

Nicht bestanden: ☐

Vorname: _____

Matrikelnummer: _____

Endnote: _____

Fakultät Agrarwissenschaften und Landschaftsarchitektur (AuL)

Klausurfragen Bio Data Science

für Pflichtmodule

im 1. & 2. Semester B.Sc./M.Sc.

(Prüfungsleistung der Wahlpflichtmodule ist eine Portfolioprüfung)

Prüfer: Prof. Dr. Jochen Kruppa-Scheetz
Fakultät für Agrarwissenschaften und Landschaftsarchitektur
j.kruppa@hs-osnabrueck.de

Wintersemester 2024/25

*„The test of a student is not how much he knows,
but how much he wants to know.“*
— Alice W. Rollins

Erlaubte Hilfsmittel

- Normaler Taschenrechner ohne Möglichkeit der Kommunikation mit anderen Geräten! Ausdrücklich kein Handy!
- Eine DIN A4-Seite als beidseitig, selbstgeschriebene, handschriftliche Formelsammlung. Keine digitalen Ausdrucke!
- **Die Verwendung eines roten Farbstiftes ist nicht gestattet! Korrekturfarbe!**
- *You can answer the questions in English without any consequences.*

Endnote

_____ von 20 Punkten sind aus den Multiple Choice Aufgaben erreicht.
_____ von 56 Punkten sind aus den Rechen- und Textaufgaben erreicht.
_____ von 76 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
72.5 - 76.0	1,0
69.0 - 72.0	1,3
65.0 - 68.5	1,7
61.5 - 64.5	2,0
57.5 - 61.0	2,3
53.5 - 57.0	2,7
50.0 - 53.0	3,0
46.0 - 49.5	3,3
42.5 - 45.5	3,7
38.0 - 42.0	4,0

Es ergibt sich eine Endnote von _____.

Multiple Choice Aufgaben

- Pro Multiple Choice Frage ist *genau* eine Antwort richtig.
- Übertragen Sie Ihre Kreuze in die Tabelle auf dieser Seite.

	A	B	C	D	E	✓
Aufgabe 1						
Aufgabe 2						
Aufgabe 3						
Aufgabe 4						
Aufgabe 5						
Aufgabe 6						
Aufgabe 7						
Aufgabe 8						
Aufgabe 9						
Aufgabe 10						

- Es sind ____ von 20 Punkten erreicht worden.

Rechen- und Textaufgaben

Aufgabe	11	12	13	14	15	16	17
Punkte	8	12	7	9	8	12	NA

- Es sind ____ von 56 Punkten erreicht worden.

Multiple Choice Aufgaben

Die Multiple Choice Aufgaben **unterliegen dem Zufall**. Die Reihenfolge der Antworten ist zufällig. Die Fragen und Antworten sind semantisch zufällig und haben somit **verschiedene Textvarianten**. Insbesondere die reinen Textaufgaben haben verschiedene Versionen. Der Text mag sich unterscheiden, die Inhalte sind aber gleich.

ANOVA

1 Aufgabe

(2 Punkte)

Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein $\eta^2 = 0.12$. Welche Aussage ist richtig?

- A** ☐ Das η^2 beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen nicht erklärt wird. Somit der Rest an nicht erklärbarer Varianz.
- B** ☐ Das η^2 ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.
- C** ☐ Die Berechnung von η^2 ist ein Wert für die Interaktion.
- D** ☐ Das η^2 wird genutzt um zu erfahren welchen Anteil der Varianz die Behandlungsbedingungen erklären.
- E** ☐ Das η^2 ist ein Wert für die Güte der ANOVA. Je kleiner desto besser. Ein η^2 von 0 bedeutet ein perfektes Modell mit keiner Abweichung. Die Varianz ist null.

2 Aufgabe

(2 Punkte)

Sie führen ein Feldexperiment durch um das Gewicht von Erdbeeren zu steigern. Die Pflanzen wachsen unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen. Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein $\eta^2 = 0.21$. Welche Aussage ist richtig?

- A** ☐ Das η^2 beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird. Daher werden 21% der Varianz erklärt.
- B** ☐ Das η^2 beschreibt den Anteil der Varianz, der von den Umweltbedingungen erklärt wird. Daher werden 21% der Varianz durch die Umweltbedingungen erklärt. Der Anteil der Varianz durch die Behandlungsgruppen ist dann 79%.
- C** ☐ Es werden 21% der Varianz durch den Versuch erklärt. Das η^2 beschreibt den Anteil der Varianz, der durch Fehler in der Versuchsdurchführung entsteht.
- D** ☐ Es werden 79% der Varianz durch die Behandlung erklärt. Das η^2 beschreibt den Anteil der Varianz, der von den unterschiedlichen Behandlungsbedingungen nicht erklärt wird.
- E** ☐ Das η^2 beschreibt den Anteil der Varianz, der durch den Forschenden entsteht. Es gilt die Regel, dass ca. 70% der Varianz eines Versuches durch die Versuchsdurchführung entstehen sollen.

3 Aufgabe

(2 Punkte)

Sie rechnen eine einfaktorielle ANOVA und erhalten eine Teststatistik. Nun müssen Sie diese Teststatistik interpretieren. Welche Aussage ist richtig?

- A** ☐ Die F-Statistik wird berechnet indem die MS der Behandlung durch die MS des Fehlers geteilt werden. Wenn die F-Statistik sich kaum von der Null unterscheidet kann die Nullhypothese nicht abgelehnt werden.
- B** ☐ Die ANOVA berechnet die F-Statistik indem die MS des Fehlers durch die MS der Behandlung geteilt werden. Wenn die F-Statistik sich der 1 annähert kann die Nullhypothese nicht abgelehnt werden.
- C** ☐ Die ANOVA berechnet die T-Statistik indem den Mittelwertsunterschied der Gruppen simultan durch die Standardabweichung der Gruppen teilt. Wenn die T-Statistik höher als 1.96 ist, kann die Nullhypothese abgelehnt werden.

- D** ☐ Wenn die F-Statistik höher ist als der kritische Wert kann die Nullhypothese nicht abgelehnt werden. Die F-Statistik ist die Differenz der MS der Behandlung durch die MS des Fehlers.
- E** ☐ Die ANOVA berechnet die T-Statistik aus der Multiplikation der MS Behandlung mit der MS der Fehler. Wenn die F-Statistik genau 0 ist, kann die Nullhypothese nicht abgelehnt werden.

4 Aufgabe

(2 Punkte)

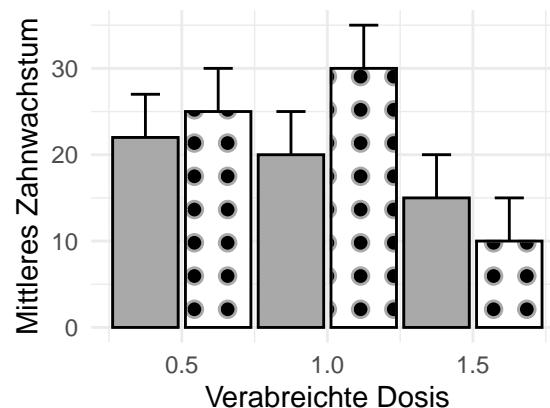
Viele statistische Verfahren nutzen eine Teststatistik um eine Aussage über den Zusammenhang zwischen der Grundgesamtheit und der Stichprobe abzubilden. Ein statistisches Testwerkzeug ist hierbei die ANOVA. Die ANOVA rechnet dabei...

- A** ☐ ... den Unterschied zwischen der Varianz über alle Behandlungsgruppen oder der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, muss sich zwischen einem der beiden Varianzquellen entschieden werden.
- B** ☐ ... den Unterschied zwischen der Varianz ausgelöst durch alle Behandlungsgruppen und der Varianz aus globalen Behandlungsgruppen der Kontrollen. Wenn die ANOVA nicht signifikant ist, muss ein Posthoc-Test ausgeschlossen werden.
- C** ☐ ... den Unterschied zwischen der Mittelwerte und der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, ist bekannt welcher Vergleich konkret unterschiedlich ist.
- D** ☐ ... den Unterschied zwischen mehreren Varianzen aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, ist nicht bekannt welcher Vergleich konkret unterschiedlich ist.
- E** ☐ ... den Unterschied zwischen der globalen Varianz und der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, ist nicht bekannt welcher Vergleich konkret unterschiedlich ist.

5 Aufgabe

(2 Punkte)

Ein Versuch wurde an 55 Tieren durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0.5, 1 und 1.5 mg/Tag) über eine von zwei Verabreichungsmethoden erhielt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung von Vitamin D auf das Zahnwachstum bei Schweine. Welche Aussage ist richtig, wenn Sie eine zweifaktorielle ANOVA rechnen?



- A** ☐ Eine mittlere bis starke Interaktion liegt vor ($p \leq 0.05$)
- B** ☐ Keine Korrelation liegt vor ($p \geq 0.05$).
- C** ☐ Die Koeffizienten sind positiv ($\beta_0 > 0$; $\beta_1 > 0$).
- D** ☐ Keine Interaktion liegt vor ($p \leq 0.05$).
- E** ☐ Das Bestimmtheitsmaß R^2 ist groß.

Deskriptive Statistik & Explorative Datenanalyse

6 Aufgabe

(2 Punkte)

Wie lautet der Mittelwert und Standardabweichung von y mit 10, 9, 6, 9 und 8.

- A ☐ Es berechnet sich 9.4 ± 2.3
- B ☐ Sie erhalten 8.4 ± 1.52
- C ☐ Es berechnet sich 8.4 ± 2.3
- D ☐ Sie erhalten 8.4 ± 0.76
- E ☐ Sie erhalten 8.4 ± 1.23

7 Aufgabe

(2 Punkte)

Berechnen Sie den Median, das 1st Quartile sowie das 3rd Quartile von y mit 18, 13, 24, 32, 22, 23, 22, 29, 30, 24 und 42.

- A ☐ Es berechnet sich 25 [23; 31]
- B ☐ Sie erhalten 24 ± 30
- C ☐ Es berechnet sich 25 [23; 29]
- D ☐ Es ergibt sich 25 ± 22
- E ☐ Sie erhalten 24 [22; 30]

8 Aufgabe

(2 Punkte)

Mit einem Dotplot können Sie sehr gut die Verteilung von Daten visualisieren. Die empfohlene Mindestanzahl an Beobachtungen ist dabei?

- A ☐ Die optimale Anzahl ist größer als hundert Beobachtungen, wobei es gerne sehr viel mehr sein können.
- B ☐ 1 Beobachtung.
- C ☐ Wir sollten zwei bis fünf Beobachtungen mindestens pro Gruppe vorliegen haben.
- D ☐ Mindestens 20 Beobachtungen.
- E ☐ Wir sollten eine Beobachtung mindestens pro Gruppe vorliegen haben.

9 Aufgabe

(2 Punkte)

Um die Varianz zu berechnen müssen wir folgende Rechenoperationen durchführen.

- A ☐ Den Mittelwert berechnen, dann die quadratischen Abstände zum Mittelwert aufsummieren und durch die Fallzahl teilen.
- B ☐ Als erstes berechnen wir den Mittelwert. Dann bilden wir die Summe der quadratischen Abstände zu dem Mittelwert. Abschließend subtrahieren wir die Fallzahl.
- C ☐ Als erstes berechnen wir den Mittelwert. Dann bilden wir die Summe der quadratischen Abstände zu dem Mittelwert. Abschließend teilen wir durch die Fallzahl. Nicht zu vergessen, am Ende dann noch die Wurzel zu ziehen.
- D ☐ Den Median berechnen, dann die quadratischen Abstände zum Median aufsummieren, dann die Wurzel ziehen.
- E ☐ Den Mittelwert berechnen und die Abstände quadrieren. Die Summe mit der Fallzahl multiplizieren.

10 Aufgabe

(2 Punkte)

Der Boxplot stellt folgende statistische Maßzahlen in einer Abbildung dar. Damit gehört der Boxplot zu einem der am meisten genutzten statistischen Verfahren zur Visualisierung von Daten.

- A ☐ Der Boxplot stellt die Mittelwerte und die Standardabweichung dar.
- B ☐ Den Mittelwert und die Varianz.
- C ☐ Den Median und die Quartile.
- D ☐ Der Boxplot stellt die Mittelwerte und die Varianz dar.
- E ☐ Den Mittelwert sowie den Median und die Streuung.

11 Aufgabe

(2 Punkte)

Nachdem Sie in einem Feldexperiment zu Leistungssteigerung von Kartoffel durchgeführt haben, berechnen Sie den Mittelwert und den Median. Der Mittelwert \bar{y} und der Median \tilde{y} unterscheiden sich nicht. Welche Aussage ist richtig?

- A ☐ Wenn sich der Mittelwert und der Median unterscheiden, liegen vermutlich Outlier in den Daten vor.
- B ☐ Da sich der Mittelwert und der Median nicht unterscheiden, liegen vermutlich Outlier in den Daten vor. Wir untersuchen den Datensatz nach auffälligen Beobachtungen.
- C ☐ Der Mittelwert und der Median sollten sich unterscheiden sein, wenn Outlier in den Daten vorliegen.
- D ☐ Da sich der Mittelwert und der Median unterscheiden, liegen vermutlich keine Outlier in den Daten vor. Wir verwenden den Datensatz so wie er ist.
- E ☐ Da sich der Mittelwert und der Median unterscheiden, ist der Datensatz nicht zu verwenden. Mittelwert und Median müssen gleich sein.

12 Aufgabe

(2 Punkte)

Sie wollen eine ANOVA im Anschluss an Ihr Feldexperiment rechnen. Dafür muss Ihr gemessener Endpunkt die Annahme einer Varianzhomogenität genügen. Zur Überprüfung können Sie folgende Visualisierung nutzen. Welche entsprechende Regel zur Abschätzung der Annahme einer Varianzhomogenität kommt zur Anwendung?

- A ☐ Nach dem Einlesen der Daten nutzen wir einen Barplot um zu schauen, ob alle Mittelwerte über alle Behandlungen in etwa gleich groß sind. Damit ist dann auch die Varianz in allen Behandlungen in etwa gleich.
- B ☐ Einen Boxplot. Der Median, dargestellt als Linie, muss in der Mitte des IQR, dargestellt durch die Box, liegen.
- C ☐ Einen Violinplot. Der Bauch der Violine muss hierbei einen höheren Wert annehmen als der Steg der Violine. Dann kann die Annahme einer Varianzhomogenität angenommen werden.
- D ☐ Einen Boxplot. Das IQR muss über alle Behandlungen zusammen mit den Whiskers ungefähr gleich aussehen.
- E ☐ Wir erstellen uns für jede Behandlung einen Dotplot und schauen, ob die Dots und damit die Varianz für jede Behandlung gleich groß sind.

13 Aufgabe

(2 Punkte)

In der Statistik müssen wir häufig überprüfen, ob unser Outcome einer bestimmten Verteilung folgt. Meistens überprüfen wir, ob eine Normalverteilung vorliegt. Folgende drei Abbildungen eignen sich im Besonderen für die Überprüfung einer Verteilungsannahme an eine Variable.

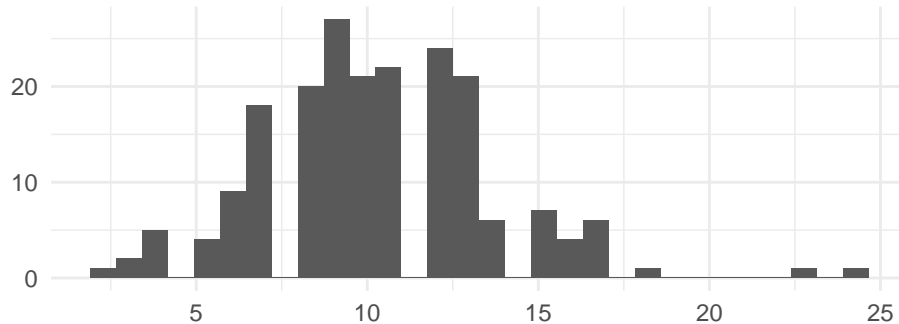
- A ☐ Violinplot, Scatterplot, Barplot
- B ☐ Violinplot, Boxplot, Densityplot

- C** ☐ Histogramm, Scatterplot, Boxplot
- D** ☐ Barplot, Mosaicplot, Violinplot
- E** ☐ Histogramm, Densityplot, Dotplot

14 Aufgabe

(2 Punkte)

In dem folgenden Histogramm von $n = 187$ Pflanzen ist welche Verteilung abgebildet?



- A** ☐ Dem Histogramm entnehmen wir eine Poisson-Verteilung.
- B** ☐ Es handelt sich um eine Binomial-Verteilung.
- C** ☐ Eine multivariate Normalverteilung.
- D** ☐ Wir haben eine Gammaverteilung vorliegen.
- E** ☐ In dem Histogramm ist eine Normalverteilung dargestellt.

Lineare Regression & Korrelation

15 Aufgabe

(2 Punkte)

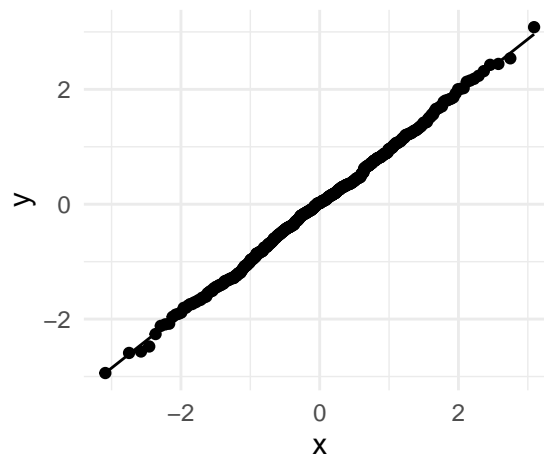
Im Allgemeinen gibt es zwei mögliche Ziele für ein Regressionsmodell. Wir können ein Vorhersagemodell oder ein kausales Modell rechnen. Welche Aussage ist für ein prädiktives Modell richtig?

- A** ☐ Ein prädiktives Modell möchte die Zusammenhänge von X auf Y modellieren. Hierbei geht es um die Effekte von X auf Y . Man sagt, wenn x_1 um 1 ansteigt ändert sich Y um einen Betrag β_1 .
- B** ☐ Es wird ein Trainingsdatensatz zum Trainieren des Modells benötigt. Der Testdatensatz dient zur Validierung. Dies gilt insbesondere für ein prädiktives Modell.
- C** ☐ Es wird ein Trainingsdatensatz zum Modellieren des Trainingsmodells benötigt. Der Testdatensatz dient rein zur Visualisierung. Dies gilt vor allem für ein prädiktives Modell.
- D** ☐ Ein prädiktives Modell wird auf einem Trainingsdatensatz trainiert und anschliessend über eine explorative Datenanalyse validiert. Signifikanzen über β_i können hier nicht festgestellt werden.
- E** ☐ Wenn ein prädiktives Modell gerechnet werden soll dann kann dies auf dem gesamten Datensatz geschehen. Das Ziel ist es einen Zusammenhang von X auf Y zu modellieren. Wie wirken sich die Einflussvariablen Y auf die gemessenen Endpunkte $X = x_1, \dots, x_p$ aus?

16 Aufgabe

(2 Punkte)

Nach der Modellierung einer Regression stellt sich die Frage, ob die Residuen approximativ einer Normalverteilung folgen. Sie können einen QQ-Plot für die visuelle Überprüfung der Annahme an die Residuen nutzen. Welche Aussage ist richtig?

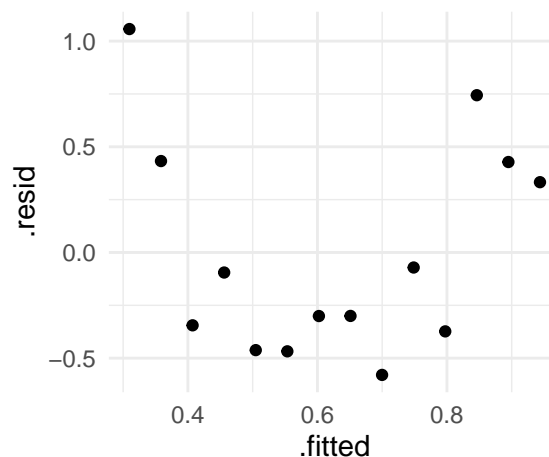


- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- B** ☐ Wir betrachten die Punkte. Wenn die Punkte einigermaßen gleichmäßig verteilt liegen, dann gehen wir von normalen Residuen aus.
- C** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden und Korrelation ist negativ.
- D** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden.
- E** ☐ Wir betrachten die Gerade und dabei insbesondere die beiden Enden der Gerade. Hier sollten die Punkte auf der Geraden liegen, dann ist die Annahme an die Normalverteilung der Residuen erfüllt. Diese Annahme ist nicht erfüllt.

17 Aufgabe

(2 Punkte)

Nach einer Regressions sollten die Residuen (.resid) gleichmäßig um die Gerade verortet sein. Was bei einer simplen Regression noch relativ einfach visuell in einem Scatterplot zu überprüfen ist. Für komplexere Modell liefert der Residual Plot die notwendigen Informationen. Welche Aussage ist richtig?



- A** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Vereinzelte Punkte liegen oberhalb bzw. unterhalb der Geraden um die 0 Linie weiter entfernt. Ein klares Muster ist zu erkennen.
- B** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Es ist ein Muster zu erkennen und wir können damit auf die Signifikanz von x_1, \dots, x_p schließen.
- C** ☐ Wir betrachten die Nulllinie und alle Punkte sollten ohne Muster gleichmäßig um die Nulllinie liegen. Da dies der Fal ist, gehen wir von keinen Ausreißern aus.

- D** ☐ Wenn wir die Nulllinie betrachten so müssen die Punkte gleichmäßig über der Nulllinie liegen. Unser Modell erfüllt somit nicht die Annahme von normalverteilten Residuen mit einem Mittelwert von > 0 und einer Streuung von s .
- E** ☐ Die Punkte müssen gleichmäßig, mit ähnlichen Abständen, in dem positiven wie auch negativen Bereich liegen. Dies ist hier klar nicht der Fall. Einzelne Ausreißer können beobachtet werden. Wir können mit dem Model so nicht rechnen und müssen erst die auffälligen Werte gesondert betrachten.

18 Aufgabe

(2 Punkte)

In den Humanwissenschaften wird der Korrelationskoeffizienten ρ sehr häufig verwendet. Daher ist es auch wichtig für andere Forschende den Korrelationskoeffizienten ρ zu verstehen. Welche Aussage zu dem Korrelationskoeffizienten ρ ist richtig?

- A** ☐ Korrelationskoeffizienten ρ liegt zwischen 0 und 1. Darüber hinaus ist der Korrelationskoeffizienten ρ einheitslos und kann als Standardisierung verstanden werden.
- B** ☐ Der Korrelationskoeffizienten ρ liegt zwischen -1 und 1. Darüber hinaus ist der Korrelationskoeffizienten ρ einheitslos und kann als standardisierte Steigung verstanden werden.
- C** ☐ Der Korrelationskoeffizienten ρ liegt zwischen -1 und 1. Darüber hinaus ist der Korrelationskoeffizienten ρ als standardisierte Steigung zu verstehen, wenn eine Standardisierung durchgeführt wurde. Diese Adjustierung nach Fischer muss am Anschluß der Berechnung der Korrelation durchgeführt werden.
- D** ☐ Der Korrelationskoeffizienten ρ ist eine standardisierte, statistische Maßzahl, die zwischen 0 und 1 liegt. Dabei ist Korrelationskoeffizienten ρ einheitslos. Eine Signifikanz kann nicht nachgewiesen werden.
- E** ☐ Der Korrelationskoeffizienten ρ wird wie das η^2 aus der ANOVA interpretiert. Der Korrelationskoeffizienten ρ beschreibt den Anteil an erklärter Varianz durch die Regression. Dabei gibt er jedoch eine Richtung an und kann auch negativ werden.

19 Aufgabe


(2 Punkte)




Nach einer simplen linearen Regression zur Untersuchung vom Einfluss der PO_2 -Konzentration in $[\mu g]$ im Wasser auf den absoluten Proteingehalt von Erbsen in $[kg]$ erhalten Sie einen β_{PO_2} Koeffizienten von 1.1×10^{-5} und einen hoch signifikanten p -Wert mit $2e - 04$. Warum sehen Sie so einen kleinen Effekt bei einer so deutlichen Signifikanz?

- A** ☐ Das Gewicht und die PO_2 -Konzentration korrelieren sehr stark, deshalb wird der β_{PO_2} Koeffizient sehr klein. Mit einer ANOVA kann für die Korrelation korrigiert werden und der Effektschätzer passt dann zum p -Wert.
- B** ☐ Die Einheit der PO_2 -Konzentration ist zu klein gewählt. Dadurch sehen wir den sehr kleinen p -Wert. Der p -Wert und die Einheit von der PO_2 -Konzentration hängen antiproportional zusammen.
- C** ☐ Die Einheit der PO_2 -Konzentration ist zu klein gewählt. Die Erhöhung der PO_2 -Konzentration um 1 Einheit führt nur zu einem sehr winzigen Anstieg von β_{PO_2} im Gewicht der Wasserlinsen. Die Einheit $[\mu g]$ muss besser gewählt werden.
- D** ☐ Die Fallzahl ist zu hoch angesetzt. Je höher die Fallzahl ist, desto kleiner ist die Teststatistik und damit ist dann auch der p -Wert sehr klein. Es sollte über eine Reduzierung der Fallzahl nachgedacht werden. Dann sollte der Effekt zum p -Wert passen.
- E** ☐ Manchmal ist die Einheit der Einflussvariable X zu groß gewählt, so dass der Anstieg von 1 Einheit in X zu einer zu großen Änderung in y führt. Daher kann der Effekt β_{PO_2} sehr klein wirken, da der p -Wert wird auf einer einheitslosen Teststatistik bestimmt wird.

20 Aufgabe

(2 Punkte)



Nachdem Sie Ihr Experiment abgeschlossen haben, stehen Sie vor der Frage wie Sie Ihre Daten modellieren sollen. In der Beispielauswertung von Ihrem Betreuenden finden Sie die Funktion $\text{lm}()$ in . Welche Aussage ist richtig?

- A** ☐ Die Funktion `lm()` in  ist der letzte Schritt für einen Gruppenvergleich. Vorher kann eine ANOVA oder aber ein multipler Vergleich in `{emmeans}` gerechnet werden. In der Funktion `lm()` werden die Gruppenvarianzen bestimmt.
- B** ☐ Die Funktion `lm()` in  wird klassischerweise für die nicht-lineare Regression genutzt. Ist die Einflussvariable X numerisch so werden die Gruppenmittelwerte geschätzt.
- C** ☐ Neben der klassischen Verwendung der Funktion `lm()` in der linearen Regression kann auch ein Gruppenvergleich gerechnet werden. Dafür müssen aber alle Faktoren aus den Daten entfernt und numerisch umgewandelt werden. Dann kann das R Paket `{emmeans}` genutzt werden um die Korrelation zu berechnen. Eine Adjustierung ist dann nicht mehr notwendig.
- D** ☐ Ist die Einflussvariable X ein Faktor so werden die Gruppenmittelwerte geschätzt und eine anschließende ANOVA sowie multipler Gruppenvergleich mit `{emmeans}` ist möglich. Die Funktion `lm()` kann dabei eigentlich weggelassen werden, wird aber traditionell gerechnet.
- E** ☐ Ist die Einflussvariable X ein Faktor so werden die Gruppenmittelwerte geschätzt und eine anschließende ANOVA sowie multipler Gruppenvergleich mit `{emmeans}` ist möglich. Dennoch muss zuerst ein lineares Modell mit der Funktion `lm()` in  gerechnet werden.

21 Aufgabe

(2 Punkte)

Welche Aussage über das *generalisierte lineare Modell (GLM)* ist richtig?

- A** ☐ Dank dem *generalisierten linearen Modell (GLM)* können auch andere Verteilungsfamilien als die Normalverteilung mit einer linearen Regression modelliert werden.
- B** ☐ Das GLM erlaubt auch nicht normalverteilte Residuen in der Schätzung der Regressionsgrade.
- C** ☐ Das GLM ist ein faktisch maschineller Lernalgorithmus, der selbstständig die Verteilungsfamilie für Y wählt.
- D** ☐ In  ist mit dem *generalisierten linearen Modell (GLM)* eine Modellierung implementiert, die die Poissonverteilung für Zähldaten oder die Binomialverteilung für 0/1-Daten modellieren kann. Weitere Modellierungen sind in  auch mit zusätzlich geladenen Paketen nicht möglich.
- E** ☐ Das *generalisierte lineare Modell (GLM)* erlaubt auch weitere Verteilungsgruppen für das X bzw. die Einflussvariablen in einer linearen Regression zu wählen.

Vermischte Themen

22 Aufgabe

(2 Punkte)

Die Randomisierung von Beobachtungen zu den Versuchseinheiten ist bedeutend in der Versuchsplanung. Welche der folgenden Aussagen ist richtig?

- A** ☐ Strukturgleichheit ist durch Randomisierung gegeben. Leider hilft die Randomisierung noch nicht um von der Stichprobe auf die Grundgesamtheit zu schließen. Deshalb wurde das Falsifikationsprinzip entwickelt.
- B** ☐ Strukturgleichheit ist durch Randomisierung gegeben. Somit kann von der Stichprobe auf die Grundgesamtheit geschlossen werden
- C** ☐ Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte. Ohne Mittelwerte keine Varianz und somit auch kein statistischer Test.
- D** ☐ Randomisierung ist die direkte Folge von Strukturgleichheit. Die Strukturgleichheit erlaubt es erst von der Stichprobe auf die Grundgesamtheit zurückzuschließen.
- E** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschließen.

23 Aufgabe

(2 Punkte)

Wenn Sie einen Datensatz erstellen, dann ist es ratsam die Spalten und die Einträge in englischer Sprache zu verfassen, wenn Sie später die Daten in R auswerten wollen. Welcher Aussage ist richtig?

- ☐ A R Pakete sind nur in englischer Sprache verfasst. Es macht keinen Sinn R daher in Deutsch zu bedienen.
- ☐ B Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Die Nutzung von englischer Sprache umgeht dieses Problem in eleganter Art.
- ☐ C Alle Funktionen und auch Anwendungen sind in R in englischer Sprache. Die Nutzung von deutschen Wörtern ist nicht schick und das ist zu vermeiden.
- ☐ D Die Spracherkennung von R ist nicht in der Lage Deutsch zu verstehen.
- ☐ E Programmiersprachen können nur englische Begriffe verarbeiten. Zusätzliche Pakete können zwar geladen werden, aber meist funktionieren diese Pakete nicht richtig. Deutsch ist International nicht bedeutend genug.

24 Aufgabe

(2 Punkte)

Bei der explorativen Datenanalyse (EDA) in R gibt es eine richtige Abfolge von Prozessschritten, auch extitCircle of life genannt. Wie lautet die richtige Reihenfolge für die Erstellung einer EDA?

- ☐ A Für eine explorativen Datenanalyse (EDA) in R müssen wir als erstes die Daten über `read_excel()` einlesen. Danach müssen wir schauen, dass wir die Zeilen richtig über `mutate()` transformiert haben. Insbesondere müssen Variablen mit kontinuierlichen Werten in einen Faktor umgewandelt werden. Am Ende nutzen wir die Funktion `ggplot()` für die eigentlich EDA.
- ☐ B Wir lesen als erstes die Daten über `read_excel()` ein, transformieren die Spalten über `mutate()` in die richtige Form und können dann über `ggplot()` uns die Abbildungen erstellen lassen. Wichtig ist, dass wir keine Faktoren sondern nur numerische Variablen vorliegen haben.
- ☐ C Die Funktionsreihenfolge ist wie folgt: `read_excel()` -> `mutate()` -> `ggplot()`. Dabei ist bei der Transformation der Daten darauf zu achten, dass die Faktoren richtig erstellt werden.
- ☐ D Die Funktionsreihenfolge ist wie folgt: `read_excel()` -> `mutate()` -> `ggplot()`. Dabei ist bei der Transformation der Daten darauf zu achten, dass keine Faktoren erstellt werden.
- ☐ E Wir lesen die Daten über eine generische Funktion `read()` ein und müssen dann die Funktion `ggplot()` nur noch installieren. Dann haben wir die Abbildungen als *.png vorliegen.

25 Aufgabe

(2 Punkte)

Sie haben das abstrakte Modell $Y \sim X$ mit X als Faktor mit zwei Levels vorliegen. Welche Aussage über $s_1^2 = s_2^2$ ist richtig?

- ☐ A Es handelt sich um ein balanciertes Design.
- ☐ B Es handelt sich um ein unbalanciertes Design.
- ☐ C Es liegt Varianzheterogenität vor.
- ☐ D Es handelt sich um abhängige Beobachtungen.
- ☐ E Es liegt Varianzhomogenität vor.

26 Aufgabe

(2 Punkte)

In einem Zuchtexperiment messen wir die Ferkel verschiedener Sauen. Die Ferkel einer Muttersau sind daher im statistischen Sinne...

- A** ☐ Untereinander unabhängig. Die Ferkel sind eigenständig und benötigen keine zusätzliche Behandlung.
- B** ☐ Untereinander stark korreliert. Die Ferkel sind von einer Mutter und somit miteinander korreliert. Dies wird in der Statistik jedoch meist nicht modelliert.
- C** ☐ Je nach Stallanlage kommt eine andere Analyse in Betracht. Eine allgemeine Aussage über Ferkel und Sauen lässt sich statistisch nicht treffen.
- D** ☐ Untereinander abhängig. Die Ferkel stammen von einem Muttertier und haben vermutlich eine ähnliche Varianzstruktur.
- E** ☐ Die Ferkel stammen von der gleichen Sau und sind somit untereinander unabhängig.

27 Aufgabe

(2 Punkte)

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Schweinen durch. Bei 3 Tieren finden Sie eine Erkrankung der Klauen vor und 8 Tiere sind gesund. Welche Aussage über den Effektschätzer Odds ratio ist richtig?

- A** ☐ Es ergibt sich ein Odds ratio von 0.27, da es sich um ein Anteil handelt. Wir berechnen den Anteil der Kranken.
- B** ☐ Das Verhältnis der Chancen Odds ratio ergibt ein Chancenverhältnis von 0.27. Wir sind an der Chance krank zu sein interessiert.
- C** ☐ Es ergibt sich ein Odds ratio von 2.67, da es sich um ein Anteil handelt.
- D** ☐ Es ergibt sich ein Odds ratio von 0.38, da es sich um eine Chancenverhältnis handelt
- E** ☐ Es ergibt sich ein Odds ratio von 0.27, da es sich um eine Chancenverhältnis handelt.

28 Aufgabe

(2 Punkte)

Historisch gesehen ergibt sich ein Problem, wenn Sie mit sehr großen Datensätzen, wie in der Bio Data Science üblich, rechnen. Warum ist es ein Problem, wenn Ihre Datensätze sehr groß werden hinsichtlich der Bewertung anhand der Signifikanz?

- A** ☐ Mehr Fallzahl in Datensätzen bedeutet mehr signifikante Ergebnisse, da in mehr Daten auch mehr Informationen beinhaltet sind. Deshalb lohnen sich riesige Datensätze, die durch die vielen signifikanten Ergebnisse auch eine Menge an relevanten Erkenntnissen liefern.
- B** ☐ Aktuell werden immer größere Datensätze erhoben. Eine erhöhte Fallzahl führt automatisch auch zu mehr signifikanten Ergebnissen, selbst wenn die eigentlichen Effekte nicht relevant sind.
- C** ☐ Aktuell werden immer größere Datensätze erhoben. Dadurch wird auch die Varianz immer höher was automatisch zu mehr signifikanten Ergebnissen führt.
- D** ☐ Riesige Datensätze haben mehr Fallzahl was zur α -Inflation führt. Durch eine Adjustierung kann dem Problem entgegengewirkt werden.
- E** ☐ Big Data ist ein Problem der parametrischen Statistik. Parameter lassen sich nur auf kleinen Datensätzen berechnen, da es sich sonst nicht mehr um eine Stichprobe im engen Sinne der Statistik handelt.

Multiple Gruppenvergleiche

29 Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.01, 0.001, 0.21, 0.03 und 0.89. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.002, 2e-04, 0.042, 0.006 und 0.178. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.05, 0.005, 1.05, 0.15 und 4.45. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.05, 0.005, 1, 0.15 und 1. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.05, 0.005, 1, 0.15 und 1. Die adjustierten p-Werte werden zu einem α -Niveau von 1% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.002, 2e-04, 0.042, 0.006 und 0.178. Die adjustierten p-Werte werden zu einem α -Niveau von 1% verglichen.

30 Aufgabe

(2 Punkte)

Die Abkürzung CLD steht für welches statistische Verfahren? Welche folgende Beschreibung der Interpretation ist korrekt?

- A** ☐ Compact letter display. Gleiche Buchstaben bedeuten, dass sich die Behandlungen unterscheiden. Daher ist das CLD sehr unintuitiv. Es wäre besser, wenn gleiche Buchstaben Gleichheit anzeigen würden. Dies ist aber leider in der statistischen Testtheorie nicht möglich.
- B** ☐ Compact letter display. Gleiche Buchstaben zeigen Gleichheit in den Behandlungen. Die Interpretation ist deshalb sehr intuitiv und einfach. Darüber hinaus ist damit das CLD auch auf einer Linie mit der Testtheorie, da wir ja auch dort die Gültigkeit der Nullhypothese nachweisen. Wir suchen ja Gleichheit.
- C** ☐ Compact letter detection. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt.
- D** ☐ Compact letter display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Teilweise ist die Interpretation des CLD herausfordernd, da wir ja nach dem Unterschied suchen.
- E** ☐ Contrast letter display. Unterschiede in den Behandlungen werden durch den gleichen Buchstaben oder Symbol dargestellt. Die Interpretation des CLD führt häufig in die Irre.

31 Aufgabe

(2 Punkte)

In Ihrer Bachelorarbeit müssen Sie einen Feldversuch auswerten. Nachdem Sie die zweifaktorielle ANOVA gerechnet haben und keine signifikante Interaktion vorliegt, wollen Sie jetzt einen Posthoc-Test rechnen. Welches R Paket nutzen Sie dafür am besten?

- A** ☐ Das R Paket {emmeans} erlaubt die Durchführung eines multiplen Gruppenvergleichs. Aus einem emmeans Objekt lässt sich leider kein CLD erstellen. Dennoch ist das Paket einfach zu bedienen und wird deshalb genutzt. Die Interpretation der statistischen Auswertung wird über einen Barplot abgebildet.
- B** ☐ Das R Paket {ggplot}. Wir erhalten hier sofort eine Visualisierung der Daten. Anhand der Visualisierung lässt sich eine explorative Datenanalyse durchführen, die gleichwertig zu einem Posthoc-Test ist.
- C** ☐ Sie nutzen das R Paket {emmeans} für die Berechnung des multiplen Gruppenvergleichs. Die Ausgabe der Funktion emmeans() erlaubt zügig über {ggplot} einen Barplot zu erstellen und dann auch das CLD zu berechnen. Sie haben alles sofort zusammen.
- D** ☐ Das R Paket {lm}. Das Paket {lm} erstellt selbstständig Konfidenzintervalle und entsprechende p-Werte. Da wir in dem Paket nicht adjustieren müssen, ist es bei Anwendern sehr beliebt.
- E** ☐ Das R Paket {hmisc} erlaubt die Durchführung eines multiplen Gruppenvergleichs aus verschiedenen Modellen heraus. Aus einem hmisc Objekt lässt sich recht einfach das CLD erstellen und so über Barplots eine schnelle Interpretation der statistischen Auswertung durchführen.

32 Aufgabe

(2 Punkte)

Bei einem multiplen Vergleich oder Posthoc Test kann es zu einer Besonderheit beim statistischen Testen kommen. Wie nennt man diese Besonderheit beim statistischen Testen und wie kann man mit ihr umgehen?

- A** ☐ Beim multiplen Testen kann es zu einer β -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 20%. Daher müssen die p-Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der p-Werte nach Bonferroni das bekannteste Verfahren ist.
- B** ☐ Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern sehr viel höher. Es kommt zu einer α -Inflation. Dagegen kann mit der Adjustierung der p-Werte nach Bonferroni vorgegangen werden.
- C** ☐ Beim multiplen Testen kann es zu Varianzheterogenität kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5%. Daher müssen die p-Werte entsprechend adjustiert werden. Das Verfahren nach Welch, bekannt aus dem t-Test, ist hier häufig anzuwenden.
- D** ☐ Die Adjustierung der p-Werte nach Bonferroni erlaubt es gegen die β -Inflation vorzugehen, die häufig beim multiplen Testen auftritt. Das globale Powerniveau liegt nicht mehr bei 80% sondern sehr viel niedriger.
- E** ☐ Beim multiplen Testen kann es zu einer α -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern weit darunter. Daher müssen die p-Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der p-Werte nach Welch das bekannteste Verfahren ist.

33 Aufgabe

(2 Punkte)

In einem Feldversuch haben Sie einen Behandlungsfaktor mit mehreren Levels vorliegen. Sie rechnen einen multiplen Vergleich. Vorher hatten Sie eine einfaktorielle ANOVA mit einem signifikanten Ergebnis vorliegen. Welche Aussage ist richtig?

- A** ☐ Beim multiplen Testen kann es zu einer Δ -Deflation kommen. Das globale Relevanzniveau liegt nicht mehr bei 5% sondern weit darunter. Daher müssen die Δ -Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der Δ -Werte nach Bonferroni das bekannteste Verfahren ist. Die Δ -Werte werden durch die Anzahl an Vergleichen geteilt.
- B** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt Δ adjustiert werden im Gegensatz zu den p-Werten.
- C** ☐ Beim multiplen Testen muss der Effekt, wie der Mittelwertsunterschied Δ aus einem t-Test, nicht adjustiert werden.
- D** ☐ Beim multiplen Testen kann es zu einer Δ -Inflation kommen. Das globale Effektniveau liegt nicht mehr bei 20%. Daher müssen die Effekte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der Effekte nach Bonferroni das bekannteste Verfahren ist.
- E** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt Δ nicht adjustiert werden. Bei einem Effekt im multiplen Testen handelt es sich um eine Wahrscheinlichkeit für das Auftreten der Nullhypothese.

Statistische Testtheorie

34 Aufgabe

(2 Punkte)

Sie haben den mathematischen Ausdruck $Pr(D|H_0)$ vorliegen, welche Aussage ist richtig?

- A** ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit nicht die Daten D zu beobachten sondern die Nullhypothese, wenn diese wahr ist.
- B** ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit der Alternativehypothese und somit $1 - Pr(H_A)$
- C** ☐ Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativehypothese überdeckt.
- D** ☐ Die Wahrscheinlichkeit für die Nullhypothese, wenn die Daten wahr sind.
- E** ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit die Daten D zu beobachten, wenn die Nullhypothese wahr ist.

35 Aufgabe

(2 Punkte)

Die Testtheorie hat einen philosophischen Unterbau. Eins der Prinzipien ist das Falsifikationsprinzip. Das Falsifikationsprinzip besagt,

- A ☐ ... dass in der Wissenschaft immer etwas falsch sein muss. Sonst gebe es keinen Fortschritt.
- B ☐ ... dass Annahmen an statistische Modelle meist falsch sind.
- C ☐ ... dass ein schlechtes Modell durch ein weniger schlechtes Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.
- D ☐ ... dass Fehlerterme in statistischen Modellen nicht verifiziert werden können.
- E ☐ ... dass ein schlechtes Modell durch ein schlechteres Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.

36 Aufgabe

(2 Punkte)

In fast allen wissenschaftlichen Disziplinen liegt der Grenzwert für das Signifikanzniveau α bei 5%. Wieso wurde dieser Konsens über die Signifikanzschwelle in dieser Form getroffen?

- A ☐ Auf einer Statistikkonferenz in Genf im Jahre 1942 wurde dieser Cut-Off nach langen Diskussionen festgelegt. Bis heute ist der Cut Off aber umstritten, da wegen dem 2. Weltkrieg viele Wissenschaftler nicht teilnehmen konnten.
- B ☐ Im Rahmen eines langen Disputs zwischen Neyman und Fischer wurde $\alpha = 5\%$ festgelegt. Leider werden die Randbedingungen und Voraussetzungen an statistische Modelle heute immer wieder ignoriert.
- C ☐ Der Wert ergab sich aus einer Auswertung von 1042 wissenschaftlichen Veröffentlichungen zwischen 1914 und 1948. Der Wert 5% wurde in 28% der Veröffentlichungen genutzt. Daher legte man sich auf diese Zahl fest.
- D ☐ Da Wissenschaftler eine Schwelle für die statistische Testentscheidung benötigen wurde α historisch gewählt. Damit ist $\alpha = 5\%$ eine Kulturkonstante.
- E ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.

37 Aufgabe

(2 Punkte)

Betrachten wir die Teststatistik aus einem abstrakteren Blickwinkel. Beim statistischen Testen wird das „*signal*“ mit dem „*noise*“ aus den Daten D zu einer Teststatistik T_D verrechnet. Welche der Formel berechnet korrekt die Teststatistik T_D ?

- A ☐ Es gilt $T_D = \frac{signal}{noise}$
- B ☐ Es gilt $T_D = signal \cdot noise$
- C ☐ Es gilt $T_D = \frac{noise}{signal}$
- D ☐ Es gilt $T_D = (signal \cdot noise)^2$
- E ☐ Es gilt $T_D = \frac{signal}{noise^2}$

38 Aufgabe

(2 Punkte)

In der Theorie zur statistischen Testentscheidung kann folgende Aussage in welche richtige Analogie gesetzt werden?

H_0 beibehalten obwohl die H_0 falsch ist

- A ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.
- B ☐ Dem α -Fehler in der Analogie eines Rauchmelder: *Alarm without fire*.
- C ☐ *Alarm with fire*, dem α -Fehler in der Analogie von Feuer.
- D ☐ In die Analogie eines Rauchmelders: *Alarm without fire police*, dem α -Fehler.
- E ☐ Dem β -Fehler mit der Analogie eines Rauchmelders: *Fire without alarm*.

39 Aufgabe

(2 Punkte)

Welche statistische Maßzahl erlaubt es Relevanz mit Signifikanz zu verbinden? Welche Aussage ist richtig?

- A ☐ Einem Konfidenzintervall. Das Konfidenzintervall bringt durch eine Visualisierung und zwei Intervallgrenzen die Möglichkeit mit, eine Relevanzschwelle neben der definierten Signifikanzschwelle zu definieren.
- B ☐ Über das Konfidenzintervall. Das Konfidenzintervall inkludiert eine Entscheidung über die Relevanz und zusätzlich kann über die Visualisierung des Konfidenzintervalls eine Signifikanzschwelle vom Forschenden definiert werden.
- C ☐ Das Δ . Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da Δ antiproportional zum p-Wert ist, bedeutet auch ein hohes Δ ein sehr kleinen p-Wert.
- D ☐ Die Teststatistik. Durch den Vergleich von T_c zu T_k ist es möglich die H_0 abzulehnen. Die Relevanz ergibt sich aus der Fläche rechts vom dem T_c -Wert.
- E ☐ Einem Konfidenzintervall. Das Konfidenzintervall bringt durch eine Visualisierung und drei Intervallgrenzen die Möglichkeit mit, eine Relevanzschwelle neben der Signifikanzschwelle und der α -Schwelle zu definieren.

40 Aufgabe

(2 Punkte)

Ein statistischer Test produziert für einen Gruppenvergleich einen p-Wert. Welche Aussage zusammen mit dem Signifikanzniveau α gleich 5% stimmt?

- A ☐ Wir schauen, ob der p-Wert kleiner ist als das Signifikanzniveau α und vergleichen somit Wahrscheinlichkeiten. Die Wahrscheinlichkeiten werden als Flächen unter der Kurve der Teststatistik dargestellt, wenn die H_0 gilt.
- B ☐ Wir machen eine Aussage über die Flächen und zwischen den Kurve der Teststatistiken der Hypothesen H_0 und H_A , wenn die H_0 gilt. Dabei werden Wahrscheinlichkeiten verglichen, die durch die Flächen unter der Kurve repräsentiert werden.
- C ☐ Wir vergleichen mit dem p-Wert und dem Signifikanzniveau α absolute Werte auf einem Zahlenstrahl und damit den Unterschied der Teststatistiken, wenn die H_0 gilt.
- D ☐ Wir vergleichen die Effekte des p-Wertes mit den Effekten der Signifikanzschwelle unter der Annahme der Nullhypothese. Dabei gilt, dass wir die Nullhypothese nur ablehnen können anhand des Falsifikationsprinzips.
- E ☐ Wir vergleichen mit dem p-Wert und dem Signifikanzniveau α Wahrscheinlichkeiten und damit die absoluten Werte auf einem Zahlenstrahl, wenn die H_0 gilt.

41 Aufgabe

(2 Punkte)

Um die Ergebnisse eines statistischen Tests und die damit verbundene Theorie besser zu verstehen, kann eine Analogie zur Wettervorhersage genutzt werden. Welche Analogie zu der Testtheorie trifft am meisten zu?

- A** ☐ In der Analogie der Regenwahrscheinlichkeit in einem bestimmten Gebiet: ein statistischer Test gibt die Wahrscheinlichkeit für ein Ereignis in einem Experiment mit den Daten D wieder und lässt sich kaum verallgemeinern.
- B** ☐ In der Analogie der Durchschnittstemperatur: Wie oft tritt ein Effekt durchschnittlich ein? Wir erhalten eine Wahrscheinlichkeit für die Effekte. Zum Beispiel, wie hoch ist die Wahrscheinlichkeit für einen Mittelwert als Durchschnitt.
- C** ☐ In der Analogie der Sonnenscheindauer: Wie lange kann mit einem entsprechenden Effekt gerechnet werden? Die Wahrscheinlichkeit für den Effekt gibt der statistische Test wieder.
- D** ☐ In der Analogie der Maximaltemperatur: Was ist der maximale Unterschied zwischen zwei Gruppen. Wir erhalten hier eine Aussage über die Spannweite und den maximalen Effekt.
- E** ☐ In der Analogie der Regenwahrscheinlichkeit: ein statistischer Test gibt die Wahrscheinlichkeit für das Auftreten eines Ereignisses wieder. Die Stärke des Effektes wird nicht wiedergegeben.

42 Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit wollen Sie eine Aussage über die untersuchte Population treffen. Dazu nutzen Sie einen statistischen Test. Können Sie eine valide Aussage treffen?

- A** ☐ Nein, wir erhalten eine Aussage. Müssen aber das Individuum im Kontext der Population adjustieren.
- B** ☐ Ja, die untersuchte Population können wir mit einem statistischen Test auswerten. Wir erhalten dann eine Aussage zur Population.
- C** ☐ Nein, es ist nicht möglich die untersuchte Population mit einem t-Test auszuwerten. Wir erhalten dann leider keine Aussage zur Population.
- D** ☐ Nein, wir können die untersuchte Population nicht mit einem t-Test auswerten. Wir erhalten keine Aussage zur Population. Wir können aber den Effekt als Quelle der Relevanz nutzen.
- E** ☐ Weder eine Aussage über die Population noch über das Individuum ist mit einem statistischen Test möglich. Wir erhalten eine Aussage über ein Experiment.

43 Aufgabe

(2 Punkte)

Welche Aussage über die *Power* ist richtig?

- A** ☐ Die Power $1 - \beta$ wird auf 80% gesetzt. Alle statistischen Tests sind so konstruiert, dass die H_A mit 80% *bewiesen wird*.
- B** ☐ Die Power $1 - \beta$ wird auf 80% gesetzt. Damit liegt die Wahrscheinlichkeit für die H_0 bei 20%.
- C** ☐ Die Power beschreibt die Wahrscheinlichkeit die H_A abzulehnen. Wir testen die Power jedoch nicht.
- D** ☐ Die Power ist nicht in der aktuellen Testtheorie mehr vertreten. Wir rechnen nur noch mit dem Fehler 1. Art.
- E** ☐ Alle statistischen Tests sind so konstruiert, dass die H_A mit 20% *bewiesen wird*. Die Power ist $1 - \beta$ mit β gleich 80% gesetzt.

44 Aufgabe

(2 Punkte)

Sie rechnen einen statistischen Test und erhalten neben dem p-Wert noch einen Effekt wiedergegeben. Welche Aussage zum Effekt ist richtig?

- A** ☐ Durch den Effekt erfahren wir die biologisch interpretierbare Ausgabe eines statistischen Tests. Zum Beispiel das η^2 aus einer ANOVA. Damit können wir die Relevanz direkt mit dem Effekt verbinden. Am Ende muss der Forschende aber entscheiden, ob der Effekt entsprechend seinen Erwartungen als bedeutet zu bewerten ist.
- B** ☐ Der Forschende muss am Anfang wissen, ob das Ergebnis eines Experiments relevant für seine Forschung ist. Dafür kann der Effekt eines statistischen Tests genutzt werden oder auch der Prähoc-Test. Damit beschreibt der Effekt den biologischen interpretierbaren Teil eines Experiments vor der Durchführung. Zum Beispiel der Unterschied zwischen zwei Mittelwerten.
- C** ☐ Durch den Effekt erfahren wir die statistische interpretierbare Ausgabe eines statistischen Tests. Zum Beispiel das η^2 aus einer ANOVA. Damit können wir die Signifikanz direkt mit dem Effekt verbinden. Am Ende muss der Forschende aber entscheiden, ob der Effekt entsprechend seinen Erwartungen als bedeutet zu bewerten ist.
- D** ☐ Der Effekt eines statistischen Tests beschreibt die biologisch interpretierbare Ausgabe eines Tests. Damit ist der Effekt direkt mit dem Begriff der Signifikanz verbunden. Die Entscheidung über die Signifikanz trifft der Forschende unabhängig von der Relevanz eines statistischen Tests.
- E** ☐ Der Effekt eines statistischen Tests beschreibt die mathematisch interpretierbare Ausgabe eines Tests. Damit ist der Effekt direkt mit dem Begriff der Signifikanz verbunden. Die Entscheidung über die Signifikanz trifft der Forschende unabhängig von der Relevanz eines statistischen Tests.

45 Aufgabe

(2 Punkte)

Welche Aussage über die Entscheidung anhand des 95%-Konfidenzintervalls gegen die Nullhypothese ist richtig?

- A** ☐ Ist T_D höher als der kritische Wert $T_{\alpha=5\%}$ dann wird die Nullhypothese H_0 abgelehnt.
- B** ☐ Anhand des 95%-Konfidenzintervalls lässt sich wie folgt eine Entscheidung treffen. Liegt der Wert in dem Signifikanzniveauintervall α dann kann die Nullhypothese abgelehnt werden.
- C** ☐ Ist $Pr(D|H_0)$ kleiner als das Signifikanzniveau α gleich 5% dann wird die Nullhypothese H_0 abgelehnt.
- D** ☐ Ist in dem 95%-Konfidenzintervall nicht die Null enthalten dann wird die Nullhypothese H_0 abgelehnt.
- E** ☐ Anhand des 95%-Konfidenzintervalls lässt sich wie folgt eine Entscheidung treffen. Liegt der Wert über oder gleich dem Signifikanzniveau α dann kann die Nullhypothese abgelehnt werden.

46 Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit müssen Sie für die statistischen Tests im Anhang Ihrer Arbeit die Hypothesen H formulieren. Welche Aussage über Hypothesen H ist richtig?

- A** ☐ Mit der Nullhypothese H_A und der Alternativhypothese H_0 gibt es zwei Hypothesen, die aber selten genutzt werden.
- B** ☐ Es gibt - bedingt durch das das Falsifikationsprinzip - ein Set von k Nullhypothesen, die iterative gegen $k - 1$ Alternativhypothesen getestet werden.
- C** ☐ Ein statistisches Hypothesenpaare gibt es. Zum einen die Nullhypothese und zum anderen die Alternativhypothese. Es ist aber nur notwendig die Alternative anzugeben, da die Nullhypothese nicht beim Testen benötigt wird.
- D** ☐ Es gibt ein statistisches Hypothesenpaar mit der Hypothese für und gegen die wissenschaftliche Fragestellung. Die Hypothesen werden H_{pro} und H_{contra} bezeichnet.
- E** ☐ Ein statistisches Hypothesenpaare gibt es. Zum einen die Nullhypothese H_0 und zum anderen die Alternativhypothese H_A oder H_1

Statistische Tests für Gruppenvergleiche

47 Aufgabe

(2 Punkte)

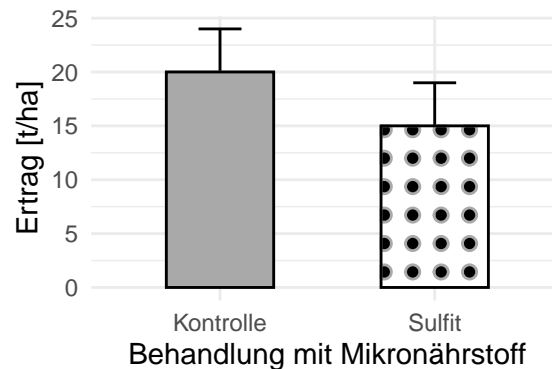
Welche Aussage über den t-Test im Allgemeinen ist richtig? Berücksichtigen Sie den Welch t-Test wie auch den Student t-Test!

- A** ☐ Der t-Test vergleicht die Mittelwerte von zwei Gruppen unter der strikten Annahme von Varianzhomogenität. Sollte keine Varianzhomogenität vorliegen, so gibt es keine Möglichkeit den t-Test in einer Variante anzuwenden.
- B** ☐ Der t-Test vergleicht zwei Gruppen indem die Mittelwerte miteinander verglichen werden.
- C** ☐ Der t-Test testet generell zu einem erhöhten α -Niveau von 20%.
- D** ☐ Der t-Test vergleicht zwei oder mehr Gruppen indem die Mittelwerte miteinander verglichen werden.
- E** ☐ Der t-Test vergleicht die Varianzen von mindestens zwei oder mehr Gruppen

48 Aufgabe

(2 Punkte)

Ein Versuch wurde in 12 Parzellen pro Gruppe durchgeführt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung des Mikronährstoff Sulfit auf den Ertrag in t/ha von Weizen im Vergleich zu einer Kontrolle. Welche Aussage ist richtig, wenn Sie einen t-Test rechnen?



- A** ☐ Nach Betrachtung des Barplots liegt ein signifikanter Unterschied vor. Der Effekt liegt bei -5.
- B** ☐ Der Effekt und die Signifikanz lassen sich nicht aus Barplots abschätzen. Höchstens der Effekt als relativer Unterschied zwischen der Höhe der Barplots. Standard ist der mediane Unterschied aus Boxplots.
- C** ☐ Die Barplots deuten auf keinen signifikanten Unterschied. Der Effekt liegt vermutlich bei -5. Wir müssen aber einen Posthoc-Test rechnen um den Effekt wirklich bestimmen zu können.
- D** ☐ Die Barplots deuten auf kein signifikanten Unterschied. Der Effekt liegt vermutlich bei -5.
- E** ☐ Die Barplots deuten auf einen signifikanten Unterschied. Der Effekt liegt vermutlich bei -5 unter einer groben Abschätzung. Wir müssen aber eine ANOVA rechnen um den Effekt wirklich bestimmen zu können.

49 Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit betrachten Sie die Effekte von einer Behandlung vor und nach der Gabe eines Vitamins. Sie müssen einen gepaarten t-Test rechnen. Welche Aussage ist richtig?

- A** ☐ Beim gepaarten t-Test kombinieren wir die Vorteile des Student t-Test für Varianzhomogenität mit den Vorteilen des Welch t-Test für Varianzheterogenität. Wir bilden dafür die Differenz der Einzelbeobachtungen.
- B** ☐ Der gepaarte t-Test nutzt die Varianz der Beobachtungen jeweils paarweise und bildet dafür eine verbundene Stichprobe. Dieser Datensatz d dient dann zur Differenzbildung.

- C** ☐ Wenn die Beobachtungen unabhängig voneinander sind, rechnen wir einen gepaarten t-Test. Messen wir wiederholt an dem gleichen Tier oder Pflanze dann bilden wir das Produkt zwischen den zwei Messpunkten.
- D** ☐ Der gepaarte t-Test wird genutzt, wenn die Differenzen der Beobachtungen verbunden sind und wir dadurch die Unabhängigkeit nicht mehr vorliegen haben.
- E** ☐ Abhängige Beobachtungen müssen gesondert in einem gepaarten t-Test modelliert werden. Wenn wiederholt an dem gleichen Tier oder Pflanze gemessen wird, dann bilden wir die Differenz zwischen den beiden Zeitpunkten. Auf den Differenzen rechnen wir den gepaarten t-Test.

50 Aufgabe

(2 Punkte)

Nach einem Experiment mit drei Weizensorten ergibt eine ANOVA ($p = 0.048$) einen signifikanten Unterschied für den Ertrag. Sie führen anschließend die paarweisen t-Tests für alle Vergleiche der verschiedenen Weizensorten durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der α -Schwelle. Sie schauen sich auch die rohen, unadjustierten p-Werte an und finden hier als niedrigsten p-Wert $p_{3-2} = 0.053$. Welche Aussage ist richtig?

- A** ☐ Die adjustierten p-Werte deuten in die richtige Richtung. Zusammen mit den nicht signifikanten rohen p-Werten ist von einem Fehler in der ANOVA auszugehen.
- B** ☐ Die ANOVA testet auf der gesamten Fallzahl. Es wäre besser die ANOVA auf der gleichen Fallzahl wie die einzelnen t-Tests zu rechnen.
- C** ☐ Es gibt einen Fehler in der Varianzstruktur. Daher kann die ANOVA nicht richtig sein und paarweise t-Tests liefern das richtige Ergebnis.
- D** ☐ Hier kommt der Effekt der steigenden Fallzahl auf die Anzahl an signifikante Ergebnisse zu tragen. Da die ANOVA auf mehr Fallzahl testet als die einzelnen paarweisen t-Tests, kann die ANOVA leichter einen signifikanten Unterscheid nachweisen. Die p-Werte sind immer etwas kleiner als bei den t-Tests.
- E** ☐ Der Fehler liegt in den t-Tests. Wenn eine ANOVA signifikant ist, dann muss zwangsweise auch ein t-Test signifikant sein.

Deskriptive Statistik & Explorative Datenanalyse

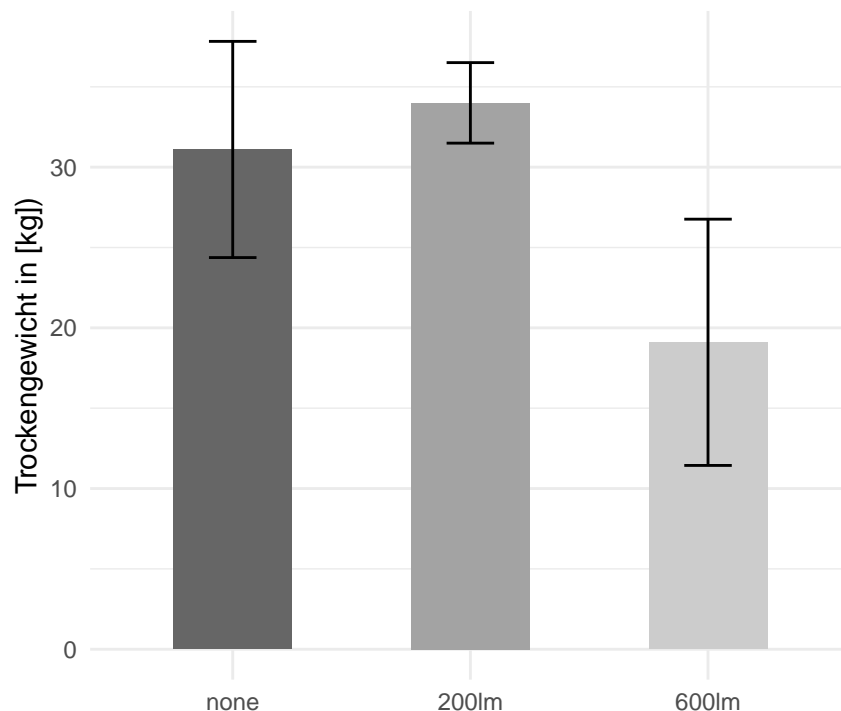
51 Aufgabe

(7 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Barplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Mark nicht gemerkt, welche statistischen Maßzahlen für einen Barplot erhoben werden müssen. Das ist in soweit doof, da nach seiner Betreuer erstmal ein Barplot nachgebaut werden soll, bevor es mit seinem Projektbericht losgeht. Dann hat er schonmal den Code vorliegen und nachher geht dann alles schneller. Die Behandlung für Maiss werden verschiedene Lichtstufen (*none*, *200lm* und *600lm*) sein. Erfasst wird als Endpunkt (Y) *Trockengewicht*. Mark soll dann *drymatter* in seiner Exceldatei eintragen.



Leider kennt sich Mark mit der Erstellung von Barplots in nicht aus. Deshalb braucht er bei der Visualisierung Ihre Hilfe!

1. Erstellen Sie eine Tabelle mit den statistischen Maßzahlen aus der obigen Abbildung der drei Barplots! Beachten Sie die korrekte Darstellungsform der statistischen Maßzahlen! **(3 Punkte)**
2. Erstellen Sie einen beispielhaften Datensatz, aus dem die drei Barplots *möglicherweise* erstellt wurden, im üblichen Format! **(2 Punkte)**
3. Erwarten Sie einen Unterschied zwischen den Behandlungen? Begründen Sie Ihre Antwort! **(2 Punkte)**

52 Aufgabe

(7 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Paula steht vor einem ersten Problem, denn wenn es nach ihrem Betreuer geht, soll sie in einem einem Gewächshausexperiment Brokoli ausgewertet. Soweit eigentlich alles passend. Die Behandlung waren verschiedene Lichtstufen (*none*, *200lm* und *600lm*). In ihrer Exceldatei hat sie den Endpunkt (Y) *Ertrag* als *yield* aufgenommen. Nun soll Paula die Daten einmal als Barplots in einer Präsentation visualisieren, damit ihrem Betreuer wieder klar wird, was sie eigentlich nochmal gemacht hat und was für ein Ergebnis in einem statistischen Test zu erwarten wäre.

treatment	yield
none	23.4
none	34.9
200lm	28.0
200lm	38.8
200lm	34.1
600lm	29.7
none	26.5
none	41.0
200lm	10.1
600lm	14.4
200lm	30.9
600lm	20.9
600lm	29.7
600lm	29.1
none	27.3

Leider kennt sich Paula mit der Erstellung von Barplots nicht aus. Deshalb braucht sie bei der Visualisierung Ihre Hilfe!

1. Zeichnen Sie in *einer* Abbildung die Barplots für die Behandlung von Brokoli! Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Beschriften Sie *einen* Barplot mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen den Behandlungen von Brokoli erwarten würden, wie sehen dann die Barplots aus? Antworten Sie mit einer Skizze der Barplots! **(1 Punkt)**

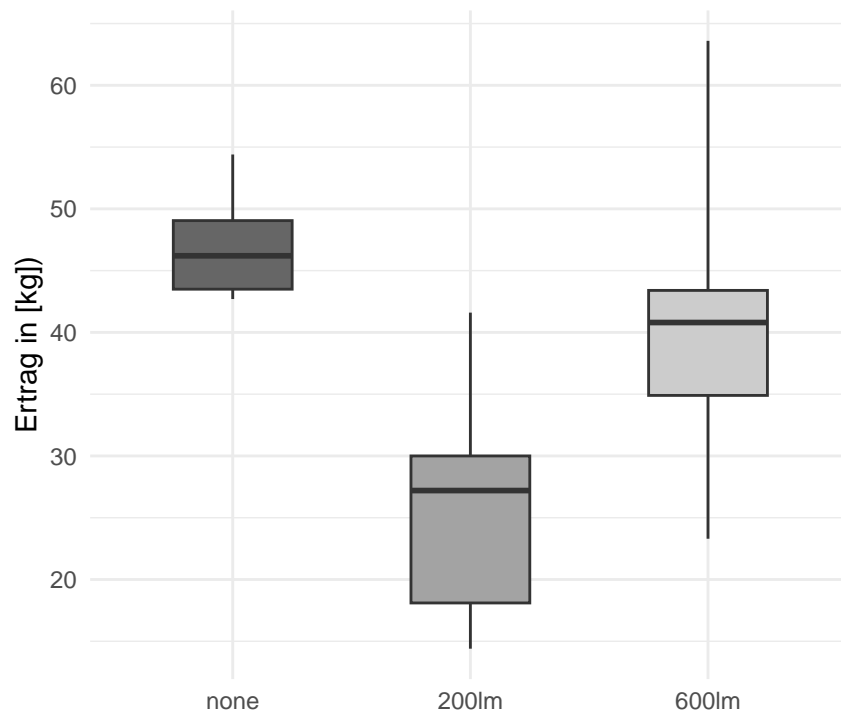
53 Aufgabe

(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Boxplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Jessica nicht gemerkt, welche statistischen Maßzahlen für einen Boxplot erhoben werden müssen. Das ist in soweit doof, da nach ihrem Betreuer erstmal ein Boxplot nachgebaut werden soll, bevor es mit ihrer Abschlussarbeit losgeht. Anhand von Boxplots lässt sich eine Aussage über die Normalverteilung von Y treffen. Dann hat sie schonmal den Code vorliegen und nachher geht dann alles schneller. Die Behandlung für Erdbeeren werden verschiedene Lichtstufen (*none*, *200lm* und *600lm*) sein. Erfasst wird als Outcome (Y) *Ertrag*. Jessica soll dann *yield* in ihrer Exceldatei eintragen.



Leider kennt sich Jessica mit der Erstellung von Boxplots in nicht aus. Deshalb braucht sie bei der Visualisierung Ihre Hilfe!

1. Erstellen Sie eine Tabelle mit den statistischen Maßzahlen aus der obigen Abbildung der drei Boxplots! Beachten Sie die korrekte Darstellungsform der statistischen Maßzahlen! **(3 Punkte)**
2. Beschriften Sie *einen* der Boxplots mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
3. Erstellen Sie einen beispielhaften Datensatz, aus dem die drei Boxplots *möglicherweise* erstellt wurden, im üblichen Format! **(2 Punkte)**
4. Erwarten Sie einen Unterschied zwischen den Behandlungen? Begründen Sie Ihre Antwort! **(2 Punkte)**

54 Aufgabe

(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Boxplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Mark nicht gemerkt, welche statistischen Maßzahlen für einen Boxplot erhoben werden müssen. Das ist in soweit doof, da nach seinem Betreuer nun Boxplots aus seinen Daten gebaut werden sollen, bevor es mit dem statistischen Testen weitergeht. Anhand von Boxplots lässt sich eine Aussage über die Varianzhomogenität über die Behandlungsgruppen treffen. Die Behandlung für Erbsen waren verschiedene Substrattypen (*torf* und *70p30n*). Erfasst wurde von Mark als Outcome (*Y*) *Ertrag*. Mark hat dann *yield* in seiner Exceldatei eintragen.

treatment	drymatter
torf	33.4
70p30n	43.2
torf	34.9
70p30n	37.6
70p30n	40.4
70p30n	47.1
torf	42.8
70p30n	43.9
70p30n	44.4
70p30n	39.7
torf	38.2
torf	34.6
70p30n	46.0
70p30n	45.7
torf	30.1
torf	38.8

Leider kennt sich Mark mit der Erstellung von Boxplots nicht aus. Deshalb braucht er bei der Visualisierung Ihre Hilfe!

1. Zeichnen Sie in *einer* Abbildung die beiden Boxplots für die zwei Behandlungen von Erbsen! Beschriften Sie die Achsen entsprechend! **(5 Punkte)**
2. Wie ist Ihr Vorgehen, wenn Sie eine *gerade* Anzahl an Beobachtungen pro Gruppe haben? **(1 Punkt)**
3. Beschriften Sie *einen* der beiden Boxplots mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
4. Wenn Sie *keinen Effekt* zwischen den Behandlungen von Erbsen erwarten würden, wie sehen dann die beiden Boxplots aus? Antworten Sie mit einer Skizze der Boxplots! **(1 Punkt)**

55 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Mark betrachtet die folgenden Daten nach einem Stallexperiment mit Fleischrinder. In dem Experiment wurden die Anzahl an gedrehten Haaren pro cm^2 gezählt. Nach der Meinung seiner Betreuerin muss als erstes geschaut werden, wie diese verteilt sind. Also welcher statistischen Verteilung die Anzahl an gedrehten Haaren pro cm^2 folgen. Dazu soll Mark ein Histogramm verwenden. Dann hätte man auch einen guten Überblick über den Endpunkt (Y).

Die Anzahl an gedrehten Haaren pro cm^2 : 1, 1, 4, 5, 2, 8, 3, 1, 3, 3, 8, 4, 3, 5, 6, 2, 0, 3, 3, 6, 5, 4, 6, 6, 5, 6, 6, 3, 3, 4, 4, 4, 4

Leider kennt sich Mark mit der Erstellung von Histogrammen überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Zeichnen Sie ein Histogramm um die Verteilung der Daten zu visualisieren! (3 Punkte)
2. Beschriften Sie die Achsen der Abbildung! (2 Punkte)
3. Ergänzen Sie die absoluten und relativen Häufigkeiten in der Abbildung! (1 Punkt)
4. Berechnen Sie aus den Daten die *Wahrscheinlichkeit* mehr als die Anzahl 5 zu beobachten! (1 Punkt)
5. Berechnen Sie aus den Daten die *Chance* mehr als die Anzahl 5 zu beobachten! (1 Punkt)

56 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In seiner Hausarbeit möchte Manuel gerne die Daten aus einem Kreuzungsexperiment mit Hühner in einem Histogramm darstellen. Das Histogramm erlaubt ihm dabei Rückschlüsse auf die Verteilung über den Endpunkt (Y) zu treffen. In seinem Experiment hat Manuel die mittleren dunklen Pigmentstörungen gezählt.

Die mittleren dunklen Pigmentstörungen: 9.3, 11.9, 9.4, 10.1, 12.3, 10.4, 9.4, 7.5, 11.4, 10.2, 8.6, 9.5, 11.5, 9, 6.8, 9.1, 9.3, 9.2, 13.5, 7.4, 8.5, 10.7, 7.8, 9.2, 8, 12.7, 10.2, 7.4

Leider kennt sich Manuel mit der Erstellung von Histogrammen überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Zeichnen Sie ein Histogramm um die Verteilung der Daten zu visualisieren! **(3 Punkte)**
2. Erläutern Sie Ihr Vorgehen um ein Histogramm für kontinuierliche Daten zu zeichnen! **(2 Punkte)**
3. Beschriften Sie die Achsen der Abbildung! **(2 Punkte)**
4. Ergänzen Sie die relativen Häufigkeiten in der Abbildung! **(1 Punkt)**

57 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Sabine möchte gerne den Zusammenhang zwischen durchschnittlicher Tagestemperatur [C/d] und Fettgehalt [%/kg] im Kontext von Hühner herausfinden. Hierfür hat Sabine ein Kreuzungsexperiment im Emsland durchgeführt. Nach einigen unvorgesehenen Ereignissen hat sie es geschafft folgende Datentabelle zu erstellen. Nun stellt sich die Frage für sie, ob es überhaupt einen Zusammenhang zwischen den gemessenen Variablen gibt. Deshalb möchte Sabine als erstes eine explorative Datenanalyse durchführen.

Fettgehalt [%/kg]	Durchschnittlicher Tagestemperatur [C/d]
19.2	23.1
16.8	18.7
22.1	26.9
20.5	21.2
17.0	25.3
20.2	24.6
21.5	25.2
19.5	24.2
19.0	27.3
18.8	22.3
20.0	24.6
19.7	22.0

Leider kennt sich Sabine mit der Erstellung einer explorativen Datenanalyse für kontinuierliche Daten überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

1. Erstellen Sie eine Visualisierung für die Datentabelle. Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Schätzen Sie eine Gerade durch die Punkte! **(1 Punkt)**
3. Beschriften Sie die Gerade mit den gängigen statistischen Maßzahlen! Geben Sie die numerischen Zahlenwerte mit an! **(3 Punkte)**
4. Wenn *ein* Effekt von x auf y vorhanden wäre, wie würde die Gerade verlaufen und welche Werte würden die statistischen Maßzahlen annehmen? **(2 Punkt)**

58 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Zwei kategoriale Variablen darzustellen ist nicht so einfach. Mark hatte erst über einen Mittelwert nachgedacht, dann aber die Idee verworfen. In einen Leistungssteigerungsversuch im Oldenburger Land hatte er sich zum einen die Behandlung Klimakontrolle [ja/nein] und zum anderen die Messung Fettgehalt erreicht [ja/nein] im Kontext von Fleischrinder angeschaut. Jetzt möchte sein Betreuer erstmal die langen Tabellen mit ja/nein in einer explorativen Datenanalyse zusammengefasst bekommen. Sonst geht es bei seinem Projektbericht nicht weiter.

Klimakontrolle	Fettgehalt reicht	er-	Klimakontrolle	Fettgehalt reicht	er-
nein	ja		ja	nein	
ja	ja		nein	nein	
ja	ja		nein	ja	
ja	nein		ja	ja	
nein	nein		nein	nein	
ja	nein		nein	nein	
ja	nein		nein	nein	
nein	ja		ja	ja	
nein	ja		ja	nein	
ja	nein		nein	nein	
ja	ja		nein	ja	
nein	nein		nein	ja	
ja	nein		ja	ja	
ja	nein		ja	ja	
ja	ja		nein	nein	
ja	nein		ja	ja	
nein	nein		nein	nein	

Leider kennt sich Mark mit der Erstellung einer explorativen Datenanalyse für kategoriale Daten überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Stellen Sie den Zusammenhang zwischen den beiden kategorialen Variablen in einer zusammenfassenden Tabelle dar! **(3 Punkte)**
2. Visualisieren Sie den Zusammenhang zwischen den beiden kategorialen Variablen! **(3 Punkte)**
3. Berechnen Sie die Verhältnisse in der Visualisierung! Welche Annahme haben Sie getroffen? **(2 Punkte)**
4. Wenn *ein* Effekt von der Behandlung vorliegen würde, wie würde die Tabelle und die Visualisierung aussehen? **(2 Punkt)**

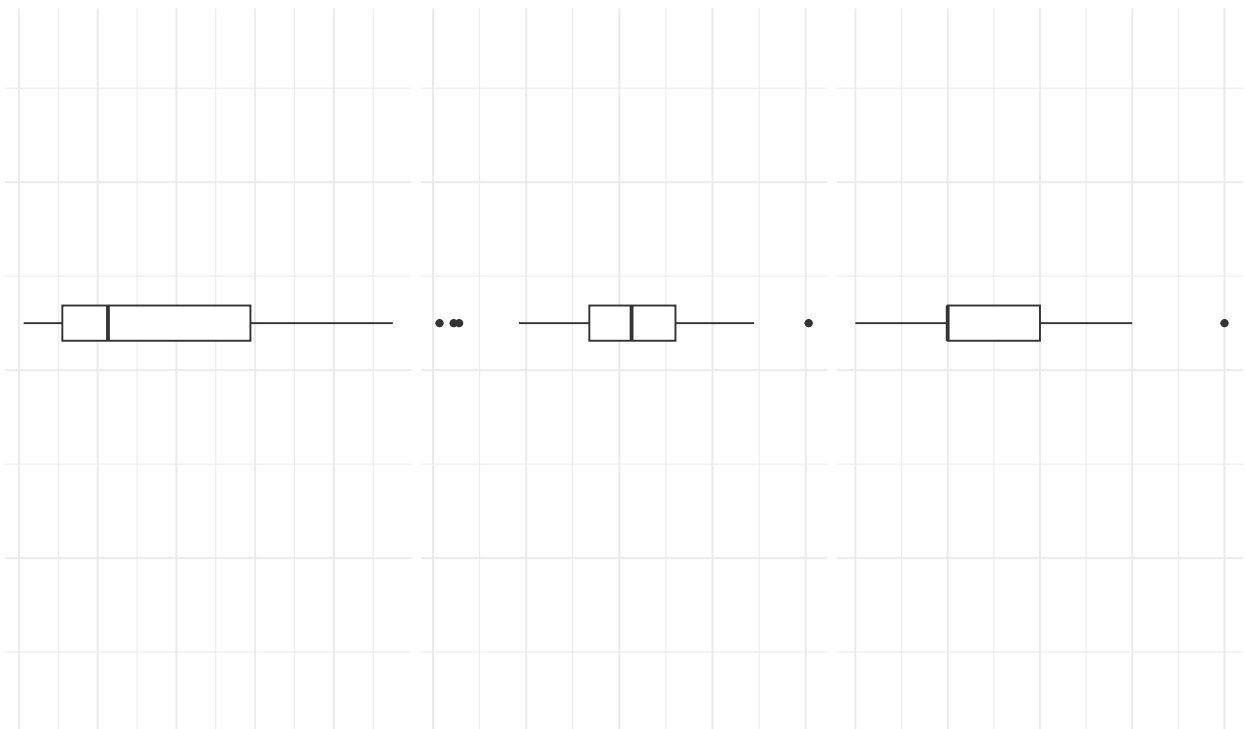
59 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



1. Zeichnen Sie über die untenstehenden Boxplots die entsprechende zugehörige Verteilung! **(3 Punkte)**
2. Zeichnen Sie unter die untenstehenden Boxplots die entsprechende zugehörige Beobachtungen als Stiche! **(3 Punkte)**
3. Wieviel Prozent der Beobachtungen fallen in das IQR? Ergänzen Sie die Abbildung entsprechend um den Bereich! **(2 Punkte)**
4. Wieviel Prozent der Beobachtungen fallen in $\pm 1s$ unter der Annahme einer Normalverteilung? Wenn möglich, ergänzen Sie die Abbildung entsprechend um den Bereich! **(2 Punkte)**



60 Aufgabe

(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



1. Skizzieren Sie 3 Normalverteilungen *in einer Abbildung* mit $\bar{y}_1 \neq \bar{y}_2 \neq \bar{y}_3$ und $s_1 = s_2 = s_3$! **(3 Punkte)**
2. Beschriften Sie die Normalverteilungen mit den entsprechenden Parametern! **(2 Punkte)**
3. Ergänzen Sie die Bereiche in der 68% und 95% der Beobachtungen fallen! Beschriften Sie die Grenzen der Bereiche mit der statistischen Maßzahl! **(2 Punkte)**
4. Liegt Varianzhomogenität oder Varianzheterogenität vor? Begründen Sie Ihre Antwort! **(2 Punkte)**

61 Aufgabe

(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



1. Skizzieren Sie in die unten stehenden, freien Abbildungen die Verteilungen, die sich nach der Abbildungsüberschrift ergeben! **(6 Punkte)**
2. Beschriften Sie die Achsen der Abbildungen entsprechend! **(1 Punkt)**
3. Achten Sie auf die entsprechende Skalierung der beiden Verteilungen in den Abbildungen! **(2 Punkte)**

$N(3, 4)$ und $N(1, 4)$



$Pois(15)$ und $Pois(3)$



62 Aufgabe

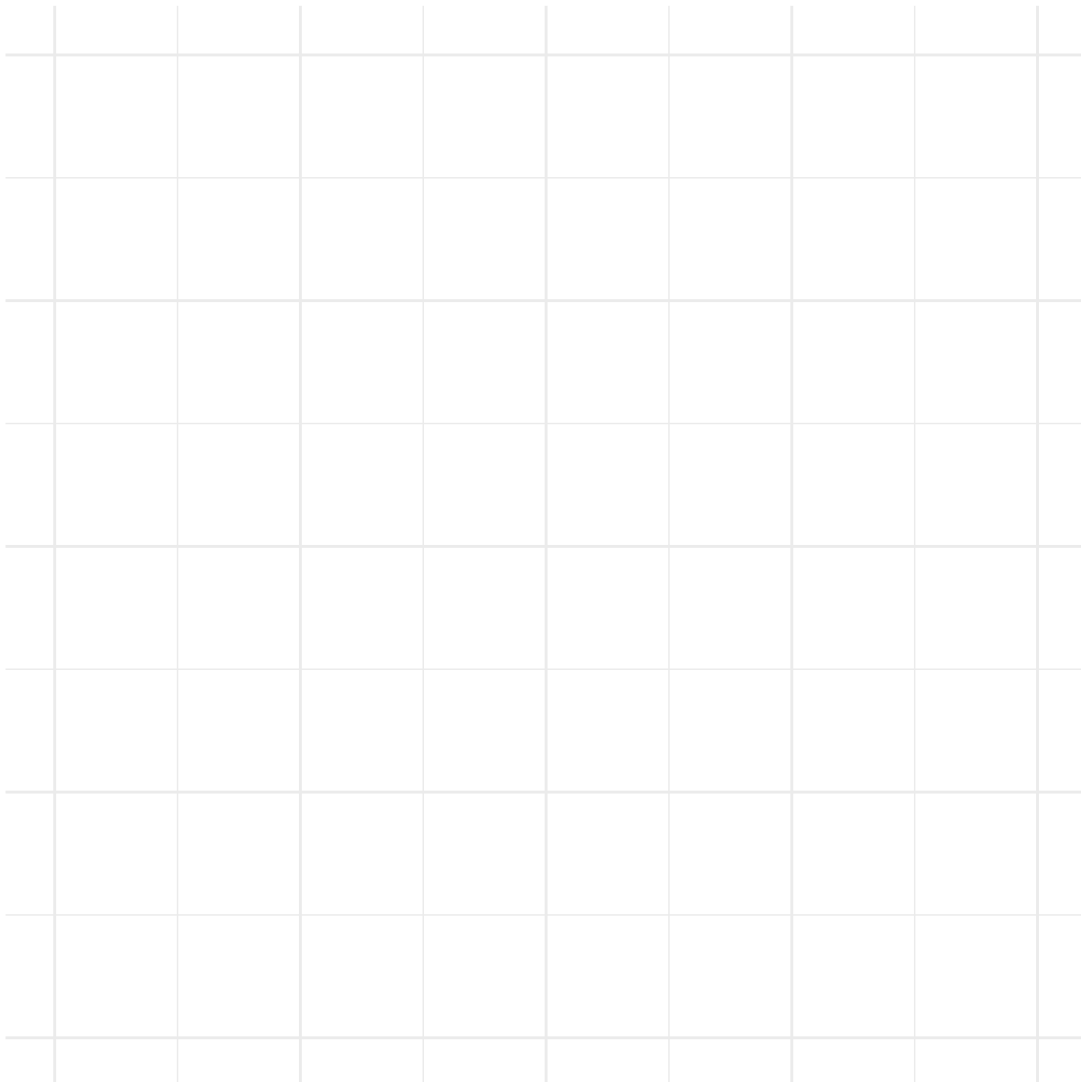
(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Sie haben folgende Zahlenreihe y vorliegen $y = \{17, 20, 22, 18, 23, 17, 17\}$.

1. Visualisieren Sie den Mittelwert von y in der untenstehenden Abbildung! **(4 Punkte)**
2. Beschriften Sie die Y und X -Achse entsprechend! **(2 Punkte)**
3. Für die Berechnung der Varianz wird der Abstand der einzelnen Werte y_i zum Mittelwert \bar{y} quadriert. Warum muss der Abstand, $y_i - \bar{y}$, in der Varianzformel quadriert werden? Erklären Sie den Zusammenhang unter Berücksichtigung der Abbildung! **(2 Punkte)**



Statistisches Testen

63 Aufgabe

(9 Punkte)



Grundlage des statistischen Testen ist das Verständnis von der Grundgesamtheit (eng. *population* oder *ground truth*) und der experimentellen Stichprobe (eng. *sample*).

1. Nennen Sie das statistische Verfahren und zwei konkrete Beispiele zur Durchführung um von einer Grundgesamtheit auf eine Stichprobe zu gelangen! **(3 Punkte)**
2. Erklären Sie den Zusammenhang zwischen Stichprobe und Grundgesamtheit an einem Schaubild! Beschriften Sie das Schaubild entsprechend! *Nutzen Sie hierfür als Veranschaulichung die Körpergröße von Männern oder Frauen aus den Gummibärchendaten!* **(3 Punkte)**
3. Erweitern Sie das Schaubild um die Entstehung von $Pr(D|H_0)$! *Nutzen Sie hierfür als Veranschaulichung zusätzlich die Gruppierungsvariable „Modul“ aus den Gummibärchendaten!* **(3 Punkte)**

64 Aufgabe

(9 Punkte)



Für ein besseres Verständnis der statistischen Testtheorie, auch Null-Ritual genannt, kann eine Visualisierung als Kreuztabelle genutzt werden.

1. Tragen Sie folgende statistische Fachbegriffe zur statistischen Testtheorie korrekt eine selbst erstellte Kreuztabelle ein! **(3 Punkte)**

(Unbekannte) Wahrheit H_0 abgelehnt β -Fehler H_0 wahr

2. Ergänzen Sie Ihre erstellte Kreuztabelle um vier weitere, passende Fachbegriffe zur statistischen Testtheorie! **(2 Punkte)**

Die Entscheidungsfindung durch einen statistischen Test kann auch durch die Analogie zu einem Feuermelder abgebildet werden. Dabei symbolisiert der Feuermelder den statistischen Test und es soll getestet werden, ob ein Feuer ausgebrochen ist.

3. In der Analogie des Feuermelders, wie lautet der α -Fehler? **(1 Punkt)**
4. In der Analogie des Feuermelders, wie lautet der β -Fehler? **(1 Punkt)**
5. Wenn der Feuermelder einmal pro Tag messen würde, wie oft würde der Feuermelder mit einem α von 5% in einem Monat Alarm schlagen? Begründen Sie Ihre Antwort! **(2 Punkte)**

65 Aufgabe

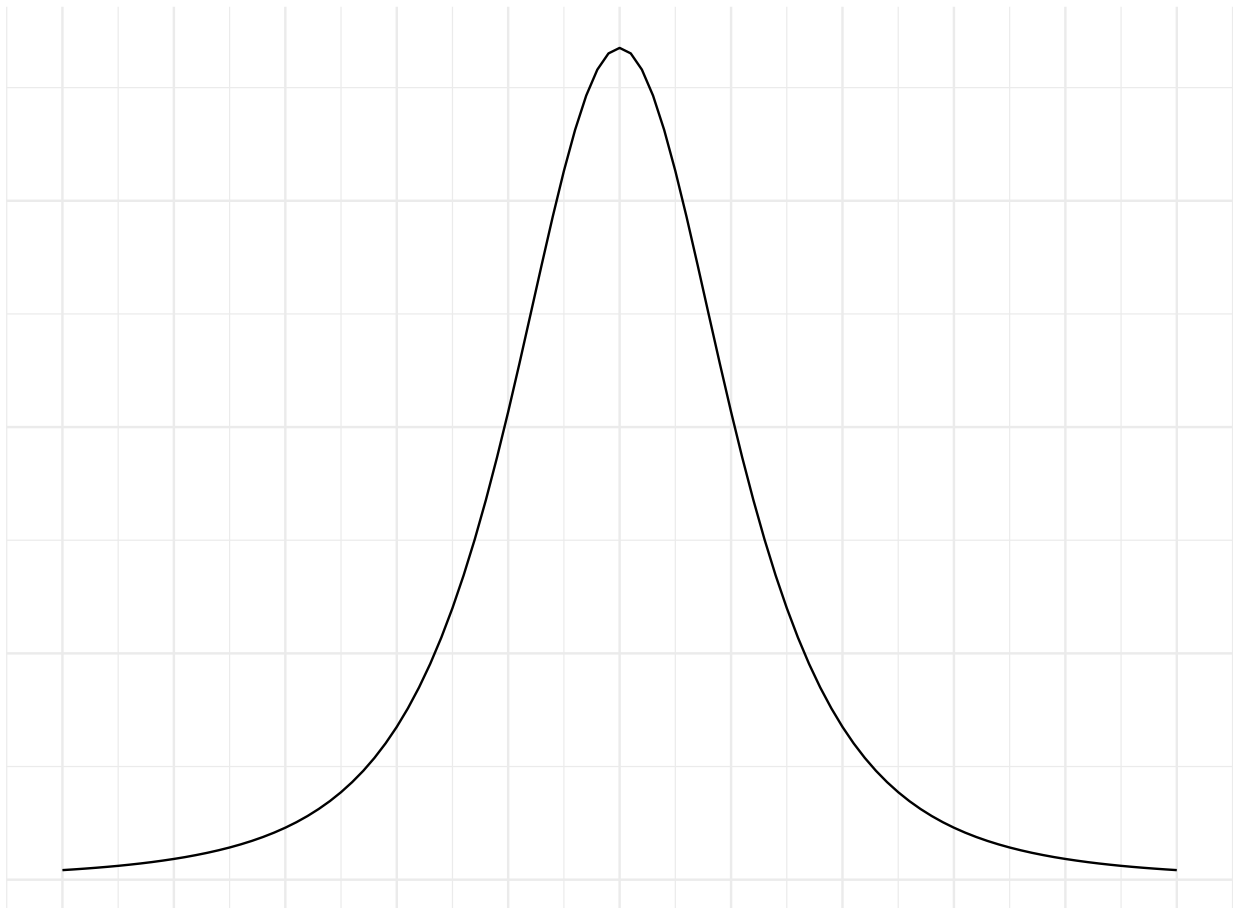
(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Abgebildet ist die t-Verteilung unter der Annahme der Gültigkeit der Nullhypothese. Beachten Sie, dass im Folgenden keine numerisch korrekte Darstellung verlangt wird! Es gilt Erkennbarkeit vor Genauigkeit!

1. Ergänzen Sie eine beschriftete x-Achse! **(1 Punkt)**
2. Ergänzen Sie „ $\bar{y}_1 = \bar{y}_2$ “! **(1 Punkt)**
3. Ergänzen Sie „ $A = 95\%$ “! **(1 Punkt)**
4. Zeichnen Sie $T_{\alpha=5\%}$ in die Abbildung! **(1 Punkt)**
5. Zeichnen Sie das Signifikanzniveau α in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Zeichnen Sie $+T_D$ in die Abbildung! **(1 Punkt)**
7. Zeichnen Sie einen signifikant p-Wert in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**





Sie rechnen einen t-Test für Gruppenvergleiche der Mittelwerte. Sie schätzen den Unterschied zwischen dem mittleren Trockengewicht nach Düngergabe zu einer unbehandelten Kontrolle.

1. Beschriften Sie die untenstehende Abbildung mit der Signifikanzschwelle! Begründen Sie Ihre Antwort! **(2 Punkte)**
2. Ergänzen Sie eine *in den Kontext passende* Relevanzschwelle! Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie in die untenstehende Abbildung sechs einzelne Konfidenzintervalle (a-f) mit den jeweiligen Eigenschaften! **(6 Punkte)**
 - (a) Ein signifikantes, relevantes 99% Konfidenzintervall.
 - (b) Ein signifikantes, nicht relevantes 95% Konfidenzintervall
 - (c) Ein signifikantes, relevantes 95% Konfidenzintervall
 - (d) Ein 95% Konfidenzintervall mit niedriger Varianz s_p in der Stichprobe als der Rest 95% der Konfidenzintervalle
 - (e) Ein nicht signifikantes, nicht relevantes 95% Konfidenzintervall
 - (f) Ein 95% Konfidenzintervall mit höherer Varianz s_p in der Stichprobe als der Rest der 95% Konfidenzintervalle



67 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Beim statistischen Testen gibt es einen Zusammenhang zwischen dem Effekt, der Streuung sowie der Fallzahl. Gegeben sei die Formel für den Student t-Test auf den die folgenden Überlegungen basieren sollen. Welche Auswirkung hat die Änderungen der jeweiligen statistischen Maßzahl des Effekts Δ , der Streuung s und der Fallzahl n auf die Teststatistik T_D , den p-Wert $Pr(D|H_0)$ sowie dem Konfidenzintervall $KI_{1-\alpha}$?

1. Visualisieren Sie den Zusammenhang zwischen der Teststatistik T_D und dem p-Wert $Pr(D|H_0)$ für sich verändernde T_D -Werte! Geben Sie dafür ein numerisches Beispiel in dem Sie drei T_D -Werte und deren Einfluss auf den p-Wert vergleichen! (3 Punkte)
2. Füllen Sie die untenstehende Tabelle aus in dem Sie die Änderung der statistischen Maßzahlen auf die Teststatistik, den p-Wert sowie das Konfidenzintervall in einem Wort oder Symbol beschreiben! (4 Punkte)

	T_D	$Pr(D H_0)$	$KI_{1-\alpha}$		T_D	$Pr(D H_0)$	$KI_{1-\alpha}$
$\Delta \uparrow$				$\Delta \downarrow$			
$s \uparrow$				$s \downarrow$			
$n \uparrow$				$n \downarrow$			

3. Visualisieren Sie ein 95%-iges Konfidenzintervall im Vergleich zu einem 90%-igen Konfidenzintervall! Begründen Sie Ihre Visualisierung anhand der Formel des Konfidenzintervalls des t-Tests mathematisch! (3 Punkte)

Der t-Test

68 Aufgabe

(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Der t-Test. Jonas erschauert. Ein mächtiges Werkzeug in den Händen desjenigen, der einen normalverteilten Endpunkt (Y) hat. Aber erstmal überhaupt den t-Test rechnen können. Wie sah das Experiment von Jonas überhaupt aus? Jonas hat ein Feldexperiment mit Erbsen durchgeführt um eine neue technische Versuchsanlage zu testen. Bei dem Pilotexperiment mit sehr geringer Fallzahl ($n_1 = n_2 = 3$) wurde die Behandlung Düngestufen (*ctrl* und *high*) an den Erbsen getestet und dabei wurde geschaut, ob der Versuch überhaupt technisch klappen könnte. Gemessen hat Jonas dann als Messwert Trockengewicht [kg/ha]. Warum der Versuch im Teutoburgerwald für seiner Hausarbeit stattfinden musste, ist ihm bis heute ein Rätsel. Egal. Gibt es jetzt einen Zusammenhang zwischen der Behandlung und Trockengewicht [kg/ha]?

treatment	weight
dose	19.2
dose	36.4
ctrl	20.5
ctrl	22.0
ctrl	21.1
dose	20.5

Leider kennt sich Jonas mit der Berechnung eines t-Tests überhaupt nicht aus. Deshalb braucht er bei der Berechnung Ihre Hilfe!

1. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
2. Bestimmen Sie die Teststatistik T_D eines Welch t-Tests! **(3 Punkte)**
3. Treffen Sie mit $T_{\alpha=5\%} = 1.64$ eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Berechnen Sie den Effekt des Welch t-Tests! **(1 Punkt)**
5. Formulieren Sie eine Antwort an Jonas über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

69 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Eric ist in der Uckermark für einen Versuch mit Brokoli. Allein diese Tatsache ist für ihn eine Erzählung wert. Für seinen Projektbericht musste er ein Freilandversuch mit Brokoli durchführen und das sollte laut seinem Betreuer an diesem Nichtort besonders gut gelingen. Ablenkung gibt es jedenfalls keine. Gar keine. Alleine sein hilft jetzt aber nur bedingt, denn seine Behandlung Lüftungssystemen und Folientunneln (*ctrl* und *tornado*) und der Messwert Trockengewicht [kg/ha] sollen mit einem t-Test ausgewertet werden. Immerhin weiß er, dass sein Messwert einer Normalverteilung folgt.

Lüftungssystemen	Trockengewicht
tornado	20.0
ctrl	28.1
tornado	29.5
ctrl	31.6
tornado	28.4
ctrl	24.3
ctrl	40.9
ctrl	23.4
tornado	29.2
ctrl	24.3
tornado	19.1
tornado	23.4
tornado	24.3
tornado	26.2
ctrl	21.6
ctrl	19.1
tornado	22.2

Leider kennt sich Eric mit der Berechnung eines t-Tests überhaupt nicht aus. Deshalb braucht er bei der Berechnung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Bestimmen Sie die Teststatistik T_D eines Student t-Tests! **(3 Punkte)**
4. Treffen Sie mit $T_{\alpha=5\%} = 1.84$ eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effekt des Student t-Tests! **(1 Punkt)**
6. Wenn Sie *einen* Unterschied zwischen den Behandlungsgruppen erwarten würden, wie groß wäre dann die Teststatistik T_D ? Begründen Sie Ihre Antwort! **(2 Punkte)**
7. Formulieren Sie eine Antwort an Eric über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

70 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Der t-Test. Jessica erschauert. Ein mächtiges Werkzeug in den Händen desjenigen, der einen normalverteilten Messwert (Y) hat. Aber erstmal überhaupt den t-Test rechnen können. Wie sah das Experiment von Jessica überhaupt aus? Jessica hat einen Versuch in einer Klimakammer mit Lauch durchgeführt. Dabei wurde die Behandlung Bewässerungstypen (*low* und *high*) an den Lauch getestet. Gemessen hat Jessica dann als Messwert Trockengewicht [kg/ha]. Warum der Versuch im Oldenburger Land für ihre Abschlussarbeit stattfinden musste, ist ihr bis heute ein Rätsel. Egal. Gibt es jetzt einen Zusammenhang zwischen der Behandlung und Trockengewicht [kg/ha]?

Bewässerungstypen	Trockengewicht
low	48.1
high	29.5
low	40.0
high	33.4
high	23.9
low	55.2
low	45.9
low	45.0
high	35.7
high	21.6
low	39.7
high	44.9
low	50.0
high	45.1
low	51.2
low	34.6
high	47.8
high	24.4
high	15.6
high	42.6

Leider kennt sich Jessica mit der Berechnung eines t-Tests überhaupt nicht aus. Deshalb braucht sie bei der Berechnung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Bestimmen Sie die Teststatistik T_D eines Welch t-Tests! **(3 Punkte)**
4. Treffen Sie mit $T_{\alpha=5\%} = 1.84$ eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie das 99% Konfidenzintervall. Welche Annahmen haben Sie getroffen? **(2 Punkte)**
6. Nennen Sie den statistischen Grund, warum Sie sich zwischen einem Student t-Test und einem Welch t-Test entscheiden müssen! **(1 Punkt)**
7. Formulieren Sie eine Antwort an Jessica über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

71 Aufgabe

(11 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Das Schlachtgewicht von Hühnern hängt maßgeblich von der Gewichtszunahme in den ersten sieben Tagen ab. Um einen Fütterungsversuch in verschiedenen Ställen zu planen würde ein Pilotversuch durchgeführt. Das Gewicht von Hühnern wurde *vor* der Behandlung mit STARTex und 1 Woche *nach* der Behandlung gemessen. Es ergibt sich die folgende Datentabelle.

animal_id	before	after
1	4	8
2	6	12
3	1	12
4	10	13
5	15	9
6	1	14
7	6	11
8	12	10
9	7	12

1. Formulieren Sie die Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik T_{calc} eines gepaarten t-Tests für den Vergleich der beiden Zeitpunkte! **(4 Punkte)**
4. Treffen Sie mit $T_{\alpha=5\%} = 1.64$ und dem berechneten T_D eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Schätzen Sie den p -Wert aus Ihrem berechneten T_{calc} ab! Begründen Sie Ihre Antwort mit einer Skizze! **(2 Punkte)**

72 Aufgabe

(10 Punkte)



Sie erhalten folgende  Ausgabe der Funktion `t.test()` zu einem Versuch mit Maiss.


```
##  
## Two Sample t-test  
##  
## data: weight [g/qm] by group  
## t = -1.5092, df = 8, p-value = 0.1697  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -9.60637 2.00637  
## sample estimates:  
## mean in group mid mean in group trt1  
##          16.2          20.0
```

1. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
2. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie eine Abbildung in der Sie T_D , $Pr(D|H_0)$, $A = 0.95$, sowie $T_{\alpha=5\%} = |2.31|$ einzeichnen! **(4 Punkte)**
4. Beschriften Sie die Abbildung entsprechend! **(1 Punkt)**
5. Berechnen Sie den Effekt der Behandlungen! **(1 Punkt)**

73 Aufgabe

(8 Punkte)



Sie erhalten folgende  Ausgabe der Funktion `t.test()` zu einem Versuch mit Erdbeeren.


```
##  
## Two Sample t-test  
##  
## data: drymatter [kg/ha] by Fe  
## t = 1.3458, df = 16, p-value = 0.1971  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -1.342226 6.008893  
## sample estimates:  
## mean in group mid mean in group low  
## 18.22222 15.88889
```

1. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
2. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie das sich ergebende 95% Konfidenzintervall! **(2 Punkte)**
4. Beschriften Sie die Abbildung und das 95% Konfidenzintervall entsprechend! **(2 Punkte)**

74 Aufgabe

(9 Punkte)



Sie erhalten folgende  Ausgabe der Funktion `t.test()` zu einem Versuch mit Erdbeeren.


```
##  
## Two Sample t-test  
##  
## data: waterintake [ml/d] by infusion  
## t = -0.79596, df = 16, p-value = 0.4377  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -5.291480 2.402591  
## sample estimates:  
## mean in group mid mean in group low  
## 17.44444 18.88889
```

1. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
2. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie die sich ergebenden Boxplot! Welche Annahmen an die Daten haben Sie getroffen? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Skizzieren Sie die sich ergebenden Barplots! **(2 Punkte)**
5. Berechnen Sie den Effekt der Behandlungen! **(1 Punkt)**

75 Aufgabe

(8 Punkte)



Sie erhalten folgende  Ausgabe der Funktion `t.test()` zu einem Versuch mit Lauch.

```
##
## Paired t-test
##
## data: before and after
## t = 3.5082, df = 4, p-value = 0.02471
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
##  0.8343655 7.1656345
## sample estimates:
## mean difference
##                4
```

1. Formulieren Sie die wissenschaftliche Fragestellung! **(2 Punkte)**
2. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie den sich ergebenden Datensatz mit $n = 4$ Beobachtungen! Die Daten müssen *nicht* die Mittelwertsdifferenz d erfüllen! **(2 Punkte)**
4. Skizzieren Sie den sich ergebenden Boxplot der Differenzen! Welche Annahmen an die Daten haben Sie getroffen? Begründen Sie Ihre Antwort! **(2 Punkte)**

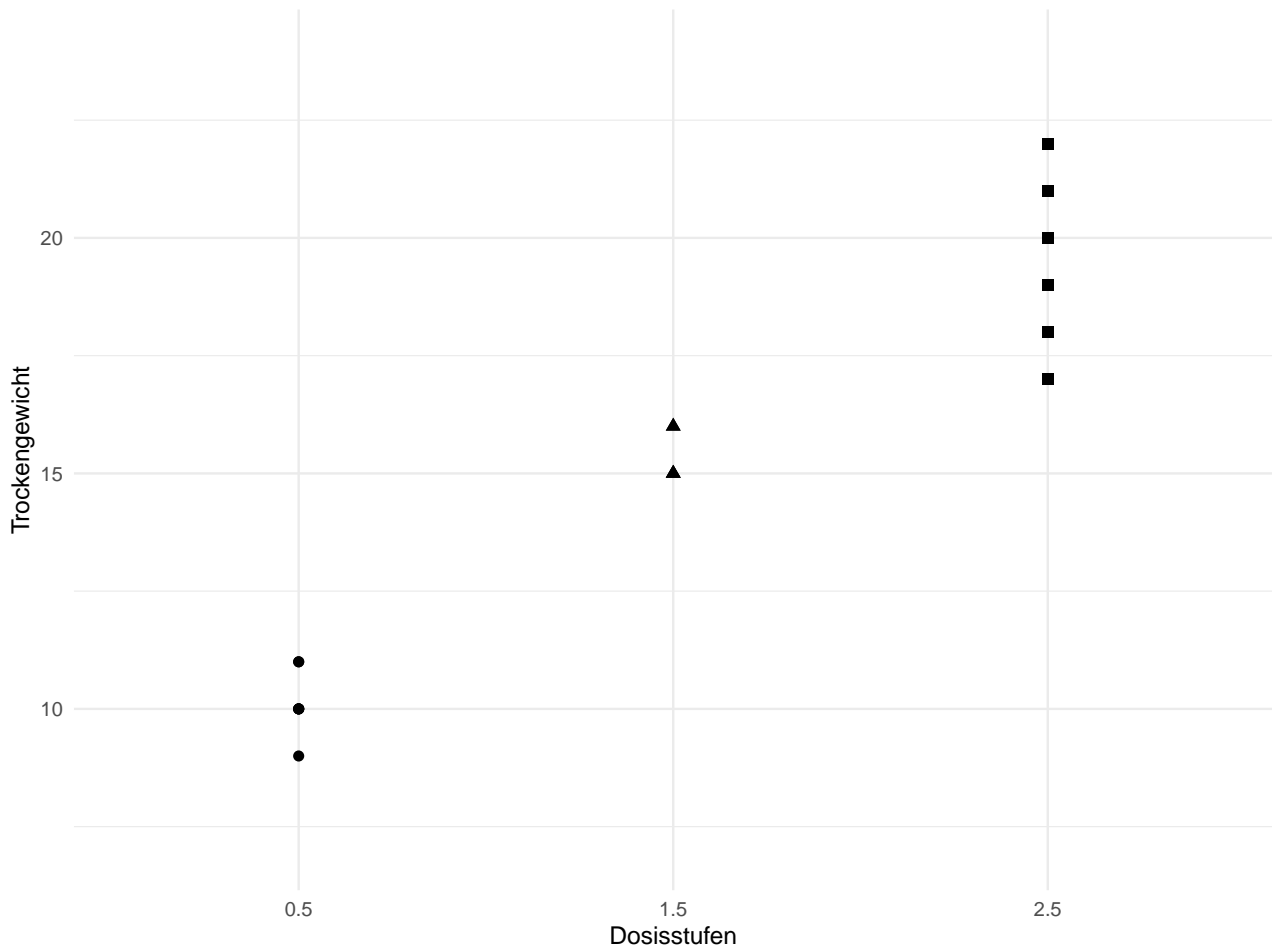
Die ANOVA

76 Aufgabe

(8 Punkte)



In einem Feldexperiment wurde der Ertrag von Erdbeeren unter drei verschiedenen Pestizid-Dosen 0.5g/l, 1.5g/l und 2.5g/l gemessen. Im Folgenden ist der Datensatz einmal visualisiert.



1. Zeichnen Sie folgende statistischen Maßzahlen in die Abbildung ein! Beschriften Sie die statistischen Maßzahlen! **(6 Punkte)**
 - Globale Mittelwert: β_0
 - Mittelwerte der einzelnen Dosen: $\bar{y}_{0.5}$, $\bar{y}_{1.5}$ und $\bar{y}_{2.5}$
 - Effekt der einzelnen Dosen: $\beta_{0.5}$, $\beta_{1.5}$, und $\beta_{2.5}$
 - Residuen oder Fehler: ϵ
2. Liegt ein *vermutlicher* signifikanter Unterschied zwischen den Dosisstufen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**

77 Aufgabe

(9 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Der Datensatz `week7_growth_tbl` enthält den Proteingehalt von Enten, die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden. Als Behandlung wurden verschiedene Nahrungszusätze in unterschiedlichen Dosen verfüttert. Als Behandlung haben Sie somit den Faktor *group* mit den Faktorstufen *ctrl*, *extreme* und *lethal* vorliegen.

1. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle mit den gegebenen Informationen von **Df** und **Sum Sq** aus! (3 Punkte)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	5.89			
error	20	58.98			
total	22				

2. Schätzen Sie den p-Wert der Tabelle mit der Information von $F_{\alpha=5\%} = 3.49$ ab. Begründen Sie Ihre Antwort! (2 Punkte)
3. Berechnen Sie den Effektschätzer η^2 . Was sagt Ihnen der Wert von η^2 aus? (2 Punkte)
4. Skizzieren Sie Körpergröße von drei Tieren pro Behandlung für eine signifikante, einfaktorielle ANOVA! (2 Punkte)

78 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Der Datensatz `fertilizer_growth_tbl` enthält das Trockengewicht der Lauchstangen, die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden. Dabei wurden die Lauchstangen unter verschiedenen Konzentrationen von einem alternativen Dünger angebaut. Als Behandlung haben Sie daher den Faktor *group* mit den Faktorstufen *lethal*, *extreme* und *pos* vorliegen.

1. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle mit den gegebenen Informationen von **Df** und **Sum Sq** aus! (3 Punkte)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	180.59			
error	22	71.97			

2. Schätzen Sie den p-Wert der Tabelle mit der Information von $F_{\alpha=5\%} = 3.44$ ab. Begründen Sie Ihre Antwort! (2 Punkte)
3. Was bedeutet ein signifikantes Ergebnis in einer einfaktoriellen ANOVA im Bezug auf die möglichen Unterschiede zwischen den Gruppen? Beziehen Sie sich auf den obigen Fragetext bei Ihrer Antwort! (2 Punkte)
4. Berechnen Sie *einen* Student t-Test mit für den *vermutlich* signifikantesten Gruppenvergleich anhand der untenstehenden Tabelle mit $T_{\alpha=5\%} = 2.03$. Begründen Sie Ihre Auswahl! (3 Punkte)


group	n	mean	sd
lethal	7	16.14	2.04
extreme	9	19.44	0.88
pos	9	13.11	2.26

5. Gegebenen der ANOVA Tabelle war das Ergebnis des t-Tests zu erwarten? Begründen Sie Ihre Antwort! (2 Punkte)

79 Aufgabe

(9 Punkte)



Der Datensatz *plant_tbl* enthält das Outcome *drymatter* für ein Feldexperiment mit Maisspflanzen, welches unter drei verschiedenen Düngerbedingungen erzielt wurden. Die Düngerbedingungen sind in dem Faktor *trt* mit den Faktorstufen *A*, *B* und *trt2* codiert. Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: drymatter
##           Df Sum Sq Mean Sq F value Pr(>F)
## trt         2  5.9289  2.96444   4.3159 0.02623
## Residuals  22 15.1111  0.68687
```

1. Stellen Sie die statistische H_0 und H_A Hypothese für die obige einfaktorielle ANOVA auf! **(2 Punkte)**
2. Interpretieren Sie das Ergebnis der einfaktoriellen ANOVA! **(2 Punkt)**
3. Berechnen Sie den Effektschätzer η^2 . Was sagt Ihnen der Wert von η^2 aus? **(2 Punkte)**
4. Skizzieren Sie eine Abbildung, der dem obigen Ergebnis der einfaktoriellen ANOVA näherungsweise entspricht! **(3 Punkte)**

80 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Der Datensatz *tooth_tbl* enthält Daten aus einer Studie zur Bewertung der Wirkung von Vitamin C auf das Zahnwachstum bei Meerschweinchen. Der Versuch wurde an verschiedenen Schweinen durchgeführt, wobei jedes Tier eine von 5 Vitamin-C-Dosen *dose* über eine von 2 Verabreichungsmethoden *supp* erhielt. Die Zahnlänge wurde als normalverteiltes Outcome gemessen.

1. Füllen Sie die unterstehende zweifaktorielle ANOVA Ergebnistabelle mit den gegebenen Informationen von **Df** und **Sum Sq** aus! **(4 Punkte)**
2. Schätzen Sie den p-Wert der Tabelle mit der Information von den $F_{\alpha=5\%}$ -Werten mit $F_{supp} = 4.26$ und $F_{dose} = 3.40$ sowie $F_{supp:dose} = 5.23$ ab. Begründen Sie Ihre Antwort! **(4 Punkte)**


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	6.49			
dose	4	360.93			
supp:dose	4	226.42			
error	40	487.09			

3. Was bedeutet ein signifikantes Ergebnis in einer zweifaktoriellen ANOVA im Bezug auf die möglichen Unterschiede zwischen den Gruppen? Beziehen Sie sich dabei einmal auf den Faktor *supp* und einmal auf den Faktor *dose*! **(2 Punkte)**
4. Was sagt der Term *supp:dose* aus? Interpretieren Sie das Ergebnis des abgeschätzten p-Wertes! **(2 Punkte)**


81 Aufgabe

(10 Punkte)



Der Datensatz *gain_weight_tbl* enthält Daten aus einer Studie zur Bewertung der Wirkung vom Vitamin Selen auf das Wachstum von Hängebauchschweinen. Der Versuch wurde an 65 Hängebauchschweinen durchgeführt, wobei jedes Tier eine von drei Selen-Dosen *dose* (0.2 ng/Tag, 1 ng/Tag und 20 ng/Tag) über eine von zwei Verabreichungsmethoden *form* erhielt (Wasser oder Festnahrung). Sie erhalten folgende Ausgabe in .

```
## Analysis of Variance Table
##
## Response: fat_perc
##      Df Sum Sq Mean Sq F value    Pr(>F)
## dose      2  78.49   39.243    6.766 0.006437
## form      1 177.07  177.067   30.529 3.022e-05
## dose:form  2 392.97  196.487   33.877 7.910e-07
## Residuals 18 104.40    5.800
```

1. Stellen Sie die statistische H_0 und H_A Hypothese für die obige zweifaktorielle ANOVA für den Faktor *dose* auf! **(2 Punkte)**
2. Interpretieren Sie das Ergebnis der zweifaktoriellen ANOVA. Gehen Sie im besonderen auf den Term *dose : form* ein! **(2 Punkte)**
3. Zeichnen Sie eine Abbildung, der dem obigen Ergebnis der zweifaktoriellen ANOVA näherungsweise entspricht! **(4 Punkte)**
4. Beschriften Sie die Abbildung entsprechend der  Ausgabe! **(2 Punkte)**

82 Aufgabe

(8 Punkte)



In der untenstehenden Tabelle ist die Formel für den F-Test aus der ANOVA und die Formel für den Student t-Test dargestellt. In der ANOVA berechnen Sie die F-Statistik F_{calc} und in dem Student t-Test die T-Statistik T_{calc} .

$$F_{calc} = \frac{MS_{treatment}}{MS_{error}} \quad T_{calc} = \frac{\bar{y}_1 - \bar{y}_2}{s_p \cdot \sqrt{2/n_g}}$$

1. Erklären Sie den konzeptionellen Zusammenhang zwischen der F_{calc} Statistik und T_{calc} Statistik! **(2 Punkte)**
2. Visualisieren Sie eine nicht signifikante F_{calc} Statistik sowie eine signifikante F_{calc} Statistik anhand von $MS_{treatment}$ und MS_{error} ! Beschriften Sie die Abbildung! **(2 Punkte)**
3. Erklären Sie an der Formel des F-Tests sowie an der Abbildung warum das Minimum der F-Statistik 0 ist! **(2 Punkte)**
4. Wenn die F-Statistik 0 ist, spricht dies eher für oder gegen die Nullhypothese? Begründen Sie Ihre Antwort! **(2 Punkte)**

83 Aufgabe

(8 Punkte)



Sie rechnen eine zweifaktorielle ANOVA und erhalten einen signifikanten Interaktionseffekt zwischen den beiden Faktoren f_1 und f_2 . Der Faktor f_1 hat drei Level. Der Faktor f_2 hat dagegen nur zwei Level.

1. Visualisieren Sie in zwei getrennten Abbildungen keine und eine starke Interaktion zwischen den Faktoren f_1 und f_2 ! **(4 Punkte)**
2. Erklären Sie den Unterschied zwischen den beiden Stärken der Interaktion! **(2 Punkte)**
3. Wenn eine signifikante Interaktion in den Daten vorliegt, wie ist dann das weitere Vorgehen bei einem Posthoc-Test? **(2 Punkte)**

84 Aufgabe

(9 Punkte)



Sie rechnen eine einfaktorielle ANOVA mit einem Faktor f_1 mit drei Leveln. Nachdem Sie die einfaktorielle ANOVA gerechnet haben, erhalten Sie einen p-Wert von 0.078 und eine F Statistik mit $F_{calc} = 1.2$. Als Sie sich die Boxplots der Behandlungen anschauen, stellen Sie fest, dass es eigentlich einen Mittelwertsunterschied zwischen dem dritten und zweiten Level geben müsste. Die IQR-Bereiche überlappen sich nicht und die Mediane liegen auch weit vom globalen Mittel entfernt.

1. Erklären Sie die Annahme der Normalverteilung und die Annahme der Varianzhomogenität für eine ANOVA an einer passenden Abbildung! **(3 Punkte)**
2. Visualisieren Sie die Berechnung von F_{calc} am obigen Beispiel! **(3 Punkte)**
3. Erklären Sie das Ergebnis der obigen einfaktoriellen ANOVA unter der Berücksichtigung der Annahmen an eine ANOVA! **(3 Punkte)**

Multiple Gruppenvergleiche

85 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In einem Experiment zur Dosiswirkung wurden verschiedene Dosisstufen mit einer Kontrollgruppe verglichen. Es wurden verschiedene t-Test für den Mittelwertsvergleich gerechnet und es ergab sich folgende Tabelle mit den rohen p-Werten (*Raw p-value*).

Vergleich	Raw p-value	Adjusted p-value	Reject Null
dose 80 - ctrl	0.002		
dose 40 - ctrl	0.060		
dose 50 - ctrl	0.080		
dose 20 - ctrl	0.012		


1. Füllen Sie die Spalte „Adjusted p-value“ mit den adjustierten p-Werten nach Bonferoni aus! **(4 Punkte)**
2. Entscheiden Sie, ob nach der Adjustierung die Nullhypothese weiter abgelehnt werden kann. Tragen Sie Ihre Entscheidung in die obige Tabelle ein. Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Erklären Sie warum die p-Werte bei multiplen Vergleichen adjustiert werden müssen! **(2 Punkte)**

86 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In einem Experiment für den Ertrag von Weizen in kg/ha mit fünf Dosisstufen (ctrl, low, mid, high und pos) erhalten Sie folgendes *Compact letter display (CLD)* als  Ausgabe aus den rohen, unadjustierten *p*-Werten.

```
## ctrl high low mid pos  
## "a" "a" "b" "b" "c"
```


1. Zeichnen Sie eine Abbildung, der sich ergebenden Barplots! **(2 Punkte)**
2. Ergänzen Sie das *Compact letter display (CLD)* zu der Abbildung! **(1 Punkt)**
3. Erklären Sie *einen* Vorteil und *einen* Nachteil des *Compact letter display (CLD)*! **(2 Punkte)**
4. Erstellen Sie eine Matrix mit den paarweisen *p*-Werten, die sich näherungsweise aus dem *Compact letter display (CLD)* ergeben würde! Begründen Sie Ihre Antwort! **(3 Punkte)**

87 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In einem Experiment für den Ertrag von Weizen in kg/ha mit vier Dosisstufen (ctrl, low, mid und high) erhalten Sie folgende Matrix als  Ausgabe mit den rohen, unadjustierten p -Werten.

```
##          ctrl          high          low          mid
## ctrl 1.0000000 0.0742321 0.0004006 0.4698843
## high 0.0742321 1.0000000 0.0000020 0.2735058
## low  0.0004006 0.0000020 1.0000000 0.0000497
## mid  0.4698843 0.2735058 0.0000497 1.0000000
```

Im Weiteren erhalten Sie folgende Informationen über die Fallzahl n , den Mittelwert $mean$ und die Standardabweichung sd in den jeweiligen Dosisstufen.

trt	n	mean	sd
ctrl	9	9.91	1.08
high	9	6.73	4.28
low	9	16.73	4.63
mid	9	8.65	3.55

1. Zeichnen Sie in eine Abbildung, die sich ergebenden Barplots! *Sortieren Sie dabei die Gruppen nach absteigender Effektstärke!* **(3 Punkte)**
2. Adjustieren Sie die rohen p -Werte nach Bonferroni. Begründen Sie Ihre Antwort! **(3 Punkte)**
3. Ergänzen Sie das *Compact letter display (CLD)* zu der Abbildung. Nutzen Sie dazu die rohen p -Werte! **(2 Punkte)**
4. Interpretieren Sie das *Compact letter display (CLD)*! **(2 Punkte)**

Der χ^2 -Test & Der diagnostische Test

88 Aufgabe

(11 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Nach einem Feldexperiment ergibt sich die folgende 2x2 Datentabelle mit einem Pestizid (ja/nein) der Marke KillAll, dargestellt in den Zeilen, und dem infizierten Pflanzenstatus (ja/nein) von Kohl, dargestellt in den Spalten. Insgesamt wurden $n = 120$ Pflanzen untersucht.

	Erkrankt (ja)	Erkrankt (nein)	
Pestizid (ja)	38	21	
Pestizid (nein)	23	38	

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die Randsummen! **(1 Punkt)**
3. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test auf der 2x2 Tafel! **(3 Punkte)**
4. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem $\chi^2_{\alpha=5\%} = 3.841$! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Skizzieren Sie die χ^2 -Verteilung, wenn die H_0 wahr ist! Ergänzen Sie $\chi^2_{\alpha=5\%}$ und χ^2_{calc} in der Abbildung! **(2 Punkte)**
6. Berechnen Sie den Effektschätzer *Cramers V*! Interpretieren Sie den Effektschätzer! **(2 Punkte)**

89 Aufgabe

(7 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Gegeben sind folgende Randsummen in einer 2x2 Kreuztabelle aus einem Experiment mit $n = 116$ Sauen. In dem Experiment wurde gemessen, ob eine Sau nach einer Behandlung mit einem Medikament (ja/nein) mehr als 30 Ferkel pro Jahr bekommen konnte (ja/nein).

	>30 Ferkel (ja)	≤30 Ferkel (nein)	
Medikament (ja)			55
Medikament (nein)			61
	58	58	116

1. Ergänzen Sie die Felder innerhalb der 2x2 Kreuztabelle in dem Sinne, dass *ein* signifikanter Effekt zu erwarten wäre! **(2 Punkte)**
2. Erklären und Begründen Sie Ihr Vorgehen an der Formel des Chi-Quadrat-Tests mit

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

Sie können dies an einem Beispiel erklären! **(2 Punkte)**

3. Was ist die Mindestanzahl an Beobachtungen je Zelle? Wenn in einer der Zellen weniger Beobachtungen auftreten, welchen Test können Sie anstatt des „normalen“ Chi-Quadrat-Tests anwenden? **(2 Punkte)**
4. Warum hat die obige Vierfeldertafel einen Freiheitsgrad von $df = 1$? **(1 Punkt)**


90 Aufgabe

(10 Punkte)



Nach einem Experiment erhalten Sie folgende 2x2 Kreuztabelle aus Ihren erhobenen Daten.

```
##           Infiziert
## Gruppe      yes no
##  Belana      6   3
##  Goldmarie   6   4
```

Aus der 2x2 Kreuztabelle erhalten Sie folgende  Ausgabe der Funktion `fisher.test()`.

```
##
## Fisher's Exact Test for Count Data
##
## data:  mat
## p-value = 1
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1442262 13.2691123
## sample estimates:
## odds ratio
##  1.313276
```

1. Formulieren Sie die wissenschaftliche Fragestellung! **(2 Punkte)**
2. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie das sich ergebende 95% Konfidenzintervall! **(2 Punkte)**
4. Beschriften Sie die Abbildung und das 95% Konfidenzintervall entsprechend! **(2 Punkte)**
5. Interpretieren Sie das *Odds ratio* im Kontext der wissenschaftlichen Fragestellung! **(2 Punkte)**

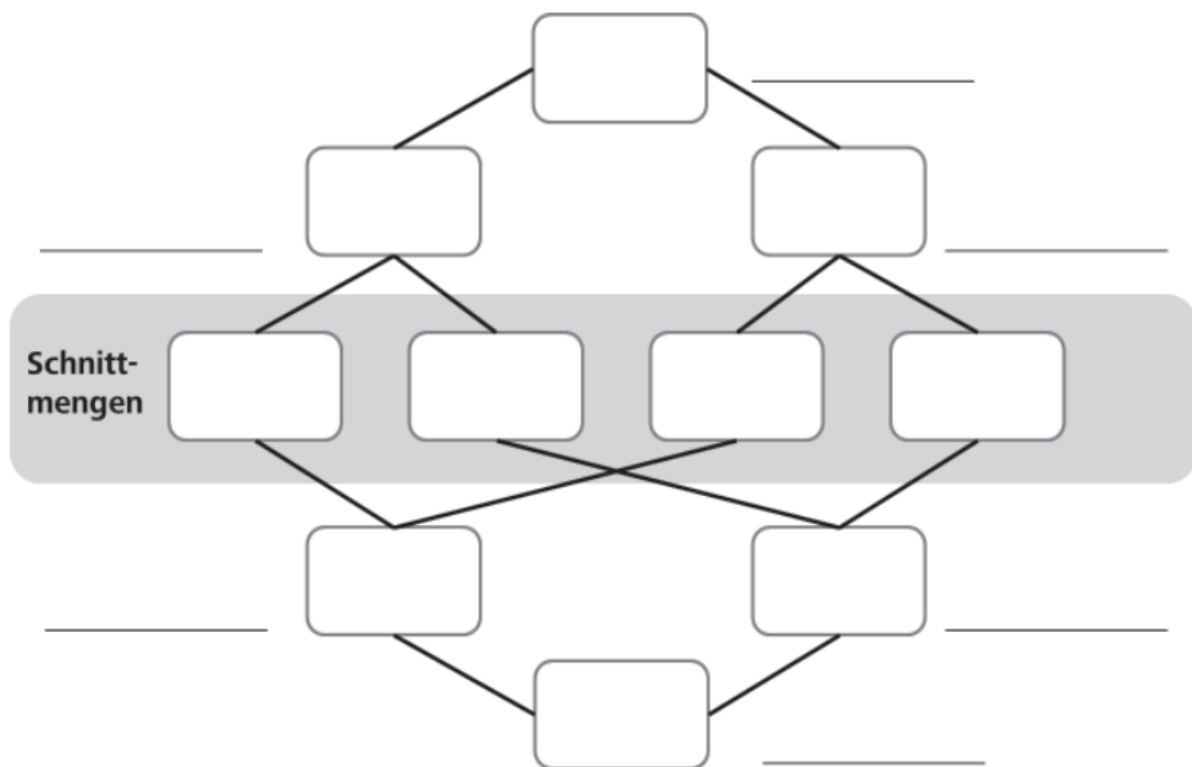
91 Aufgabe

(11 Punkte)



Die Prävalenz von Klauenseuche bei Wollschweinen wird mit 3% angenommen. In 85% der Fälle ist ein Test positiv, wenn das Wollschwein erkrankt ist. In 8% der Fälle ist ein Test positiv, wenn das Wollschwein *nicht* erkrankt ist und somit gesund ist. Sie werten 4000 Wollschweine mit einem diagnostischen Test auf Klauenseuche aus.

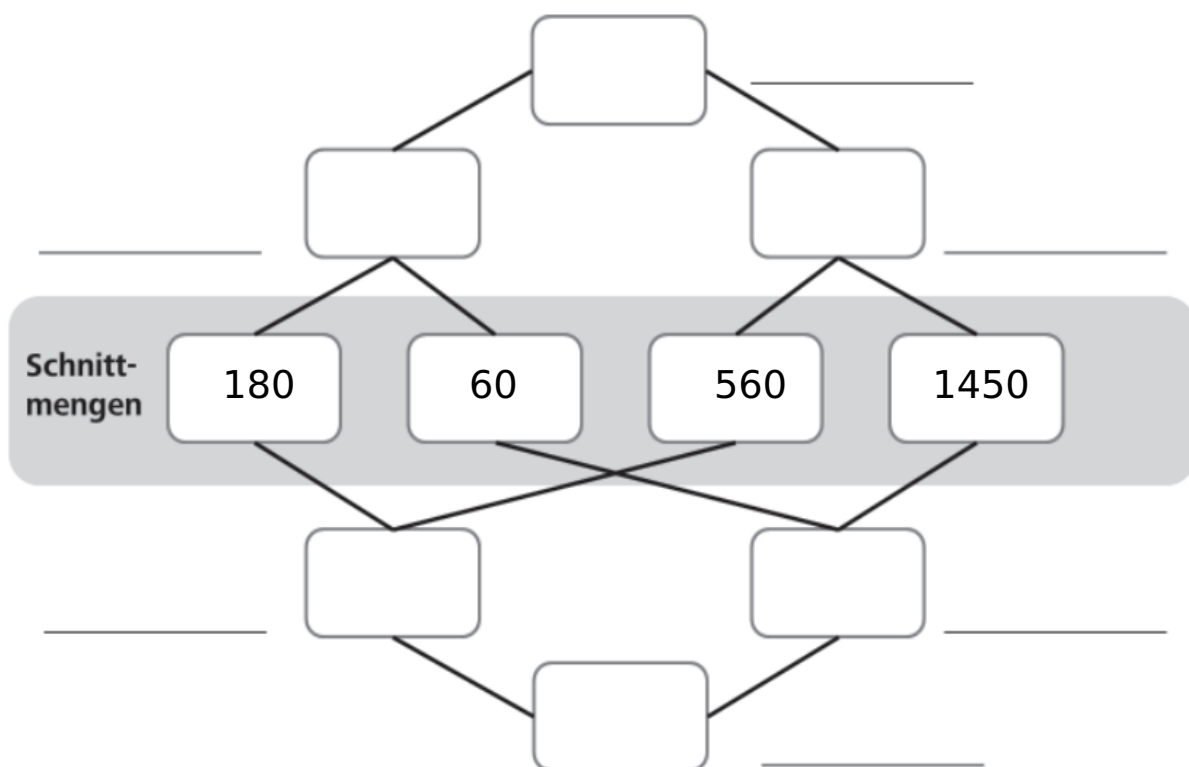
1. Füllen und beschriften Sie den untenstehenden Doppelbaum! Beschriften Sie auch die Äste des Doppelbaumes, mit denen Ihnen bekannten Informationen! **(8 Punkte)**
2. Berechnen Sie die Wahrscheinlichkeit $Pr(K^+|T^+)$! **(2 Punkte)**
3. Was sagt Ihnen die Wahrscheinlichkeit $Pr(K^+|T^+)$ aus? **(1 Punkt)**





Folgender diagnostischer Doppelbaum nach der Testung auf Klauenseuche bei Fleckvieh ist gegeben.

1. Füllen und beschriften Sie den untenstehenden Doppelbaum! **(4 Punkte)**
2. Berechnen Sie die Wahrscheinlichkeit $Pr(K^+|T^+)$! **(2 Punkte)**
3. Berechnen Sie die Prävalenz für Klauenseuche! **(2 Punkte)**
4. Berechnen Sie die Sensitivität und Spezifität des diagnostischen Tests für Klauenseuche! Erstellen Sie dafür zunächst eine 2x2 Kreuztabelle aus dem ausgefüllten Doppelbaum! **(4 Punkte)**



Nicht parametrische Tests

93 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Die Anzahl an Nematoden wurde vor und nach einer Behandlung mit einem bioaktiven Dünger gezählt. Es ergibt sich folgende Datentabelle.

Vorher	Nachher	Differenz	Vorzeichen	Rang	Positiv Rang	Negativ Rang
13	12					
6	12					
10	7					
15	10					
8	11					
10	14					
9	11					
11	13					
10	15					
9	14					
11	11					
6	17					
12	13					
9	11					
7	13					

1. Ergänzen Sie die obige Tabelle mit den notwendigen Informationen, die Sie benötigen um einen Wilcoxon-Vorzeichen-Rang-Test zu rechnen! **(4 Punkte)**
2. Bestimmen Sie die Teststatistik W_D mit $W_D = \min(T_-; T_+)$ und berechnen Sie den erwarteten Wert $\mu_W = \frac{n_{10} \cdot (n_{10} + 1)}{4}$! **(2 Punkte)**
3. Berechnen Sie anschließend den z-Wert mit $z = \frac{W_D - \mu_W}{15.93}$! **(2 Punkte)**
4. Liegt mit einer Signifikanzschwelle von $z_{\alpha=5\%} = 1.96$ ein Unterschied zwischen den beiden Zeitpunkten vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie die Effektstärke mit $r = |\frac{z}{\sqrt{n}}|$ und interpretieren Sie die Effektstärke! **(2 Punkte)**

94 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Nach einer Behandlung mit RootsGoneX wurde die mittlere Anzahl an Wurzeln an der invasiven Lupine (*Lupinus polyphyllus*) gezählt. Es ergab sich folgender Datensatz an mittleren Wurzelanzahl.

Treatment	Count
RootsGoneX	12.6
Kontrolle	-0.3
RootsGoneX	11.1
Kontrolle	-1.3
Kontrolle	3.5
RootsGoneX	13.0
RootsGoneX	13.3
RootsGoneX	11.7
Kontrolle	2.0
RootsGoneX	10.8
Kontrolle	2.5

Rechnen Sie einen Mann-Whitney-U-Test auf den obigen Daten.

1. Bestimmen Sie hierfür U_D mit $U_D = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1!$ (4 Punkte)

2. Geben Sie eine Aussage über die Signifikanz von U_D durch $z = \frac{U_D - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$ und dem kritischen

Wert von $z_{\alpha=5\%} = 1.96$. Begründen Sie Ihre Antwort! (2 Punkte)

3. Berechnen Sie die Effektstärke mit $r = |\frac{z}{\sqrt{n}}|$ und interpretieren Sie die Effektstärke! (2 Punkte)

95 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Die Anzahl an Blüten der Vanillepflanze pro Box wurde nach der Gabe von zusätzlichen Phosphorlösung (Kontrolle, Dosis 20 und Dosis 40) bestimmt. Es ergeben sich folgende nach der Anzahl der Blüten geordnete Daten.

Treatment	Count	Rang Kontrolle	Rang Dosis 20	Rang Dosis 40
Dosis 20	13.7			
Dosis 40	10.1			
Kontrolle	7.8			
Dosis 40	11.3			
Kontrolle	10.5			
Dosis 40	10.3			
Dosis 20	10.3			
Dosis 40	10.5			
Kontrolle	9.1			
Kontrolle	8.7			
Dosis 20	12.6			
Dosis 20	10.7			
Dosis 20	10.4			
Kontrolle	7.1			
Dosis 40	8.8			
Kontrolle	11.6			
Dosis 40	4.9			

Rechnen Sie einen Kruskal-Wallis-Test auf den obigen Daten.

- Bestimmen Sie hierfür H_D mit $H_D = \frac{12}{n(n+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} \right) - 3(n+1)$! **(6 Punkte)**
- Geben Sie eine Aussage über die Signifikanz von H_D durch den kritischen Wert von $H_{\alpha=5\%} = 5.99$! **(1 Punkt)**
- Wie lautet die statistische Nullhypothese die Sie mit dem Kruskal-Wallis-Test überprüfen? **(1 Punkt)**
- Was sagt ein signifikantes Ergebnis des Kruskal-Wallis-Test in Bezug auf die einzelnen Gruppenvergleiche aus? **(1 Punkt)**
- Nennen Sie das statistische Verfahren, welches Sie als Posthoc Test nach einem signifikanten Kruskal-Wallis-Test durchführen würden! **(1 Punkt)**

Lineare Regression & Korrelation

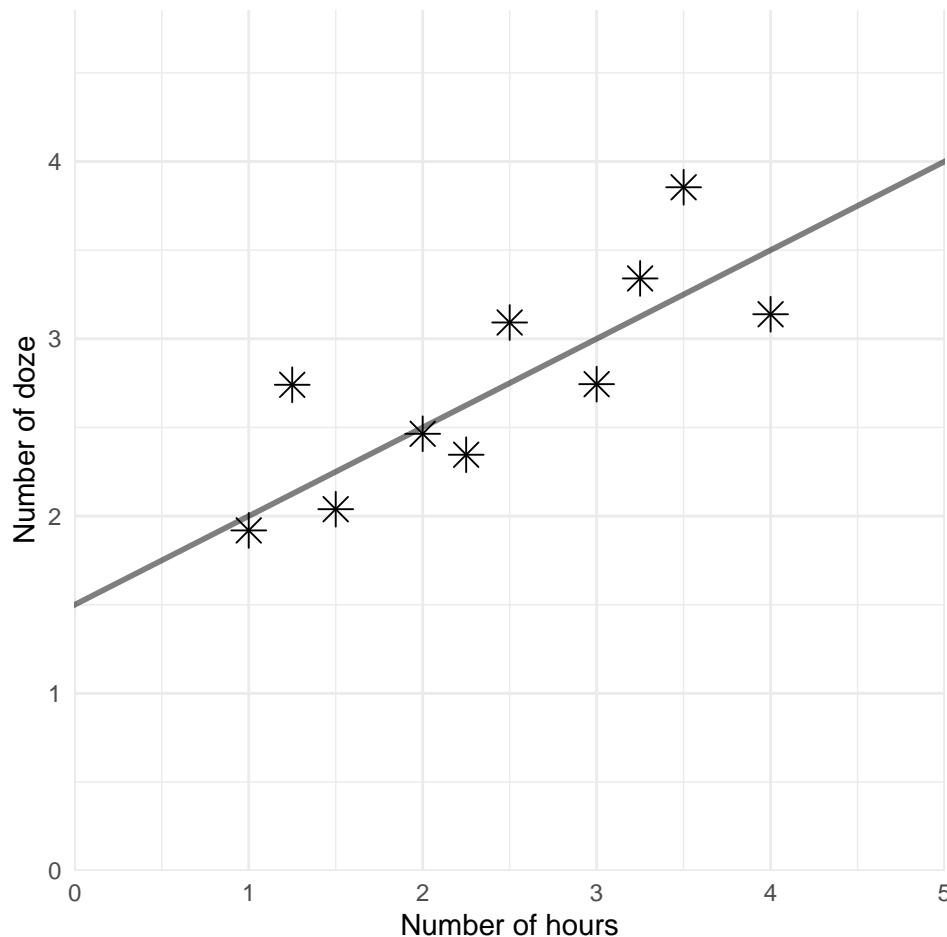
96 Aufgabe

(7 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In einer Studie zur „Arbeitssicherheit auf dem Feld“ wurde gemessen wie viele Stunden auf einem Feld gefahren wurden und wie oft der Fahrer dabei drohte einzunicken. Es ergab sich folgende Abbildung.




1. Erstellen Sie die Regressionsgleichung aus der obigen Abbildung in der Form $y \sim \beta_0 + \beta_1 \cdot x$! **(2 Punkte)**
2. Beschriften Sie die Grade mit den Parametern der linearen Regressionsgleichung! **(2 Punkte)**
3. Liegt ein Zusammenhang zwischen der Anzahl an gefahrenen Runden und der Müdigkeit vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Wenn kein Zusammenhang zu beobachten wäre, wie würde die Grade aussehen? **(1 Punkt)**

97 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In einem Stallexperiment mit $n = 60$ Ferkeln wurde der Gewichtszuwachs in kg unter ansteigender Lichteinstrahlung in nm gemessen. Sie erhalten den  Output einer simplen Gaussian linearen Regression sieben Wochen nach der ersten Messung.

term	estimate	std.error	t statistic	p-value
(Intercept)	3.47	1.73		
light	1.85	0.17		

1. Zeichnen Sie die Grade aus der obigen Tabelle in ein Koordinatenkreuz! **(1 Punkt)**
2. Beschriften Sie die Abbildung und die Gerade mit den statistischen Kenngrößen! **(2 Punkte)**
3. Formulieren Sie die Regressionsgleichung! **(2 Punkte)**
4. Berechnen Sie die t Statistik für *(Intercept)* und *light*! **(2 Punkte)**
5. Schätzen Sie den p-Wert für *(Intercept)* und *light* mit $T_{\alpha=5\%} = 1.96$ ab. Was sagt Ihnen der p-Wert aus? Begründen Sie Ihre Antwort! **(3 Punkte)**

98 Aufgabe

(10 Punkte)



Sie rechnen eine lineare Regression um nach einem Feldexperiment den Zusammenhang zwischen Trockengewicht kg/m^2 (*weight*) und Wassergabe l/m^2 (*water*) bei Erdbeerpflanzen zu bestimmen. Sie erhalten folgende R Ausgabe.

```
##
## Call:
## lm(formula = weight ~ water, data = data_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.40  -2.15   0.10    2.10   6.60
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.400      1.631  10.669 5.23e-06
## waterB         10.000      2.307   4.336 0.00249
##
## Residual standard error: 3.647 on 8 degrees of freedom
## Multiple R-squared:  0.7015, Adjusted R-squared:  0.6641
## F-statistic: 18.8 on 1 and 8 DF,  p-value: 0.002493
```

1. Ist die Annahme der Normalverteilung an das Outcome *water* erfüllt? Begründen Sie die Antwort! **(2 Punkte)**
2. Wie groß ist der Effekt der Wassergabe? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Liegt ein signifikanter Effekt vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Erklären Sie *kurz* den Begriff R-squared! Was sagt Ihnen der Wert 0.7015 aus? **(2 Punkte)**
5. Schreiben Sie das Ergebnis der R Ausgabe in zwei Sätzen auf, der die Information zum Effekt und der Signifikanz enthält! **(2 Punkte)**

99 Aufgabe

(9 Punkte)

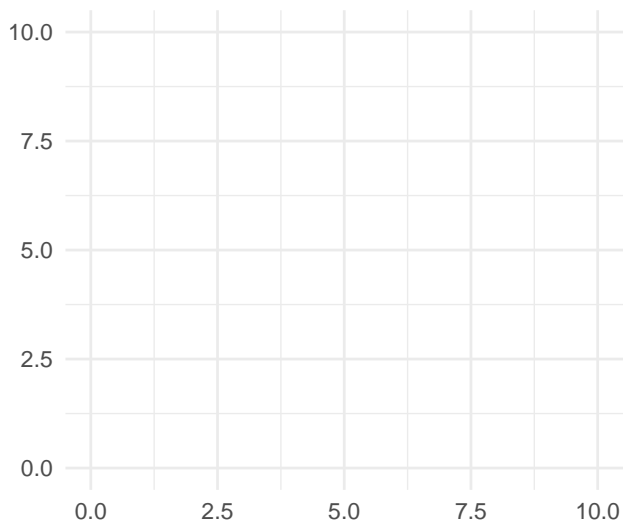


Im folgenden sehen Sie drei leere Scatterplots. Füllen Sie diese Scatterplots nach folgenden Anweisungen.

1. Zeichnen Sie für die angegebene ρ -Werte eine Gerade in die entsprechende Abbildung! **(3 Punkte)**
2. Zeichnen Sie für die angegebenen R^2 -Werte die entsprechende Punktwolke um die Gerade. **(3 Punkte)**
3. Sie rechnen ein statistisches Modell. Was sagen Ihnen die R^2 -Werte über das jeweilige Modell? **(3 Punkte)**

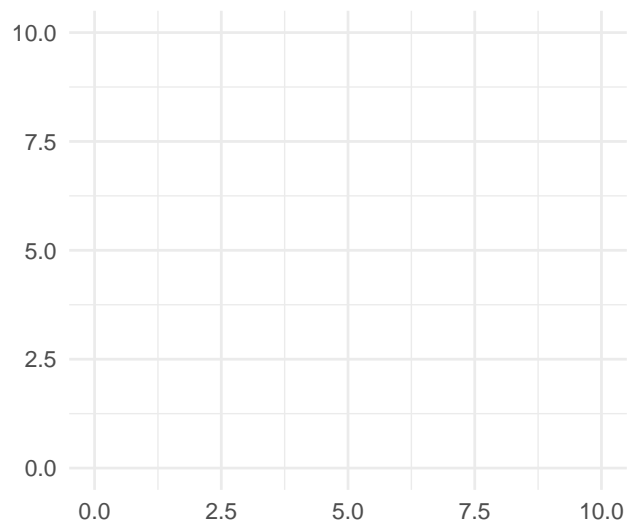
Pearsons $\rho = 0.5$

$R^2 = 1$



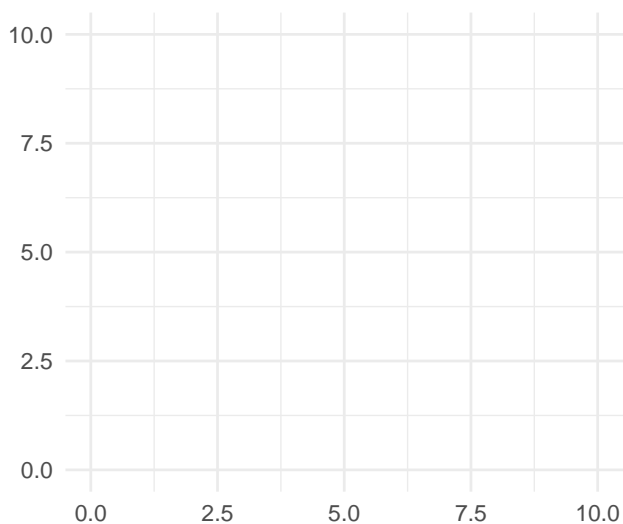
Pearsons $\rho = 0.75$

$R^2 = 0.75$



Pearsons $\rho = -0.75$

$R^2 = 0.5$



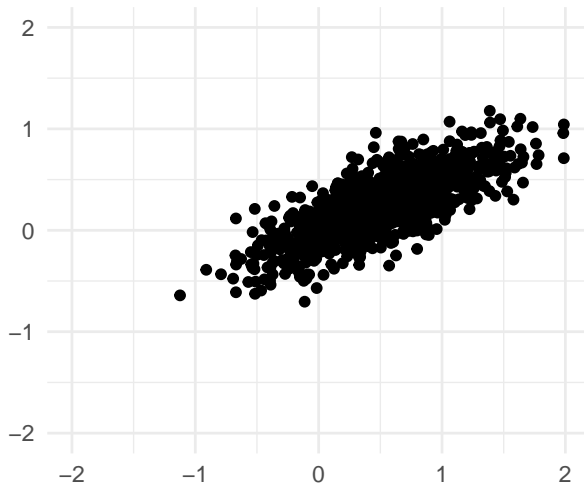


Im folgenden sehen Sie vier Scatterplots. Ergänzen Sie die Überschriften der jeweiligen Scatterplots.

1. Schätzen Sie die ρ -Werte in der entsprechenden Abbildung! **(4 Punkte)**
2. Schätzen Sie die R^2 -Werte in der entsprechenden Punktwolke um die Gerade! **(4 Punkte)**
3. Sie rechnen ein statistisches Modell. Was sagen Ihnen die R^2 -Werte über das jeweilige Modell? **(1 Punkt)**

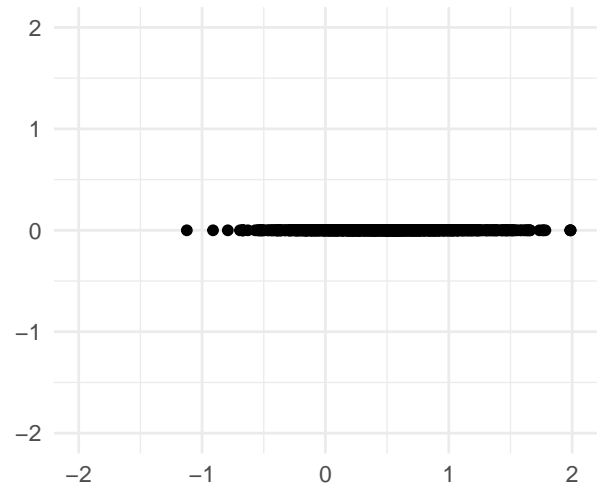
Pearsons $\rho =$

$R^2 =$



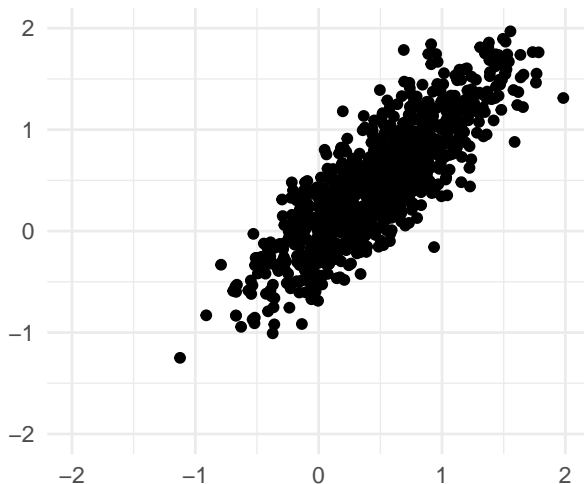
Pearsons $\rho =$

$R^2 =$



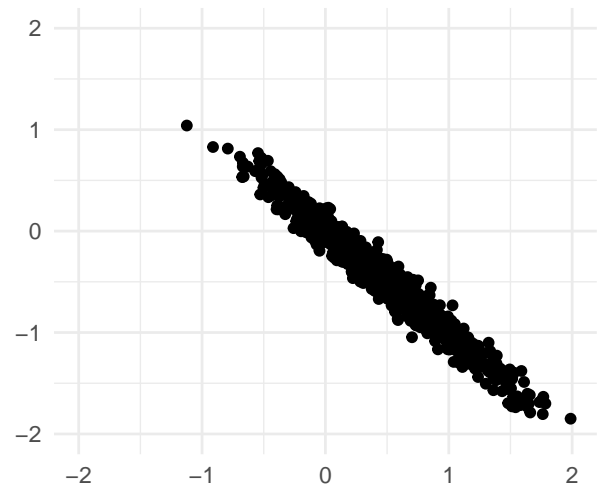
Pearsons $\rho =$

$R^2 =$



Pearsons $\rho =$


$R^2 =$



101 Aufgabe

(9 Punkte)



Sie erhalten folgende  Ausgabe der Funktion `cor.test()`.

```
##  
## Kendall's correlation  
##  
## data: weight and fertilizer  
## t = -11.326, df = 8, p-value = 3.326e-06  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.9931495 -0.8752247  
## sample estimates:  
## cor  
## -0.9702053
```

1. Formulieren Sie die wissenschaftliche Fragestellung! **(2 Punkt)**
2. Nennen Sie die zwei Eigenschaften des Korrelationskoeffizienten! Erklären Sie *eine* der Eigenschaften an einem Beispiel! **(2 Punkte)**
3. Interpretieren Sie den Korrelationskoeffizienten hinsichtlich des Effekts und der Signifikanz! Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Visualisieren Sie das 95% Konfidenzintervall! Beschriften Sie die Abbildung! **(2 Punkte)**
5. Sind die Variablen `weight` and `fertilizer` normalverteilt? Begründen Sie Ihre Antwort! **(1 Punkt)**



Sie rechnen eine lineare Regression um nach einem Feldexperiment den Zusammenhang zwischen Trockengewicht kg/m^2 (*drymatter*) und Wassergabe l/m^2 (*water*) bei Spargel zu bestimmen. Sie erhalten folgende Datentabelle.

.id	drymatter	water	.fitted	.resid
1	26.9	12.2	28.1	
2	19.4	5.3	19.6	
3	24.5	9.5	24.8	
4	29.9	13.5	29.8	
5	25.7	9.3	24.5	
6	25.5	9.9	25.3	
7	31.3	14.3	30.9	
8	25.1	10.3	25.8	
9	26.8	10.7	26.3	

1. Ergänzen Sie die Werte in der Spalte *.resid* in der obigen Tabelle. Geben Sie den Rechenweg und Formel mit an! **(4 Punkte)**
2. Zeichnen Sie den sich aus der obigen Tabelle ergebenden Residualplot. Beschriften Sie die Abbildung! **(4 Punkte)**
3. Gibt es auffällige Werte anhand des Residualplots? Begründen Sie Ihre Antwort! **(2 Punkte)**

103 Aufgabe



(12 Punkte)

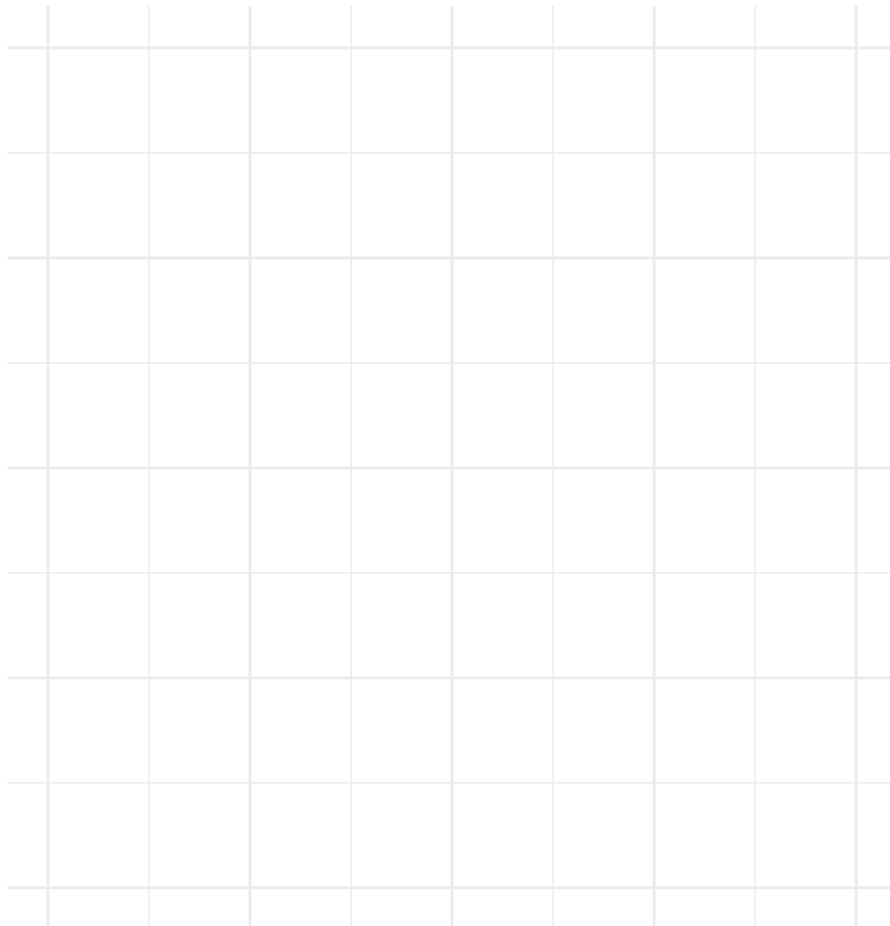


1. Zeichnen Sie in die drei untenstehenden, leeren Abbildungen die Zeile des Regressionskreuzes der Normalverteilung. Wählen Sie die Beschriftung der y-Achse sowie der x-Achse entsprechend aus! **(6 Punkte)**
2. Ergänzen Sie die jeweiligen statistischen Methoden zu der Abbildung! **(2 Punkte)**
3. Welchen Effektschätzer erhalten Sie aus der entsprechend linearen Regression bzw. den Gruppenvergleich? Geben Sie ein Beispiel! **(2 Punkte)**
4. Wenn Sie keinen Effekt erwarten, welchen *Zahlenraum* nimmt dann der Effektschätzer ein? Geben Sie ein Beispiel! **(2 Punkte)**



Ein Feldexperiment wurde mit $n = 200$ Pflanzen durchgeführt. Folgende Einflussvariablen (x) wurden erhoben: dosage, rainfall und S. Als mögliche Outcomevariablen stehen Ihnen nun folgende gemessene Endpunkte zu Verfügung: drymatter, yield, count, quality_score und dead.

1. Wählen Sie ein Outcome was zu der Verteilungsfamilie *Gaussian* gehört! **(1 Punkt)**
2. Schreiben Sie das Modell in der Form $y \sim x$ wie es in  in der Funktion `glm()` üblich ist *ohne Interaktionsterm*! **(3 Punkte)**
3. Schreiben Sie das Modell in der Form $y \sim x$ wie es in  üblich ist und ergänzen Sie *einen* Interaktionsterm nach Wahl! **(1 Punkt)**
4. Zeichnen Sie eine *schwache* Interaktion in die Abbildung unten für den Endpunkt *yield*. Ergänzen Sie eine aussagekräftige Legende. Wie erkennen Sie eine Interaktion? Begründen Sie Ihre Antwort! **(4 Punkte)**



Experimentelles Design

105 Aufgabe

(10 Punkte)



Nach einem erfolgreichen Pilotversuch zur Wirksamkeit von Pestiziden bei Maiss in einem Feldexperiment wollen Sie nun den Versuch eine Nummer größer anlegen. Dafür entscheiden Sie sich für ein faktorielles Versuchsdesign. In Ihrem Hauptversuch stellt die Wirksamkeit von Pestiziden den ersten Faktor mit insgesamt 4 Leveln dar. Der zweite Faktor mit dem Block beinhaltet 3 Level.

Im ersten Schritt überlegen Sie ein einfaches experimentelles Design zu probieren. Sie entscheiden sich für das *Latin square design*.

1. Skizzieren Sie das *Latin square design* für Ihren Versuch! **(4 Punkte)**
2. Skizzieren Sie eine Datentabelle für den Versuch mit zwei Wiederholungen! **(2 Punkte)**
3. Skizzieren Sie eine Abbildung mit Barplots und einem angenommenen normalverteilten Outcome! **(4 Punkte)**

Mathematik

106 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Herodot – der Schimmel aus Ivenack Während der Besetzung Mecklenburgs durch die Franzosen kamen Napoleon die Geschichten des berühmten Apfelschimmels Herodot aus Ivenack zu Gehör. Herodot lief zwar niemals Rennen, war aber eines der berühmtesten Pferde dieser Zeit. Napoleon selbst gab den Auftrag, diesen Schimmel durch die Armee nach Frankreich zu bringen. Der Legende nach sollen Arbeiter den Schimmel im hohlen Stamm einer 1000-jährigen Eiche aus Ivenack vor den Franzosen versteckt haben. Doch Herodot verrät sein Versteck durch lautes Wiehern, woraufhin die französische Armee den Schimmel beschlagnahmte und nach Frankreich führte¹.

Forschungsfrage: "Konnten die Ivenacker den Apfelschimmel Herodot vor dem Zugriff von Napoleon in der 1000-jährigen Eiche verstecken?"

Gehen Sie von einem radialen Wachstum der 1000-jährigen Eiche von 0.9mm pro Jahr aus. Es ist bekannt, dass die Eiche im Jahr 2022 einen Umfang von 10.5m in Bruthöhe hatte.

1. Wie groß war der Durchmesser in m der Eiche im Jahr 1840 als Herodot in der Eiche versteckt werden sollte? **(2 Punkte)**
2. Skizzieren Sie in einer Abbildung einen linearen Zusammenhang und einen exponentiellen Zusammenhang für das Wachstum der 1000-jährigen Eiche. Erklären Sie die Auswirkungen der Entscheidung für linear oder exponentiell auf Ihre Berechnungen! **(2 Punkte)**

Herodot hatte eine Schulterhöhe von 195cm , eine Breite von 90cm sowie eine Länge von 220cm .

3. Berechnen Sie das effektive Volumen von Herodot in m^3 , welches Herodot in der 1000-jährigen Eiche einnehmen würde! **(1 Punkt)**

Es wurde berichtet, dass sich Herodot in der 1000-jährigen Eiche *mühsam* um die eigene Achse drehen konnte.

4. Berechnen Sie die Dicke der Eichenwand in cm ! Verdeutlichen Sie Ihre Berechnungen an einer aussagekräftigen Skizze für Pferd und Eiche! **(2 Punkte)**
5. Unter einer Dicke der Eichenwand von 25cm bricht die Eiche zusammen. Beantworten Sie die Forschungsfrage! Begründen Sie Ihre Antwort! **(1 Punkt)**

¹Die Quelle der Inspiration für die Aufgabe war eine Fahrt an die Ostsee und folgender Artikel: [Entdecke das erste Nationale Naturmonument Deutschlands - Ivenacker Eichen und Hutewald](#)

107 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Von Töpfen auf Tischen In einem Experiment wollen Sie die Wuchshöhe von 120 Stockrosen bestimmen. Bevor Sie überhaupt mit dem Experiment beginnen können, gibt es aber ein paar Abschätzungen über die Kosten und den Aufwand zu treffen. Zum einen müssen Sie die Stockrosen einpflanzen und müssen dafür Substrat bestellen. Zum anderen müssen Sie die Stockrosen auch bewegen und in ein Gewächshaus platzieren. Die Töpfe für die Keimung haben einen Durchmesser von 9cm und eine Höhe von 7cm. Der Kubikmeterpreis für Torf liegt bei 260 EUR.

1. Skizzieren Sie den Versuchsplan auf *drei* Tischen im Gewächshaus! **(2 Punkte)**
2. Berechnen Sie die benötigte Anzahl an Pflanztöpfen, wenn Sie Randpflanzen mit berücksichtigen wollen! **(1 Punkt)**
3. Welche *Tischfläche* in m^2 gegeben der Anzahl an Pflanztöpfen inklusive Randpflanzen benötigen Sie im Gewächshaus am Anfang der Keimungsphase? **(3 Punkte)**
4. Berechnen Sie die benötigte Menge an Torf in Liter l , die Sie für das Befüllen der Pflanztöpfe benötigen! Gehen Sie von *einem Zylinder* für die Pflanztöpfe aus! **(3 Punkte)**
5. Berechnen Sie die Kosten in EUR für Ihre Torfbestellung! **(1 Punkt)**

108 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Solar- & Biogasanlagen Um die Energiekosten Ihres Betriebes zu senken, wollen Sie eine Solaranlage auf den Schweinestall montieren lassen. Sie messen Ihren Stall und finden folgende Maße wieder. Die vordere Seite des Schweinestall hat eine Höhe h_v von $6m$. Die hintere Seite des Schweinestall hat eine Höhe h_b von $9.5m$. Der Schweinestall hat eine Tiefe t von $12m$ und eine Breite b von $70m$.

1. Skizzieren Sie den Schweinestall auf dem die Solaranlage montiert werden soll! Ergänzen Sie die Angaben für die Höhen h_v , h_b , die Tiefe t und die Breite b des Stalls! **(2 Punkte)**
2. Berechnen Sie die Fläche der schrägen, neuen Solaranlage auf dem Schweinestall! **(3 Punkte)**

Ebenfalls planen Sie eine neue Biogasanlage für Ihren Betrieb. Der neue Methantank hat einen Radius r von $1.5m$. Leider gibt es ein paar bauliche Beschränkungen auf dem Grundstück. Ihr Fundament des zylindrischen Methantanks kann nur ein Gewicht von maximal $12t$ aushalten bevor der Tank wegbricht. Sie rechnen eine Sicherheitstoleranz von 25% ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei $-80^\circ C$ eine Dichte von $220kg/m^3$. Bei $-100^\circ C$ hat Methan eine Dichte von $270kg/m^3$. Sie betreiben Ihre Anlage bei $-88^\circ C$.

3. Extrapolieren Sie die effektive Dichte des Methans in Ihrem Methantank! Welche Annahme haben Sie getroffen? **(1 Punkt)**
4. Berechnen Sie wie viel Kubikmeter m^3 Sie in den Methantank füllen können, bevor das Fundament nachgibt! **(2 Punkte)**
5. Berechnen Sie die maximale Höhe h_{max} in m für den gefüllten Methantank mit dem Radius r , bevor das Fundament wegbricht! **(2 Punkte)**

109 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Stichworte: Riesenfaultier • Evolution der Avocado • Bluetooth • Blauzahn • Colonia Dignidad • ODESSA • Rattenlinie • Adolf Eichmann

Aligatorenbirnen und Blaubeeren “Sind Sie ein Riesenfaultier oder warum kaufen Sie so viele Aligatorenbirnen?”, spricht es hinter Ihnen. Irritiert drehen Sie sich um und blicken in das puderrote Gesicht von einem Mädchen mit Zöpfen und Zahnspange. “Wieso?”, entfährt es Ihnen und Sie bereuen sogleich die Frage. Sofort werden Sie zu einem Whiteboard voller roter Schnüre geschliffen und müssen folgenden mathematischen untermauerten Argumenten im Rewe über sich ergehen lassen. Da kommen Sie nicht mehr raus, also können Sie auch gleich mitmachen. Das Problem liegt in Chile². Tja, die Deutschen und Südamerika.

Zuerst werden Ihre Fähigkeiten getestet, der Mathematik folgen zu können. Oder berechnen Sie gerade den Einkauf von einem Mädchen mit Zöpfen und Zahnspange?

1. Wenn 7 Blaubeerschalen 13.93 Euro kosten, wie viel kosten 9 Schalen? **(1 Punkt)**
2. Wenn Sie die 9 Blaubeerschalen gekauft haben, wie viele Aligatorbirnen zu je 2.59 EUR können Sie sich dann noch für 100 EUR leisten? **(1 Punkt)**

Das Whiteboard beinhaltet folgende Liste mit Informationen zum Wasserverbrauch bei der Produktion von Gemüse aus Chile. Seltsam, was man so alles in einem Rewe über Gemüse erfährt.

- Ein Kilo Strauchtomaten benötigt 160l Wasser. Eine Strauchtomate wiegt 100 - 120g.
- Ein Kilo Salat benötigt 130l Wasser. Ein Salatkopf wiegt 310 - 490g.
- Ein Kilo Avocado benötigt 1100l Wasser. Eine Avocado wiegt 130 - 410g.
- Ein Kilo Blaubeeren benötigt 830l Wasser. Eine Blaubeere wiegt 3.1 - 3.7g.

3. Berechnen Sie den Wasserverbrauch für die Produktion für jeweils eine Strauchtomate, einem Salat, einer Avocado und einer Blaubeeren. Stellen Sie das Ergebnis als Tabelle dar! **(2 Punkte)**

Chile exportiert im großem Ausmaß Blaubeeren und Avocados. In dem Exportjahr 2021 blieben die Erträge von Blaubeeren mit 7.7×10^4 t in dem prognostizierten Rahmen. Die Menge steigerte sich um 7.2%. Die Exporte für Avocados stiegen in dem gleichen Zeitraum um 17.3% auf 1.9×10^5 t.

4. Wie viele Kubikmeter Wasser hat Chile in dem Exportjahr 2020 exportiert? **(2 Punkte)**

Chile ist eines der wenigen Länder der Welt, die ihr Wasser komplett privatisiert haben. Derzeit sind nur drei Prozent des Wassers des Landes für den häuslichen Verbrauch vorgesehen. In den Dörfern der Anbauregionen versorgen Tankwagen die Bevölkerung jede Woche mit Wasser, es gibt etwa 6l Wasser pro Kopf für den täglichen Bedarf. In *Deutschland* liegt der Verbrauch bei 8 - 17 Liter pro Spülmaschinenlauf und 10 - 15 Liter pro Minute Duschen.

5. Mit der rationierten Wassermenge aus Chiles Anbaugebieten können Sie in *Deutschland* wie oft Ihren Bedarf stillen? **(1 Punkt)**

Das alles hätten Sie nicht von einem Mädchen mit Zöpfen und Zahnspange erwartet. Ganz schön viele Informationen wurden da zusammengetragen.

6. Nennen Sie eine *Datenquelle* im Internet, wo Sie mehr Informationen zu landwirtschaftlichen Daten oder klimatischen, wirtschaftlichen und gesellschaftlichen Daten erhalten! **(1 Punkt)**

²Die Quelle der Inspiration für die Aufgabe waren folgende Reportagen: “Bis zum letzten Tropfen” in AMNESTY – Magazin der Menschenrechte vom August 2021 und “Wasserknappheit in Chile: Eine Folge der Privatisierung?” in Die Welternährung dem Fachjournal der Welthungerhilfe vom April 2022.

110 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Stichworte: Kardaschow-Skala • Dyson-Sphäre • Hohlerde • Entropie • Proton $r_p = 1.7 \times 10^{-15}$ • Wasserstoff $r_H = 5.3 \times 10^{-11}$

Die Dampfnudelerde “Was für einen Unsinn!“, rufen Sie. Jetzt haben Sie kostbaren Schlaf prokrastiniert um einem Ernährungswissenschaftler auf YouTube über die Erde als Dampfnudel zu lauschen. Irgendwie passt es dann doch mit der Analogie. Übermüdet müssen Sie darüber nachdenken, warum vor 65 Millionen Jahren die Dinosaurier - so groß sie auch waren - nicht von der Schwerkraft zu Boden gerissen wurden. Hat der Dampfplauderer etwa recht und war die Schwerkraft vor Millionen von Jahren eine andere? Sind deshalb alle Lebewesen auf der Erde *heutzutage* so viel kleiner, weil die Schwerkraft größer ist als damals? War die Erde kleiner und hatte weniger Masse? Oder ist es nur ein Rechenfehler wie bei der Theorie der Hohlerde von Edmond Halley aus dem 17.-18. Jahrhundert? Müde reiben Sie sich die Augen. So wird es nichts mehr mit dem Schlafen, dann können Sie auch mal etwas rechnen³.

Betrachten wir die Schwerkraft oder Gewichtskraft, die auf Lebewesen damals und heute gewirkt haben soll. Nehmen Sie für die Fallbeschleunigung g der Erde *heutzutage* einen Wert von 9.78 m/s^2 an. Im Weiteren hat die Erde einen ungefähren Durchmesser von $1.1956 \times 10^4 \text{ km}$ und eine mittlere Dichte ρ von 5.86 g/cm^3 . Das Gewicht von einem heute lebenden Waldelefanten mit 2.7t liegt bei 6t und das Gewicht von einem Tyrannosaurus rex (T. rex) bei 4.5t bis 8t.

1. Welchen Durchmesser müsste die Erde vor 65 Millionen Jahren gehabt haben, wenn Dinosaurier und Elefanten die gleiche Gewichtskraft \vec{F}_G damals und heute erfahren hätten? *Beantworten Sie die Frage anhand der folgenden Teilaufgaben!*
 - a) Berechnen Sie die Fallbeschleunigung von vor 65 Millionen Jahren unter der obigen Annahme gleich wirkender Gewichtskraft \vec{F}_G auf Elefant und Dinosaurier! **(1 Punkt)**
 - b) Berechnen Sie Masse der heutigen Erde! **(2 Punkte)**
 - c) Schließen Sie über die Masse auf den Durchmesser der Erde vor 65 Millionen Jahren! **(2 Punkte)**
2. Beantworten Sie die Eingangsfrage mit 1-2 Antwortsätzen! **(1 Punkt)**

Die Distanz zwischen Sonne und Erde entspricht 1.03 astronomische Einheiten (AE). Die Einheit 1 AE wird mit $1.5 \times 10^8 \text{ km}$ angegeben. Der *massebehaftete* Sonnenwind besteht aus 85% Wasserstoffkernen mit einer molaren Masse von 1.05g/mol, 12% Heliumkernen mit 4.11g/mol sowie 3% weiteren Atomkernen mit 89.32g/mol. Die Teilchendichte bei Eintritt in die Erdatmosphäre liegt zwischen 0.4 bis 100 Teilchen cm^{-3} pro Sekunde mit einer mittleren Teilchendichte von 5 cm^{-3} pro Sekunde.

Lösen Sie den folgenden Aufgabenteil mit einer aussagekräftigen Skizze!

4. Berechnen Sie die Anzahl an massebehafteten Teilchen des Sonnenwindes, die die gesamte Erde pro Sekunde treffen! **(2 Punkte)**
5. Berechnen Sie die Anzahl an massebehafteten Teilchen des Sonnenwindes, die die Sonne pro Sekunde in alle Richtungen aussendet! **(2 Punkte)**
6. Berechnen Sie die Masse, die die Erde pro Jahr durch die *massebehafteten* Teilchen des Sonnenwind zunimmt! **(2 Punkte)**

³Die Quelle der Inspiration für die Aufgabe war folgender Artikel: "Skeptische Anmerkungen — Die Erde als Dampfnudel" in Der Humanistische Pressedienst

111 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Entschuldigung, ist das Ihre Feder in meinem Auge? So hört man häufiger höfliche Gänse in Mastställen sagen. Das ist natürlich etwas ungünstig, den dann kommt es zu Picken und Kannibalismus. Denn wenn der Nachbar nervt, dann muss zu Maßnahmen gegriffen werden. Kennt jeder aus einer mittelmäßigen Wohngemeinschaft. Das wollen wir aber als vorsorgliche Gänse-Halter:innen nicht⁴. Betrachten wir also einmal das Platzangebot (eng. *space allowance*, abk. *SA*) der Gänse für vier Tätigkeiten und versuchen die notwendige Fläche zu optimieren. Wie immer gibt es dafür eine mathematische Formel:

$$SA = \sum_{i=1}^n (A_i \times PB_i) \quad A_i = \pi \times (r_i + R_i)^2$$

mit

- *SA* dem benötigten Platzangebot aller aufsummierten Verhalten *i*.
- *A_i* dem benötigten Platz für ein Verhalten *i*.
- *PB_i* dem Anteil des Auftretens eines Verhaltens *i*.
- *r_i* dem Radius Gans plus dem benötigten Radius für das Verhalten *i*.
- *R_i* dem notwendigen Abstand zu den Nachbarn für das Verhalten *i*.
- *i* dem Verhalten: (1) dustbathing, (2) walking, (3) foraging incl. scratching und (4) wingflapping.

In der folgenden Tabelle 1 sind die Werte für *r_i*, *R_i* und *PB_i* für ein spezifisches Verhalten *i* aus drei wissenschaftlichen Veröffentlichungen dargestellt.

	Aldridge et al. (2021)	Baxter et al. (2022)	Jacobs et al. (2019)
dustbathing	46cm; 28cm; 4.1%	40cm; 32cm; 5.2%	42cm; 29cm; 1.8%
walking	37cm; 24cm; 0.4%	34cm; 20cm; 0.4%	36cm; 28cm; 0.9%
foraging incl. scratching	34cm; 24cm; 1.2%	45cm; 21cm; 0.5%	33cm; 36cm; 7.6%
wingflapping	37cm; 17cm; 0.8%	37cm; 22cm; 0.8%	41cm; 23cm; 0.1%

1. Erstellen Sie eine zusammenfassende Tabelle mit den mittleren Werten für *r*, *R* und *PB* aus der obigen Tabelle 1 für die jeweiligen Verhalten! **(3 Punkte)**
2. Ergänzen Sie eine Spalte mit dem benötigten Platz *A* für das jeweilige Verhalten, welches sich aus den mittleren Werten ergibt! **(1 Punkt)**
3. Berechnen Sie das benötigte Platzangebot *SA* für alle betrachteten Verhalten! **(1 Punkt)**
4. Skizzieren Sie die Werte *r_i*, *R_i* und *A_i* für zwei nebeneinander agierende Gänse für ein Verhalten *i*. Nutzen Sie hierfür vereinfachte geometrische Formen! **(2 Punkte)**
5. Sie entnehmen der Literatur folgende Aussage zur Verteilung der Gänse in der Fläche *A*: „Assuming, that the animals will optimally and equally distribute in an area *A*, we observe a small part, which is not covered. This area is called ω and is calculated with $\omega = \frac{A}{0.9069}$.“ Veranschaulichen Sie die Fläche ω in einer aussagekräftigen Abbildung! **(1 Punkt)**
6. Ein Tier braucht Platz für sich selbst. Berechnen Sie nun die Körperfläche *a*, die ein Tier einnimmt. Welche Annahmen haben Sie für die Berechnung der Körperfläche getroffen? **(2 Punkte)**

⁴Die Quelle der Inspiration für die Aufgabe war der folgende wissenschaftliche Artikel: [EFSA Panel on Animal Health and Welfare, et al. \(2023\) Welfare of broilers on farm. EFSA Journal 21.2.](#)

112 Aufgabe

(8 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Nelken von den Molukken In der Ausstellung „Europa und das Meer“ im Deutschen Historischen Museum in Berlin gab es folgendes Zitat über die Probleme der frühen Hochseeschifffahrt.

»Ohne ausreichende Zufuhr von Vitamin C stellen sich nach 45 Tagen die ersten Symptome ein; die ersten Toten sind nach 72 Tagen zu beklagen; nach 100 Tagen rafft die Skorbut eine ganze Schiffsbesatzung dahin«

Ferdinand Magellan stach im Jahre 1519 in See um eine Passage durch den südamerikanischen Kontinent zu finden. Zu seiner Flotte gehörten fünf Schiffe - das Flaggschiff Trinidad, die San Antonio, die Victoria, die Concepción und die Santiago - mit einer Besatzung von insgesamt 237 Mann.

1. Stellen Sie den Verlauf der Anzahl an Matrosen auf einem Schiff der Flotte in der Form einer Überlebenszeitkurve dar! Beschriften Sie die Achsen entsprechend! **(2 Punkte)**
2. Schätzen Sie die Überlebenswahrscheinlichkeit nach 95 Tagen aus Ihrer Abbildung ab! **(1 Punkt)**

Der Chronist an Bord der Trinidad, Antonio Pigafetta, schrieb in seinem Bericht „[...] Um nicht Hungers zu sterben, aßen wir das Leder, mit dem die große Rahe zum Schutz der Taue umwunden war.“ Insbesondere die Mannschaft der Concepción erlitt große Verluste durch die Skorbut bei der Überquerung des Pazifiks, da durch Erkundungsfahrten weniger Zeit blieb, um wilden Sellerie aufzunehmen. Wilder Sellerie enthält $7000\mu\text{g}/150\text{mg}$ Vitamin C. Der Bedarf liegt bei 120mg pro Tag für Männer.

3. Berechnen Sie die notwendige Menge in kg an aufzunehmenden wilden Sellerie auf die Concepción für die ununterbrochene Fahrt von drei Monate und 24 Tage über den Pazifik! **(3 Punkte)**
4. Skizzieren Sie die Überlebenszeitkurve für die Concepción im Vergleich zu der Überlebenszeitkurve der Trinidad! Beschriften Sie die Achsen! **(2 Punkte)**

113 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Event Horizon – Am Rande des Universums Die Sonne hat eine aktuelle, angenommene Masse von $2 \times 10^{30} \text{ kg}$. Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 5000m kollabiert, wird die Sonne 30% der aktuellen Masse verloren haben. Ein Lichtteilchen mit der Masse m_f und der Fluchtgeschwindigkeit v_f will dem schwarzen Loch entkommen. Sie haben folgende Formeln für die kinetische Energie des Lichtteilchens E_{kin} und der Gravitationsenergie des schwarzen Lochs E_{grav} gegeben⁵.

$$E_{kin} = \frac{1}{2} m_f v_f^2 \quad E_{grav} = \frac{G m_s m_f}{r_s}$$

mit

- m_f , gleich der Masse [kg] des fliehenden Objektes
- m_s , gleich der Masse [kg] des stationären Objekts
- r_s , gleich dem Radius [m] des stationären Objekts
- G , gleich der Gravitationskonstante mit $5.974 \cdot 10^{-11} \text{ m}^3 (\text{kg} \cdot \text{s}^2)^{-1}$

Im Folgenden wollen wir uns mit der Frage beschäftigen, ob das Lichtteilchen der Gravitation des schwarzen Lochs entkommen kann.

1. Geben Sie die Formel für die Fluchtgeschwindigkeit v_f an! **(1 Punkt)**
2. Überprüfen Sie Ihre umgestellte Formel nach v_f anhand der Einheiten! **(1 Punkt)**
3. Berechnen Sie die notwendige Fluchtgeschwindigkeit v_f des Lichtteilchens mit den angegebenen Informationen! **(2 Punkte)**
4. Gehen Sie von einer Lichtgeschwindigkeit von $2.9 \times 10^8 \text{ m/s}$ aus. Kann das Lichtteilchen der Gravitation des schwarzen Lochs entkommen? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Stellen Sie den Zusammenhang zwischen dem sich verringernden Radius r des schwarzen Lochs bei gleichbleibender Masse m_s und der notwendigen Fluchtgeschwindigkeit v_f in einer Abbildung dar! **(2 Punkte)**
6. Ein Elefant und ein Lolli stürzen aus großer und gleicher Höhe in ein schwarzes Loch. Welches der beiden Objekte überschreitet zuerst den Ereignishorizont des schwarzen Loches? Begründen Sie Ihre Antwort mathematisch! **(2 Punkte)**

⁵Die Quelle der Inspiration für die Aufgabe war ein Montagnachtfilm: [Event Horizon – Am Rande des Universums](#)

114 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Das Fermi Paradoxon Der Kernphysiker Enrico Fermi diskutierte 1950 auf dem Weg zum Mittagessen im Los Alamos National Laboratory mit seinen Kollegen angebliche UFO-Sichtungen und fragte schließlich: „Where is everybody?“. Warum seien weder Raumschiffe anderer Weltraumbewohner noch andere Spuren extraterrestrischer Technik zu beobachten? Wie lange würde eine außerirdische Zivilisation benötigen um die gesamte Milchstraße zu besuchen, wenn das maximale Reisetempo die Geschwindigkeit der Voyager 1 Sonde wäre?⁶

Wir treffen folgende Annahmen. Eine außerirdische Zivilisation schickt vier Voyager 1 ähnliche Sonden mit der Geschwindigkeit von $6.0523 \times 10^4 \text{ km/h}$ los um sich auf den erreichten Planeten selbst zu replizieren. Nach 750 Jahren ist die Replikation abgeschlossen und wiederum vier Sonden werden ausgesendet. Gehen Sie von 6.23 Lichtjahren als mittlerer Abstand der Sterne in der Milchstraße aus. Es gibt 10^{11} Sterne in der Milchstraße. Nehmen Sie eine Lichtgeschwindigkeit von $2.7 \times 10^8 \text{ m/s}$ an.

1. Skizzieren Sie in einer Abbildung die ersten drei Schritte der Vervielfältigung der Sonden in der Galaxie! Beschriften Sie die Abbildung mit der Dauer und der Anzahl an Sonden für jeden Schritt der Vervielfältigung! **(4 Punkte)**
2. Berechnen Sie die theoretische Anzahl an Vervielfältigungsschritten die benötigt werden um mit *einem einzigen Vervielfältigungsschritt* die gesamten Sterne der Milchstraße mit Sonden zu besuchen! **(2 Punkte)**
3. Berechnen Sie die Dauer, die eine außerirdische Zivilisation annäherungsweise benötigt um die gesamten Sterne der Milchstraße mit Sonden zu besuchen! **(2 Punkte)**
4. Bei einem vermuteten Alter der Erde von 4.6×10^9 Jahren, wie oft war dann eine Sonde einer außerirdischen Zivilisation schon zu Besuch? Korrigieren Sie Ihre Antwort mit dem Wissen, dass sich die Kontinentalplatten einmal alle 10^8 Jahre vollständig im Erdinneren umgewandelt haben! **(2 Punkte)**

⁶Die Quelle der Inspiration für die Aufgabe war folgender Wikipediaeintrag: [Fermi-Paradoxon](#)

115 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Pyramiden bauen Es stehen die mecklemburgischen Pyramidentage an und Sie sind auf abenteuerlichen Wegen für den Bau der Pyramiden zuständig. Zu allem Überfluss handelt es sich auch noch eine *Reenactment* Veranstaltung. Thema der diesjährigen Pyramidentage sind die Pyramiden von Meroe, die den Königen und Königinnen des historischen Reiches von Kusch in Nubien, dem heutigen Sudan, als Grabstätten dienten. Die Pyramiden in Meroe fallen durch ihren steilen Winkel von 71 Grad im Vergleich zu den ägyptischen Pyramiden mit 60 Grad auf. Die durchschnittliche Seitenlänge der Grundfläche einer Pyramide beträgt 32 Königsellen. Eine Königselle misst 52.4cm.

Lösen Sie diese Aufgabe mit Hilfe einer Skizze der Pyramide. Bezeichnen Sie Seiten und die Winkel der Pyramide entsprechend!

1. Bei der Königspyramide von Meroe soll eine Seitenlänge der Grundfläche 32 Königsellen lang sein. Welche Höhe der Königspyramide in m ergibt sich? **(1 Punkt)**
2. Die Außenflächen der Pyramide soll begrünt werden. Für die Bepflanzung muss eine 5cm dicke Torfschicht auf die Pyramide aufgebracht werden. Berechnen Sie die ungefähre Menge an benötigten Torf in m^3 ! **(2 Punkte)**

Wie in jedem guten *Reenactment* gibt es viel Oberschicht, aber nur 3 Sklaven, die Ihnen bei dem Befüllen der Pyramide mit Schutt zu Seite stehen. Leider haben Ihre Sklaven zu allem Überfluss auch noch chronische Rückenschmerzen entwickelt, als sie von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 80% aus. In eine Schubkarre passen 110 Liter.

3. Wie oft müssen Ihre maladen Sklaven die Rampe mit der Schubkarre zur Spitze der Pyramide hochfahren um die Pyramide mit Schutt zu füllen? **(1 Punkt)**
4. Berechnen Sie die Länge der Rampe zur Spitze der Pyramide mit einem Anstellwinkel von 9° ! **(2 Punkte)**
5. Wie weit reicht Ihre Rampe vom Fuß der Pyramide in die mecklemburgische Landschaft? **(2 Punkte)**

Bei der Besichtigung der Pyramide teilt Ihnen der leicht übergewichtige Pharaο (Nebenberuf *Versicherungsverteter*) mit, das die Pyramide zu flach sei und somit nicht in die mecklemburgische Landschaft passen würde. Sie müssen nochmal ran.

6. Die Grundfläche der Pyramide ändert sich nicht. Berechnen Sie die Änderung der Höhe in Königsellen, wenn sich der Anstellwinkel der Pyramide um 5° ändert! **(2 Punkte)**

116 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Geocaching – Von Satelliten und Plastikdosen Es ist Wochenende und das Wetter ist *sweet*. Sie schwingen sich auf Ihr Cachermobil um mit 18km/h, geleitet von modernster Satellitentechnologie und einem Supercomputer aus dem Jahr 2000 in Ihren Händen, Plastikdosen in der Natur und an sehenswerten Orten zu finden. Sie wollen diesmal endlich die abwärts Terrainchallenge durchführen. Die Reihenfolge der Caches nach Terrainwertung gibt daher die von Ihnen abzufahrenden Orte vor. Die Terrain- und Schwierigkeitswertungen laufen von 1 (leichteste Wertung) bis 5 (schwierigste Wertung) in 0.5 Schritten. Folgende Informationen zu den Orten und den entsprechenden Caches stehen Ihnen für Ihre Planung der Route zu Verfügung⁷.

Ort	Cache	Wertung (S T G)
A	GCRPYVI	2.5 5.0 Klein
B	GCI9W0Y	2.0 1.5 Klein
C	GCBLVDG	3.5 4.0 Mikro
D	GCVDPZT	3.0 2.0 Klein
E	GCFEQVJ	1.5 3.5 Klein

Im Weiteren sind Ihnen folgende Informationen zu den Entfernungen der Orte zugänglich. Der Entfernungsvektor \overrightarrow{AB} ist 4km. Im Weiteren ist Ihnen der Entfernungsvektor \overrightarrow{CB} mit 6.5km bekannt. Der Entfernungsvektor \overrightarrow{BE} ist das 1.3-fache des Entfernungsvektor \overrightarrow{CB} . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr 25° südlich. Wenn Sie von dem Ort C den Ort B anpeilen, so liegt der Ort D ungefähr 60° östlich. Vom Ort B betrachtet, bilden die Orte C und D einen rechten Winkel am Ort B. Der Ort B liegt auf gerader Linie zwischen den Orten C und E. Somit liegt der Ort E südlich von B. Die Strecke zwischen A und E ist nicht passierbar. Sie starten an dem Ort D Ihre Cachertour.

1. Lösen Sie diese Aufgabe mit Hilfe einer aussagekräftigen Skizze der Orte und Caches. Bezeichnen Sie die Strecken und die Winkel Ihrer Skizze entsprechend! **(2 Punkte)**
2. Welche Strecke in *km* legen Sie bei der Bewältigung der abwärts Terrainchallenge zurück? **(5 Punkte)**
3. Gehen Sie von einer zusätzlichen Suchzeit in Stunden für die Caches an den jeweiligen Orten zur reinen Reisezeit mit Ihrem Cachermobil aus. Die Suchzeit in Stunden für *jeden einzelnen* Cache wird durch die Funktion

$$\text{Suchzeit} = 0.05 + 0.18 \cdot \text{Schwierigkeit}$$

beschreiben. Wie lange in Stunden benötigen Sie um die abwärts Terrainchallenge zu erfüllen? **(3 Punkte)**

4. An der höchsten Schwierigkeit müssen Sie angeln. Ihre Angel ist ausgefahren 5m lang. Erreichen Sie einen Cache in der Höhe von 6.8m? Berechnen Sie dazu Ihre maximale mögliche Angelhöhe! Welche Annahmen mussten Sie treffen um die Aufgabe zu lösen? **(2 Punkte)**

⁷Die Quelle der Inspiration für die Aufgabe war folgende Tätigkeit: [Geocaching – Mach mit bei der weltweit größten Schatzsuche](#).

117 Aufgabe

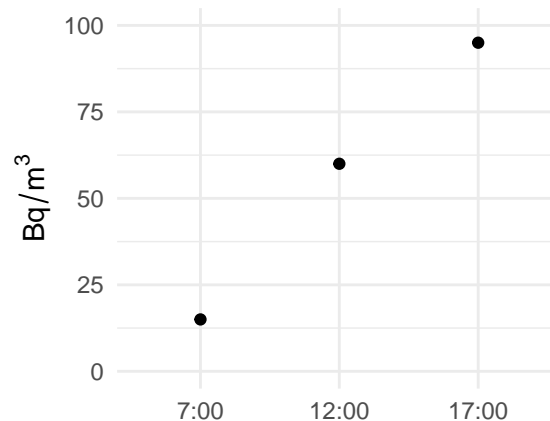
(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Stichworte: Brot aus Luft • Walöl • Haber-Bosch-Verfahren • 1. Weltkrieg • 40% N im menschlichen Körper • Positivist

Die atmende Wand und Brot aus Luft Als Kellerkind vom Dorf wollen das Ausmaß der Radonbelastung in ihrem Kellerzimmer bestimmen und lüften daher nicht. Spart dann auch Energie und lüften wird sowieso überschätzt. Während einer Messperiode von 7:00 Uhr bis 17:00 bestimmen Sie dreimal automatisch die Radonbelastung in Ihrem Kellerraum in Bq/m^3 . Es ergibt sich folgende Abbildung⁸.



1. Wie lange dauert es in Stunden bis Sie eine kritische Belastung von $400 Bq/m^3$ in Ihrem ungelüfteten Kellerraum erreicht haben? **(2 Punkte)**

Radon zerfällt mit einer Halbwertszeit von 3.5d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 140d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

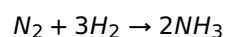
2. Wie lange dauert es in Stunden bis Ihre kritische Radonbelastung von $400 Bq/m^3$ auf unter $90 Bq/m^3$ gefallen ist? **(4 Punkte)**

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	78.1	28.4	
Sauerstoff	21.3	16.5	
Kohlenstoffdioxid	0.035	12.5	

3. Rechnen Sie die Volumenprozent (Vol-%) der Umgebungsluft in die entsprechenden ppm-Werte um und ergänzen Sie die berechneten ppm-Werte in die Tabelle! **(1 Punkt)**

Während Sie Ihr etwas pappiges Toastbrot mampfen kommt Ihnen die Dokumentation über Brot aus Luft in den Sinn. Sie denken darüber ein wenig nach. Für die Umwandlung von Stickstoff N_2 mit Wasserstoff H_2 zu Ammoniak NH_3 gilt folgende Reaktionsgleichung⁹:



Ein Mol eines beliebigen Gases hat bei normalen Umweltbedingungen ein Volumen von 22.4 Liter.

4. Welche Masse an Ammoniak in Kilogramm kg können Sie aus einem Kubikmeter m^3 Luft unter normalen Umweltbedingungen gewinnen? **(2 Punkte)**
5. Wieviel Ammoniak in mol erhalten Sie aus einem Kubikmeter Luft? **(1 Punkt)**

⁸Die Quelle der Inspiration für die Aufgabe war folgender Artikel: [Atmende Wand](#)

⁹Die Quelle der Inspiration für die Aufgabe war folgender Artikel: [Haber-Bosch-Verfahren – Brot aus Luft](#)

118 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Armee der Finsternis Ihr Studentenjob war nach Ladenschluss bei Penny die Regale einzuräumen. Dabei ist Ihnen in der Auslage der Sonderangebote das Necronomicon¹⁰ in die Hände gefallen. Nun sind Sie eine Magierin der Zeichen geworden! Also eigentlich können Sie nur Mathe und das dämliche Necronomicon hat Sie in die Vergangenheit geschleudert... aber gut, was tut man nicht alles im Jahr 558 n. Chr. für den neuen Lehnsherren Henry dem Roten. Sie bauen natürlich einen Schrottkugelturm um sich den Horden der Finsternis mit genug Schrott erwehren zu können! Ihnen stehen zwei mächtige magische Formeln zur Unterstützung zu Verfügung.

$$E_{kin} = \frac{1}{2} \cdot m \cdot v^2 \quad E_{pot} = m \cdot g \cdot h$$

mit

- m , gleich der Masse [kg] des Objekts
- h , gleich der Höhe [m] des ruhenden Objekts
- v , gleich der Geschwindigkeit [m/s] des Objekts
- g , gleich der Erdbeschleunigung mit $9.81 \frac{m}{s^2}$

Als erstes müssen Sie die Höhe des zu bauenden Schrottkugelturmes bestimmen. Hierfür ist wichtig zu wissen, dass sich die Bleitropfen mit einem Gewicht von $20mg$ zu gleichförmigen Bleikugeln bei einer Geschwindigkeit von $14m/s$ bilden.

1. Wie hoch müssen Sie den Schrottkugelturm bauen lassen, damit sich runde Bleikugeln durch die Fallgeschwindigkeit von $14m/s$ bilden? **(3 Punkte)**

Ihre erstellten Schrottkugeln sind leider zu groß und somit sind zu wenige Schrottkugeln in einer Ladung. Damit können Sie die Armee der Finsternis nicht aufhalten. Die Sachlage müssen Sie einmal mathematisch untersuchen.

2. Nennen Sie die beiden geometrischen Formen aus denen sich näherungsweise ein Tropfen zusammensetzt! Erstellen Sie eine beschriftete Skizze des Tropfens! **(2 Punkte)**
3. Sie messen eine Länge des Tropfens von $3.2mm$. Die Löcher im Sieb erlauben ein Tropfendurchmesser von $1.7mm$. Welchen Durchmesser in mm haben Ihre produzierten Bleikugeln? **(3 Punkte)**

Sie haben jetzt die 6.1×10^4 Bleikugeln zusammen. Blei hat eine Dichte von $15.1g/cm^3$.

4. Wie schwer in Kilogramm kg sind die 6.1×10^4 produzierten Bleikugeln, die Sie jetzt auf die Burgmauer transportieren müssen? **(1 Punkt)**

Am Ende müssen Sie noch die Produktion von dem Bleischrott im Turm optimieren.

5. Wie groß in cm^2 ist Ihr quadratisches Sieb am oberen Ende des Turms, wenn Sie pro Fall ca. 1100 Bleikugeln produzieren wollen und die Bleikugel im Fall $0.5cm$ Abstand haben müssen? **(1 Punkt)**

¹⁰Ein wirklich gefährliches Buch ist: *Du bist genug: Vom Mut, glücklich zu sein* von Fumitake Koga und Ichiro Kishimi

119 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Armee der Kaninchen Leider hat es mit Ihrer Faultierpension in Down Under nicht geklappt. War vielleicht auch nicht so die beste Idee... aber dafür haben Sie eine Neue! Oder wie es Mike Tyson zugeschrieben wird: „Ich wurde nie niedergeschlagen, ich war immer am Aufstehen!“. Daher machen Sie jetzt einen Großhandel mit Kaninchenfleisch und damit dem teuersten Fleisch in Australien auf. Moment, hopsen hier nicht, seit Thomas Austin im Jahr 1860 ungefähr 26 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Dem wollen wir mal mathematisch nachgehen!¹¹

Forscherinnen fand folgende Sättigungsfunktion für das jährliche Wachstum der gesamten Kaninchenpopulation im westlichen Australien.

$$f(t) = 9 \times 10^9 - 1.4 \times 10^9 \cdot 2^{-0.15 \cdot t + 3.1}$$

1. Skizzieren Sie die Sättigungsfunktion *annäherungsweise* in einer Abbildung! **(1 Punkt)**
2. Wie viele Kaninchen können nach der Sättigungsfunktion maximal im westlichen Australien leben? Ergänzen Sie den Wert in Ihrer Abbildung! **(2 Punkte)**
3. Wie viele Millionen Kaninchen leben nach der Sättigungsfunktion nach 6 Jahren auf dem australischen Kontinent? **(1 Punkt)**

Um den Kaninchen Einhalt zu gebieten wurde das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV) in 12 Kaninchen ausgebracht. Da die Kaninchen keine Maßnahmen gegen die Ausbreitung vornehmen können, verläuft die Ausbreitung mit einem wöchentlichen Wachstumsfaktor von 1.8 nach folgender Formel.

$$N(t) = N(0) \cdot a^t$$

3. Wie viele Wochen benötigen die Viren um theoretisch die gesamte Kaninchenpopulation nach 4 Jahren Wachstum zu durchseuchen? **(1 Punkt)**

Das Myxoma Virus und das RHDV töteten 98.5% der Kaninchenpopulation innerhalb weniger Wochen.

4. Wie lange in Jahren dauert es bis eine Kaninchenpopulation nach einer Viruspandemie wieder auf 70% der gesättigten Kaninchenpopulation angewachsen ist? **(2 Punkte)**

Thomas Austin entließ die Kaninchen im äußersten Norden von Australien. Australien hat eine West-Ost-Ausdehnung von 4200km und eine Nord-Süd-Ausdehnung von knapp 3700km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 8.8km pro Jahr aus.

5. Wie lange dauert es in Jahren bis die Kaninchen jeden Ort in Australien erreicht haben? *Lösen Sie die Aufgabe unter der Verwendung einer schematischen Skizze!* **(2 Punkte)**

Eine jährliche Impfung gegen das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV) kosten 10\$ pro Tier und der durchführende Arzt verlangt ca. 40\$ pro Tier.

6. In Ihrem Stall leben 800 Mastkaninchen. Mit welchen jährlichen Zusatzkosten für die Impfungen der Kaninchen müssen Sie daher kalkulieren? **(1 Punkt)**

¹¹Die Quelle der Inspiration für die Aufgabe war der folgendes YouTube Video: [Incredible Stories – Why don't they eat wild rabbits in Australia? They have millions of them! The reason is surprising...](#)

120 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Lüneburger Heide. Unendliche Weiten. Wir schreiben das Jahr 2024. Dies sind die Abenteuer der Kuh Fridolin und Ihnen. Grünes Gras unter Ihren Füßen und ein strammer Wind im Gesicht, egal wohin Sie schauen. Ein schmatzendes Geräusch ertönt unter Ihnen. Sie sinnieren, sollten Sie Ihre weiten Graslandschaften jetzt schon düngen? Dafür benötigen Sie die *Grünlandtemperatur*! Die Grünlandtemperatur (GLT) ist die Summe aller positiven Tagesmitteltemperaturen seit Jahresbeginn. Ab einer GLT von 200° kann mit der Stickstoffdüngung begonnen werden. Sie sehen nicht ein, Geld für einen Agrarmetrologen zu bezahlen. Also rechnen Sie mit folgenden Informationen zu Monatsmultiplikatoren des GLT-Wertes: Januar mit $0.6\times$, Februar mit $0.7\times$ und März mit $1.1\times$. Sie haben noch im letzten Jahr folgende Temperaturen gemessen.

Datum	C°
01. Jan 2023	0.3
01. Feb 2023	1.2
01. Mrz 2023	3.5
01. Apr 2023	4.3

1. Erstellen Sie eine Skizze aus den Informationen aus der Temperaturtabelle! **(1 Punkt)**
2. Stellen Sie die linearen Funktionen $f_1(t)$, $f_2(t)$ und $f_3(t)$ aus der obigen Temperaturtabelle auf! **(1 Punkt)**
3. Bestimmen Sie die Stammfunktionen $F_1(t)$, $F_2(t)$ und $F_3(t)$ für Ihre linearen Funktionen aus der obigen Temperaturtabelle! **(1 Punkt)**
4. Osterglocken beginnen ab einer GLT von 210°C zu blühen. An welchem Tag im 1. Quartal des Jahres 2023 war dies der Fall? *Ignorieren Sie ein eventuelles Schaltjahr in Ihrer Berechnung.* **(4 Punkte)**

Auf dem Weg zu Ihrer Jonagoldplantage wurden Sie mit Ihrem Trecker von einer Gruppe elektrifizierter Rentner abgedrängt. Der Trecker muss wieder aus dem Graben! Fridolin und die elektrifizierten Rentner ziehen an zwei, separaten Seilen. Dabei zieht Fridolin mit 230N . Die elektrifizierten Rentner bringen eine Kraft von 140N auf.

Lösen Sie diese Aufgabe mit Hilfe einer aussagekräftigen Skizze der Kraftvektoren. Bezeichnen Sie die Kraftvektoren und die Winkel Ihrer Skizze entsprechend!

5. Im ersten Versuch legen Sie das Seil für Fridolin lotrecht über einen Ast oberhalb des Treckers. Die Rentner ziehen in einer geraden Linie über die Böschung hinweg am anderen Seil. Welche Kraft wird aufgebracht? **(2 Punkte)**
6. Im zweiten Versuch ziehen Fridolin und die Rentner mit einem 50° Winkel mit ihrem Seil an dem Trecker. Welche Kraft wird aufgebracht? **(2 Punkte)**
7. Mit welcher Beschleunigung ziehen Sie den 1.2t schweren Trecker *jeweils* aus dem Graben, wenn $F = m \cdot a$ gilt? **(1 Punkt)**

121 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



In der Kartonagenfabrik Wenn man sich zu spät anmeldet, dann ist die Exkursion nicht so toll. Also geht es mit Rektor Skinner und Mrs. Krabappel in die Kartonagenfabrik. Wie schon im vorherigen Semester... In der Kartonagenfabrik angekommen erfahren Sie, dass die Kartons zum Versand von Nägeln nicht hier zusammengebaut werden sondern das sich die Endfertigung in Flint, Michigan befindet. Unter anderem wird dort der berühmte *Doppelt gewellte, 8-mal-gefaltete, 0.7mm, 60-cm-Karton* durch Falzung hergestellt. Beim letzten Mal war Rektor Skinner die Stimmung zu schlecht und deshalb geht es erst nach Hause, wenn ein paar Aufgaben gelöst sind. Martin gefällt das. An dem Vorrat an Zigaretten von Mrs. Krabappel meinen Sie wenig Zuversicht zu erkennen.

Jetzt heißt es Kartons optimieren. Der nun zu optimierende, flache Karton hat eine Länge von 60cm und eine Breite von 20cm. Die Kartonagenmaschine in Flint soll dann einen quadratischen Eckenausschnitt der Länge x falzen.

1. Erstellen Sie eine Skizze des Karton**blatt**rohlings! Beschriften Sie die Skizze mit den entsprechenden Längenangaben **(1 Punkt)**
2. Berechnen Sie die Falztiefe x für ein maximales Volumen des flachen Kartons! **(3 Punkte)**
3. Welches Volumen in Liter ergibt sich mit der von Ihnen berechneten Falztiefe x ? **(1 Punkt)**
4. Sie wollen noch einen bündig mit dem Boden abschließenden Deckel für den Karton stanzen lassen. Wie groß ist die Fläche des Kartondeckel**blatt**rohlings in cm^2 ? **(2 Punkte)**

Rektor Skinner möchte sich gerne wieder in seinem Vorgarten aufhalten und nicht die ganze Zeit von Bart mit Erdnüssen beworfen werden. Deshalb möchte er einen geräumigen Teil seines Vorgartens einzäunen. Ein Teil der Umzäunung bildet seine Vorderhauswand. Wegen Lieferschwierigkeiten stehen Rektor Skinner nur 100m Zaun zu Verfügung. Sie wollen nun die maximale Fläche des abgeschirmten Vorgartens in Abhängigkeit der Seitenlängen bei der Verwendung von 100m Zaun bestimmen!

5. Welche Seitenlängen für den Zaun ergeben sich für die maximale Fläche des abgeschirmten Vorgartens? **(2 Punkte)**
6. Berechnen Sie die Fläche des abgeschirmten Vorgartens! **(1 Punkt)**

122 Aufgabe

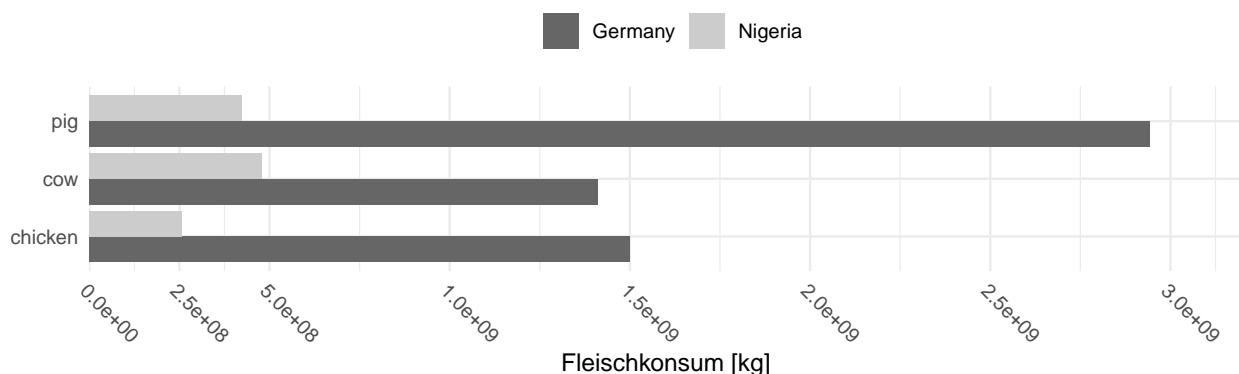
(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



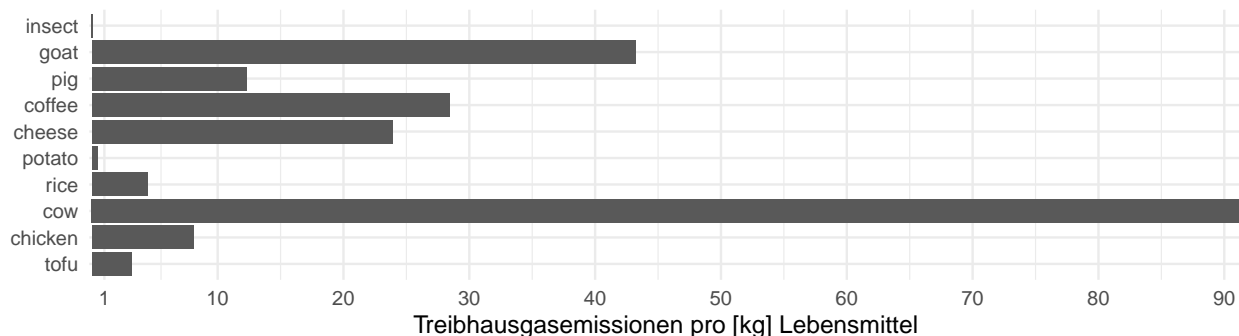
Ein Pfund Insekten, bitte! Das wird wohl häufiger gehört werden, wenn wir die Menschheit mit Proteinen ausreichend ernähren wollen¹². Schauen wir uns dazu einmal den Vergleich Deutschland zu Nigeria an. Nigeria hat eine der am schnellsten wachsenden Bevölkerungen der Welt und wird vermutlich im Jahr 2100 zu den Top 5 der bevölkerungsreichsten Länder zählen. Im Jahr 2021 leben ca. 8.2×10^7 Menschen in Deutschland und ca. 1.84×10^8 Menschen in Nigeria. Mit den Informationen wollen wir anfangen und dann eine Prognose für den Fleischkonsum im Jahr 2050 zu treffen.

Im folgenden ist Abbildung des Fleischkonsums im Jahr 2021 in Deutschland und Nigeria in [kg] einmal dargestellt.



1. Stellen Sie den Fleischkonsum in Deutschland und Nigeria im Jahr 2021 *pro Kopf* in einer aussagekräftigen Tabelle dar! **(2 Punkte)**
2. Ergänzen Sie in der Tabelle eine Spalte in der Sie für den Fleischkonsum in Nigeria auf Deutschland normieren, daher ins Verhältnis Nigeria/Deutschland, setzen! **(1 Punkt)**

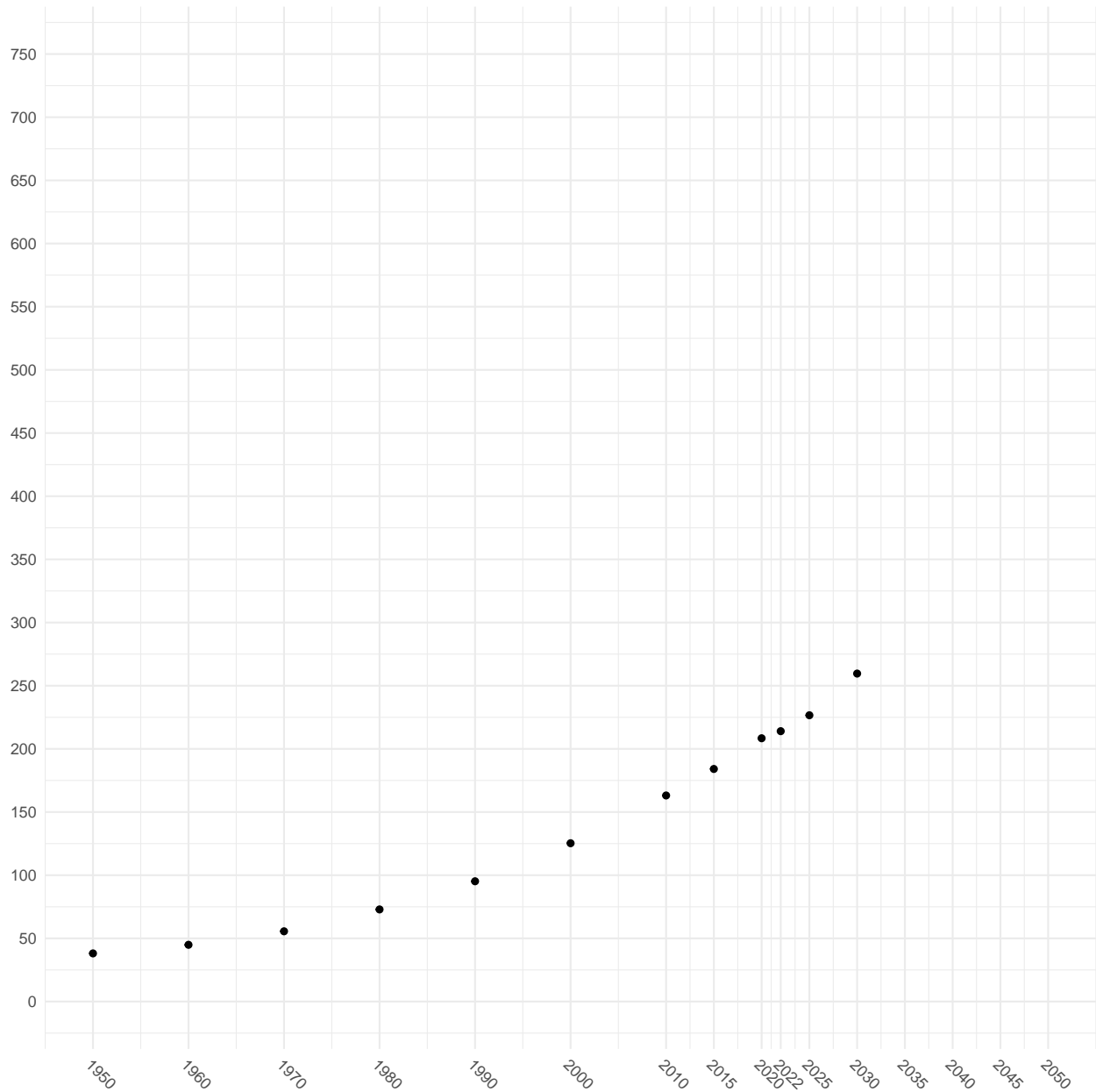
In der nächsten Abbildung finden Sie die CO₂-Emission in [kg] nach Lebensmittel, die durch die Produktion entsteht, abgebildet.



3. Stellen Sie in einer Tabelle die Treibhausgasemissionen an CO₂ pro Kopf, die durch den Fleischkonsum in Deutschland und Nigeria im Jahr 2021 entstehen, dar! Ergänzen Sie auch hier das Verhältnis Nigeria zu Deutschland! **(2 Punkte)**

¹²Die Quelle der Inspiration für die Aufgabe war der folgende Artikel aus dem Spiegel: [Acht Milliarden - sind wir bald zu viele Menschen auf der Erde?](#)

In der folgenden Abbildung sehen Sie die Bevölkerungsentwicklung [Millionen] in Nigeria von 1950 bis ins Jahr 2030 fortgeführt.



4. Schätzen Sie graphisch die zu erwartende Bevölkerung [Millionen] in Nigeria im Jahr 2050, die sich anhand der Informationen aus der Abbildung ergibt!
 - a) Ohne Berücksichtigung der Covid-19-Pandemie! **(1 Punkt)**
 - b) Unter Berücksichtigung der Covid-19-Pandemie! **(1 Punkt)**
5. Berechnen Sie den geschätzten Fleischkonsum von Nigeria im Jahr 2050 unter der Annahme 60%-iger Angleichung der Lebensbedingungen zu Deutschland im Jahr 2021! **(1 Punkt)**
6. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an CO_2 in Nigeria im Jahr 2050 im Vergleich zum Jahr 2021, der sich durch den angeglichenen Fleischkonsum ergibt! **(1 Punkt)**
7. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an CO_2 in Nigeria, wenn die gesamte Proteinaufnahme durch Insekten ersetzt würde! **(1 Punkt)**

123 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Tödliche Seuche AIDS – Die rätselhafte Krankheit Irritiert legen Sie die historische Ausgabe des Spiegels aus den 80zigern beiseite. Sie sind bei Ihrem Augenarzt und wollen einen AIDS-Test machen lassen. Woanders leider keinen Termin gekriegt... Immerhin denken Sie und Ihre Partnerin über Nachwuchs nach und da geht es eben nur durch ungeschützten Sex. Was wissen Sie nun aber über AIDS und dem diagnostischen AIDS-Test, den Sie nun machen werden?

Die Prävalenz von AIDS bei einem Menschen in Europa wird mit 0.5% angenommen. In 96% der Fälle ist ein HIV-Test positiv, wenn der Patient erkrankt ist. In 1% der Fälle ist ein HIV-Test positiv, wenn der Patient *nicht* erkrankt ist und somit gesund ist. Sie stutzen. Wie wahrscheinlich ist es denn eigentlich an AIDS erkrankt zu sein (K^+), wenn Sie einen positiven AIDS-Test vorliegen haben (T^+)? Gehen Sie für die folgenden Berechnungen von $n = 2 \times 10^4$ Patienten mit einem diagnostischen Test für AIDS aus. Sie nehmen sich also einen Kuli und fangen an auf der historischen Ausgabe des Spiegels zu rechnen¹³.

1. Welche Wahrscheinlichkeit Pr wollen Sie berechnen? **(1 Punkt)**
2. Zeichnen Sie einen Häufigkeitsdoppelbaum zur Bestimmung der gesuchten Wahrscheinlichkeit Pr ! **(2 Punkte)**
3. Beschriften Sie den Häufigkeitsdoppelbaum, mit denen Ihnen bekannten Informationen zu der AIDS Erkrankung und dem AIDS-Test! **(1 Punkt)**
4. Füllen Sie den Häufigkeitsdoppelbaum mit den sich ergebenden, absoluten Patientenzahlen n aus! **(2 Punkte)**
5. Berechnen Sie die gesuchte Wahrscheinlichkeit Pr ! **(1 Punkt)**

Bei dem folgenden Arztgespräch erfahren Sie, dass beim diagnostischen Testen *True Positives (TP)*, *True Negatives (TN)*, *False Positives (FP)* und *False Negatives (FN)* auftreten. Das verstehen Sie so noch nicht und deshalb stellen Sie für sich den Zusammenhang in einer 2x2 Kreuztabelle dar.

6. Tragen Sie TP , TN , FP und FN in eine 2x2 Kreuztabelle ein. Beschriften Sie die Tabelle entsprechend! **(1 Punkt)**
7. Berechnen Sie die Sensitivität und Spezifität des diagnostischen Tests für AIDS! Füllen Sie dafür die 2x2 Kreuztabelle mit den Informationen aus dem Häufigkeitsdoppelbaum aus! **(2 Punkte)**
8. Was beschreibt die Sensitivität und die Spezifität im Bezug auf die Gesunden und Kranken? Stellen Sie beide diagnostische Maßzahlen als Wahrscheinlichkeiten Pr dar! **(2 Punkte)**

¹³Die Quelle der Inspiration für die Aufgabe war der folgende wissenschaftlicher Artikel: Binder et al. (2022) Von Baumdiagrammen über Doppelbäume zu Häufigkeitsnetzen – kognitive Überlastung oder didaktische Unterstützung? Journal für Mathematik-Didaktik, 1-33

124 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Network-Marketing oder Schneeballschlacht! Eine Möglichkeit, leicht Geld zu verdienen, ist es anderen Menschen für Geld zu versprechen, wie man leicht reich werden kann. Am besten natürlich ohne viel Aufwand und ortsunabhängig. Dann wollen wir mal loslegen. Schnell ein YouTube-Werbevideo gedreht und auf geht es mit unserem Network-Marketing. Aber Moment, wie funktioniert Network-Marketing eigentlich und was hat das alles mit einer Schneeballschlacht zu tun? Wir wollen hier einmal in die Untiefen des „passiven Einkommens“ abtauchen¹⁴.

Das Jahr 2022 war das erfolgreichste Jahr in der Geschichte von KH Gesund und Schön Components (KH-GSC). Das Unternehmen steigerte den Umsatz um rund 24 Prozent von 290 Millionen Euro im Jahr 2021. Doch wie viel kommt bei den Partnern an? Laut KH-GSC habe das Unternehmen 3.6×10^5 aktive Partner.

1. Berechnen Sie zuerst den Umsatz der Firma KH-GSC im Jahr 2022! **(1 Punkt)**
2. Wie viel von dem Umsatz im Jahr 2022 wird im Durchschnitt von jedem aktiven Partner erwirtschaftet? **(1 Punkt)**
3. Welche *monatlicher* Umsatz ergibt sich dadurch im Durchschnitt für jeden aktiven Partner bei einer direkten Provision von 25%? **(1 Punkt)**

Ihr zu vermarkendes Produkt, hinter dem Sie voll stehen, kostet 50EUR pro Einheit im Direktverkauf. Die direkte Provision für die erste Stufe beträgt 20%. Für die zweite, dritte und vierte Stufe betragen die indirekten Provisionen jeweils 3.25%, 2.25% und 1.75%. Jeder Ihrer angeworbenen „Partner“ wirbt wiederum vier Partner für sich selbst an. Pro Monat werden im Schnitt zwei Einheiten vom Produkt verkauft. Sie wollen nun 4100EUR im Monat *passiv* – also durch indirekte Provisionen – erwirtschaften.

4. Ergänzen Sie die folgende Tabelle mit den obigen Informationen! **(2 Punkte)**

Stufe	Anzahl Partner	Umsatz/Stufe	Provision
1	Sie selber		
2			
3			
4			

5. Wie viele Partner müssen Sie auf der 2 Stufe anwerben um Ihr passives Einkommen durch indirekte Provision zu erreichen? Wie viele Menschen arbeiten am Ende indirekt für Sie? Stellen Sie den Zusammenhang graphisch dar! **(3 Punkte)**

Sie mussten zum Einstieg bei KH-GSC Einheiten des Produkts für 1500EUR kaufen. Diese Einheiten können Sie nur direkt verkaufen. Leider mussten Sie den Kauf über einen Kredit über 5.5% p.a. über 36 Monate finanzieren.

6. Berechnen Sie die Gesamtsumme, die Sie als Kredit abbezahlen müssen! **(2 Punkte)**
7. Wie viele Einheiten müssen Sie pro Monat verkaufen um die anfallenden Zinsen durch die direkte Provision zu erwirtschaften? **(1 Punkt)**
8. Wie lange in Monaten benötigen Sie um den Kredit durch die direkte Provision abzubezahlen? **(1 Punkt)**

¹⁴Die Quellen der Inspiration für die Aufgabe waren folgendes YouTube Video: [Simplicissimus – Die meistgesuchte Betrügerin der Welt](#) und der Artikel: [Deutschlandfunk Kultur – Die Illusion, schnell reich zu werden](#)

125 Aufgabe

(10 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Höhlen & Drachen Nachdem Sie sich begeistert in der Serie *Stranger Thinks* verloren haben, wollen Sie bei einem Ihrer Freunde einmal *Höhlen & Drachen* ausprobieren. Um Geld zu sparen, das Zeug kostet echt, wurde etwas an den Regeln gebastelt. Schnell stellen Sie fest, dass hier ganz schön viele unterschiedliche Würfel durch die Gegend fliegen. Daher müssen Sie sich jetzt einiges an Fragen stellen.

In dem Spiel haben Sie nun auf einmal 6 zwölfseitige Würfel ($6d12$) zum würfeln in der Hand. Wenn Sie eine 12 würfeln, haben Sie einen Erfolg.

1. Berechnen Sie die Wahrscheinlichkeit *genau* 4 Erfolge zu erzielen! **(2 Punkte)**
2. Berechnen Sie die Wahrscheinlichkeit keinen Erfolg zu erzielen! **(1 Punkt)**

Sie betrachten nun aufmerksam die ausufernden Ausrüstungstabellen. Ihnen wird aber geholfen und Sie müssen sich jetzt nur zwischen der Axt oder dem Schwert entscheiden.

3. Würden Sie die Axt mit zwei achtseitigen Würfeln ($2d8$) als Schaden oder das Schwert mit einem vierseitigen Würfel plus 2 ($1d4+2$) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

Jetzt wird es immer wilder, da Sie sich jetzt überlegen müssen, wie wahrscheinlich es ist, dass Ihr Rettungswurf gegen den zaubernden Hexer funktioniert. Sie haben folgende Wahrscheinlichkeiten gegeben. Die Wahrscheinlichkeit für das Ereignis A , der Rettungswurf ist erfolgreich, ist $Pr(A) = 0.6$, die Wahrscheinlichkeit für das Ereignis B , der Zauberwurf des Hexers ist erfolgreich, ist $Pr(B) = 0.9$. Sie haben mitgezählt und festgestellt, dass in 40 von 100 Fällen Ihr Rettungswurf bei einem erfolgreichen Zauber funktioniert hat.

4. Erstellen Sie eine 2×2 Kreuztabelle mit den Ereignissen A und B sowie den Gegenereignissen \bar{A} und \bar{B} mit einem $\Omega = 100$. Beachten Sie hierbei die entsprechenden Wahrscheinlichkeiten für die Ereignisse A und B ! **(2 Punkte)**
5. Bestimmen Sie $Pr(A \cap B)$! **(1 Punkt)**
6. Erstellen Sie ein Baumdiagramm mit den passenden Informationen aus der 2×2 Kreuztabelle! **(2 Punkte)**
7. Bestimmen Sie Wahrscheinlichkeit $Pr(A|B)$, dass Ihr Rettungswurf gelingt, wenn der Hexer erfolgreich gezaubert hat! **(1 Punkt)**

126 Aufgabe

(12 Punkte)

Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Retrocheck im TV „Und hier ist sie wieder, die Show der fantastischen Preise. Seien Sie mit dabei, wenn es wieder heißt: Der Preis ist heiß!“, ertönt es und Sie fragen sich, ob Sie nicht doch lieber bezahlter Gast bei Barbara Salesch hätten sein sollten. Aber Sie brauchen das Geld und jetzt heißt es Spielschows farmen! Erstmal eine Kaffeemaschine von Mitropa gewinnen. Ein Kandidat gewinnt die Kaffeemaschine von Mitropa, wenn nicht alle Kandidaten überbieten (eng. *outbid*). Mit Ihnen bilden Elke und Helmut das Team der drei Kandidaten.

Name	$P(\text{win})$	$P(\text{outbid})$
Elke	0.2	0.076
Helmut	0.4	0.02

1. Mit welcher Wahrscheinlichkeit gewinnen Sie die Kaffeemaschine von Mitropa, wenn keiner der Kandidaten überbietet? **(1 Punkt)**
2. Wenn Ihre Überbietungswahrscheinlichkeit $P(\text{outbid})$ bei 0.08 liegt, mit welcher Wahrscheinlichkeit gewinnt *keiner* die Kaffeemaschine von Mitropa? **(1 Punkt)**

Glücksrad für Arme auf der Kirmes! Leider hat es für Maren Gilzer nicht gereicht. Deshalb sind Sie jetzt auf der Kirmes und spielen mit Catwoman um das große Geld. Das Glücksrad hat 24 Felder. Sie drehen das Glücksrad zweimal. Auf 10 Feldern gewinnen Sie 4000EUR sonst 1000EUR. Ganz schön viel Geld und ganz schön zwielichtig hier...

3. Skizzieren Sie das Glücksrad und ergänzen Sie die Wahrscheinlichkeiten! **(1 Punkt)**
4. Zeichnen Sie das zugehörige Baumdiagramm für das zweimalige Drehen! Ergänzen Sie die Wahrscheinlichkeiten und die entsprechenden Ereignisse **(2 Punkte)**
5. Mit welcher Wahrscheinlichkeit gewinnen Sie 5000EUR? **(1 Punkt)**

Nach Ihrem Fiebertraum reisen Sie im Zug nach Köln um bei „Geh aufs Ganze!“ mitzuspielen. Sie schaffen es tatsächlich ins Finale und können als Hauptgewinn ein Auto hinter einer der drei Türen gewinnen.

6. Bevor die Show beginnt, wird das Auto hinter eine zufällig bestimmte Tür gestellt. Mit welcher Wahrscheinlichkeit wird jeweils eine der drei Türen ausgewählt? Zeichnen Sie ein Baumdiagramm! **(1 Punkt)**
7. Mit welcher Wahrscheinlichkeit wählen Sie sofort die Tür mit dem Auto? Erweitern Sie das Baumdiagramm entsprechend! **(1 Punkt)**
8. Der Moderator öffnet nun eine der nicht gewählten Türen, aber natürlich nicht die mit dem Auto. Mit welcher Wahrscheinlichkeit steht das Auto hinter der anderen Tür? Erweitern Sie das Baumdiagramm entsprechend! **(2 Punkte)**
9. Lösen Sie nun das „Ziegenproblem“! Berechnen Sie dazu die Wahrscheinlichkeiten der einzelnen Pfade. Lohnt sich ein Wechsel der anfangs gewählte Tür? Begründen Sie Ihre Antwort mathematisch! **(2 Punkte)**

Angewandte Nutztier- und Pflanzenwissenschaften (M.Sc.)

Im Rahmen der Klausur zu dem Modul Biostatistik werden auch Fragen nur für die Studierenden des Schwerpunktes Nutztierwissenschaften gestellt. Im Folgenden daher eine lose Sammlung von möglichen Fragen zu diesem Themenkomplex.

127 Aufgabe

(6 Punkte)

Vergleichen Sie die Standardabweichung mit dem Standardfehler und grenzen Sie die beiden Kennzahlen voneinander ab.

128 Aufgabe

(8 Punkte)

Ihnen liegt folgendes Varianzanalysemodell mit der üblichen Beschreibung zur Auswertung des Merkmals fett- und eiweißkorrigierte Milchleistung pro Kuh und Jahr in kg vor:

$$Y_{ijkl} = \mu + Var_i + EKA_j + VarEKA_{ij} + V_k + b(L_{ij} - L) + e_{ijkl}$$

mit

- Y_{ijkl} : l-te Beobachtung
- μ : Populationsmittel
- Var_i : fixer Effekt der i-ten Variante (i: Kontrolle, Versuchsgruppe 1, Versuchsgruppe 2)
- EKA_j : fixer Effekt der j-ten Erstkalbealtergruppe (j: $EKA \leq 25$ Monate, $EKA > 25$ Monate)
- $VarEKA_{ij}$: fixer Effekt der Interaktion Variante x Erstkalbealtergruppe
- V_k : zufälliger Effekt des Vaters
- $b(L_{ij} - L)$: lineare Kovariable Laktationsnummer
- e_{ijkl} : zufälliger Restfehler

Erläutern Sie anhand dieses Beispiels die Begriffe fixer Effekt, Interaktion, zufälliger Effekt und Kovariable und grenzen Sie diese Begriffe voneinander ab.

129 Aufgabe

(6 Punkte)

Wie bestimmen Sie die richtige Stichprobengröße? Welche Kennzahlen / statistische Maßzahlen benötigen Sie dabei und nennen Sie die Voraussetzungen.