overset{i.i.d}{\sim} F$), dann sagt der zentrale Grenzwertsatz aus, dass die Varianz $\mathrm{Var}(X_1) = \sigma_F^2$ bei grossen n gegen 0 geht.

c) Damit der zentrale Grenzwertsatz gilt, müssen alle Zufallsvariablen X_i und X_j $(i \neq j)$ unabhängig sein.

d) Die Lebensdauer von normalen Batterien in Stirnlampen kann als unabhängig angenommen werden. Johannes möchte nun eine 5 stündige Höhlenexpedition unternehmen und fragt sich, wie viele Batterien er für die Stirnlampe mitnehmen sollte. Eine Batterie hält im Schnitt eine halbe Stunde $(0.5h)$ mit einer Standardabweichung von $0.1h$. Johannes nimmt 12 Batterien mit auf die Expedition. Die Wahrscheinlichkeit, dass er die 5 stündige Expedition mit Licht bewältigen kann ist grösser als 95%.

18. Beurteilen Sie folgende Aussagen.

a) Die folgende Funktion $f(x)$ ist eine Dichtefunktion:

$$f(x) = \begin{cases} x & \text{falls } -1 \leq x < 0 \\ 1 & \text{falls } 0 \leq x < 2 \\ 0 & \text{sonst.} \end{cases}$$

b) Die Wartezeit in Minuten für die Essensausgabe in der Mensa um 12:00 kann mit einer Uniformverteilung $X \sim Uniform(4, 12)$ beschrieben werden. Die Wahrscheinlichkeit mehr als 8 Minuten zu warten liegt bei 50% (d.h. $P(X > 8) = 0.5$).

c) Die Wartezeit in Stunden in der Notfallaufnahme kann gut mit der Exponentialverteilung $Exp(1/2)$ beschrieben werden. Im Schnitt muss ein Patient 2 Stunden warten.

d) In einer Sägerei schneidet eine Maschine Bretter mit der gleichen Länge. Die kleinen Abweichungen in cm beim Schneideprozess kann man mit einer Standard-Normalverteilung $(X \sim \mathcal{N}(0, 1))$ beschreiben. Die Wahrscheinlichkeit ist 69%, dass die Maschine ein Brett mehr als $0.5cm$ zu kurz schneidet.

19. Beurteilen Sie folgende Aussagen.

a) Der Momentenschätzer für den Erwartungswert einer normalverteilten Zufallsvariabeln ist das arithmetische Mittel.

b) Der Maximum Likelihood Schätzer für die Erfolgswahrscheinlichkeit π bei einer Binomialverteilung mit $n = 1$ ist $\hat{\pi}_{MLE} = x_1$, wobei x_1 der beobachtete Datenpunkt ist.

c) Für alle Verteilungen gilt, dass der Maximum Likelihood Schätzer **nie** identisch zum Momentenschätzer ist.

d) Falls $X \sim Poisson(\lambda)$ mit $x = 2$ der beobachte Datenpunkt, dann hat der Momentenschätzer von λ einen Wert von 2^2.

20. Beurteilen Sie folgende Aussagen.

a) Herr Meier wird jeden Sonntag nach Fussballspielen in einem Stadion gesichtet. Meier ist also ein grosser Fussballfan.

b) Sei $Y := 2 - X$ und X zwei Zufallsvariabeln. Dann ist die Korrelation $Corr(X, Y) = -1$.

c) Bei einer Losbude kann man aus einer verdeckten Schüssel 2 Kugeln (ohne zurücklegen) ziehen. In der Schüssel befinden sich 2 schwarze und 8 weisse Kugeln. Zieht man beide schwarzen Kugeln gewinnt man. Die Wahrscheinlichkeit zu gewinnen ist 2.22%.

d) Die konkurrenzierende Losbude bietet ein ähnliches Spiel an wie die Losbude aus Unteraufgabe c) an. In der Schüssel dieser zweiten Losbude hat es 2 schwarze, 1 rote und 7 blaue Kugeln. Wenn man die zwei schwarzen Kugeln zieht (ohne zurrücklegen), dann hat man gewonnen. Die Wahrscheinlichkeit zu gewinnen ist höher als 2.22%.

Gruppe B

t-Test

1. In einer Studie wird ein neues Medikament zur Senkung des Augeninnendrucks getestet. Es wurden 50 Probanden über 4 Wochen hinweg mit dem Medikament behandelt. Dafür wurde bei jedem Proband das Medikament in das linke Auge getropft. Das rechte Auge wurde jeweils unbehandelt gelassen. Der Augeninnendruck des rechten bzw. des linken Auges pro Proband nach der Behandlung ist in der folgenden (unvollständigen) Tabelle aufgeführt:

Rechtes Auge x_i	15	21	17	...	29
Linkes Auge y_i	14	19	24	...	16

Wir nehmen an, dass die Messungen der jeweiligen Probanden unabhängig voneinander sind, sowie einer Normalverteilung folgen mit Erwartungswert μ_x (rechtes Auge), beziehungsweise μ_y (linkes Auge) und gleicher Standardabweichung σ.

Beurteilen Sie folgende Aussagen.

a) Es handelt sich um ungepaarte Stichproben.

b) Bei einem gepaarten t-Test können die Stichprobengrössen in beiden Gruppen ungleich sein.

c) Wir können bei gepaarten Stichproben sowohl einen gepaarten t-Test wie auch einen ungepaarten t-Test verwenden.

d) Der z-Test setzt voraus, dass die wahre Standardabweichung bekannt ist.

2. Aus den Daten der Augenstudie (vorherige Aufgabe) wurden folgende Kennzahlen ermittelt: $\bar{x} = 20.2$, $\bar{y} = 17.6$ und die geschätzte Varianz von $X - Y$ ist $\widehat{Var} = 4.5^2$. Wir führen nun einen gepaarten t-Test durch. Wir legen den Test folgendermassen an: $H_0 : \mu_x = \mu_y,\ H_A : \mu_x > \mu_y$.

a) Die Macht des Tests für eine einseitige Alternativhypothese ist immer grösser als für $H_A : \mu_x \neq \mu_y$.

b) Der beobachtete Wert der Teststatistik liegt zwischen 3.5 und 3.55.

c) Unter H_0 folgt die Teststatistik einer t_{49}-Verteilung.

d) Angenommen, die Teststatistik wäre 2.64 und der Verwerfungsbereich des Tests wäre $[1.69, \infty)$ für das Signifikanzniveau $\alpha = 5\%$, dann wird die Nullhypothese auf dem 5%-Signifikanzniveau verworfen.

Gruppe B

3. Beurteilen Sie folgende Aussagen.

a) Um den Wilcoxon-Test anwenden zu können, müssen die Datenpunkte i.i.d. sein und von einer symmetrischen Verteilung stammen.

b) Angenommen wir würden die Differenzen $\{x_1 - y_1, ..., x_{50} - y_{50}\}$ aus der Augenstudie betrachten. Nun führen wir auf diesem Datensatz einen Vorzeichen-Test durch, um eine Senkung des Augeninnendrucks durch das Medikament zu überprüfen. Die Verteilung der entsprechenden Teststatistik unter der Nullhypothese ist nun $Binomial(49, 0.5)$.

c) Wir führen einen ungepaarten t-Test auf den Daten der Augen-Studie durch mit $H_0 : \mu_x = \mu_y$ und $H_A : \mu_x \neq \mu_y$. Angenommen $\bar{x} = 20.2$, $\bar{y} = 17.6$, $\hat{\sigma}_x = 3.9$ und $\hat{\sigma}_y = 4.7$. Die Teststatistik des ungepaarten t-Tests liegt zwischen 3.0 und 3.05.

d) Angenommen ein t-Test liefert einen P-Wert, der es uns erlaubt, die Nullhypothese auf dem 5%-Signifikanzniveau zu verwerfen. Dann können wir die Nullhypothese immer auch auf dem 1%-Signifikanzniveau verwerfen.

4. Beurteilen Sie folgende Aussagen.

a) Beim Vorzeichen-Test muss die wahre Standardabweichung der Zufallsvariablen bekannt sein.

b) Die Nullhypothese des Vorzeichen-Tests lautet $H_0 : \mu = \mu_0$, wobei μ der Mittelwert und μ_0 ein vorgegebener, bekannter Wert ist.

c) Bei einem Einstichproben t-Test ($H_0 : \mu = 0$, $H_A : \mu > 0$, 30 Beobachtungen insgesamt) ist der P-Wert 2.5%. Daher muss der beobachtete Wert der Teststatistik 2.06 gewesen sein.

d) Wir führen einen Einstichproben t-Test ($H_0 : \mu = 0$, $H_A : \mu \neq 0$, 24 Beobachtungen insgesamt) durch. Angenommen wir haben $\bar{x} = 15.4$ und $\hat{\sigma}_x = 1.8$. Ein exaktes zweiseitiges 95%-Vertrauensintervall für μ ist $[14.64, 16.16]$.

Gruppe B

Lineare Regression

5. Lukas, der Sprengexperte, hat bei verschiedenen Sprengungen in Silber- und Goldminen die verwendetete Menge TNT und das abgesprengte Volumen gemessen. Er möchte das abgesprengte Volumen (Variable `volumen`), gemessen in Kubikzentimetern cm^3, in Abhängigkeit der verwendeten Menge TNT (Variable `tnt`), gemessen in Gramm g, beschreiben. Hierfür verwendet er ein lineares Modell:

$$\texttt{volumen}_i = \beta_0 + \beta_1 \cdot \texttt{tnt}_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2).$$

Der (unvollständige) Regressionsoutput sieht wie folgt aus:

```
Residuals:
   Min     1Q Median     3Q    Max
 -43.9  -12.8    1.0   12.1   47.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.6384        ???    0.77     0.44
tnt           1.2423     0.0159     ???   <2e-16 ***
---
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.9 on 62 degrees of freedom
Multiple R-squared:  0.99,        Adjusted R-squared:  0.99
F-statistic: 6.12e+03 on 1 and 62 DF,  p-value: <2e-16
```

Beurteilen Sie folgende Aussagen.

a) Die lineare Regression wurde basierend auf $n = 60$ Datenpunkten berechnet.

b) Der geschätzte Standardfehler des Parameters $\hat{\beta}_0$ ist 4.73.

c) $H_0 : \beta_1 = 0$ wird auf dem 1% Niveau verworfen.

d) Ein zweiseitiges 99%-Vertrauensintervall für β_1 ist gegeben durch $[1.20, 1.28]$. (Das entsprechende Quantil der t-Verteilung ist 2.67)

6. Mit Hilfe des geschätzten Modells aus der Sprengstudie (vorherige Aufgabe) wollen wir nun Vorhersagen machen. Beurteilen Sie folgende Aussagen.

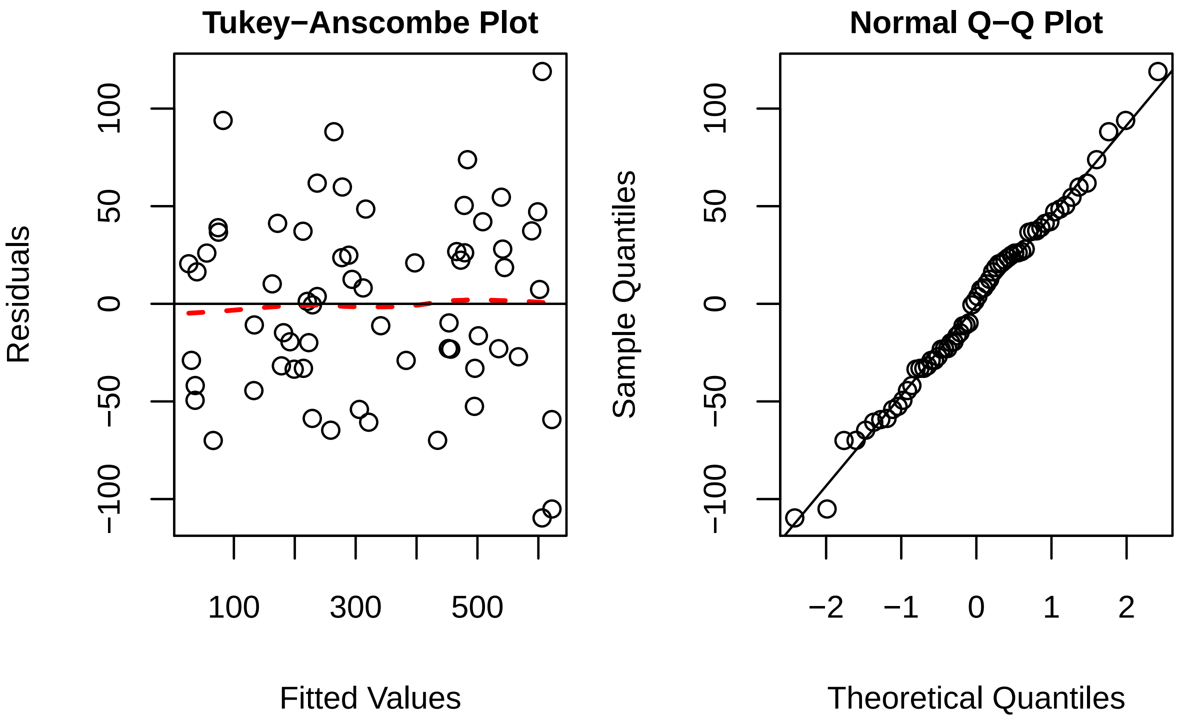
a) Wenn Lukas die Menge an eingesetztem TNT um $10\,g$ reduziert, dann reduziert sich das erwartete Sprengvolumen um $0.124\,cm^3$.

b) Angenommen, Lukas hat mit dem Modell ein erwartetes Sprengvolumen von $35\,cm^3$ vorhergesagt. Dafür muss Lukas $33\,g$ TNT einsetzen.

c) Falls Lukas $0.30\,kg$ TNT verwendet, kann er im Mittel ein Volumen von $4.01\,cm^3$ sprengen.

d) Das 95%-Vorhersageintervall für das Sprengvolumen beschreibt den Bereich, in welchem sich das erwartete Sprengvolumen mit 95% Wahrscheinlichkeit befindet.

7. Betrachten Sie die nachfolgenden Plots, welche die Residuen aus dem Regressionsmodell der Sprengstudie zeigen.



Beurteilen Sie folgende Aussagen.

a) Der Mittelwert der Residuen ist in etwa 0.

b) Die Normalitätsannahme ist erfüllt.

c) Es gibt zwei starke Ausreisser, daher sind die Parameterschätzungen der linearen Regression verfälscht.

d) Angenommen, wir verwenden den Logarithmus des Sprengvolumen als Zielvariable ($\log(\texttt{volumen}_i) = \beta_0 + \beta_1 \texttt{tnt}_i + \varepsilon_i$). Dann ist dieses Regressionsmodell kein lineares Modell.

Gruppe B

Gemischte Fragen

8. Beurteilen Sie folgende Aussagen.

a) Seien $E(X) = 1$, $E(Y) = 2$ und $E(Z) = 3$, wobei X, Y und Z unabhängig sind. Dann gilt $E[X + Y] = E[Z - X + Y - 1]$.

b) Sei $\mathrm{Var}(X) = 1$ und $\mathrm{Var}(Y) = 2$, wobei X und Y unabhängig sind. Dann gilt $\mathrm{Var}[2X] = \mathrm{Var}[Y]$.

c) Das arithmetische Mittel der Daten $\{0, -1, 2, 3, 4\}$ ist 2.

d) Die (empirisch) standardisierten Daten von $\{1, 2, 3\}$ sind $\{-1, 0, 1\}$.

9. Beurteilen Sie folgende Aussagen.

a) Falls $P(A) = 0.1$ und $P(B) = 0.9$. Dann gilt immer, dass $P(A \cup B) = 1$.

b) Es gilt $odds(E^c) = \frac{1}{odds(E)}$ für ein Ereignis E und sein Komplement E^c.

c) Falls $P(E) < P(F)$, dann gilt $odds(E) > odds(F)$.

d) Falls $A \cap B = \varnothing$ und $P(A) = P(B) = 1/8$. Dann ist $odds(A \cup B) = 1/3$.

10. Beurteilen Sie folgende Aussagen.

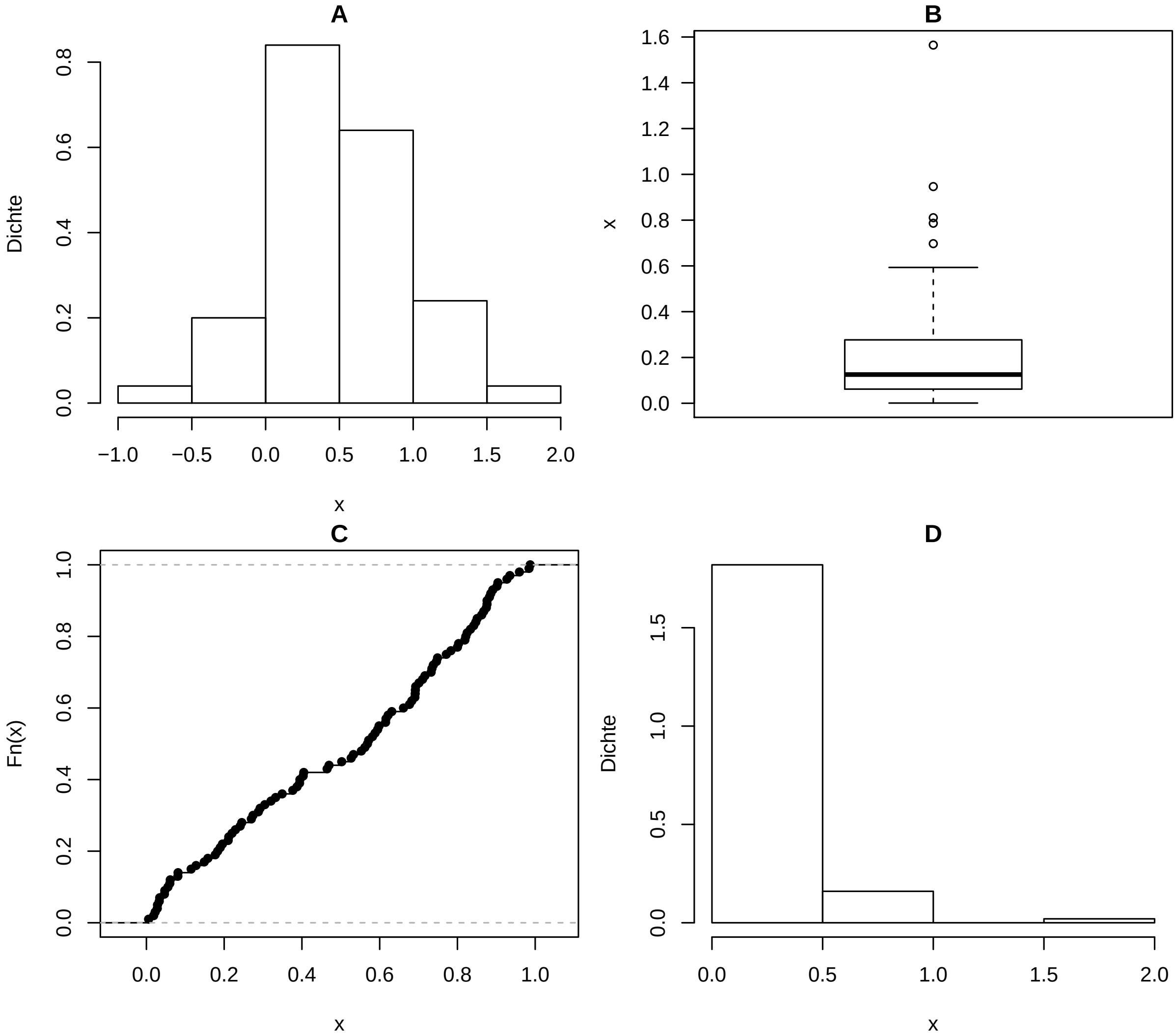
a) Falls $P(A|B) = 0.1$ und $P(B) = 0.2$, dann muss $P(A \cap B) = 0.02$ sein.

b) Falls $P(A \cup B) = 0.3$ und $P(A) = P(B) = 0.2$, dann ist $P(A|B) = 0.5$

c) Falls $P(A) = P(B) = 0.1$ mit A und B unabhängig, dann gilt $P(A|B) = 0$.

d) Falls $P(A) \neq 0$ und $P(B) \neq 0$, wobei A und B disjunkt sind, dann können A und B **nicht** auch unabhängig sein.

11. Betrachten Sie die nachfolgenden Plots.



Beurteilen Sie folgende Aussagen.

a) Das Histogramm in Plot **A** und die empirische Verteilungsfunktion in Plot **C** beschreiben symmetrische Verteilungen.

b) Das Histogramm in Plot **A** und die empirische Verteilungsfunktion in Plot **C** stammen **nicht** von denselben Daten.

c) Der Boxplot in Plot **B** und das Histogram in Plot **D** können **nicht** von denselben Daten stammen.

d) Im Boxplot in Plot **B** sind 25% der Daten grösser als 0.6.

Gruppe B

12. Es seien $X_1, ..., X_n \sim F$, wobei F eine Verteilung ist mit Erwartungswert $E[X_i] = \mu$ und endlicher Varianz σ_F^2. Beurteilen Sie folgende Aussagen.

a) Falls $X_1 = \ldots = X_n$ gilt, kann der zentralen Grenzwertsatz anwendet werden und die Varianz von $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ geht gegen 0.

b) Falls Unabhängigkeit gilt ($X_1, ..., X_n \overset{i.i.d}{\sim} F$), dann sagt der zentrale Grenzwertsatz aus, dass die Varianz $\mathrm{Var}(X_1) = \sigma_F^2$ bei grossen n gegen 0 geht.

c) Damit der zentrale Grenzwertsatz gilt, müssen alle Zufallsvariablen X_i und X_j $(i \neq j)$ unabhängig sein.

d) Die Lebensdauer von normalen Batterien in Stirnlampen kann als unabhängig angenommen werden. Johannes möchte nun eine 5 stündige Höhlenexpedition unternehmen und fragt sich, wie viele Batterien er für die Stirnlampe mitnehmen sollte. Eine Batterie hält im Schnitt eine halbe Stunde ($0.5h$) mit einer Standardabweichung von $0.1h$. Johannes nimmt 12 Batterien mit auf die Expedition. Die Wahrscheinlichkeit, dass er die 5 stündige Expedition mit Licht bewältigen kann ist grösser als 95%.

13. Beurteilen Sie folgende Aussagen.

a) Die folgende Funktion $f(x)$ ist eine Dichtefunktion:

$$f(x) = \begin{cases} x & \text{falls } -1 \leq x < 0 \\ 1 & \text{falls } 0 \leq x < 2 \\ 0 & \text{sonst.} \end{cases}$$

b) Die Wartezeit in Minuten für die Essensausgabe in der Mensa um 12:00 kann mit einer Uniformverteilung $X \sim Uniform(4, 12)$ beschrieben werden. Die Wahrscheinlichkeit mehr als 8 Minuten zu warten liegt bei 50% (d.h. $P(X > 8) = 0.5$).

c) Die Wartezeit in Stunden in der Notfallaufnahme kann gut mit der Exponentialverteilung $Exp(1/2)$ beschrieben werden. Im Schnitt muss ein Patient 2 Stunden warten.

d) In einer Sägerei schneidet eine Maschine Bretter mit der gleichen Länge. Die kleinen Abweichungen in cm beim Schneideprozess kann man mit einer Standard-Normalverteilung ($X \sim \mathcal{N}(0, 1)$) beschreiben. Die Wahrscheinlichkeit ist 69%, dass die Maschine ein Brett mehr als $0.5cm$ zu kurz schneidet.

14. Beurteilen Sie folgende Aussagen.

a) Der Momentenschätzer für den Erwartungswert einer normalverteilten Zufallsvariabeln ist das arithmetische Mittel.

b) Der Maximum Likelihood Schätzer für die Erfolgswahrscheinlichkeit π bei einer Binomialverteilung mit $n = 1$ ist $\hat{\pi}_{MLE} = x_1$, wobei x_1 der beobachtete Datenpunkt ist.

c) Für alle Verteilungen gilt, dass der Maximum Likelihood Schätzer **nie** identisch zum Momentenschätzer ist.

d) Falls $X \sim Poisson(\lambda)$ mit $x = 2$ der beobachte Datenpunkt, dann hat der Momentenschätzer von λ einen Wert von 2^2.

15. Beurteilen Sie folgende Aussagen.

a) Herr Meier wird jeden Sonntag nach Fussballspielen in einem Stadion gesichtet. Meier ist also ein grosser Fussballfan.

b) Sei $Y := 2 - X$ und X zwei Zufallsvariabeln. Dann ist die Korrelation $Corr(X, Y) = -1$.

c) Bei einer Losbude kann man aus einer verdeckten Schüssel 2 Kugeln (ohne zurücklegen) ziehen. In der Schüssel befinden sich 2 schwarze und 8 weisse Kugeln. Zieht man beide schwarzen Kugeln gewinnt man. Die Wahrscheinlichkeit zu gewinnen ist 2.22%.

d) Die konkurrenzierende Losbude bietet ein ähnliches Spiel an wie die Losbude aus Unteraufgabe c) an. In der Schüssel dieser zweiten Losbude hat es 2 schwarze, 1 rote und 7 blaue Kugeln. Wenn man die zwei schwarzen Kugeln zieht (ohne zurrücklegen), dann hat man gewonnen. Die Wahrscheinlichkeit zu gewinnen ist höher als 2.22%.

Gruppe B

Binomialverteilung und -test

16. Angenommen, die Zufallsvariablen X_i für $i \in \{1,2,3,4\}$ sind unabhängig und identisch verteilt mit Verteilung $P(X_i = 0) = 0.3$ und $P(X_i = 1) = 0.7$. Dann gilt ...

a) Für alle $i \in \{1,2,3,4\}$ hat X_i die Verteilung $Binomial(1, 0.3)$.

b) $E(X_1) = 0.7$.

c) $E(X_1 + X_2 + X_3 + X_4) = 2.8$.

d) Angenommen, $X \sim Bernoulli(\pi)$ und $Y \sim Binomial(n, \pi)$. Dann gilt allgemein, dass $n\,\mathrm{Var}(X) = \mathrm{Var}(Y)$.

17. Michael tritt einen Ball gegen eine Torwand. Wir nehmen an, dass seine Schüsse unabhängig voneinander sind. Die Zufallsvariable X_i beschreibt das Ereignis, dass Michael im $i$ten Versuch trifft ($X_i = 1$) oder verfehlt ($X_i = 0$). Im Schnitt trifft Michael in 10% der Fälle. Insgesamt schiesst er 100 mal auf die Torwand. Sei die Anzahl Treffer Y definiert als $Y = X_1 + X_2 + X_3 + ... + X_{100}$. Dann gilt ...

a) Die Anzahl der fehlgeschlagenen Versuche berechnet man als $100 - Y$.

b) Y ist binomialverteilt mit $n = 100$ und $\pi = 0.1$.

c) Es gilt $P(Y < 50) > P(Y \geq 50)$.

d) Von den 100 Versuchen erwarten wir, dass Michael 10 mal trifft.

18. Bei einem Binomialtest ($n = 8,\ \alpha = 0.05$) mit $H_0 : \pi = 0.5$ und $H_A : \pi > 0.5$ und Testatistik X, ist der beobachtete Wert der Teststatistik $x = 6$. Beurteilen Sie folgende Aussagen.

a) Der Verwerfungsbereich ist $K = \{7, 8\}$.

b) Der P-Wert der Daten ist 0.145.

c) Der P-Wert ist gleich $P(X \in \{6, 7, 8\})$.

d) Je grösser der Fehler 1.Art, desto grösser ist die Macht.

Gruppe B

19. Beurteilen Sie folgende Aussagen.

a) Die Faustregeln für die Normalapproximation einer Binomialverteilung sagen aus, dass bei einem Wert von $\pi = 0.5$ mindestens $n = 11$ Stichproben vorhanden sein müssen.

b) Wenn X_1 und X_2 bernoulliverteilt mit gleichem Parameter $\pi = 0.4$ sind, dann ist $X_1 + X_2$ binomialverteilt mit $n = 2$ und $\pi = 0.4$.

c) Wenn Y_1 und Y_2 i.i.d. $Binomial(n, \pi)$ verteilt sind, dann gilt $Y_1 + Y_2 \sim Binomial(2n, \pi)$ verteilt.

d) Auf einem Parkplatz stehen 14 rote Autos und 7 blaue Autos. Jemand bestimmt zufällig 5 Autos, welche abgeschleppt werden sollen. Die Anzahl der blauen Autos unter den 5 abgeschleppten Autos kann mit einer Binomialverteilung genau modelliert werden.

20. Bei einem Binomialtest ($n = 100, \alpha = 0.05$) mit $H_0 : \pi = 0.5$ und $H_A : \pi \neq 0.5$ wurde der Verwerfungsbereich $K = \{0, ..., 39\} \cup \{61, ..., 100\}$ konstruiert. Sei X die Teststatistik. Der beobachtete Wert der Teststatistik ist $x = 71$. Beurteilen Sie folgende Aussagen.

a) Unter H_0 ist die Wahrscheinlichkeit $P(X = 71) = 0.05$.

b) $P_{H_0}(X \in K) \leq 0.05$.

c) Angenommen $Y \sim Binomial(n = 100, \pi = 0.6)$. Die Macht für die konkrete Alternative ist gegeben durch $P(Y \in K)$.

d) Der P-Wert in diesem Fall ist ungefähr $2P_{H_0}(X \geq 71)$.

Gruppe C

Gemischte Fragen

1. Beurteilen Sie folgende Aussagen.

a) Seien $E(X) = 1$, $E(Y) = 2$ und $E(Z) = 3$, wobei X, Y und Z unabhängig sind. Dann gilt $E[X + Y] = E[Z - X + Y - 1]$.

b) Sei $\mathrm{Var}(X) = 1$ und $\mathrm{Var}(Y) = 2$, wobei X und Y unabhängig sind. Dann gilt $\mathrm{Var}[2X] = \mathrm{Var}[Y]$.

c) Das arithmetische Mittel der Daten $\{0, -1, 2, 3, 4\}$ ist 2.

d) Die (empirisch) standardisierten Daten von $\{1, 2, 3\}$ sind $\{-1, 0, 1\}$.

2. Beurteilen Sie folgende Aussagen.

a) Falls $P(A) = 0.1$ und $P(B) = 0.9$. Dann gilt immer, dass $P(A \cup B) = 1$.

b) Es gilt $odds(E^c) = \frac{1}{odds(E)}$ für ein Ereignis E und sein Komplement E^c.

c) Falls $P(E) < P(F)$, dann gilt $odds(E) > odds(F)$.

d) Falls $A \cap B = \varnothing$ und $P(A) = P(B) = 1/8$. Dann ist $odds(A \cup B) = 1/3$.

3. Beurteilen Sie folgende Aussagen.

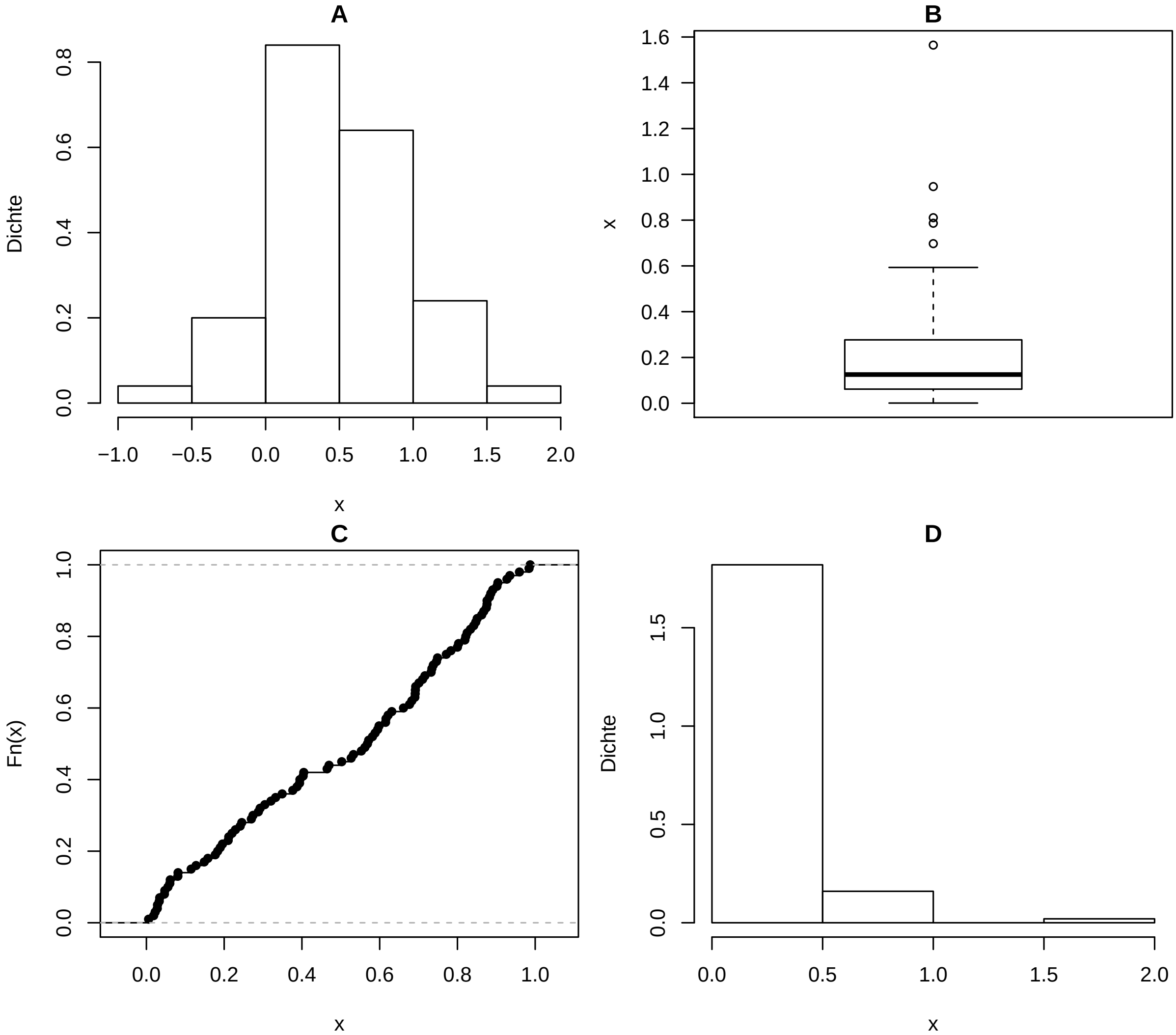
a) Falls $P(A|B) = 0.1$ und $P(B) = 0.2$, dann muss $P(A \cap B) = 0.02$ sein.

b) Falls $P(A \cup B) = 0.3$ und $P(A) = P(B) = 0.2$, dann ist $P(A|B) = 0.5$

c) Falls $P(A) = P(B) = 0.1$ mit A und B unabhängig, dann gilt $P(A|B) = 0$.

d) Falls $P(A) \neq 0$ und $P(B) \neq 0$, wobei A und B disjunkt sind, dann können A und B **nicht** auch unabhängig sein.

4. Betrachten Sie die nachfolgenden Plots.



Beurteilen Sie folgende Aussagen.

a) Das Histogramm in Plot **A** und die empirische Verteilungsfunktion in Plot **C** beschreiben symmetrische Verteilungen.

b) Das Histogramm in Plot **A** und die empirische Verteilungsfunktion in Plot **C** stammen **nicht** von denselben Daten.

c) Der Boxplot in Plot **B** und das Histogram in Plot **D** können **nicht** von denselben Daten stammen.

d) Im Boxplot in Plot **B** sind 25% der Daten grösser als 0.6.

5. Es seien $X_1, ..., X_n \sim F$, wobei F eine Verteilung ist mit Erwartungswert $E[X_i] = \mu$ und endlicher Varianz σ_F^2. Beurteilen Sie folgende Aussagen.

a) Falls $X_1 = \ldots = X_n$ gilt, kann der zentralen Grenzwertsatz anwendet werden und die Varianz von $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ geht gegen 0.

b) Falls Unabhängigkeit gilt ($X_1, ..., X_n \overset{i.i.d}{\sim} F$), dann sagt der zentrale Grenzwertsatz aus, dass die Varianz $\text{Var}(X_1) = \sigma_F^2$ bei grossen n gegen 0 geht.

c) Damit der zentrale Grenzwertsatz gilt, müssen alle Zufallsvariablen X_i und X_j $(i \neq j)$ unabhängig sein.

d) Die Lebensdauer von normalen Batterien in Stirnlampen kann als unabhängig angenommen werden. Johannes möchte nun eine 5 stündige Höhlenexpedition unternehmen und fragt sich, wie viele Batterien er für die Stirnlampe mitnehmen sollte. Eine Batterie hält im Schnitt eine halbe Stunde ($0.5h$) mit einer Standardabweichung von $0.1h$. Johannes nimmt 12 Batterien mit auf die Expedition. Die Wahrscheinlichkeit, dass er die 5 stündige Expedition mit Licht bewältigen kann ist grösser als 95%.

6. Beurteilen Sie folgende Aussagen.

a) Die folgende Funktion $f(x)$ ist eine Dichtefunktion:

$$f(x) = \begin{cases} x & \text{falls } -1 \leq x < 0 \\ 1 & \text{falls } 0 \leq x < 2 \\ 0 & \text{sonst.} \end{cases}$$

b) Die Wartezeit in Minuten für die Essensausgabe in der Mensa um 12:00 kann mit einer Uniformverteilung $X \sim Uniform(4, 12)$ beschrieben werden. Die Wahrscheinlichkeit mehr als 8 Minuten zu warten liegt bei 50% (d.h. $P(X > 8) = 0.5$).

c) Die Wartezeit in Stunden in der Notfallaufnahme kann gut mit der Exponentialverteilung $Exp(1/2)$ beschrieben werden. Im Schnitt muss ein Patient 2 Stunden warten.

d) In einer Sägerei schneidet eine Maschine Bretter mit der gleichen Länge. Die kleinen Abweichungen in cm beim Schneideprozess kann man mit einer Standard-Normalverteilung ($X \sim \mathcal{N}(0, 1)$) beschreiben. Die Wahrscheinlichkeit ist 69%, dass die Maschine ein Brett mehr als $0.5cm$ zu kurz schneidet.

7. Beurteilen Sie folgende Aussagen.

a) Der Momentenschätzer für den Erwartungswert einer normalverteilten Zufallsvariabeln ist das arithmetische Mittel.

b) Der Maximum Likelihood Schätzer für die Erfolgswahrscheinlichkeit π bei einer Binomialverteilung mit $n = 1$ ist $\hat{\pi}_{MLE} = x_1$, wobei x_1 der beobachtete Datenpunkt ist.

c) Für alle Verteilungen gilt, dass der Maximum Likelihood Schätzer **nie** identisch zum Momentenschätzer ist.

d) Falls $X \sim Poisson(\lambda)$ mit $x = 2$ der beobachte Datenpunkt, dann hat der Momentenschätzer von λ einen Wert von 2^2.

8. Beurteilen Sie folgende Aussagen.

a) Herr Meier wird jeden Sonntag nach Fussballspielen in einem Stadion gesichtet. Meier ist also ein grosser Fussballfan.

b) Sei $Y := 2 - X$ und X zwei Zufallsvariabeln. Dann ist die Korrelation $Corr(X, Y) = -1$.

c) Bei einer Losbude kann man aus einer verdeckten Schüssel 2 Kugeln (ohne zurücklegen) ziehen. In der Schüssel befinden sich 2 schwarze und 8 weisse Kugeln. Zieht man beide schwarzen Kugeln gewinnt man. Die Wahrscheinlichkeit zu gewinnen ist 2.22%.

d) Die konkurrenzierende Losbude bietet ein ähnliches Spiel an wie die Losbude aus Unteraufgabe c) an. In der Schüssel dieser zweiten Losbude hat es 2 schwarze, 1 rote und 7 blaue Kugeln. Wenn man die zwei schwarzen Kugeln zieht (ohne zurrücklegen), dann hat man gewonnen. Die Wahrscheinlichkeit zu gewinnen ist höher als 2.22%.

Gruppe C

Binomialverteilung und -test

9. Angenommen, die Zufallsvariablen X_i für $i \in \{1, 2, 3, 4\}$ sind unabhängig und identisch verteilt mit Verteilung $P(X_i = 0) = 0.3$ und $P(X_i = 1) = 0.7$. Dann gilt ...

a) Für alle $i \in \{1, 2, 3, 4\}$ hat X_i die Verteilung $Binomial(1, 0.3)$.

b) $E(X_1) = 0.7$.

c) $E(X_1 + X_2 + X_3 + X_4) = 2.8$.

d) Angenommen, $X \sim Bernoulli(\pi)$ und $Y \sim Binomial(n, \pi)$. Dann gilt allgemein, dass $n \operatorname{Var}(X) = \operatorname{Var}(Y)$.

10. Michael tritt einen Ball gegen eine Torwand. Wir nehmen an, dass seine Schüsse unabhängig voneinander sind. Die Zufallsvariable X_i beschreibt das Ereignis, dass Michael im $i$ten Versuch trifft ($X_i = 1$) oder verfehlt ($X_i = 0$). Im Schnitt trifft Michael in 10% der Fälle. Insgesamt schiesst er 100 mal auf die Torwand. Sei die Anzahl Treffer Y definiert als $Y = X_1 + X_2 + X_3 + ... + X_{100}$. Dann gilt ...

a) Die Anzahl der fehlgeschlagenen Versuche berechnet man als $100 - Y$.

b) Y ist binomialverteilt mit $n = 100$ und $\pi = 0.1$.

c) Es gilt $P(Y < 50) > P(Y \geq 50)$.

d) Von den 100 Versuchen erwarten wir, dass Michael 10 mal trifft.

11. Bei einem Binomialtest ($n = 8,\ \alpha = 0.05$) mit $H_0 : \pi = 0.5$ und $H_A : \pi > 0.5$ und Testatistik X, ist der beobachtete Wert der Teststatistik $x = 6$. Beurteilen Sie folgende Aussagen.

a) Der Verwerfungsbereich ist $K = \{7, 8\}$.

b) Der P-Wert der Daten ist 0.145.

c) Der P-Wert ist gleich $P(X \in \{6, 7, 8\})$.

d) Je grösser der Fehler 1.Art, desto grösser ist die Macht.

12. Beurteilen Sie folgende Aussagen.

a) Die Faustregeln für die Normalapproximation einer Binomialverteilung sagen aus, dass bei einem Wert von $\pi = 0.5$ mindestens $n = 11$ Stichproben vorhanden sein müssen.

b) Wenn X_1 und X_2 bernoulliverteilt mit gleichem Parameter $\pi = 0.4$ sind, dann ist $X_1 + X_2$ binomialverteilt mit $n = 2$ und $\pi = 0.4$.

c) Wenn Y_1 und Y_2 i.i.d. $Binomial(n, \pi)$ verteilt sind, dann gilt $Y_1 + Y_2 \sim Binomial(2n, \pi)$ verteilt.

d) Auf einem Parkplatz stehen 14 rote Autos und 7 blaue Autos. Jemand bestimmt zufällig 5 Autos, welche abgeschleppt werden sollen. Die Anzahl der blauen Autos unter den 5 abgeschleppten Autos kann mit einer Binomialverteilung genau modelliert werden.

13. Bei einem Binomialtest $(n = 100, \alpha = 0.05)$ mit $H_0 : \pi = 0.5$ und $H_A : \pi \neq 0.5$ wurde der Verwerfungsbereich $K = \{0, ..., 39\} \cup \{61, ..., 100\}$ konstruiert. Sei X die Teststatistik. Der beobachtete Wert der Teststatistik ist $x = 71$. Beurteilen Sie folgende Aussagen.

a) Unter H_0 ist die Wahrscheinlichkeit $P(X = 71) = 0.05$.

b) $P_{H_0}(X \in K) \leq 0.05$.

c) Angenommen $Y \sim Binomial(n = 100, \pi = 0.6)$. Die Macht für die konkrete Alternative ist gegeben durch $P(Y \in K)$.

d) Der P-Wert in diesem Fall ist ungefähr $2P_{H_0}(X \geq 71)$.

Gruppe C

t-Test

14. In einer Studie wird ein neues Medikament zur Senkung des Augeninnendrucks getestet. Es wurden 50 Probanden über 4 Wochen hinweg mit dem Medikament behandelt. Dafür wurde bei jedem Proband das Medikament in das linke Auge getropft. Das rechte Auge wurde jeweils unbehandelt gelassen. Der Augeninnendruck des rechten bzw. des linken Auges pro Proband nach der Behandlung ist in der folgenden (unvollständigen) Tabelle aufgeführt:

Rechtes Auge x_i	15	21	17	...	29
Linkes Auge y_i	14	19	24	...	16

Wir nehmen an, dass die Messungen der jeweiligen Probanden unabhängig voneinander sind, sowie einer Normalverteilung folgen mit Erwartungswert μ_x (rechtes Auge), beziehungsweise μ_y (linkes Auge) und gleicher Standardabweichung σ.

Beurteilen Sie folgende Aussagen.

a) Es handelt sich um ungepaarte Stichproben.

b) Bei einem gepaarten t-Test können die Stichprobengrössen in beiden Gruppen ungleich sein.

c) Wir können bei gepaarten Stichproben sowohl einen gepaarten t-Test wie auch einen ungepaarten t-Test verwenden.

d) Der z-Test setzt voraus, dass die wahre Standardabweichung bekannt ist.

15. Aus den Daten der Augenstudie (vorherige Aufgabe) wurden folgende Kennzahlen ermittelt: $\bar{x} = 20.2$, $\bar{y} = 17.6$ und die geschätzte Varianz von $X - Y$ ist $\widehat{Var} = 4.5^2$. Wir führen nun einen gepaarten t-Test durch. Wir legen den Test folgendermassen an: $H_0 : \mu_x = \mu_y, \ H_A : \mu_x > \mu_y$.

a) Die Macht des Tests für eine einseitige Alternativhypothese ist immer grösser als für $H_A : \mu_x \neq \mu_y$.

b) Der beobachtete Wert der Teststatistik liegt zwischen 3.5 und 3.55.

c) Unter H_0 folgt die Teststatistik einer t_{49}-Verteilung.

d) Angenommen, die Teststatistik wäre 2.64 und der Verwerfungsbereich des Tests wäre $[1.69, \infty)$ für das Signifikanzniveau $\alpha = 5\%$, dann wird die Nullhypothese auf dem 5%-Signifikanzniveau verworfen.

16. Beurteilen Sie folgende Aussagen.

a) Um den Wilcoxon-Test anwenden zu können, müssen die Datenpunkte i.i.d. sein und von einer symmetrischen Verteilung stammen.

b) Angenommen wir würden die Differenzen $\{x_1 - y_1, ..., x_{50} - y_{50}\}$ aus der Augen-studie betrachten. Nun führen wir auf diesem Datensatz einen Vorzeichen-Test durch, um eine Senkung des Augeninnendrucks durch das Medikament zu überprüfen. Die Verteilung der entsprechenden Teststatistik unter der Nullhypothese ist nun $Binomial(49, 0.5)$.

c) Wir führen einen ungepaarten t-Test auf den Daten der Augen-Studie durch mit $H_0 : \mu_x = \mu_y$ und $H_A : \mu_x \neq \mu_y$. Angenommen $\bar{x} = 20.2$, $\bar{y} = 17.6$, $\hat{\sigma}_x = 3.9$ und $\hat{\sigma}_y = 4.7$. Die Teststatistik des ungepaarten t-Tests liegt zwischen 3.0 und 3.05.

d) Angenommen ein t-Test liefert einen P-Wert, der es uns erlaubt, die Nullhypothese auf dem 5%-Signifikanzniveau zu verwerfen. Dann können wir die Nullhypothese immer auch auf dem 1%-Signifikanzniveau verwerfen.

17. Beurteilen Sie folgende Aussagen.

a) Beim Vorzeichen-Test muss die wahre Standardabweichung der Zufallsvariablen bekannt sein.

b) Die Nullhypothese des Vorzeichen-Tests lautet $H_0 : \mu = \mu_0$, wobei μ der Mittelwert und μ_0 ein vorgegebener, bekannter Wert ist.

c) Bei einem Einstichproben t-Test ($H_0 : \mu = 0,\ H_A : \mu > 0$, 30 Beobachtungen insgesamt) ist der P-Wert 2.5%. Daher muss der beobachtete Wert der Teststatistik 2.06 gewesen sein.

d) Wir führen einen Einstichproben t-Test ($H_0 : \mu = 0,\ H_A : \mu \neq 0$, 24 Beobachtungen insgesamt) durch. Angenommen wir haben $\bar{x} = 15.4$ und $\hat{\sigma}_x = 1.8$. Ein exaktes zweiseitiges 95%-Vertrauensintervall für μ ist $[14.64, 16.16]$.

Lineare Regression

18. Lukas, der Sprengexperte, hat bei verschiedenen Sprengungen in Silber- und Goldminen die verwendetete Menge TNT und das abgesprengte Volumen gemessen. Er möchte das abgesprengte Volumen (Variable `volumen`), gemessen in Kubikzentimetern cm^3, in Abhängigkeit der verwendeten Menge TNT (Variable `tnt`), gemessen in Gramm g, beschreiben. Hierfür verwendet er ein lineares Modell:

$$\texttt{volumen}_i = \beta_0 + \beta_1 \cdot \texttt{tnt}_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2).$$

Der (unvollständige) Regressionsoutput sieht wie folgt aus:

```
Residuals:
   Min     1Q Median     3Q    Max
 -43.9  -12.8    1.0   12.1   47.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.6384        ???    0.77     0.44
tnt           1.2423     0.0159     ???   <2e-16 ***
---
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.9 on 62 degrees of freedom
Multiple R-squared:  0.99,        Adjusted R-squared:  0.99
F-statistic: 6.12e+03 on 1 and 62 DF,  p-value: <2e-16
```

Beurteilen Sie folgende Aussagen.

a) Die lineare Regression wurde basierend auf $n = 60$ Datenpunkten berechnet.

b) Der geschätzte Standardfehler des Parameters $\hat{\beta}_0$ ist 4.73.

c) $H_0 : \beta_1 = 0$ wird auf dem 1% Niveau verworfen.

d) Ein zweiseitiges 99%-Vertrauensintervall für β_1 ist gegeben durch $[1.20, 1.28]$. (Das entsprechende Quantil der t-Verteilung ist 2.67)

19. Mit Hilfe des geschätzten Modells aus der Sprengstudie (vorherige Aufgabe) wollen wir nun Vorhersagen machen. Beurteilen Sie folgende Aussagen.

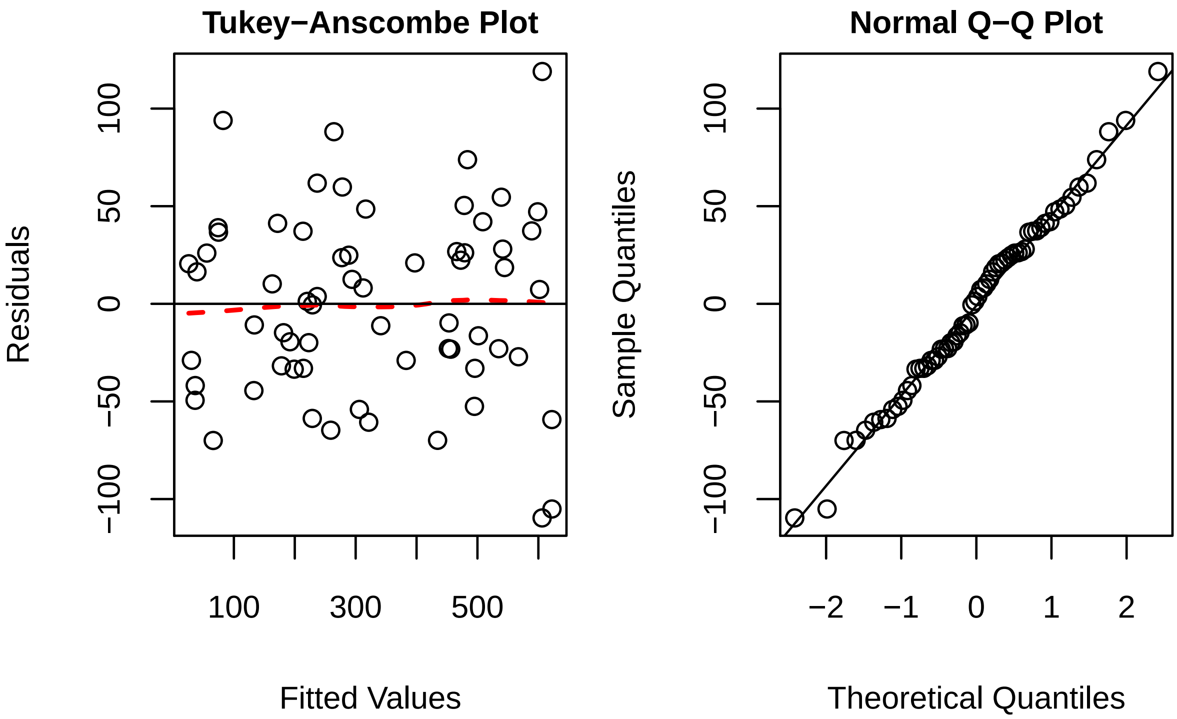
a) Wenn Lukas die Menge an eingesetztem TNT um $10\,g$ reduziert, dann reduziert sich das erwartete Sprengvolumen um $0.124\,cm^3$.

b) Angenommen, Lukas hat mit dem Modell ein erwartetes Sprengvolumen von $35\,cm^3$ vorhergesagt. Dafür muss Lukas $33\,g$ TNT einsetzen.

c) Falls Lukas $0.30\,kg$ TNT verwendet, kann er im Mittel ein Volumen von $4.01\,cm^3$ sprengen.

d) Das 95%-Vorhersageintervall für das Sprengvolumen beschreibt den Bereich, in welchem sich das erwartete Sprengvolumen mit 95% Wahrscheinlichkeit befindet.

20. Betrachten Sie die nachfolgenden Plots, welche die Residuen aus dem Regressionsmodell der Sprengstudie zeigen.



Beurteilen Sie folgende Aussagen.

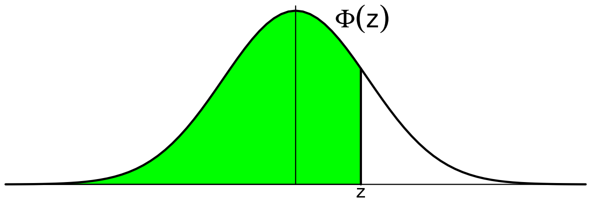
a) Der Mittelwert der Residuen ist in etwa 0.

b) Die Normalitätsannahme ist erfüllt.

c) Es gibt zwei starke Ausreisser, daher sind die Parameterschätzungen der linearen Regression verfälscht.

d) Angenommen, wir verwenden den Logarithmus des Sprengvolumen als Zielvariable ($\log(\texttt{volumen}_i) = \beta_0 + \beta_1 \texttt{tnt}_i + \varepsilon_i$). Dann ist dieses Regressionsmodell kein lineares Modell.

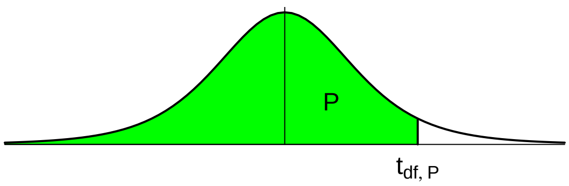
Tabelle der Kumulativen Normalverteilung $\Phi(z) = \mathrm{P}\left[Z \leq z\right], \; Z \sim \mathcal{N}(0,1)$



Bsp.: $\mathrm{P}\left[Z \leq 1.96\right] = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Perzentile der t-Verteilung



Bsp.: $t_{9;\ 0.975} = 2.262$

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576