

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 she or he knows,  
but how much the student wants to know.“*  
— Alice W. Rollins



Alex studiert im 3. Semester und wiederholt das Modul, da er im ersten Jahr andere Prioritäten für sich gesetzt hat. Das musste sein, da er sich ziemlich im Abitur verausgabt hat. Darüber hinaus war die WG auch eher auf Party angelegt. Alex hofft jetzt über Pünktlichkeit wieder in die Bahn zu kommen. Dafür steht er jetzt immer um 5 Uhr auf! Freunde von ihm beschreiben ihn eher als extrovertiert. Er kennt Paula noch aus der Schulzeit und er überlegt, ob nicht beide Mal nach Mallorca sollten.

Nach zwei Semestern Studium an der Universität Osnabrück war es dann Jessica doch viel zu theoretisch. Etwas angewandtes sollte es sein, wo sie auch eine Fähigkeit lernt, die Frau nutzen kann. Deshalb hat sich Jessica an der Hochschule eingeschrieben. Hoffentlich lernt sie etwas nützliches, wo andere für Geld geben würden. Wer nützlich ist, ist wertvoll. Ihr Traum ist ja eine Hundeschule aufzumachen. Die großen Parties hat sie immer gemieden. Sie ist lieber mit ihrer Hündin im Teutoburgerwald.



Das ist jetzt der letzte Versuch mit einem Studium. Wenn es nicht klappt dann überlegt Jonas das **Programm der IHK zu Ausbildungsvermittlung** zu nutzen. Aber eine Runde gibt er sich noch. Struktur ist eigentlich das Wichtigste und diesmal hat er sich alle Altklausuren der Fachschaft besorgt. Dann ist er auch noch gleich der Fachschaft beigetreten um mehr Informationen abzugreifen. Und er versucht nicht seine Zeit mit Alex zu verdaddeln oder in der Fachschaft bei einem Bier oder so...

Nächstes Semester geht es nach Kanada davon hat er schon auf der Berufsschule geträumt. Deshalb konzentriert er sich sehr auf die Prüfungen. Ein Schiff ist im Hafen sicher, aber dafür ist es nicht gebaut worden. Das **International Faculty Office** der Fakultät Agrarwissenschaften und Landschaftsarchitektur hat super geholfen, aber es waren einiges an Unterlagen. Jetzt hofft er, dass Tina dann doch noch mitkommt. Aber sonst macht er das eben alleine. Obwohl das eher nicht so seine Art ist. Vielleicht sollte er sich mal einen Tipp bei Tina holen, sie wirkt sehr entschlossen.



Nach der Ausbildung wollte Nilufar eigentlich gleich anfangen zu arbeiten, aber nach einem Praktikum und der Probezeit stellte sie fest, dass es ohne einen Hochschulabschluss schwer wird Führungsverantwortung zu übernehmen. Mit Menschen kann sie schon immer und dann auch eigene Projekte mit anderen verwirklichen, das ist doch was. Mit dem notwendigen Abschluss sollte der Start um so einfacher sein. Dann ist sie keine Befehlsempfängerin mehr sondern gibt die Marschrichtung vor. Schon jetzt koordiniert Nilufar das Studium von anderen.

Paula möchte die Welt zu einem besseren Ort machen. Wenn da nicht die anderen Mitmenschen wären. Paula muss das Modul nochmal wiederholen, da es dann am Ende des Semesters zu viel für sie wurde. Eine Lerngruppe hätte geholfen, aber das ist dann gar nicht so einfach eine zu finden. Zwar kennt sie schon Nilufar, aber Nilufar ist ihr manchmal zu forsch. Ihr schwant aber, dass alleine das Studium sehr schwer werden wird. Das Abitur war schon so ein Lernhorror, das möchte sie nicht nochmal. Alex sieht sie da als Vorbild.



Sommer, Sonne, Natur. Das ist es was Steffen mag. Raus in die Komune und die Natur genießen. Leider hat Steffen noch andere Bedürfnisse, die ein Einkommen benötigen. Da Studierende mehr verdienen, würde dann in Teilzeit auch mehr rausspringen. Wenn er dann privat was anbauen kann, dann spart er gleich doppelt. Leider sind viele seiner Kommilitonen total verkrampte Karrieristen. Es geht nur ums Äußere. Dabei verliert sich Steffen gerne im Prozess. Das hat auch seinen Schulabschluss etwas verzögert. Steffen lässt sich eben Zeit.

Wille war es, die es Tina hierher gebracht hat und Wille wird es sein, die Tina dann auch zum Abschluß treibt. Nach einem Rückschlag muss Tina jetzt einige Module wiederholen, damit sie dann auch fertig wird. Ab und zu ist sie schwach gewesen und das hat dann Zeit gekostet. Das Tina es dann manchmal übertreibt, weiß sie nur zu gut, aber irgendwie muss die Kontrolle ja erhalten bleiben? Insbesondere, wenn sie mal wieder die Nacht durchgefeiert hat, verachtet Tina sich. Dann baut Nilufar sie dann bei einem Tee wieder auf.



Für Yuki war es nicht einfach. Teilweise war die Krankheit sehr hinderlich, dann war es Yuki selber. Dann muss man auch wieder auf die Beine kommen und es dauert eben seine Zeit. Aber immerhin hat Yuki es jetzt den Abschluss gekriegt und hat einen Studienplatz. Jetzt heißt es in den Rhythmus kommen und schauen, was noch so passiert. Immerhin hat Yuki schon eine kleine Gruppe gefunden, in der Yuki dann Hilfe findet. Ist aber auch sehr unübersichtlich so ein Studium. Steffen ist immer super entspannt.

### 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 69 Punkten sind aus den Rechen- und Textaufgaben erreicht.  
\_\_\_\_\_ von 89 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
85.0 - 89.0	1,0
80.5 - 84.5	1,3
76.5 - 80.0	1,7
72.0 - 76.0	2,0
67.5 - 71.5	2,3
63.0 - 67.0	2,7
58.5 - 62.5	3,0
54.5 - 58.0	3,3
50.0 - 54.0	3,7
44.5 - 49.5	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	✓
<b>Aufgabe 1</b>						
<b>Aufgabe 2</b>						
<b>Aufgabe 3</b>						
<b>Aufgabe 4</b>						
<b>Aufgabe 5</b>						
<b>Aufgabe 6</b>						
<b>Aufgabe 7</b>						
<b>Aufgabe 8</b>						
<b>Aufgabe 9</b>						
<b>Aufgabe 10</b>						

- Es sind \_\_\_\_ von 20 Punkten erreicht worden.

## Rechen- und Textaufgaben

<b>Aufgabe</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>
<b>Punkte</b>	10	10	10	10	9	10	10

- Es sind \_\_\_\_ von 69 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 Textvarianten. Die Semeantik mag sich unterscheiden, die Inhalte sind aber gleich.

### ANOVA

#### 1. Aufgabe

(2 Punkte)

Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein  $\eta^2 = 0.52$ . Welche Aussage ist richtig?

- A** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird. Das  $\eta^2$  ist damit mit dem  $R^2$  aus der linearen Regression zu vergleichen.
- B** ☐ Die Berechnung von  $\eta^2$  ist ein Wert für die Interaktion.
- C** ☐ Das  $\eta^2$  ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.
- D** ☐ Das  $\eta^2$  ist ein Wert für die Güte der ANOVA. Je kleiner desto besser. Ein  $\eta^2$  von 0 bedeutet ein perfektes Modell mit keiner Abweichung. Die Varianz ist null.
- E** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen nicht erklärt wird. Somit der Rest an nicht erklärbarer Varianz.

#### 2. Aufgabe

(2 Punkte)

Sie führen ein Feldexperiment durch um das Gewicht von Maiss zu steigern. Die Pflanzen wachsen unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen. Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein  $\eta^2 = 0.22$ . Welche Aussage ist richtig?

- A** ☐ Es werden 22% der Varianz durch den Versuch erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der durch Fehler in der Versuchsdurchführung entsteht.
- B** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Umweltbedingungen erklärt wird. Daher werden 22% der Varianz durch die Umweltbedingungen erklärt. Der Anteil der Varianz durch die Behandlungsgruppen ist dann 78%.
- C** ☐ Es werden 78% der Varianz durch die Behandlung erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den unterschiedlichen Behandlungsbedingungen nicht erklärt wird.
- D** ☐ Mit dem  $\eta^2$  lässt sich auf die Qualität der Randomisierung und damit der Strukturgleichheit zwischen der Grundgesamtheit und der Stichprobe schließen. Es gilt dabei die Regel, dass ein  $\eta^2$ -Wert von 1 zu bevorzugen ist.
- E** ☐ Es werden 22% der Varianz durch die Behandlung erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den unterschiedlichen Behandlungsbedingungen erklärt wird.

#### 3. Aufgabe

(2 Punkte)

Die einfaktorielle ANOVA ist ein Standardverfahren in der agrawissenschaftlichen Forschung wenn es um den Vergleich von Behandlungsgruppen geht. Welche der folgenden Aussage zu der Berechnung der Teststatistik der einfaktoriellen ANOVA ist richtig?

- A** ☐ Die ANOVA berechnet die F-Statistik indem die MS des Fehlers durch die MS der Behandlung geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese 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 kleiner als der kritische Wert ist kann die Nullhypothese nicht abgelehnt werden. Die F-Statistik ist der Quotient der MS der Behandlung durch die MS des Fehlers.
- E** ☐ 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.

#### 4. Aufgabe

(2 Punkte)

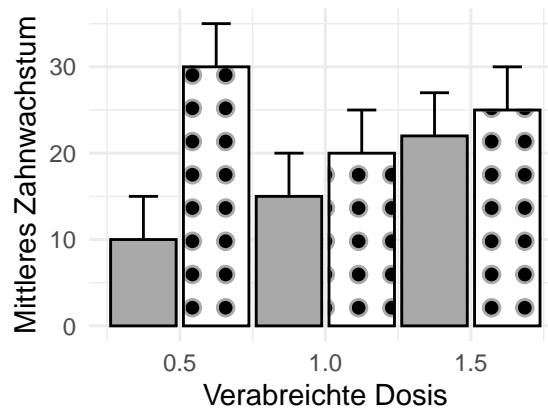
Viele statistische Verfahren nutzen eine Teststatistik um eine Aussage über den Zusammenhang zwischen der Grundgesamtheit und der Stichprobe abzubilden. Ein statistisches Testwerkzeug ist hierbei die ANOVA. Die ANOVA rechnet dabei...

- A** ☐ ... den Unterschied zwischen der F-Statistik anhand der Varianz der Gruppen. Wenn die F-Statistik exakt 0 ist, kann die Nullhypothese abgelehnt werden.
- B** ☐ ... 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.
- C** ☐ ... den Unterschied zwischen zwei paarweisen Mittelwerten aus verschiedenen Behandlungsgruppen. Wenn die signifikant ist, ist daher bekannt welcher Vergleich konkret unterschiedlich ist.
- D** ☐ ... 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.
- 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)

In einer Studie zur Bewertung der Wirkung von Vitamin D auf das Zahnwachstum bei Kanarienvögel entstand folgende Abbildung. Der Versuch wurde an 81 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. Welche Aussage ist im Bezug auf eine zweifaktorielle ANOVA richtig?



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

## Deskriptive Statistik & Explorative Datenanalyse

### 6. Aufgabe

(2 Punkte)

Wie lautet der Mittelwert und Standardabweichung von  $y$  mit 14, 16, 7, 10 und 9.

- A ☐ Sie erhalten 11.2 +/- 1.85
- B ☐ Es ergibt sich 10.2 +/- 6.85
- C ☐ Es berechnet sich 12.2 +/- 13.7
- D ☐ Es berechnet sich 11.2 +/- 13.7
- E ☐ Sie erhalten 11.2 +/- 3.7

### 7. Aufgabe

(2 Punkte)

Berechnen Sie den Median, das 1<sup>st</sup> Quartile sowie das 3<sup>rd</sup> Quartile von  $y$  mit 11, 13, 23, 14, 11, 5, 18, 7, 3, 6 und 51.

- A ☐ Es ergibt sich 11 +/- 6
- B ☐ Sie erhalten 11 [4; 16]
- C ☐ Es berechnet sich 12 [7; 17]
- D ☐ Es ergibt sich 11 [6; 18]
- E ☐ Es ergibt sich 15 +/- 6

### 8. Aufgabe

(2 Punkte)

Die empfohlene Mindestanzahl an Beobachtungen für die Visualisierung mit einem Boxplot sind...

- A ☐ Wir brauchen fünf oder mehr Beobachtungen.
- B ☐ 1 Beobachtung.
- C ☐ Die untere Grenze liegt bei einer Beobachtung.
- D ☐ Die optimale Anzahl ist größer als hundert Beobachtungen, wobei es gerne sehr viel mehr sein können.
- E ☐ 2-5 Beobachtungen.

### 9. Aufgabe


(2 Punkte)

Um die Standardabweichung zu berechnen müssen wir folgende Rechenoperationen durchführen.

- A ☐ Wir berechnen erst den Mittelwert und dann die absoluten Abstände zu dem Mittelwert. Diese quadratischen Abstände summieren wir auf und teilen am Ende durch die Fallzahl.
- B ☐ Den Mittelwert berechnen und die Abstände quadrieren. Die Summe mit der Fallzahl multiplizieren.
- C ☐ Wir berechnen erst den Mittelwert und dann die quadratischen Abstände zu dem Mittelwert. Diese quadratischen Abstände summieren wir auf und teilen am Ende durch die Fallzahl. Als letzten Schritt ziehen wir die quadratische Wurzel.
- D ☐ Den Mittelwert berechnen, dann die absoluten Abstände zum Mittelwert aufsummieren
- E ☐ Als erstes berechnen wir den Mittelwert. Dann bilden wir die Summe der quadratischen Abstände zu dem Mittelwert. Abschließend teilen wir durch die Fallzahl.

## 10. Aufgabe

(2 Punkte)

Nachdem Sie eine ANOVA und die paarweisen t-Tests über das  Paket {emmeans} durchgeführt haben, müssen Sie Ihre Daten nochmal zur Überprüfung visualisieren. Sie entscheiden sich für den Boxplot. Welche statistischen Maßzahlen stellt der Boxplot dar?

- A** ☐ Der Boxplot stellt den Median und die Quartile dar.
- B** ☐ Der Boxplot stellt den Median und die Streuung dar.
- C** ☐ Durch die Abbildung des Boxplot erhalten wir die Informationen über die Mittelwerte und die Varianz.
- D** ☐ Der Boxplot stellt die Mittelwerte und die Standardabweichung dar.
- E** ☐ Den Mittelwert und die Varianz.

## 11. Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit zuLauch finden Sie auf einmal seltsame Daten. Jedenfalls kommt Ihnen das so vor. Daher berechnen Sie den Mittelwert und den Median. Der Mittelwert  $\bar{y}$  und der Median  $\tilde{y}$  unterscheiden sich. Welche Aussage ist richtig?

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

Ihre Betreuung der Abschlussarbeit fragt überraschend in der letzten Besprechung, ob Ihre Messwerte einer Normalverteilung genügen. Sonst könnten Sie ja gar nicht einen t-Test rechnen. Da Ihnen die Zeit wegrennt, entscheiden Sie sich für eine schnelle Visualisierung im Anhang. Welche Visualisierung nutzen Sie und welche Regel kommt zur Abschätzung einer Normalverteilung zur Anwendung?

- A** ☐ In einer explorativen Datenanalyse nutzen wir den Violinplot. Dabei sollte der Bauch am Rand liegen. Dann können wir von einer Normalverteilung ausgehen.
- B** ☐ In einer explorativen Datenanalyse nutzen wir den Boxplot. Dabei sollte der Median als dicke Linie in der Mitte der Box liegen. Dann können wir von einer Normalverteilung ausgehen.
- C** ☐ Nach dem Einlesen der Daten nutzen wir einen Boxplot um zu schauen, ob alle Boxen über alle Behandlungen in etwa gleich groß sind. Damit ist dann auch das IQR in allen Behandlungen in etwa gleich.
- D** ☐ Einen Barplot. Die Mittelwerte müssen alle auf einer Höhe liegen. Die Fehlerbalken haben hier keine Informationen.
- E** ☐ Einen Violinplot. Der Bauch der Violine muss hierbei einen höheren Wert annehmen als der Steg der Violine. Dann kann die Annahme einer Normalverteilung angenommen werden.



### 13. Aufgabe

(2 Punkte)

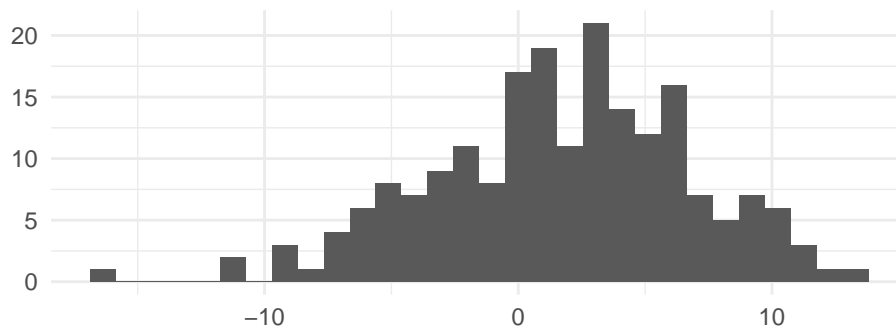
In der Statistik müssen wir häufig überprüfen, ob unser Outcome einer bestimmten Verteilung folgt. Meistens überprüfen wir, ob eine Normalverteilung vorliegt. Folgende drei Abbildungen eignen sich im Besonderen für die Überprüfung einer Verteilungsannahme an eine Variable.

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

### 14. Aufgabe

(2 Punkte)

Sie haben  $n = 194$  Pflanzen geerntet und wollen sich nun die Verteilung der Pflanzen einmal in einem Histogramm anschauen. Welche Verteilung ist dargestellt?



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

## Lineare Regression & Korrelation

### 15. Aufgabe

(2 Punkte)

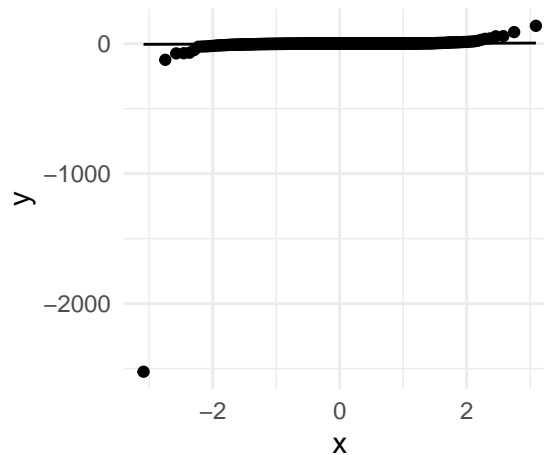
Sie haben das Modell  $Y \sim X$  vorliegen und wollen nun ein kausales Modell rechnen. Welche Aussage ist richtig?

- A ☐ Wenn ein kausales Modell gerechnet werden soll, dann muss zum einen ein Trainingsdatensatz sowie ein Testdatensatz definiert werden. Dabei ist der Trainingsdatensatz meist 1/10 und der Testdatensatz 1/3 der Fallzahl groß. Der Testdatensatz dient zur Validierung.
- B ☐ Wenn ein kausales 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  $X$  auf den gemessenen Endpunkt  $Y$  aus?
- C ☐ Es wird ein Trainingsdatensatz zum Modellieren des Trainingsmodells benötigt. Der Testdatensatz dient rein zur Visualisierung. Dies gilt vor allem für ein kausales Modell.
- D ☐ Wenn ein kausales Modell gerechnet werden soll, dann muss zum einen ein Trainingsdatensatz sowie ein Testdatensatz definiert werden. Dabei ist der Trainingsdatensatz meist 2/3 und der Testdatensatz 1/3 der Fallzahl groß. Der Testdatensatz dient zur Validierung.
- E ☐ Ein kausales Modell schließt grundsätzlich lineare Modell aus. Es muss ein Graph gefunden werden, der alle Punkte beinhaltet. Erst dann kann das  $R^2$  berechnet werden.

## 16. Aufgabe

(2 Punkte)

Sie rechnen in eine linearen Regression und erhalten folgenden QQ Plot um die Annahme der normalverteilten Residuen zu überprüfen. Welche Aussage ist richtig?

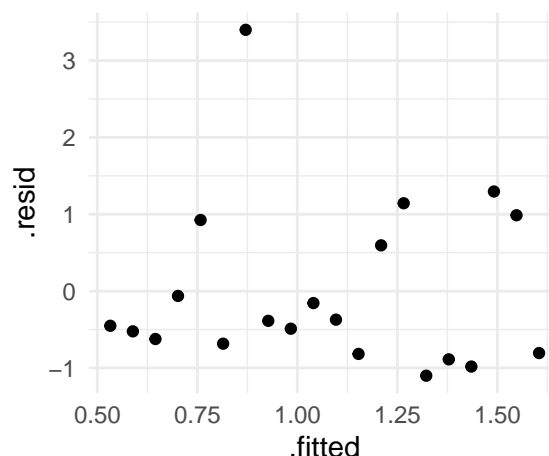


- A** ☐ Wir betrachten die Gerade, die durch die einzelnen Punkte laufen sollte. Wenn die 95% der Punkte von der Geraden getroffen werden, dann gehen wir von normalverteilten Residuen aus.
- B** ☐ Wir betrachten die Gerade und dabei insbesondere die beiden Enden der Gerade in dem IQR, also dem ersten und dritten Quartile. Hier sollten die Punkte auf der Geraden liegen, dann ist die Annahme an die Normalverteilung der Residuen erfüllt.
- C** ☐ Wir betrachten insbesondere die beiden Enden der Gerade. Der Rest ist mehr oder minder egal, dann ist die Annahme an die Normalverteilung der Residuen erfüllt.
- D** ☐ 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.
- E** ☐ Wir betrachten die Gerade. Wenn die Punkte einigermaßen gleichmäßig um die Gerade verteilt liegen, dann gehen wir von normalverteilten Residuen aus. Dies ist hier nicht der Fall. Wir haben keine normalverteilten Residuen vorliegen.

## 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 Punkte müssen gleichmäßig in dem negativen Bereich liegen. Dies ist hier klar nicht der Fall. Einzelne Ausreißer können beobachtet werden. Die Analyse ist gescheitert.

- B** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Vereinzelt Punkte liegen oberhalb bzw. unterhalb der Geraden um die 0 Linie weiter entfernt. Ein klares Muster ist zu erkennen.
- C** ☐ Wenn wir die Nulllinie betrachten so liegen die Punkte nicht gleichmäßig über und unter der Nulllinie. Unser Modell erfüllt nicht die Annahme von normalverteilten Residuen mit einem Mittelwert von 0 und einer Streuung von  $s^2$ .
- D** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Diagonalen. Damit ist das Modell erfolgreich geschätzt worden.
- E** ☐ Wenn die Punkte gleichmäßig in dem positiven wie auch negativen Bereich ohne ein klares Muster liegen, dann hat unsere Modellierung geklappt. Wir können mit dem Modell weitermachen.

## 18. Aufgabe

(2 Punkte)

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

- A** ☐ Korrelationskoeffizienten  $\rho$  liegt zwischen 0 und 1. Darüber hinaus ist der Korrelationskoeffizienten  $\rho$  einheitslos und kann als Standardisierung verstanden werden.
- B** ☐ Der Korrelationskoeffizienten  $\rho$  zeigt keinen Zusammenhang zwischen zwei Variablen  $x$  und  $y$  bei einem Wert von 0. Einen negativen Zusammenhang Richtung -1 und somit auch einen positiven Zusammenhang Richtung 1. Je größer die Zahl allgemein, desto stärker der Effekt.
- C** ☐ Der Korrelationskoeffizienten  $\rho$  liegt zwischen -1 und 1. Darüber hinaus ist der Korrelationskoeffizienten  $\rho$  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  $\rho$  liegt zwischen -1 und 1. Darüber hinaus ist der Korrelationskoeffizienten  $\rho$  einheitslos und kann als standardisierte Steigung verstanden werden.
- E** ☐ Der Korrelationskoeffizienten  $\rho$  wird wie das  $\eta^2$  aus der ANOVA interpretiert. Der Korrelationskoeffizienten  $\rho$  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)




In einer linearen Regression kann es vorkommen, dass der Effekt repräsentiert durch den  $\beta$  Koeffizienten nicht so richtig von der Größenordnung zu dem  $p$ -Wert passen will. So liefert eine Untersuchung des Einflusses von der  $PO_2$ -Konzentration in  $[\mu g]$  im Wasser auf das Wachstum in  $[kg]$  an Spitzkohl folgende Effekte und  $p$ -Werte:  $2e-04$  als  $p$ -Wert und einen  $\beta_{PO_2}$  Koeffizienten von  $1.1 \times 10^{-5}$ . Welche Aussage ist richtig?

- A** ☐ 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.
- B** ☐ 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  $\beta_{PO_2}$  im Gewicht der Wasserlinsen. Die Einheit  $[\mu g]$  muss besser gewählt werden.
- C** ☐ Die Fallzahl ist zu klein angesetzt. Je kleiner die Fallzahl ist, desto höher ist die Teststatistik und damit auch der  $p$ -Wert kleiner. Wir brauchen also mehr Fallzahl um den geringen Effekt noch signifikant zu kriegen.
- D** ☐ Wenn der Effekt  $\beta_{PO_2}$  winzig ist, dann kann es an einer falsch gewählten Einheit liegen. Der Anstieg von einer Einheit in  $X$  führt ja zu einer Änderung von  $\beta_{PO_2}$  in  $x$ . Wir müssen daher die Einheit von  $y$  entsprechend anpassen.
- 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  $\beta_{PO_2}$  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 `lm()` in . Welche Aussage ist richtig?

- A** ☐ Ist die Einflussvariable  $X$  numerisch so werden die Gruppenmittelwerte geschätzt und eine anschließende ANOVA sowie multipler Gruppenvergleich mit `{emmeans}` ist möglich.
- B** ☐ 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.
- C** ☐ Die Funktion `lm()` in  ist der erste Schritt für einen Gruppenvergleich. Danach kann eine ANOVA oder aber ein multipler Vergleich in `{emmeans}` gerechnet werden. In der Funktion `lm()` werden die Gruppenmittelwerte bestimmt.
- D** ☐ 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.
- E** ☐ Die Funktion `lm()` in  wird klassischerweise für die nicht-lineare Regression genutzt. Ist die Einflussvariable  $X$  numerisch so werden die Gruppenmittelwerte geschätzt.

## 21. Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit haben Sie neben den klassischen normalverteilten Endpunkte, wie Trockengewicht und Wuchshöhe noch den Infektionsstatus und Zählraten erhoben. Um diese nicht normalverteilten Endpunkte auszuwerten nutzen Sie das *generalisierte lineare Modell (GLM)*. Welche Aussage ist richtig?

- A** ☐ Dank dem *generalisierten linearen Modell (GLM)* können auch andere Verteilungsfamilien als die Normalverteilung mit einer linearen Regression modelliert werden.
- B** ☐ Dank dem *generalisierten linearen Modell (GLM)* können auch andere Verteilungsfamilien – außer die Normalverteilung – mit einer linearen Regression modelliert werden. Dafür werden alle Verteilungen in eine Normalverteilung überführt und anschließend standardisiert.
- C** ☐ Das GLM ist eine allgemeine Erweiterung der linearen Regression auf die Normalverteilung.
- D** ☐ Das *generalisierte lineare Modell (GLM)* erlaubt auch weitere Verteilungsgruppen für das  $X$  bzw. die Einflussvariablen in einer linearen Regression zu wählen.
- E** ☐ Das GLM erlaubt auch nicht normalverteilte Residuen in der Schätzung der Regressionsgrade.

## 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** ☐ Randomisierung sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschließen.
- C** ☐ Randomisierung erlaubt erst die Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich. Dadurch lässt sich erst ein Experiment auswerten.
- D** ☐ Randomisierung ist die direkte Folge von Strukturgleichheit. Die Strukturgleichheit erlaubt es erst von der Stichprobe auf die Grundgesamtheit zurückzuschließen.
- E** ☐ Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte. Ohne Mittelwerte keine Varianz und somit auch kein statistischer Test.

### 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 Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Daher ist die Nutzung in Deutsch in den AGBs von R untersagt.
- ☐ B Alle Funktionen und auch Anwendungen sind in R in englischer Sprache. Die Nutzung von deutschen Wörtern ist nicht schick und das ist zu vermeiden.
- ☐ C Im Allgemeinen haben Programmiersprachen Probleme mit Umlauten und Sonderzeichen, die in der deutschen Sprache vorkommen. Eine Nutzung der englischen Sprache umgeht dieses Problem auf einfache Art.
- ☐ D Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- ☐ E Die Spracherkennung von R ist nicht in der Lage Deutsch zu verstehen.

### 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 Wir lesen die Daten ein und mutieren die Daten. Dabei ist wichtig, dass wir nicht das Paket tidyverse nutzen, da dieses Paket veraltet ist. über die Funktion library(tidyverse) entfernen wir das Paket von der Analyse.
- ☐ B Wir transformieren die Spalten über mutate() in ein tibble und können dann über ggplot() uns die Abbildungen erstellen lassen. Dabei beachten wir das wir keine Faktoren in den Daten haben.
- ☐ C Für eine explorativen Datenanalyse (EDA) in R müssen wir als erstes die Daten über read\_excel() einlesen. Danach müssen wir schauen, dass wir die Spalten richtig über mutate() transformiert haben. Insbesondere müssen Variablen mit Kategorien in einen Faktor umgewandelt werden. Am Ende nutzen wir die Funktion ggplot() für die eigentlich EDA.
- ☐ D 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.
- ☐ E Die Funktionsreihenfolge ist wie folgt: read\_excel() -> mutate() -> ggplot(). Dabei ist bei der Transformation der Daten darauf zu achten, dass keine Faktoren erstellt werden.

### 25. Aufgabe

(2 Punkte)

Sie haben das abstrakte Modell  $Y \sim X$  mit  $X$  als Faktor mit zwei Leveln vorliegen. Welche Aussage über  $n_1 < n_2$  ist richtig?

- ☐ A Es liegt Varianzhetrogenität vor.
- ☐ B Es handelt sich um abhängige Beobachtungen.
- ☐ C Es handelt sich um ein balanciertes Design.
- ☐ D Es handelt sich um ein unbalanciertes Design.
- ☐ 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 abhängig, wenn die Mütter ebenfalls miteinander verwandt sind. Erst die Abhängigkeit 2. Grades wird in der Statistik modelliert.
- ☐ B Untereinander unabhängig. Die Ferkel sind eigenständig und benötigen keine zusätzliche Behandlung.

- C** ☐ Die Ferkel stammen vom gleichen Muttertier und haben vermutlich eine ähnlichere Varianzstruktur als die Ferkel von anderen Sauen. Die Ferkel sind untereinander über die Mutter abhängig.
- D** ☐ Untereinander unabhängig. Sollten die Mütter verwandt sein, so ist die Varianzstruktur ähnlich und muss modelliert werden.
- E** ☐ Untereinander stark korreliert. Die Ferkel sind von einer Mutter und somit miteinander korreliert. Dies wird in der Statistik jedoch meist nicht modelliert.

## 27. Aufgabe

(2 Punkte)

In einer Studie wollen Sie den Effektschätzer Odds ratio berechnen. Sie finden in Ihrem Experiment zur Behandlung von Klaueninfektionen bei Rinder in 4 Tieren Erkrankung der Klauen vor. 7 Tiere sind gesund. Welche Aussage ist richtig?

- A** ☐ Der Anteil der Kranken wird berechnet. Da es sich um ein Anteil handelt ergibt sich ein Odds ratio von 0.36.
- B** ☐ Es ergibt sich ein Odds ratio von 0.36, da es sich um eine Chancenverhältnis handelt.
- C** ☐ Da es sich um ein Chancenverhältnis handelt ergibt sich ein Odds ratio von 2.75.
- D** ☐ Der Anteil der Gesunden wird berechnet. Da es sich um ein Anteil handelt ergibt sich ein Odds ratio von 0.36.
- E** ☐ Da es sich um ein Chancenverhältnis handelt ergibt sich ein Odds ratio von 0.57.

## 28. Aufgabe

(2 Punkte)

In der Bio Data Science wird häufig mit sehr großen Datensätzen gerechnet. Historisch ergibt sich nun ein Problem bei der Auswertung der Daten und deren Bewertung hinsichtlich der Signifikanz. Welche Aussage ist richtig?

- A** ☐ Aktuell werden zu grosse Datensätze für die gängige Statistik gemessen. Daher wendet man maschinelle Lernverfahren für kausale Modelle an. Hier ist die Relevanz gleich Signifikanz.
- B** ☐ Relevanz und Signifikanz haben nichts miteinander zu tun. Daher gibt es auch keinen Zusammenhang zwischen hoher Fallzahl ( $n > 10000$ ) und einem signifikanten Test. Ein Effekt ist immer relevant und somit signifikant.
- C** ☐ Aktuell werden immer größere Datensätze erhoben. Eine erhöhte Fallzahl führt automatisch auch zu mehr signifikanten Ergebnissen, selbst wenn die eigentlichen Effekte nicht relevant sind.
- D** ☐ 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.
- 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: 0.02, 0.01, 0.03 und 0.42. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.005, 0.0025, 0.0075 und 0.105. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.08, 0.04, 0.12 und 1.68. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.005, 0.0025, 0.0075 und 0.105. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.

- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.08, 0.04, 0.12 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.08, 0.04, 0.12 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.

### 30. Aufgabe

(2 Punkte)

Sie rechnen einen PostHoc-Test. Nun sollen Sie ein CLD erstellen. Was bedeutet dieser Fachbegriff und welche folgende Beschreibung der Interpretation ist korrekt?

- A** ☐ Compact letter detection. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt.
- B** ☐ 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.
- C** ☐ Compact line display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Früher wurden keine Buchstaben sondern eine durchgezogene Linie verwendet. Bei mehr als drei Gruppen funktioniert die Linie aber graphisch nicht mehr.
- D** ☐ 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.
- E** ☐ 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.

### 31. Aufgabe

(2 Punkte)

Der multiple Vergleich als Posthoc-Test nach einer ANOVA ist in den Agrarwissenschaften heutzutage Standard. Welches R Paket wird häufig für den multiplen Vergleich genutzt? Welche Beschreibung der Eigenschaften ist korrekt?

- A** ☐ 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.
- B** ☐ 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.
- C** ☐ Da Sie für Ihre Bachelorarbeit einen Barplot mit CLD brauchen nutzen Sie das R Paket {emmeans} welches Ihnen schnell die notwendigen Informationen liefert um einen Barplot zu erstellen. Die Berechnung eines CLD ist hierbei auch einfach.
- D** ☐ 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.
- 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  $\beta$ -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** ☐ Beim multiplen Testen kann es zu einer  $\alpha$ -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern sehr viel höher. 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.
- C** ☐ Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern sehr viel niedriger, bei ca. 1%. Es kommt zu einer  $\alpha$ -Hyperinflation. Dagegen kann mit der Adjustierung der p-Werte nach Bonferroni vorgegangen werden.
- D** ☐ Die Adjustierung der p-Werte nach Bonferroni erlaubt es gegen die  $\beta$ -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  $\alpha$ -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 Ihrer Bachelorarbeit werten Sie einen einfaktoriellen Versuch aus. Dafür rechnen Sie in  zunächst eine ANOVA und schließen dann einen multiplen Vergleich mit t-Tests an. Welche Aussage über die Effekte in Ihrem Versuch ist richtig?

- A** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  adjustiert werden im Gegensatz zu den p-Werten.
- B** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  nach Bonferroni adjustiert werden. Dafür wird der Effekt mit der Anzahl an Vergleichen  $k$  multipliziert. Dies geschieht analog zu den p-Werten.
- C** ☐ Beim multiplen Testen kann es zu einer  $\Delta$ -Deflation kommen. Das globale Relevanzniveau liegt nicht mehr bei 5% sondern weit darunter. Daher müssen die  $\Delta$ -Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der  $\Delta$ -Werte nach Bonferroni das bekannteste Verfahren ist. Die  $\Delta$ -Werte werden durch die Anzahl an Vergleichen geteilt.
- D** ☐ Beim multiplen Testen muss der Effekt, wie der Mittelwertsunterschied  $\Delta$  aus einem t-Test, nicht adjustiert werden.
- E** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  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)

Welche Aussage zum mathematischen Ausdruck  $Pr(D|H_0)$  ist richtig?

- A** ☐ Die Wahrscheinlichkeit der Daten unter der Nullhypothese in der Grundgesamtheit.
- B** ☐ Die Wahrscheinlichkeit für die Nullhypothese, wenn die Daten wahr sind.
- C** ☐  $Pr(D|H_0)$  stellt die Wahrscheinlichkeit die Teststatistik  $T$  zu beobachten dar, wenn die Nullhypothese falsch ist.
- D** ☐  $Pr(D|H_0)$  ist die Wahrscheinlichkeit die Daten  $D$  zu beobachten, wenn die Nullhypothese wahr ist.
- E** ☐ Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativhypothese überdeckt.



### 35. Aufgabe

(2 Punkte)

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

- A** ☐ ... dass Modelle meist falsch sind und selten richtig.
- B** ☐ ... dass ein schlechtes Modell durch das Falsifikationsprinzip durch ein weniger schlechtes Modell ersetzt wird.
- C** ☐ ... dass Annahmen an statistische Modelle meist falsch sind.
- D** ☐ ... dass ein minderwertes Modell durch ein minderwertiges Modell ersetzt wird. Es gilt das Verifikationsprinzip nach Karl Popper.
- E** ☐ ... dass in der Wissenschaft immer etwas falsch sein muss. Sonst gebe es keinen Fortschritt.

### 36. Aufgabe

(2 Punkte)

Das Signifikanzniveau  $\alpha$  wird auch Fehler 1. Art genannt und liegt bei 5%. Warum wurde der Grenzwert von 5% als Signifikanzschwelle gewählt?

- A** ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.
- B** ☐ Da Wissenschaftler eine Schwelle für die statistische Testentscheidung benötigen wurde  $\alpha$  historisch gewählt. Damit ist  $\alpha = 5\%$  eine Kulturkonstante.
- C** ☐ 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.
- D** ☐ 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.
- E** ☐ 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.

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

### 38. Aufgabe

(2 Punkte)

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

$H_0$  ablehnen obwohl die  $H_0$  gilt

- A ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.
- B ☐ In die Analogie eines Rauchmelders: *Alarm without fire police*, dem  $\alpha$ -Fehler.
- C ☐ *Alarm without fire*, dem  $\alpha$ -Fehler in der Analogie eines Rauchmelders.
- D ☐ In die Analogie eines Rauchmelders: *Alarm with fire*.
- E ☐ Dem  $\beta$ -Fehler mit der Analogie eines Rauchmelders: *Fire without alarm*.

### 39. Aufgabe

(2 Punkte)

Sie sollen in Ihrer Abschlussarbeit die Relevanz und die Signifikanz in einer statistischen Maßzahl vereinen. Welche Aussage ist richtig?

- A ☐ Das  $\Delta$ . Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da  $\Delta$  antiproportional zum  $p$ -Wert ist, bedeutet auch ein hohes  $\Delta$  ein sehr kleinen  $p$ -Wert.
- B ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- C ☐ Einem Konfidenzintervall. Das Konfidenzintervall bringt durch eine Visualisierung und drei Intervallgrenzen die Möglichkeit mit, eine Relevanzschwelle neben der Signifikanzschwelle und der  $\alpha$ -Schwelle zu definieren.
- D ☐ Über das Konfidenzintervall. Das Konfidenzintervall beitet eine Entscheidung über die Signifikanz und zusätzlich kann über die Visualisierung des Konfidenzintervalls eine Relevanzschwelle definiert werden.
- E ☐ 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.

### 40. Aufgabe

(2 Punkte)

Welche Aussage über den  $p$ -Wert und dem Signifikanzniveau  $\alpha$  gleich 5% ist richtig?

- A ☐ Wir machen ein 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.
- B ☐ Wir vergleichen mit dem  $p$ -Wert und dem Signifikanzniveau  $\alpha$  absolute Werte auf einem Zahlenstrahl und damit den Unterschied der Teststatistiken, wenn die  $H_0$  gilt.
- C ☐ Wir schauen, ob der  $p$ -Wert kleiner ist als das Signifikanzniveau  $\alpha$  und vergleichen somit Wahrscheinlichkeiten. Die Wahrscheinlichkeiten werden als Flächen unter der Kurve der Teststaistik dargestellt, wenn die  $H_0$  gilt.
- D ☐ Wir machen eine Aussage über die individuelle Wahrscheinlichkeit des Eintretens der Nullhypothese  $H_0$ . Der  $p$ -Wert wird mit dem Signifikanzniveau verglichen und bewertet.
- E ☐ Wir schauen, ob der  $p$ -Wert größer ist als das Signifikanzniveau  $\alpha$  und vergleichen somit Wahrscheinlichkeiten. Die Wahrscheinlichkeiten werden als Flächen unter der Kurve der Teststaistik dargestellt, wenn die  $H_A$  gilt.

#### 41. Aufgabe

(2 Punkte)

Um die Testtheorie besser zu verstehen, mag es manchmal sinnvoll sein ein Beispiel aus dem Alltag zu wählen. Die Ergebnisse der Analyse durch einen statistischen Test können auch in grobe Analogie zur Wettervorhersage gebracht werden. Welche Aussage trifft am ehesten 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 Maximaltemperatur: Was ist der maximale Unterschied zwischen zwei Gruppen. Wir erhalten hier eine Aussage über die Spannweite und den maximalen Effekt.
- C** ☐ In der Analogie der Wahrscheinlichkeit für Regen: ein statistischer Test erlaubt die Wahrscheinlichkeit für ein Ereignis abzuschätzen. Die Stärke des Effektes können wir nicht bestimmen.
- D** ☐ In der Analogie des Niederschlags oder Regenmenge: ein statistischer Test gibt die Stärke eines Effektes wieder. Zum Beispiel, wie hoch ist der Mittelwertsunterschied.
- E** ☐ In der Analogie der Sonnenscheindauer: Wie lange kann mit einem entsprechenden Effekt gerechnet werden? Die Wahrscheinlichkeit für den Effekt gibt der statistische Test wieder.

#### 42. Aufgabe

(2 Punkte)

In Ihrer Forschungsarbeit wollen Sie eine Aussage über die untersuchte Population treffen. Dazu nutzen Sie einen statistischen Test. Erhalten Sie eine valide Aussage aus einem statistischen Test?

- A** ☐ Nein, wir erhalten eine Aussage. Müssen aber das Individuum im Kontext der Population adjustieren.
- B** ☐ Nein, es ist nicht möglich die untersuchte Population mit einem t-Test auszuwerten. Wir erhalten dann leider keine Aussage zur Population.
- C** ☐ Ja, es ist möglich die untersuchte Population mit einem t-Test auszuwerten. Wir erhalten dann eine Aussage zur Population.
- D** ☐ Ja, wir erhalten nur eine Aussage zu zwei Individuen. Ein statistischer Test liefert Informationen zu einem Individuum im Vergleich zu einem anderen Individuum.
- E** ☐ 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.

#### 43. Aufgabe

(2 Punkte)

In der statistischen Testtheorie gibt es den Begriff *Power*. Was sagt der statistische Begriff *Power* aus?

- A** ☐ Es gilt  $\alpha + \beta = 1$  und somit liegt  $\beta$  meist bei 95%.
- B** ☐ Die Power ist nicht in der aktuellen Testthorie mehr vertreten. Wir rechnen nur noch mit dem Fehler 1. Art.
- C** ☐ Alle statistischen Tests sind so konstruiert, dass die  $H_A$  mit 80% *bewiesen wird*. Die Power ist  $1 - \beta$  mit  $\beta$  gleich 20% gesetzt.
- D** ☐ Alle statistischen Tests sind so konstruiert, dass die  $H_A$  mit 20% *bewiesen wird*. Die Power ist  $1 - \beta$  mit  $\beta$  gleich 80% gesetzt.
- E** ☐ Die Power wird berechnet und ist keine Eigenschaft des Tests. Die Power wird auf 80% gesetzt und beschreibt mit welcher Wahrscheinlichkeit  $H_0$  *bewiesen wird*

#### 44. Aufgabe

(2 Punkte)

Welche Aussage über den Effekt eines statistischen Tests ist richtig?

- A** ☐ Der Effekt eines statistischen Tests beschreibt die biologisch interpretierbare Ausgabe eines Tests. Modernen Algorithmen liefern keine Effekte mehr sondern nur noch bedingte Wahrscheinlichkeiten. Der Effekt spielt in der modernen Statistik keine Rollen mehr.
- B** ☐ Der Effekt eines statistischen Tests beschreibt die biologisch interpretierbare Ausgabe eines Tests. Zum Beispiel den mittleren Unterschied zwischen zwei Gruppen aus einem t-Test. Damit ist der Effekt direkt mit dem Begriff der Relevanz verbunden. Die Entscheidung über die Relevanz trifft der Forschende unabhängig von der Signifikanz eines statistischen Tests.
- C** ☐ 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.
- D** ☐ Durch den Effekt erfahren wir die statistische interpretierbare Ausgabe eines statistischen Tests. Zum Beispiel das  $\eta^2$  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 bedeutend zu bewerten ist.
- 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 p-Wertes gegen die Nullhypothese ist richtig?

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

#### 46. Aufgabe

(2 Punkte)

Wenn Sie im Allgemeinen einen statistischen Test rechnen, dann kommen Sie um eine statistische Hypothese  $H$  nicht herum. Welche Aussage über statistische Hypothesen ist richtig?

- A** ☐ Mit der Nullhypothese  $H_0$  und der Alternativhypothese  $H_A$  oder  $H_1$  gibt es zwei Hypothesen.
- B** ☐ Es gibt - bedingt durch das das Falsifikationsprinzip - ein Set von  $k$  Nullhypothesen, die iterative gegen  $k - 1$  Alternativhypothesen getestet werden.
- C** ☐ Die Hypothesen  $H_0$  und  $H_A$  sind rein prosarischer Natur und bilden keinen mathematischen Hintergrund ab. In der Statistik wird die wissenschaftliche Fragestellung getestet. Daher stehen auch die verständlichen Hypothesen im Mittelpunkt der biologischen Interpretation.
- D** ☐ Mit der Nullhypothese  $H_A$  und der Alternativhypothese  $H_0$  gibt es zwei Hypothesen, die aber selten genutzt werden.
- E** ☐ Es gibt ein Hypothesenset bestehend aus  $k$  Hypothesen. Meistens wird die Nullhypothese  $H_0$  und die Alternativhypothese  $H_A$  verwendet. Wegen des Falsifikationsprinzips ist es wichtig, die bekannte falsche und unbekannte richtige Hypothese mit in das Set zu nehmen.

## Statistische Tests für Gruppenvergleiche

### 47. Aufgabe

(2 Punkte)

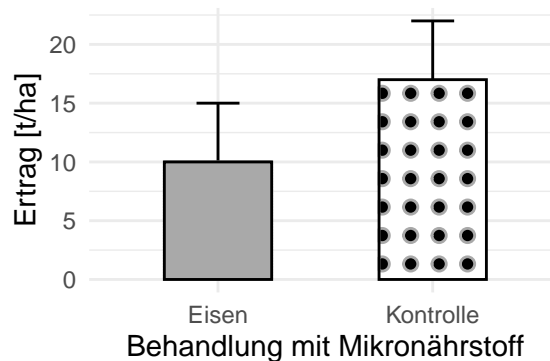
Welche Aussage über den t-Test im Allgemeinen ist richtig? Berücksichtigen Sie den Welch t-Test wie auch den Student t-Test!

- A** ☐ Der t-Test vergleicht zwei Gruppen indem die Mittelwerte miteinander verglichen werden.
- B** ☐ Der t-Test testet generell zu einem erhöhten  $\alpha$ -Niveau von 20%.
- C** ☐ 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.
- 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 13 Parzellen pro Gruppe durchgeführt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung des Mikronährstoff Eisen auf den Ertrag in t/ha von Weizen im Vergleich zu einer Kontrolle. Welche Aussage ist richtig, wenn Sie einen t-Test rechnen?



- A** ☐ Es liegt ein signifikanter Unterschied vor. Der Effekt liegt bei -0.7.
- B** ☐ Nach Betrachtung des Barplots liegt kein signifikanter Unterschied vor. Der Effekt kann nicht bei einem t-Test aus Barplots bestimmt werden.
- C** ☐ Der Test deutet auf einen signifikanten Unterschied hin. Der Effekt liegt vermutlich bei -7.
- D** ☐ Die Barplots deuten auf keinen signifikanten Unterschied. Der Effekt liegt vermutlich bei -7 unter einer groben Abschätzung.
- E** ☐ 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.

### 49. Aufgabe

(2 Punkte)

Welche Aussage über den gepaarten t-Test für verbundene Stichproben ist richtig?

- A** ☐ Abhängige Beobachtungen müssen gesondert in einem gepaarten t-Test modelliert werden. Wenn wiederholt an dem gleichen Tier oder Pflanze gemessen wird, dann bilden wir den Quotienten zwischen den beiden Zeitpunkten. Auf den Quotienten rechnen wir den gepaarten t-Test.
- B** ☐ 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.

- C** ☐ 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.
- D** ☐ Der gepaarte t-Test wird gerechnet, wenn die Beobachtungen abhängig voneinander sind. Wir messen jede Beobachtung nur einmal und berechnen dann die Differenz zu dem Mittel der anderen Beobachtungen.
- E** ☐ 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.

## 50. Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit passen die Ergebnisse einer ANOVA und eines multiplen Vergleiches nicht zusammen. Nach einem Experiment mit vier Maissorten ergibt eine ANOVA ( $p = 0.049$ ). Sie führen anschließend die paarweisen t-Tests für alle Vergleiche durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der  $\alpha$ -Schwelle. Sie schauen sich auch die rohen, unadjustierten p-Werte an und finden hier als niedrigsten p-Wert  $p_{3-2} = 0.052$ . 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** ☐ Das ist kein Wunder. Die ANOVA testet auf der gesamten Fallzahl und die paarweisen t-Tests verlieren immer eine oder mehr Gruppen als Fallzahl. Mit steigender Fallzahl sind mehr signifikante Unterschiede zu erwarten. Die p-Werte unterscheiden sich numerisch auch kaum.
- C** ☐ Der Fehler liegt in den t-Tests. Wenn eine ANOVA signifikant ist, dann muss zwangsweise auch ein t-Test signifikant sein.
- D** ☐ Es gibt einen Fehler in der Varianzstruktur. Daher kann die ANOVA nicht richtig sein und paarweise t-Tests liefern das richtige Ergebnis.
- E** ☐ Hier kommt der Effekt der steigenden Fallzahl auf die Anzahl an signifikante Ergebnisse zu tragen. Da die ANOVA auf weniger Fallzahl testet als die paarweisen t-Tests, kann die ANOVA schwerer einen signifikanten Unterscheid nachweisen.

## Teil I.

# Deskriptive Statistik & Explorative Datenanalyse

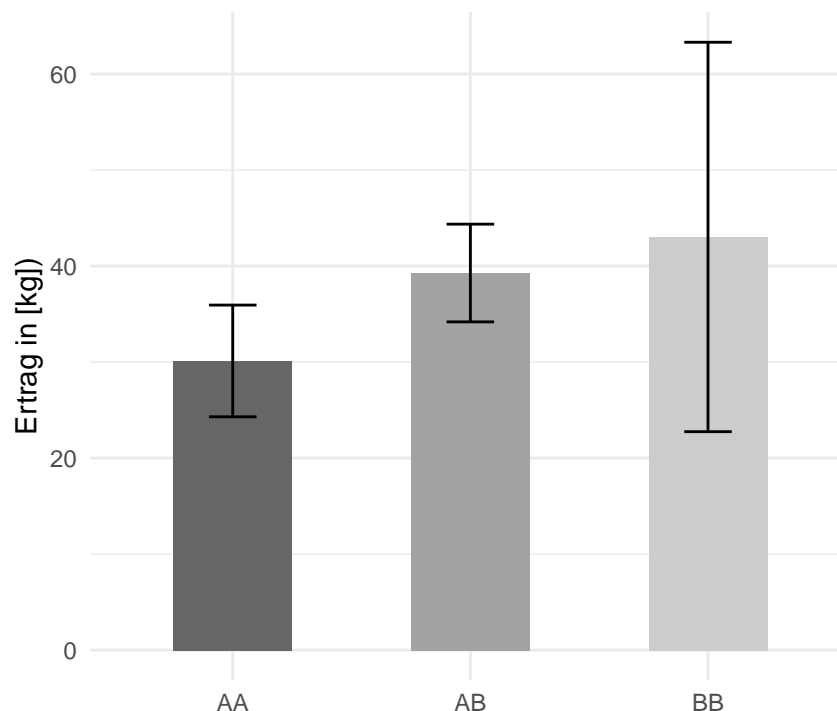
## 51. Aufgabe

(8 Punkte)

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



**Zerforschen des Barplots** Jonas steht vor einem ersten Problem, denn wenn es nach seiner Betreuerin geht, soll er in einer Klimakammer Erdbeeren auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Am Ende dann doch besser Stricken. Wunderbar. Eine echte Ablenkung für Jonas. Das heißt erstmal überlegen für Jonas. Aus den Boxen wummert Iron Maiden und sein Mund ist verklebt von Snickers. 'Herrlich', denkt Jonas. Die Behandlung werden verschiedene Genotypen (AA, AB und BB) sein. In seiner Exceldatei wird er den Endpunkt (Y) *Ertrag* als *yield* aufnehmen. Vorab soll Jonas aber einmal die folgenden Barplots seiner Betreuerin nachbauen, damit er den R Code schonmal für später vorliegen hat. Damit geht das Problem schon los. Wenn die Erschöpfung nicht wäre, ja dann wäre wohl vieles möglich für Jonas! Aber so..



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

1. Formulieren Sie die wissenschaftliche Fragestellung! (1 Punkt)
2. Erstellen Sie eine Tabelle mit den statistischen Maßzahlen der drei Barplots! Beachten Sie die korrekte Darstellungsform der statistischen Maßzahlen! (3 Punkte)
3. Erstellen Sie einen beispielhaften Datensatz im R üblichen Format, aus dem die drei Barplots möglicherweise erstellt wurden! (2 Punkte)
4. Kann Jonas einen Unterschied zwischen den Behandlungen erwarten? Begründen Sie Ihre Antwort! (2 Punkte)

## 52. Aufgabe

(8 Punkte)

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



**Visualisierung des Barplots** Jessica steht vor einem ersten Problem, denn wenn es nach ihrem Betreuer geht, soll sie in einem Gewächshausexperiment Erbsen auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Warhammer. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Jessica denkt gerne über Warhammer nach. Die Behandlung waren verschiedene Lüftungssysteme und Folientunnel (*ctrl*, *storm* und *tornado*). In ihrer Exceldatei hat sie den Outcome (Y) *Frischegewicht* als *freshmatter* aufgenommen. Nun soll Jessica 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. Wäre da nicht noch etwas. Wenn der Mangel nicht wäre, ja dann wäre wohl vieles möglich für Jessica! Aber so.. Aber egal. Einfach mal raus um Rad zu fahren. Ohne Ziel und Uhr. Das ist was für Jessica.

treatment	freshmatter
storm	30.6
tornado	35.0
ctrl	34.4
tornado	33.9
tornado	37.6
storm	27.6
tornado	35.4
ctrl	30.6
storm	22.4
ctrl	27.3
ctrl	25.4
storm	20.0
ctrl	32.3

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

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




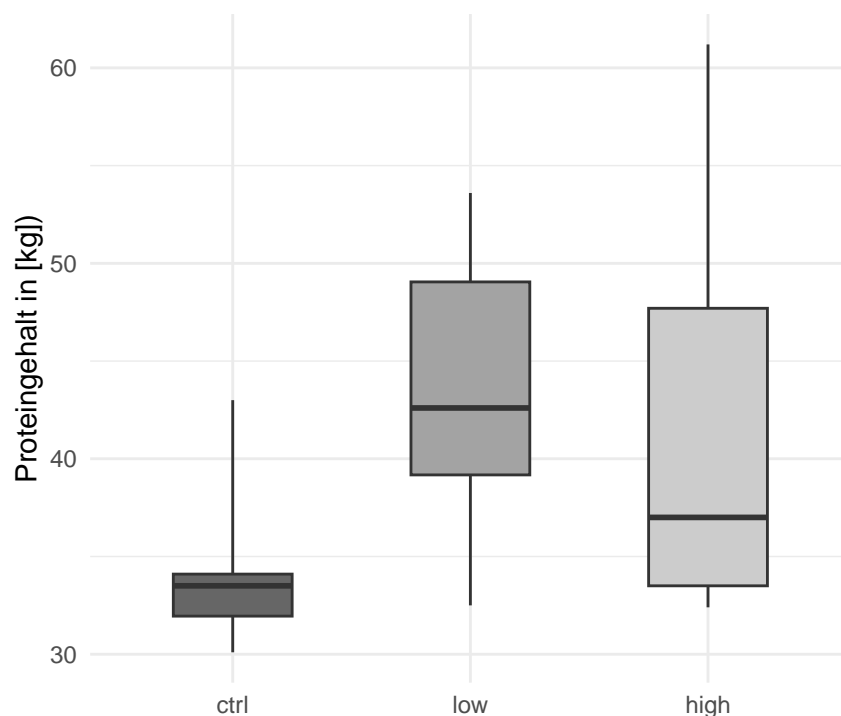
### 53. Aufgabe


(9 Punkte)


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



**Zerforschen des Boxplots** Boxplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Mark nicht gemerkt, welche statistischen Maßzahlen für einen Boxplot erhoben werden müssen. Besser wäre was anderes gewesen. Geocaching. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Mark denkt gerne über Geocaching nach. Das ist in soweit doof, da nach seiner Betreuerin erstmal ein Boxplot nachgebaut werden soll, bevor es mit seinem Projektbericht losgeht. Dann hat er schonmal den  Code vorliegen und nachher geht dann alles schneller. Na dann mal los. Mark schafft sich die nötige Stimmung. Mark streichelt liebevoll der Hamster. Der Kopf ist in seinem Schloß vergraben um den Klang von Andrea Berg zu dämpfen. In der Behandlung für Erdbeeren werden verschiedene Düngestufen (*ctrl*, *low* und *high*) sein. Erfasst wird als Endpunkt (Y) *Proteingehalt*. Mark soll dann *protein* in seiner Exceldatei eintragen. Aber nur in passender Atmosphäre! Hm, lecker Marzipankugeln und dazu dann im Hintergrund Columbo laufen lassen.



Leider kennt sich Mark mit der Erstellung von Boxplots 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 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. Kann Mark einen Unterschied zwischen den Behandlungen erwarten? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 54. Aufgabe

(9 Punkte)

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



**Visualisierung des Boxplots** Jonas steht vor einem ersten Problem, denn wenn es nach seinem Betreuer geht, soll er in einem Freilandversuch Kartoffeln auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Jonas liebt Stricken. Darin kann er sich wirklich verlieren und immer wieder neu begeistern. Die Behandlung waren verschiedene Lüftungssystemen und Folientunneln (*ctrl* und *tornado*). In seiner Exceldatei hat er den Messwert (Y) *Frischegewicht* als *freshmatter* aufgenommen. Nun soll Jonas die Daten einmal als Boxplots in einer Präsentation visualisieren, damit seinem Betreuer wieder klar wird, was er eigentlich nochmal gemacht hat und was für ein Ergebnis in einem statistischen Test zu erwarten wäre. Anhand von Boxplots lässt sich eine Aussage über die Varianzhomogenität über die Behandlungsgruppen treffen. Wäre da nicht noch etwas. Eine echte Herausforderung für ihn war schon immer die Erschöpfung gewesen. Ein leidiges Lied. Aber egal. Um zu Schwimmen geht Jonas dann später nochmal raus. Echte Entspannung.

treatment	drymatter
ctrl	47.0
ctrl	50.6
tornado	35.1
ctrl	48.2
tornado	30.5
ctrl	72.0
tornado	34.5
ctrl	36.1
ctrl	40.0
tornado	32.0
tornado	40.8
tornado	30.7
tornado	31.9
ctrl	39.4
ctrl	53.6
ctrl	54.3

Leider kennt sich Jonas 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 Kartoffeln! 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 Kartoffeln 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!



**Visualisierung des Histogramm für kategoriale Daten** 'Hm...', Snickers und Iron Maiden. Das ist und bleibt die beste Kombination zum Nachdenken für Jonas. Jonas betrachtet die folgenden Daten nach einem Feldexperiment mit Lauch. In dem Experiment wurden die Knötchen gezählt. Nach der Meinung seinem Betreuer muss als erstes geschaut werden, wie diese verteilt sind. Also welcher statistischen Verteilung die Knötchen folgen. Dazu soll Jonas ein Histogramm verwenden. Dann hätte man auch einen guten Überblick über den Endpunkt (Y). Es wäre einfacher, wenn da nicht noch was wäre. Wenn die Erschöpfung nicht wäre, ja dann wäre wohl vieles möglich für Jonas! Aber so.. Wenn Iron Maiden ertönt, dann sucht das Meerschweinchen schleunigst Schutz unter dem Sofa. Jonas schüttelt den Kopf.

Die Knötchen: 4, 6, 6, 4, 2, 1, 0, 7, 3, 8, 3, 3, 5, 2, 5, 5, 7, 3, 2, 4, 3, 7, 1, 3, 7, 5, 1, 6, 7, 7, 6, 3, 9

Leider kennt sich Jonas 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 6 zu beobachten! **(1 Punkt)**
5. Berechnen Sie aus den Daten die *Chance* mehr als die Anzahl 6 zu beobachten! **(1 Punkt)**

## 56. Aufgabe

(8 Punkte)

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



**Visualisierung des Histogramm für kontinuierliche Daten** In einem Gespräch mit ihrer Betreuerin wird Tina gebeten seine Daten aus einem Kreuzungsexperiment mit Schweinen in einem Histogramm darzustellen. Aus den Boxen wummert Tocotronic und ihr Mund ist verklebt von Katjes. 'Herrlich', denkt Tina. In ihrem Experiment hat er die mittlere Anzahl an weißen Blutkörperchen erst fotografiert und dann ausgezählt. Laut ihrer Betreuerin soll das Histogramm helfen, die Verteilung der die mittlere Anzahl an weißen Blutkörperchen zu bestimmen. Es wäre einfacher, wenn da nicht noch was wäre. Wenn die Wut nicht wäre, ja dann wäre wohl vieles möglich für Tina! Aber so.. Wenn Tocotronic ertönt, dann sucht die Spinne schleunigst Schutz unter dem Sofa. Tina schüttelt den Kopf.

Die mittlere Anzahl an weißen Blutkörperchen: 8, 12, 11.5, 12.5, 11.5, 11.3, 8.9, 12.6, 3.9, 13.9, 13.4, 9.7, 6.6, 12.4, 11.1, 12.9, 5.9, 10.3, 7, 10, 10.4, 6.5

Leider kennt sich Tina mit der Erstellung von Histogrammen überhaupt nicht aus. Deshalb braucht sie 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!



**Visualisierung des Scatterplots** Mark liest laut: 'Wenn zwei kontinuierliche Variablen vorliegen, können diese in einer explorativen Datenanalyse...'. Mark stoppt. Mark schmeißt noch eine Handvoll Marzipankugeln in seinen Rachen. Im Hintergrund klirrt leise der Spiegel zum Sound von Andrea Berg. Was waren noch gleich kontinuierliche Variablen? In seiner Hausarbeit hatte er ein Freilandversuch im Teutoburgerwald durchgeführt. Dabei ging es um den Zusammenhang zwischen Chlorophyllgehalt (SPAD-502Plus) [SPAD] und durchschnittlichen Anteil an Ton [%/l] im groben Kontext von Erbsen. Nun stellt sich die Frage für ihn, ob es überhaupt einen Zusammenhang zwischen den gemessenen Variablen gibt. Dafür war eine explorative Datenanalyse gut! Wenn die Unsicherheit nicht wäre, ja dann wäre wohl vieles möglich für Mark! Aber so.. Dann was anderes. Wenn Columbo läuft, dann ist der Hamster nicht mehr da. Aber jetzt braucht er mal Entspannung!

Chlorophyllgehalt (SPAD-502Plus) [SPAD]	Durchschnittlichen Anteil an Ton [%/l]
18.1	31.6
10.8	18.8
12.3	20.5
12.3	23.5
19.0	25.1
13.6	24.2
17.9	26.6
13.0	21.3
10.2	17.6
17.2	24.9
15.0	26.6

Leider kennt sich Mark mit der Erstellung einer explorativen Datenanalyse für kontinuierliche Daten überhaupt nicht aus. Deshalb braucht er 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 *kein* 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!



**Visualisierung des Mosaicplots** Irgendwie komisch, wenn sie Matrix anmacht, dann ist das Minischwein eigentlich sofort vor dem Bildschirm und starrt hinein. Aber Ablenkung hilft nur begrenzt. 'Uff!', denkt sich Yuki. Jetzt hat sie doch tatsächlich zwei kategoriale Variablen in ihrer Hausarbeit gemessen. Zum einen die Behandlung Klimakontrolle [ja/nein] und zum anderen die Messung Fettgehalt erreicht [ja/nein] im Kontext von Zandern. Hierfür hat sie ein Stallexperiment im Emsland durchgeführt. Jetzt möchte Yuki die Daten einmal in einer explorativen Datenanalyse darstellen. Danach kann sie dann über den passenden statistischen Test nachdenken. Dabei unterstützt ihre Betreuerin diesen Ansatz bevor es in der Datenanalyse weiter geht. So schön wie so gut. Wenn die Faulheit nicht wäre, ja dann wäre wohl vieles möglich für Yuki! Aber so..

Fettgehalt erreicht	er-	Klimakontrolle	Fettgehalt erreicht	er-	Klimakontrolle
nein		ja	ja		nein
nein		ja	ja		nein
nein		ja	nein		ja
ja		ja	nein		ja
nein		ja	ja		nein
nein		nein	ja		nein
ja		nein	nein		nein
nein		ja	nein		ja
ja		nein	ja		nein
nein		ja	ja		nein
ja		nein	nein		nein
ja		nein	nein		ja
ja		nein	nein		ja
nein		ja	ja		ja
nein		ja	ja		ja

Leider kennt sich Yuki mit der Erstellung einer explorativen Datenanalyse für kategoriale Daten überhaupt nicht aus. Deshalb braucht sie 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)**

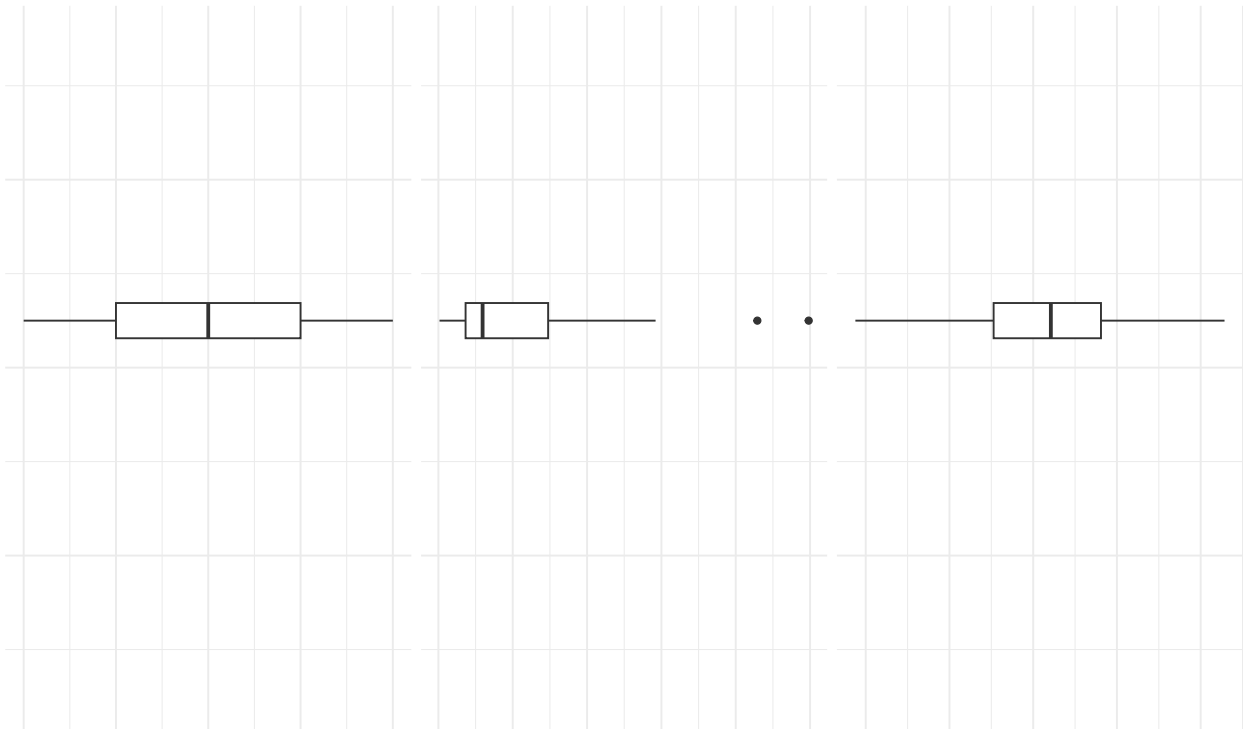
## 59. Aufgabe

(10 Punkte)

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



**Visualisierung von Verteilungen** 'Was soll das denn jetzt schon wieder sein? Drei Boxplot, die auf der Seite liegen?', entfährt es Nilufar und schaut dabei Jessica an. 'Keine Ahnung. Es ist bestimmt wieder so ein Lernziel mit der Verteilung und so.', meint Jessica sichtlich genervt und mampft noch ein paar Schokobons. 'Du weißt doch wie es heißt, *Frei ist, wer missfallen kann*.<sup>1</sup>', merkt Nilufar nickend an. Die beiden schauen angestrengt auf die drei Boxplots. Das Ziel ist es zu verstehen, wie eine Verteilung anhand eines Boxplots bewertet werden kann. Jessica und der Mangel machen die Sache nicht einfacher.



Jetzt brauchen Nilufar und Jessica Ihre Hilfe bei der Abschätzung einer Verteilung anhand von Boxplots um ihre Arbeit dann in diesem Semester noch abschließen zu können.

1. Zeichnen Sie über die Boxplots die entsprechende zugehörige Verteilung! **(3 Punkte)**
2. Zeichnen Sie unter die Boxplots die entsprechende zugehörige Beobachtungen als Stiche! **(3 Punkte)**
3. Wie viel Prozent der Beobachtungen fallen in das IQR? Ergänzen Sie die Abbildung entsprechend um den Bereich! **(2 Punkte)**
4. Wie viel Prozent der Beobachtungen fallen in  $\bar{y} \pm 1s$  und  $\bar{y} \pm 2s$  unter der Annahme einer Normalverteilung? **(2 Punkte)**

<sup>1</sup>Oschmann, A. (2024) Mädchen stärken: Stärken fördern, Selbstwert erhöhen und liebevoll durch Krisen begleiten. Goldegg Verlag

## 60. Aufgabe

(10 Punkte)

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



**Visualisierung der Normalverteilung** 'Jetzt haben wir schon überall geschaut und ich finde die verdammte Ratte nicht. Wo ist die den normalerweise? Und wenn du jetzt einen blöden Witz über die Aufgabe und normal machst, dann gehe ich.', faucht Paula Alex an. 'Ui, alles gut. Vielleicht ein paar Gummibärchen zur Entspannung?', entgegnet Alex. Manchmal macht der Perfektionismus Paula zu einem anderen Menschen, der sie nicht sein will. Da rennt dann auch die Ratte vor ihr weg. Jetzt sollen die beiden diese Aufgabe zur Normalverteilung lösen. Es geht um verschiedene Normalverteilungen und zu verstehen, wie die Parameter der Normalverteilung funktionieren. Anscheinend hängen Normalverteilungen vom Mittelwert  $\bar{y}$  und der Standardabweichung  $s$  ab.

Jetzt brauchen Paula und Alex Ihre Hilfe bei der Abschätzung einer Verteilung um ihre Arbeit dann in diesem Semester noch abschließen zu können.

1. Skizzieren Sie zwei Normalverteilungen mit  $\bar{y}_1 \neq \bar{y}_2$  und  $s_1 \neq s_2$ ! **(3 Punkte)**
2. Beschriften Sie die Normalverteilungen mit den statistischen Maßzahlen! **(2 Punkte)**
3. Liegt Varianzhomogenität oder Varianzheterogenität vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. In welchen Bereich fallen 68% bzw. 95% der Beobachtungen in einer Normalverteilung? Ergänzen Sie die Bereiche in einer Normalverteilung! **(2 Punkte)**
5. Ergänzen Sie unter einer der Normalverteilungen den entsprechenden Boxplot! **(1 Punkt)**



## 61. Aufgabe

(10 Punkte)

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

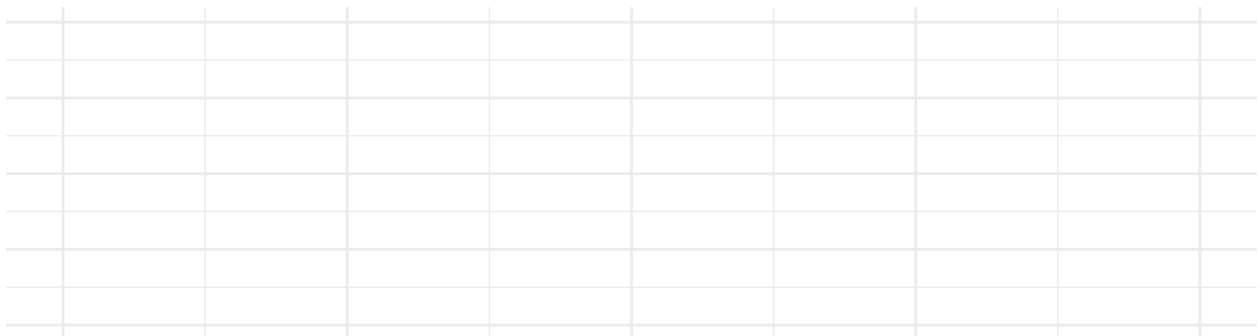


**Visualisierung der Normalverteilung und der Poissonverteilung** 'Wer hat sich denn sowas ausgedacht? Verteilungen?', entfährt es Paula und schaut dabei Jessica an. In ihrer Hand zerdrückt sie Schokobons von Jessica. 'Wir sollen eine Normalverteilung mit einem Mittelwert von  $\bar{y}_1 = 2$  und einer Standardabweichung von  $s_1 = 1$  zeichnen. Sowie eine weitere Normalverteilung mit einem Mittelwert von  $\bar{y}_2 = 0$  und einer Standardabweichung von  $s_2 = 1$ . Keine Ahnung wie das geht. Darunter sollen dann noch eine Poissonverteilung mit einem Mittelwert von  $\lambda_1 = 2$  sowie einer weiteren Poissonverteilung mit einem Mittelwert von  $\lambda_2 = 20$  gezeichnet werden.', meint Jessica sichtlich eingeschüchtert und rettet noch ein paar Schokobons in seinem Mund. Paula und der Perfektionismus machen die Suche nach der Lösung nicht einfacher. Im Hintergrund spielt viel zu laut David Bowie, die diesmal Paula ausgewählt hat und die Ratte schon in die Küche vertrieben hat, wo es den Biomüll mampft.

### Zwei Normalverteilungen



### Zwei Poissonverteilungen



Jetzt brauchen Paula und Jessica Ihre Hilfe bei der Abschätzung einer Verteilung um ihre Arbeit dann in diesem Semester noch abschließen zu können.

1. Skizzieren Sie die zwei Normalverteilungen und zwei Poissonverteilungen! **(4 Punkte)**
2. Achten Sie auf die entsprechende Skalierung in den jeweiligen Abbildungen! **(2 Punkte)**
3. Ergänzen Sie unter einer Normalverteilung den entsprechenden Boxplot! **(1 Punkt)**
4. Ergänzen Sie unter einer Poissonverteilung den entsprechenden Boxplot! **(1 Punkt)**
5. Geben Sie ein Beispiel für ein Outcome  $y$ , welches einer Normalverteilung folgt! **(1 Punkt)**
6. Geben Sie ein Beispiel für ein Outcome  $y$ , welches einer Poissonverteilung folgt! **(1 Punkt)**

## Teil II.

# Statistisches Testen & statistische Testtheorie

### 62. Aufgabe

(9 Punkte)



**Grundgesamtheit und experimentelle Stichprobe** 'Schnapspraline?', räuspert sich das Känguruh und schaukelt in der Wippe. Mark und Yuki schauen erschrocken auf. 'Ähm, das ist hier eine Klausuraufgabe...', merkt Mark mit leicht schrägen Blick an. 'Ich mache hier ein Praktikum und schreibe Teile der Aufgaben.', gähnt das Känguruh. 'Also, ich glaube das ist so nicht gedacht. Und das sind die Schnapspralinen für den Geburtstag meiner Oma!', ruft Yuki. 'Pillepalle! Meins, deins, das sind doch alles bürgerliche Kategorien!', entgegnet das Känguruh und liest von einem zerknitterten Stück Papier ab: 'Was ist der Unterschied zwischen dem Einen und dem Anderen. Steht alles unten in den Fragen. Einfach selber lesen...'.  
Leider kennen sich Mark und Yuki mit der Grundgesamtheit und der Stichprobe überhaupt nicht aus. Daher sind Sie gefragt!

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)**

## 63. Aufgabe

(9 Punkte)



**Das Nullritual - Die statistische Testtheorie** 'Fechten ist der beste Sport, den es gibt.', meint Paula. Nilufar entgegnet, 'Ich empfehle ja immer allen Kicken.' Die beiden sind im Zoo und diskutieren, ob Pinguine Knie haben. Eigentlich wollten beide nochmal die statistische Testtheorie durchgehen, sind dann aber auf abenteuerlichen Wege im Zoo gelandet. Paula starrt wie hypnotisiert auf einen strullenden Elefanten und stoppt die Zeit.<sup>2</sup> 'Du bist so peinlich.', entfährt es Nilufar.

Leider kennen sich Paula und Nilufar mit statistischen Testtheorie, auch Null-Ritual genannt, überhaupt nicht aus. Geschweige denn mit der Visualisierung als Kreuztabelle.

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

20%    Testentscheidung    5%    Richtige Entscheidung

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

---

<sup>2</sup>Yang, P. J., et al. (2014). Duration of urination does not change with body size. Proceedings of the National Academy of Sciences, 111(33), 11932-11937.

## 64. Aufgabe

(9 Punkte)

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

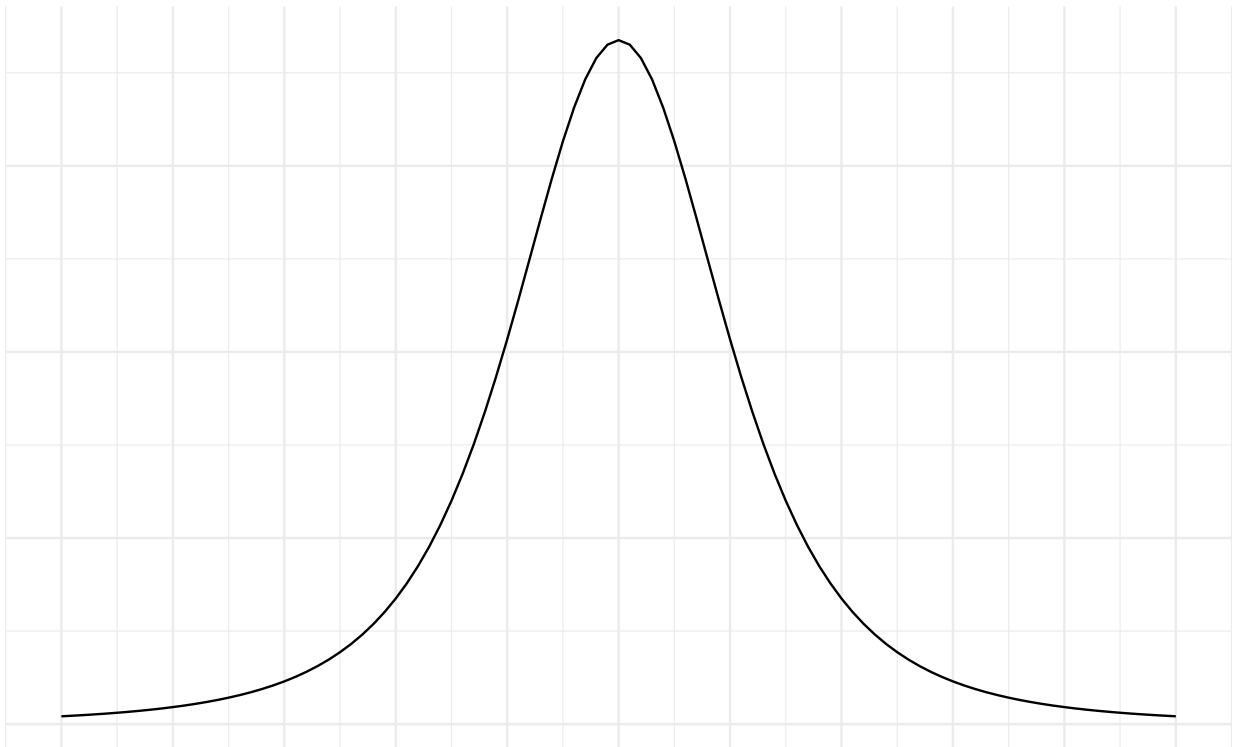


**Visualisierung der Teststatistik  $T_D$  und dem p-Wert** 'Kannst du mir nochmal an einer Visualisierung erklären, wie der Zusammenhang zwischen der Teststatistik aus den Daten  $T_D$  und dem p-Wert ist? Ich habe hier zig Fachbegriffe, kriege die aber nicht zusammen...', fragt Yuki nachdrücklich Nilufar. Das hilft aber nur bedingt, denn Nilufar hat wenig geschlafen und träumt zu den Klängen von London Grammar. Yuki hatte den ganzen Abend mit Nilufar über die Faulheit diskutiert und nun sind beide voll neben der Spur. So wird es nichts mit der Klausur. Yuki mampft noch ein paar Reese's Peanut Butter Cups und nickt ein. Jetzt brauchen die beiden gesondert Hilfe!

Leider kennen sich Yuki und Nilufar mit der Visualisierung der Teststatistik  $T_D$  und dem p-Wert überhaupt nicht aus und brauchen daher Ihre Hilfe!

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 „ $A = 0.95$ “! **(1 Punkt)**
4. Zeichnen Sie  $T_{\alpha=5\%}$  in die Abbildung! **(1 Punkt)**
5. Zeichnen Sie das Signifikanzniveau  $\alpha$  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)**



## 65. Aufgabe

(10 Punkte)



**Visualisierung des 95% Konfidenzintervalls** 'So, was haben wir gemacht? Wir haben einen t-test für den Vergleich der Mittelwerte gerechnet.', meint Jonas. Tina schaut fragend. 'Hatten wir nicht alles zu einer Kontrolle verglichen? Das war doch so!', ruft Tina laut aus. 'Wir haben doch als Messwert *Energieverbrauch der Klimakammer* erhoben.', stellt Jonas fest. Jetzt haben beide das Problem, die möglichen 95% Konfidenzintervalle zu interpretieren.

Leider kennen sich Jonas und Tina mit der Visualisierung des 95% Konfidenzintervall überhaupt nicht aus.

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 95% Konfidenzintervall mit höherer Varianz  $s_p$  in der Stichprobe als der Rest der 95% Konfidenzintervalle
  - (b) Ein nicht signifikantes, nicht relevantes 95% Konfidenzintervall
  - (c) Ein signifikantes, relevantes 95% Konfidenzintervall
  - (d) Ein signifikantes, relevantes 99% Konfidenzintervall.
  - (e) Ein 95% Konfidenzintervall mit niedriger Varianz  $s_p$  in der Stichprobe als der Rest 95% der Konfidenzintervalle
  - (f) Ein signifikantes, nicht relevantes 95% Konfidenzintervall



## 66. Aufgabe

(10 Punkte)

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



**Zusammenhang zwischen dem Effekt, der Streuung sowie der Fallzahl** An einem heißen Juliabend haben sich Nilufar und Steffen zum Lernen verabredet. Eine große Kanne Eistee und Berge von Takis Blue Heat schmelzen in der Sonne und warten darauf gegessen zu werden. Nilufar liest laut vor:

*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  $\Delta$ , der Streuung  $s$  und der Fallzahl  $n$  auf die Teststatistik  $T_D$ , den p-Wert  $Pr(D|H_0)$  sowie dem Konfidenzintervall  $KI_{1-\alpha}$ ?*

Steffen hebt die Augenbraue. 'Ich bleibe dabei. Wir sollten erstmal Harry Potter schauen, bis es kühler ist. Den Film habe ich doch extra mitgebracht!' Nilufar ist der Idee nicht abgeneigt und auch das Huhn kommt auf die Terasse um dabei zu sein.

Leider kennen sich Nilufar und Steffen mit dem Zusammenhang zwischen dem Effekt, der Streuung sowie der Fallzahl überhaupt nicht aus.

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 99%-igen Konfidenzintervall! Begründen Sie Ihre Visualisierung anhand der Formel des Konfidenzintervalls des t-Tests mathematisch! **(3 Punkte)**

# Teil III.

## Der Student t-Test, Welch t-Test & gepaarter t-Test

### 67. Aufgabe

(9 Punkte)

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



**Berechnung des Student t-Test oder Welch t-Test** Nilufar ist im Wendland für einen Pilotexperiment mit sehr geringer Fallzahl ( $n_1 = n_2 = 3$ ) mit Fleischrindern. Allein diese Tatsache ist für sie eine Erzählung wert. Eine echte Herausforderung für sie war schon immer die Erwartung gewesen. Ein leidiges Lied. Für ihre Abschlussarbeit musste sie ein Stallexperiment mit Fleischrindern durchführen und das sollte laut ihrem Betreuer an diesem Ort besonders gut gelingen, da man hier gut neue technische Anlagen und Behandlungen fernab der Bevölkerung testen könne. Zeugen gibt es hier jedenfalls keine. Gar keine. Alleine sein hilft jetzt aber nur bedingt, denn ihre Behandlung Flüssignahrung (*ctrl* und *fLOW*) und der Messwert Fettgehalt [%/kg] sollen mit einem t-Test ausgewertet werden. Immerhin weiß sie, dass ihr Messwert einer Normalverteilung folgt. Hm..., was entspannendes wäre gut. Um zu Kicken geht Nilufar dann später nochmal raus. Echte Entspannung.

treatment	weight
ctrl	16.4
dose	26.6
ctrl	11.0
dose	15.3
dose	27.6
ctrl	16.1

Leider kennt sich Nilufar 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. Bestimmen Sie die Teststatistik  $T_D$  eines Student t-Tests! **(3 Punkte)**
3. Treffen Sie mit  $T_{\alpha=5\%} = 2.04$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Berechnen Sie den Effekt des Student t-Tests! **(1 Punkt)**
5. Formulieren Sie eine Antwort an Nilufar über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

## 68. Aufgabe

(12 Punkte)

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



**Berechnung des Student t-Test** 'Der t-Test testet ein normalverteiltes Outcome (Y).', liest Paula laut. Das hilft jetzt auch nur bedingt weiter. Wenn der Perfektionismus nicht wäre, ja dann wäre wohl vieles möglich für Paula! Aber so.. Laut ihrer Betreuerin ist zwar ihr Messwert Fettgehalt [%/kg] normalverteilt, aber wie rechnet sie jetzt einen t-Test? Für ihrer Hausarbeit musste sie ein Stallexperiment mit Zandern im Wendland durchführen. Als wäre das nicht schon anstrengend genug gewesen. Jetzt soll sie auch noch testen, ob die Behandlung Bestandsdichte (*Verordnung* und *Gesteigert*) ein signifikantes Ergebnis liefert. Hm..., was entspannendes wäre gut. Auf seinem Second Screen läuft Jagd auf roter Oktober und Paula schaufelt Smarties. Nicht effizient, aber gut.

Bestandsdichte	Fettgehalt
Verordnung	32.3
Gesteigert	60.8
Gesteigert	27.2
Gesteigert	36.3
Verordnung	36.5
Verordnung	34.8
Gesteigert	51.1
Verordnung	29.7
Verordnung	25.3
Verordnung	22.0
Gesteigert	47.0
Gesteigert	44.7
Verordnung	19.4
Gesteigert	37.4
Verordnung	36.1
Gesteigert	55.9
Verordnung	32.9
Verordnung	20.5
Gesteigert	40.0
Verordnung	12.3
Gesteigert	48.8
Gesteigert	56.4

Leider kennt sich Paula 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 Student t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 2.68$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effekt des Student t-Tests! **(1 Punkt)**
6. Wenn Sie *keinen* Unterschied zwischen den Behandlungsgruppen erwarten würden, wie groß wäre dann der Effekt? Begründen Sie Ihre Antwort! **(2 Punkte)**
7. Formulieren Sie eine Antwort an Paula über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**



## 69. Aufgabe

(12 Punkte)

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



**Berechnung des Welch t-Test** Das Emsland, unendliche Weiten. Wir schreiben das Jahr 2024. Dies sind die Abenteuer von Mark, der mit seiner 1 Mann starken Besatzung 12 Wochen lang unterwegs ist, um neue Welten zu erforschen, neues Leben und neue Zivilisationen. 'Oder nennen wir es Ödnis und Verzweiflung', denkt Mark. Für seiner Hausarbeit ist Mark ins Nichts gezogen. Mark und die Unsicherheit, eine unendliche Geschichte mit kniffligen Wendungen. Was macht er nun? Mark hat einen Leistungssteigerungsversuch mit Hühnern durchgeführt. Die Behandlung Ernährungszusatz (*ctrl* und *fedX*) wurde an Hühnern getestet. Gemessen hat er dann als ein normalverteiltes Outcome (*Y*) Gewichtszuwachs in der 1LW. Jetzt soll er seiner Betreuerin nach testen, ob die Behandlung Ernährungszusatz (*ctrl* und *fedX*) ein signifikantes Ergebnis liefert. Hm..., was entspannendes wäre gut. 'Hm...', Marzipankugeln und Andrea Berg. Das ist und bleibt die beste Kombination zum Nachdenken für Mark.

Ernährungszusatz	Gewichtszuwachs
ctrl	24.0
ctrl	22.7
ctrl	23.3
fedX	24.9
fedX	34.7
fedX	15.7
fedX	29.8
ctrl	29.9
fedX	46.4
ctrl	20.3
ctrl	20.9
fedX	28.0
fedX	26.0
fedX	42.5
fedX	10.9
fedX	7.3
ctrl	31.0
ctrl	25.4
fedX	20.7

Leider kennt sich Mark 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 Welch t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 1.96$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie das 95% 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 Mark über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

## 70. Aufgabe

(11 Punkte)

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



**Berechnung des gepaarten t-Test** Jessica und Paula haben sich dazu entschieden zusammenzuarbeiten. Das sollte alles etwas einfacher machen. Jeder hat zwar ein getrenntes Themenfeld aber den Hauptversuch machen beide gemeinsam. Das hat sich schonmal als gute Idee soweit herausgestellt. In einer Abschlussarbeit sollen beide herausfinden, ob es einen Zusammenhang zwischen Ausgeizen (*ctrl* und 28d) und Chlorophyllgehalt (SPAD-502Plus) [SPAD] gibt. Die Besonderheit ist hierbei, dass die Messungen an der gleichen Beobachtung stattfinden. Beide messen also zweimal an den gleichen Erdbeeren. Hier muss dann wohl auf einen normalverteilten Messwert ( $Y$ ) ein gepaarter t-Test gerechnet werden. Jessica schaut etwas flehentlich zu Paula. Eine echte Herausforderung für sie war schon immer der Mangel gewesen. Ein leidiges Lied.. Steffen denkt derweil angestrengt an White Lies und wippt leicht mit dem Fuß.

ID	treatment	freshmatter
1	ctrl	42.3
3	28d	7.3
7	28d	23.7
8	ctrl	44.6
6	28d	22.5
7	ctrl	22.1
2	28d	22.9
9	ctrl	14.3
3	ctrl	41.8
4	28d	16.8
5	28d	26.1
4	ctrl	45.7
5	ctrl	28.2
8	28d	21.0
2	ctrl	48.2
6	ctrl	17.7
1	28d	42.0

Leider kennen sich Jessica und Paula mit der Berechnung eines gepaarten t-Tests überhaupt nicht aus. Deshalb brauchen sie beide 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 gepaarten t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 1.96$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Schätzen Sie den  $p$ -Wert des gepaarten t-Tests ab! Begründen Sie Ihre Antwort mit einer Skizze! **(2 Punkte)**
6. Formulieren Sie eine Antwort an Jessica über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

## 71. Aufgabe

(10 Punkte)

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



**Interpretation des t-Tests in R - die Teststatistik und der p-Wert** 'Wir waren im Wendland um Lauch in einem Versuch in einer Klimakammer zu messen.', Jessica legt das Dokument auf den Tisch und schaut Jonas und Nilufar fragend an. Beide schauen fragend zurück. Gäbe es die Erwartung nicht, dann wäre es für Nilufar irgendwie einfacher hier zu helfen. Echt unangenehm. Die beiden sind zu Jessica gekommen, da sie sich nicht mit R auskennen und daher Hilfe bei der Interpretation des t-Tests brauchen. Im Hintergrund wummert David Bowie und leere Schokobons Packungen stapeln sich auf dem Boden. 'Kein Problem', sagt Jessica und streichelt langsam die Hündin. 'Aber worum es in dem Versuch geht, lässt sich nur aus dem Text in seiner Hand erahnen.' merkt sie an. Vielleicht hilft da ja die Ausgabe des t-Tests in R weiter. Draußen geht blutrot die Sonne unter.

```
##
## Two Sample t-test
##
## data: Chlorophyllgehalt by Bewässerungstypen
## t = -2.4871, df = 16, p-value = 0.02429
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -27.808376 -2.216624
## sample estimates:
## mean in group low mean in group high
## 28.1500 43.1625
```

Helfen Sie Jessica bei der Interpretation des t-Tests! Sonst geht es auch für Jonas und Nilufar nicht weiter.

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Skizzieren Sie eine Abbildung in der Sie  $T_D$ ,  $Pr(D|H_0)$ ,  $A = 0.95$ , sowie  $T_{\alpha=5\%} = |2.12|$  einzeichnen! **(4 Punkte)**
5. Beschriften Sie die Abbildung! **(1 Punkt)**
6. Berechnen Sie den Effekt des t-Tests! **(1 Punkt)**

## 72. Aufgabe

(10 Punkte)

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



**Interpretation des t-Tests in R - das 95% Konfidenzintervall** Almería. Spanien. Sonne und Strand. Steffen und Mark haben ihren gemeinsamen Auslandsaufenthalt sichtlich genossen. Dann hatte sich auch noch angeboten ihre Abschlussarbeit gemeinsam in Almería durchzuführen. Es hätte sogar noch besser funktioniert, wenn Yuki nicht die Faulheit ein paar Mal im Weg gestanden hätte und Mark nicht das Problem gehabt hätte der Perfektionismus zu händeln. Nun müssen jetzt alle Daten in R ausgewertet werden, da R international der Standard in der Datenauswertung ist und die Betreuer in Spanien nur R können. Während beide Yuki Oliven mit Reese's Peanut Butter Cups füttern, hoffen Steffen und Mark mehr Informationen von Yuki über die seltsame R Ausgabe des t-Tests. Immerhin erinnern beide sich an die Behandlung Bestandsdichte (*Verordnung* und *Gesteigert*) und das es um Fleischrindern ging. Im Hintergrund wummert London Grammar und Fotos zeigen Yuki mit dem Hobby Orchideen.

```
##
## Two Sample t-test
##
## data: Schlachtgewicht by Bestandsdichte
## t = -2.5785, df = 17, p-value = 0.01953
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -17.604428 -1.760017
## sample estimates:
## mean in group Verordnung mean in group Gesteigert
## 34.04000 43.72222
```

Helfen Sie Yuki bei der Interpretation des t-Tests! Sonst geht es auch für Steffen und Mark nicht weiter.

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

### 73. Aufgabe

(9 Punkte)

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



**Interpretation des t-Tests in R - die Visualisierung** 'Mit dem R Paket {emmeans} können wir gleich die Gruppenvergleiche rechnen und uns das *compact letter displac*' ausgeben lassen!', verkündet Steffen sichtlich stolz. Ein paar Mal hat sie schon die Romantik gehindert weiterzumachen. 'Nach Meinung des Betreuers soll es aber nur erstmal ein t-Test sein. Und die Ausgabe ist schon wirr genug.', merkt Alex an. Alex und Tina sind bei Steffen um sich in R helfen zu lassen. Im Hintergrund wummert Taylor Swift. Tina streichelt zur Beruhigung die Schlange von Steffen. Die beiden waren 2 Monate in der Uckermark um einen Versuch mit Erdbeeren in einem Feldexperiment durchzuführen. Ziel war es das Outcome (Y) Frischegewicht [kg/ha] zu bestimmen. Steffen überlegt, ob er die beiden nicht noch auf den Film *Harry Potter* einlädt oder dann doch lieber raus geht um zu Ringen? Vielleicht will ja Tina mit. Besser als der Film.

```
##  
## Two Sample t-test  
##  
## data: Frischegewicht by Substrattypen  
## t = 2.2758, df = 16, p-value = 0.03696  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## 0.5984699 16.8765301  
## sample estimates:  
## mean in group torf mean in group 70p30n  
## 40.3875 31.6500
```

Helfen Sie Steffen bei der Interpretation des t-Tests! Sonst geht es auch für Alex und Tina nicht weiter.

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

## 74. Aufgabe

(10 Punkte)

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



**Interpretation des gepaarten t-Tests in R** Es gibt ja immer die Möglichkeit sich Hilfe zu holen. Das geht natürlich auch immer in einer Abschlussarbeit. Deshalb arbeiten Jessica und Nilufar gemeinsam an einer Abschlussarbeit. Das macht dann auch die Analyse ihres Hauptversuches einfacher. Zwar hat jeder von ihnen noch ein Subthema, aber auch da kann man sich ja helfen. Das hilft dann teilweise nur bedingt. Eine echte Herausforderung für sie war schon immer der Mangel gewesen. Ein leidiges Lied. In dem Hauptversuch wurde Folgendes von den beiden gemacht. Jessica und Nilufar haben sich Kartoffeln angeschaut. Dabei geht um Zusammenhang zwischen Ausgeizen (*ctrl* und *28d*) und Proteingehalt [g/kg]. Jetzt sollen beide einen gepaarten t-Test rechnen. Leider kennen sich beide nicht sehr gut in R aus. Aber wenigstens haben beide eine Menge an Schokobons und in der Wohnung wummert David Bowie.

```
##
## Paired t-test
##
## data: Proteingehalt by Ausgeizen
## t = -0.68803, df = 8, p-value = 0.5109
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -13.828462 7.472906
## sample estimates:
## mean difference
## -3.177778
```

Jetzt brauchen Jessica und Nilufar Ihre Hilfe bei der Berechnung eines gepaarten t-Tests in R um ihre Arbeit dann in diesem Semester noch abschließen zu können.

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Skizzieren Sie das sich ergebende 95% Konfidenzintervall! **(2 Punkte)**
5. Interpretieren Sie den Effekt des gepaarten t-Tests! **(2 Punkte)**
6. Skizzieren Sie den sich ergebenden Boxplot der Differenzen! Welche Annahmen an die Daten haben Sie getroffen? Begründen Sie Ihre Antwort! **(2 Punkte)**

## Teil IV.

# Die einfaktorielle & zweifaktorielle ANOVA

### 75. Aufgabe

(11 Punkte)

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



**Visualisierung der einfaktoriellen ANOVA** Nilufar und Yuki schauen sich etwas entnervt an. Gemeinsam schreiben die beiden ihre Abschlussarbeit und sollen nun als erstes einmal die Daten visualisieren damit abgeschätzt werden kann, ob überhaupt signifikante Ergebnisse zu erwarten sind. Die beiden waren im Wendland um ein Stallexperiment mit Hühnern durchzuführen. Dabei haben Nilufar und Yuki den Messwert Protein/Fettrate [%/kg] unter der Behandlung Elterlinie (*Standard*, *Yray* und *Xray*) ermittelt. Kennengelernt haben sich die beiden auf einem Konzert von London Grammar. Später wird noch Matrix geguckt. Yuki befürwortet das!

Elterlinie	Protein/Fettrate
Yray	40
Yray	39
Yray	41
Yray	40
Yray	40
Standard	46
Xray	36
Standard	44
Xray	35
Xray	35
Standard	43
Xray	35
Standard	48
Yray	40
Standard	44
Xray	36
Xray	35
Standard	40
Xray	36
Standard	47

Leider kennen sich Nilufar und Yuki mit Darstellung einer einfaktoriellen ANOVA überhaupt nicht aus.

1. Erstellen Sie eine Visualisierung der Datentabelle! Beschriften Sie die Abbildung! **(2 Punkte)**
2. Benennen Sie die Visualisierung mit dem korrekten, statistischen Fachbegriff! **(1 Punkt)**
3. Zeichnen Sie folgende statistischen Maßzahlen passend ein!
  - Globale Mittelwert:  $\beta_0$  **(1 Punkt)**
  - Mittelwerte der einzelnen Behandlungsstufen:  $\bar{y}_{0.5}$ ,  $\bar{y}_{1.5}$  und  $\bar{y}_{2.5}$  **(1 Punkt)**
  - Mittelwertsdifferenz der einzelnen Behandlungsstufen:  $\beta_{0.5}$ ,  $\beta_{1.5}$  und  $\beta_{2.5}$  **(1 Punkt)**
  - Residuen oder Fehler:  $\epsilon$  **(1 Punkt)**
4. Liegt ein *vermutlicher* signifikanter Unterschied vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Schätzen Sie die Effekte der Behandlungsstufen! **(2 Punkte)**

## 76. Aufgabe

(9 Punkte)

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



**Ergebnistabelle der einfaktoriellen ANOVA** Tina und Jonas schauen sich etwas entnervt an. Gemeinsam schreiben die beiden ihre Abschlussarbeit und sollen nun als erstes einmal die Daten mit einer einfaktoriellen ANOVA auswerten damit abgeschätzt werden kann, ob überhaupt signifikante Ergebnisse in den multiplen Gruppenvergleichen zu erwarten sind. Da hilft das Meerschweinchen von Jonas auch nur bedingt. Die beiden waren im Wendland um ein Feldexperiment mit Maiss durchzuführen. Dabei haben Tina und Jonas den Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] unter der Behandlung Genotypen (AA, AB und BB) ermittelt. Nachher wollen sich beide noch mit dem Hobby Stricken von Jonas beschäftigen. Kennt Tina noch nicht, klingt aber interessant.

Leider kennen sich Tina und Jonas mit Berechnung einer einfaktoriellen ANOVA überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe, das Meerschweinchen reicht als Hilfe nicht aus!

1. Formulieren Sie die wissenschaftliche Fragestellung! (1 Punkt)
2. Formulieren Sie das statistische Hypothesenpaar! (1 Punkt)
3. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus! (3 Punkte)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Genotypen</b>	2	469.32			
<b>error</b>	17	420.48			
<b>Total</b>	19				

4. Schätzen Sie den p-Wert der Tabelle mit  $F_{\alpha=5\%} = 3.59$  ab. Begründen Sie Ihre Antwort! (2 Punkte)
5. Berechnen Sie den Effektschätzer  $\eta^2$ . Was sagt Ihnen der Wert von  $\eta^2$  aus? (2 Punkte)



## 77. Aufgabe

(12 Punkte)

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



**Die einfaktoriellen ANOVA und der Student t-Test** Steffen und Jessica schauen sich etwas entnervt an. Gemeinsam schreiben die beiden ihre Abschlussarbeit und sollen nun als erstes einmal die Daten mit eine einfaktoriellen ANOVA auswerten damit abgeschätzt werden kann, ob überhaupt signifikante Ergebnisse in den multiplen Gruppenvergleichen zu erwarten sind. Deshalb erstmal Schokobons mampfen, die Jessica mitgebracht hat. Nun möchte erstmal ihre Betreuung der Arbeit eine ANOVA Tabelle sehen. Was immer da auch drin zu erkennen sein mag. Steffen schaut Jessica sehen erstmla gar nichts. Die beiden waren im Wendland um ein Stallexperiment mit Milchvieh durchzuführen. Dabei haben Steffen und Jessica den Messwert Schlachtgewicht [kg] unter der Behandlung Lüftungssystem (*keins*, *storm*, *tornado* und *thunder*) ermittelt. Später wollen die beiden dann noch raus um Rad zu fahren.

Leider kennen sich Steffen und Jessica mit Berechnung einer einfaktoriellen ANOVA überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus! **(3 Punkte)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lüftungssystem	3	613.11			
Error	17	450.7			

4. Schätzen Sie den p-Wert der Tabelle mit  $F_{\alpha=5\%} = 3.2$  ab. Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Was bedeutet ein signifikantes Ergebnis in einer einfaktoriellen ANOVA? **(1 Punkt)**
6. Berechnen Sie *einen* Student t-Test für den *vermutlich* signifikantesten Gruppenvergleich anhand der untenstehenden Tabelle mit  $T_{\alpha=5\%} = 2.03$ . Begründen Sie Ihre Auswahl! **(3 Punkte)**

Lüftungssystem	Fallzahl (n)	Mittelwert	Standardabweichung
keins	5	17.4	5.55
storm	5	8.2	4.60
tornado	6	16.5	5.79
thunder	5	4.6	4.34

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

## 78. Aufgabe

(9 Punkte)

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



**Die einfaktorielle ANOVA in R** Yuki schaut sich fragend in der Bibliothek um. Yuki hatte gehofft, dass jemand hier sein würde, den er kennt und sich mit R auskennt. Wird aber enttäuscht. Yuki war im Wendland um ein Feldexperiment mit Brokkoli durchzuführen. Nun möchte sein Betreuer seinem Projektbericht erstmal eine ANOVA sehen und die Ergebnisse präsentiert bekommen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Genotypen (*AA*, *AB* und *BB*) und dem Messwert Trockengewicht [kg/ha] gibt. Und eigentlich wollte Yuki doch noch zum Sport! Um zu Boldern geht Yuki dann später nochmal raus. Echte Entspannung.

```
## Analysis of Variance Table
##
## Response: Trockengewicht
##      Df Sum Sq Mean Sq F value Pr(>F)
## Genotypen  2  72.025   36.012   3.0829  0.067
## Residuals 21 245.308   11.681
```

Leider kennen sich Yuki mit Berechnung einer einfaktoriellen ANOVA überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Interpretieren Sie das Ergebnis der einfaktoriellen ANOVA! **(2 Punkte)**
4. Berechnen Sie den Effektschätzer  $\eta^2$ . Was sagt Ihnen der Wert von  $\eta^2$  aus? **(2 Punkte)**
5. Skizzieren Sie eine Abbildung, der dem obigen Ergebnis der einfaktoriellen ANOVA näherungsweise entspricht! **(3 Punkte)**

## 79. Aufgabe

(12 Punkte)

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



**Ergebnistabelle der zweifaktoriellen ANOVA** In einen Versuch in einer Klimakammer wurden Lauch mit dem Behandlung Genotypen (AA, AB und BB) sowie der Behandlung Düngestufen (*ctrl*, und *high*) untersucht. Es wurde als Messwert Trockengewicht [kg/ha] bestimmt. Jessica ahnte schon, dass es komplexer wird, als sie mit ihrem Projektbericht angefangen hat. Das es jetzt aber so kompliziert wird, hätte sie jetzt aber auch nicht gedacht. Jessica kratzt sich am Kopf. Jessica mampft aus Frust noch eine Handvoll Schokobons. Eventuell muss sie dann doch nochmal Hilfe in der statistischen Beratung holen. Jetzt versucht sie es aber erstmal selber. Und eigentlich wollte Jessica doch noch ihrem Hobby nachgehen! Warhammer. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Jessica denkt gerne über Warhammer nach.

Leider kennen sich Jessica mit Berechnung einer zweifaktoriellen ANOVA überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! (1 Punkt)
2. Formulieren Sie das statistische Hypothesenpaar! (1 Punkt)
3. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus! (3 Punkte)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Genotypen</b>	3	130.33			
<b>Düngestufen</b>	1	3.75			
<b>Genotypen:Düngestufen</b>	3	69.98			
<b>Error</b>	18	214.25			

4. Schätzen Sie den p-Wert der Tabelle ab. Begründen Sie Ihre Antwort! (3 Punkte)

	$F_{\alpha=5\%}$
<b>Genotypen</b>	4.26
<b>Düngestufen</b>	3.40
<b>Genotypen:Düngestufen</b>	5.23

5. Was bedeutet ein signifikantes Ergebnis in einer zweifaktoriellen ANOVA? (2 Punkte)
6. Was sagt der Term *Genotypen:Düngestufen* aus? Interpretieren Sie das Ergebnis! (2 Punkte)

## 80. Aufgabe

(10 Punkte)

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



**Die zweifaktorielle ANOVA in R** In ein Freilandversuch wurden Spargel mit der Behandlung Lichtstufen (*none*, *200lm*, *400lm* und *600lm*) sowie der Behandlung Lüftungssystemen und Folientunneln (*ctrl*, und *tornado*) untersucht. Es wurde als Messwert Proteingehalt [g/kg] bestimmt. Jetzt starrt Alex mit auf die R Ausgabe einer zweifaktoriellen ANOVA. Leider starrt seine Betreuerin in der gleichen Art Alex zurück an. Das wird ein langer Nachmittag, denkt er sich und kreuselt seinen Mund. 'Und was machen wir jetzt?' entfährt es ihm überrascht entnervt. Immerhin war geht es ja um der Projektbericht. Alex hätte doch nichts mit Spargel machen sollen. Spargel – was soll das auch bedeutendes sein? Eigentlich wollte Alex nachher noch einen Film schauen. Das Verrückte ist, dass die Katze Alien wirklich liebt. Das ist Alex sehr recht, denn er braucht Entspannung.

```
## Analysis of Variance Table
##
## Response: Proteingehalt
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Lichtstufen    2 364.67  182.337   21.723 1.589e-05
## Lüftungssystemen 1  97.97   97.974   11.672 0.0030772
## Lichtstufen:Lüftungssystemen 2 248.00 124.001   14.773 0.0001597
## Residuals     18 151.09    8.394
```

Leider kennt sich Alex mit Berechnung einer zweifaktoriellen ANOVA überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Interpretieren Sie das Ergebnis der einfaktoriellen ANOVA! **(3 Punkte)**
4. Zeichnen Sie eine Abbildung, der dem obigen Ergebnis der zweifaktoriellen ANOVA näherungsweise entspricht! **(5 Punkte)**

## 81. Aufgabe

(12 Punkte)

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



**Zusammenhang zwischen der ANOVA und dem t-Test** Paula schaut konzentriert auf die Formeln der ANOVA und des t-Tests. In ihrem Experiment wurde als Messwert Proteingehalt [g/kg] bestimmt. Eine echte Herausforderung für sie war schon immer der Perfektionismus gewesen. Ein leidiges Lied. Dann wäre es nicht noch komplizierter. In ein Gewächshausexperiment wurden Maiss mit der Behandlung Bewässerungstypen (*ctrl*, *low*, *mid* und *high*) sowie der Behandlung Genotypen (*AA*, und *BB*) untersucht. Beide Verfahren müssen etwas miteinander zu tun haben und ihr Betreuer möchte das jetzt auch noch verstehen. Im Hintergrund läuft leise Jagd auf roter Oktober auf ihrem Second Screen. Immerhin hat sie die beiden Formeln vorliegen.

### Gegebene Formeln

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

Leider kennen sich Paula mit dem Zusammenhang zwischen der ANOVA und dem t-Test nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

1. Welche statistische Maßzahl testet der t-Test, welche die ANOVA? Begründen Sie Ihre Antwort! **(2 Punkte)**
2. Erklären Sie den Zusammenhang zwischen der  $F_D$  Statistik und  $T_D$  Statistik! **(2 Punkte)**
3. Visualisieren Sie in einer 2x2 Tafel den Zusammenhang von  $MS_{treatment}$  und  $MS_{error}$ ! **(2 Punkte)**
4. Beschriften Sie die erstellte 2x2 Tafel mit signifikant und nicht signifikant! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Nennen Sie das numerische Minimum der F-Statistik  $F_D$ ! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Wenn die F-Statistik  $F_D$  minimal ist, welche Aussage erhalten Sie über die Nullhypothese? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 82. Aufgabe

(11 Punkte)

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



**Interaktion in der zweifaktoriellen ANOVA** In ein Feldexperiment wurden Spargel mit der Behandlung Bewässerungstypen (*ctrl*, *low*, *mid* und *high*) sowie der Behandlung Lichtstufen (*none*, und *600lm*) untersucht. Es wurde als Messwert Frischgewicht [kg/ha] bestimmt. Jetzt starrt Tina mit auf die R Ausgabe einer zweifaktoriellen ANOVA. Leider starrt ihr Betreuer in der gleichen Art Tina zurück an. Es liegt anscheinend eine signifikante Interaktion vor? 'Das wird ein langer Nachmittag.', denkt sie sich und kreuselt ihren Mund. 'Und was machen wir jetzt?' entfährt es ihr überrascht entnervt. Immerhin war geht es ja um ihre Hausarbeit. Tina hätte doch nichts mit Spargel machen sollen. Spargel – was soll das auch bedeutendes sein? Eigentlich wollte Tina nachher noch einen Film schauen. Wenn Indiana Jones läuft, dann ist die Spinne nicht mehr da. Aber jetzt braucht sie mal Entspannung!

Leider kennen sich Tina und ihr Betreuer mit der zweifaktoriellen ANOVA überhaupt nicht aus. Geschweige denn mit der Interpretation einer Interaktion. Deshalb braucht sie bei der Erstellung Ihre Hilfe, sonst wird es heute Abend mit seinem Hobby Astronomie nichts mehr!

1. Visualisieren Sie folgende mögliche Interaktionen zwischen den Behandlungen! Beschriften Sie die Abbildung! **(4 Punkte)**
  - a) Keine Interaktion liegt vor.
  - b) Eine schwache Interaktion liegt vor.
  - c) Eine starke Interaktion liegt vor.
2. Erklären Sie den Unterschied zwischen den verschiedenen Interaktionen! **(2 Punkte)**
3. Welche statistische Maßzahl betrachten Sie für die Bewertung der Interaktion? **(1 Punkt)**
4. Skizzieren Sie die notwendigen Funktionen in R für eine Post-hoc Analyse! **(2 Punkte)**
5. Wenn eine signifikante Interaktion in den Daten vorliegt, wie ist dann das weitere Vorgehen? Berücksichtigen Sie auch die Funktion `emmeans()`! **(2 Punkte)**

### 83. Aufgabe

(11 Punkte)

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



**Zusammenhang zwischen der ANOVA und dem Post-hoc-Test** 'Mit der einfaktoriellen ANOVA lassen sich flott die Gruppen in einer Behandlungen vergleichen, wenn wir normalverteilte Daten und Varianzhomogenität vorliegen haben!', seine Betreuerin scheint die einfaktoriellen ANOVA zu verstehen. Warum jetzt er jetzt nochmal alles wiederkauen muss, wird Yuki echt nicht so klar. Wenn es doch so klar ist? 'Wir haben jetzt bei der ANOVA einen p-Wert mit 0.052 raus sowie eine F-Statistik  $F_D$  mit 1.61 berechnet. Nach den Boxplots müsste sich eigentlich ein Unterschied zwischen *Yray* und *ctrl* ergeben. Der Unterschied ist in {emmeans} auch signifikant mit einem p-Wert von 0.045. Wie kann das sein?', fragt Yuki etwas provokant und dreht London Grammar leiser. Yuki war im Emsland und hatte dort ein Kreuzungsexperiment mit Puten durchgeführt. Die Komune wo er untergekommen war, war cool gewesen. Dort gab es selbstgemachte Reese's Peanut Butter Cups aus Vollkorn! Nur jetzt muss eben das Experiment fertig ausgewertet werden. Yuki hatte eine Behandlungen Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) auf Puten angewendet. Gemessen wurde der Messwert (*Y*) Protein/Fettrate [%/kg]. Dabei wurden die Daten *D* erhoben. Jetzt muss das hier zu einem Ende kommen! Yuki hat schon genug Probleme. Wenn die Faulheit nicht wäre, dann wäre es einfacher.

#### Gegebene Formeln

$$MS_{treatment} = \frac{SS_{treatment}}{df_{treatment}} \quad MS_{error} = \frac{SS_{error}}{df_{error}} \quad F_D = \frac{MS_{treatment}}{MS_{error}}$$

Leider kennen sich Yuki und seine Betreuerin mit der Interpretation einer ANOVA überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe und die Zeit wird knapp.

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Was bedeutet eine signifikante ANOVA für die beobachteten Daten *D*? **(1 Punkt)**
4. Visualisieren Sie den Unterschied zwischen Varianzhomogenität und Varianzheterogenität anhand der Daten *D*! Beschriften Sie die Abbildung! **(2 Punkte)**
5. Visualisieren Sie für die Daten *D* die Verletzung der Annahme der Varianzhomogenität der ANOVA unter zu Hilfenahme von Boxplots! Beschriften Sie die Abbildung! **(2 Punkte)**
6. Welche Auswirkung hat die Verletzung der Annahme der Varianzhomogenität für die Teststatistik  $F_D$  der ANOVA? Begründen Sie Ihre Antwort! **(2 Punkte)**
7. Erklären Sie abschließend die Diskrepanz zwischen den Ergebnis der ANOVA und dem paarweisen Gruppenvergleich in {emmeans}! **(2 Punkte)**

## Teil V.

# Multiple Gruppenvergleiche

### 84. Aufgabe

(12 Punkte)

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



**Adjustierung multipler Vergleiche** Jonas und Yuki untersuchen gemeinsam in ihrer Abschlussarbeit den Messwert Frischgewicht [kg/ha] in Lauch. Es ist einiges schiefgelaufen, wie es immer so passiert. Hauptsächlich waren es Katzen, auch wenn man erstmal bei dem Messwert nicht unbedingt an Katzen denken würde. Aber das ist eine andere Geschichte. Jetzt wollen Jonas und Yuki ihre Ergebnisse nochmal mit einer Studie von Meyer et al. (2021) vergleichen und schauen, ob was ähnliches rauskommt. Angeschaut wurde sich als Behandlung Bewässerungstypen (*none*, *ctrl*, *low*, *mid* und *high*). Es ergab sich dann die folgende Tabelle der rohen  $p$ -Werte für die Vergleiche zu Meyer et al. (2021).

Rohen $p$ -Werte	Adjustierte $p$ -Werte	Nullhypothese ablehnen?
0.0020		
0.0700		
0.0120		
0.2300		
0.0012		

Leider kennen sich Jonas und Yuki mit der Adjustierung von  $p$ -Werten und dem Signifikanzniveau  $\alpha$  überhaupt nicht aus. Deshalb brauchen die beiden bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Füllen Sie die Spalte *Adjustierte  $p$ -Werte* nach der Bonferoni-Methode aus! **(2 Punkte)**
4. Entscheiden Sie, ob nach der Adjustierung die Nullhypothese abgelehnt werden kann! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Wie ist Ihr Vorgehen, wenn Sie anstatt der  $p$ -Werte das Signifikanzniveau  $\alpha$  adjustieren? **(2 Punkte)**
6. Erklären Sie warum die  $p$ -Werte oder das Signifikanzniveau  $\alpha$  bei multiplen Vergleichen adjustiert werden müssen! **(2 Punkte)**
7. Würden Sie die Adjustierung der  $p$ -Werte oder die Adjustierung des Signifikanzniveaus  $\alpha$  vorziehen? Begründen Sie Ihre Antwort! **(2 Punkte)**



## 85. Aufgabe

(10 Punkte)

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



**Visualisierung des Compact Letter Displays (CLD)** Jessica sitzt schon etwas länger bei ihrer Betreuerin. So langsam macht Jessica sich Gedanken, ob sie nicht doch mal anmerken sollte, dass sie von CLD noch nie was gehört hat. Aber noch kann gelauscht werden, ein Ende ist erstmal nicht in Sicht! Jessica hatte in ihrem Projektbericht ein Feldexperiment durchgeführt. Deshalb sitzt sie hier. Also eigentlich nein, deshalb nicht. Jessica will fertig werden. Hat sie sich doch mit Düngestufen (*ctrl*, *low*, *mid* und *high*) und Frischegewicht [kg/ha] schon eine Menge angeschaut. Jessica beugt sich leicht nach vorne. Nein, doch keine Pause. Weiter warten auf eine Lücke im Fluss...

Behandlung	Compact letter display
ctrl	a
low	b
mid	a
high	a

Leider kennen sich Jessica mit dem *Compact letter display (CLD)* überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Zeichnen Sie die sich anhand des *Compact letter display (CLD)* ergebenden Barplots! **(2 Punkte)**
4. Ergänzen Sie das *Compact letter display (CLD)* zu den Barplots! **(1 Punkt)**
5. Erklären Sie *einen* Vorteil und *einen* Nachteil des *Compact letter display (CLD)*! **(2 Punkte)**
6. Erstellen Sie eine Matrix mit den paarweisen *p*-Werten eines Student t-Tests, die sich näherungsweise aus dem *Compact letter display (CLD)* ergeben würde! Begründen Sie Ihre Antwort! **(3 Punkte)**

## 86. Aufgabe

(12 Punkte)

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



**Berechnung des Compact Letter Displays (CLD) anhand von t-Tests** Steffen hatte in seine Abschlussarbeit ein Stallexperiment durchgeführt. Soweit so gut. Dabei hat er sich mit Fleischrindern beschäftigt. Angeblich der neueste heiße Kram... aber das ist wiederum was anderes. So richtig mitgenommen hat Steffen das Thema dann doch nicht. Hat er sich doch mit Ernährungszusatz (*ctrl*, *fedX*, *proteinX* und *getIt*) und Schlachtgewicht [kg] schon eine Menge an Daten angeschaut. Nach seine Betreuerin soll er nun ein CLD bestimmen. Weder weiß er was ein CLD ist, noch war sein erster Gedanke mit Köln und die LGBTQ Community richtig... Als erstes solle er die Gruppen nach absteigender Effektstärke sortieren. Was immer das jetzt bringen soll.

Ernährungszusatz	Fallzahl (n)	Mittelwert	Standardabweichung
ctrl	9	5.59	3.22
fedX	8	5.92	1.09
proteinX	8	5.04	0.87
getIt	8	5.72	2.00

Leider kennen sich Steffen mit dem *Compact letter display (CLD)* überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Zeichnen Sie die sich ergebenden Barplots! **(1 Punkt)**
4. Berechnen Sie die Matrix der *p*-Werte anhand von Student t-Tests! **(4 Punkte)**
5. Ergänzen Sie das *Compact letter display (CLD)* zu den gezeichneten Barplots! Begründen Sie Ihre Antwort! **(4 Punkte)**
6. Interpretieren Sie das *Compact letter display (CLD)* für Steffen und Yuki! **(1 Punkt)**

## 87. Aufgabe

(10 Punkte)

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



**Berechnung des Compact Letter Displays (CLD) anhand der Matrix der p-Werte** 'Oh, nee!', ruft Mark aus und rollt entnervt mit seinen Augen. Mark hatte seine gesamte Analyse in SPSS gerechnet. Das war ja auch alles in Ordnung. Abbildungen haben geklappt und auch die statistischen Tests gingen dann irgendwie doch. Aber das CLD nicht. Mark findet einfach keine Möglichkeit ein CLD in SPSS zu erhalten. Aber seine Betreuerin möchte unbedingt ein CLD. Sonst wird es mit der Abgabe nichts. Dabei hatte er schon wirklich eine Menge gemacht! Mark hatte sich zwei Variablen mit Bewässerungstypen (*ctrl*, *low*, *mid* und *high*) und Frischgewicht [kg/ha] in ein Gewächshausexperiment mit Maiss angeschaut. Wo kriegt er jetzt ein CLD her? Dann eben per Hand aus der Matrix der *p*-Wert. Mark stöhnt...

	<b>ctrl</b>	<b>low</b>	<b>mid</b>	<b>high</b>
<b>ctrl</b>	1.0000000	0.0066727	0.6831850	0.3225004
<b>low</b>	0.0066727	1.0000000	0.0018799	0.0010204
<b>mid</b>	0.6831850	0.0018799	1.0000000	0.5122003
<b>high</b>	0.3225004	0.0010204	0.5122003	1.0000000

Leider kennen sich Mark mit dem *Compact letter display (CLD)* überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Zeichnen Sie die sich anhand der Matrix der *p*-Werte ergebenden Barplots! **(2 Punkte)**
4. Ergänzen Sie das *Compact letter display (CLD)*! Begründen Sie Ihre Antwort! **(4 Punkte)**
5. Interpretieren Sie das *Compact letter display (CLD)* für Mark und Yuki! **(2 Punkte)**

## Teil VI.

# Der Chi-Quadrat-Test & Der diagnostische Test

### 88. Aufgabe

(10 Punkte)

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



**Den Chi-Quadrat-Test berechnen** Am Ende war es für Tina in ihrem Projektbericht dann doch kein normalverteiltes Outcome. Das was jetzt etwas doof, da er sich auf eine ANOVA gefreut hatte. Dann noch schnell Tocotronic auf das Ohr und los gehts. Prinzipiell ginge das auch irgendwie, aber nun möchte ihre Betreuerin gerne einen  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Tina hatte sich in einen Leistungssteigerungsversuch  $n = 114$  Beobachtungen von Zandern angeschaut. Dabei hat sie als Behandlung *Klimakontrolle* [ja/nein] bestimmt und zum anderen die Variable *Schlachtgewicht im Zielbereich* [ja/nein] ermittelt. Jetzt muss Tina mal schauen, wie sie das jetzt rechnet. Eigentlich wollte Tina nachher noch zum Sport. Einfach mal raus um zu Boxen. Ohne Ziel und Uhr. Das ist was für Tina.

	38	11	
	27	38	

Leider kennt sich Tina mit der Berechnung eines  $\chi^2$ -Test für kategoriale Daten überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Visualisieren Sie den Zusammenhang zwischen den beiden kategorialen Variablen! **(2 Punkte)**
4. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test! **(2 Punkte)**
5. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem  $\chi^2_{\alpha=5\%} = 3.83$ ! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Skizzieren Sie in einer Abbildung die  $\chi^2$ -Verteilung, wenn die  $H_0$  wahr ist! Ergänzen Sie  $\chi^2_{\alpha=5\%}$  und  $\chi^2_D$  in der Abbildung! Beachten Sie folgenden Informationen zur  $\chi^2$ -Verteilung. Die  $\chi^2$ -Verteilung hat ein Maxima bei  $\chi^2 = 4$  sowie ein Minima bei  $\chi^2 = 8$ . **(2 Punkte)**

## 89. Aufgabe

(10 Punkte)

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



**Den Chi-Quadrat-Test in einem Fragebogen** Jonas hatte sich gleich von Beginn an in seinem Projektbericht für eine Umfrage im Marketing interessiert. Jetzt geht es um den Haupt- und Nebenerwerb von Erlebnishöfen in Norddeutschland. Viele Höfe haben angefangen auch Großkatzen zu halten, damit mehr Kunden auf die Höfe kommen. Für den Verband der Großkatzenbesitzer e.V. möchte er nun einen Fragebogen zur Zukunftsfähigkeit Schritt für Schritt auswerten. Dabei teilt er zuerst die Antwortenden in die beiden Gruppen 'Höfe mit Großkatzen [ja]' und 'Höfe mit Großkatzen [nein]' ein. Daraufhin möchte er für folgende Frage *f3verband* einmal auswerten, ob es einen Unterschied zwischen den beiden Höfen mit oder ohne Großkatzen gibt.

*Halten Sie die Verbandsarbeit für die Verbreitung von Großkatzen im ländlichen Raum für sinnvoll?*

Jonas kratzt sich an seinem Kopf. Wie soll man eine Tabelle mit so vielen Zahlen sinnvoll auswerten? Schnell noch ein paar Snickers einwerfen und los gehts!

<i>f3verband</i>	trifft gar nicht zu	trifft nicht zu	weder noch	trifft zu	trifft voll zu	
ja	14	23	11	8	4	
nein	12	29	13	10	1	

Leider kennt sich Jonas mit der Berechnung eines  $\chi^2$ -Test auf einer Frage in einem Fragebogen überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Berechnen Sie die Teststatistik  $\chi^2_D$  eines Chi-Quadrat-Test! *Ignorieren Sie Zellbelegungen kleiner gleich fünf in der Berechnung von  $\chi^2_D$ !* **(2 Punkte)**
4. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem  $\chi^2_{\alpha=5\%} = 18.21$ ! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Visualisieren Sie die 2x5 Kreuztabelle *ohne* die Berücksichtigung der Antwortkategorie 'weder noch'! **(2 Punkte)**
6. Berechnen Sie den Effektschätzer *Cramers V* auf der 2x5 Kreuztabelle! **(1 Punkt)**
7. Interpretieren Sie den berechneten Effektschätzer *Cramers V* unter Berücksichtigung der Fragestellung! **(1 Punkt)**

## 90. Aufgabe

(10 Punkte)

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



**Den Chi-Quadrat-Test mit Effektmaß berechnen** 'Der  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet.', liest Tina in ihrer Mitschrift. So richtig helfen tut ihr das jetzt eherlichweise dann doch nicht. Dann noch schnell Tocotronic auf das Ohr und los gehts. Tina hatte sich in ein Feldexperiment  $n = 184$  Beobachtungen von Spargel angeschaut. Dabei hat er als Behandlung *Mechanische Bearbeitung [ja/nein]* bestimmt und zum anderen die Variable *Chlorophyllgehalt unter Zielwert [ja/nein]* ermittelt. Am Ende möchte dann ihre Betreuerin gerne einen  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Eigentlich wollte Tina nachher noch zum Sport. Um zu Boxen geht Tina dann später nochmal raus. Echte Entspannung.

	56	41	
	43	44	

Leider kennt sich Tina mit der Berechnung eines  $\chi^2$ -Test für kategoriale Daten überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test! **(2 Punkte)**
4. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem  $\chi^2_{\alpha=5\%} = 4.56$ ! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effektschätzer *Cramers V* auf der 2x2 Kreuztabelle! **(1 Punkt)**
6. Welchen Wertebereich kann der Effektschätzer *Cramers V* annehmen? Wann liegt kein Effekt und wann ein starker Effekt vor? **(2 Punkte)**
7. Interpretieren Sie den berechneten Effektschätzer *Cramers V* unter Berücksichtigung der Fragestellung! **(1 Punkt)**

## 91. Aufgabe

(10 Punkte)

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



**Den Chi-Quadrat-Test konzeptionell verstehen** Mark hat sich ein Herz gefasst und war für seiner Abschlussarbeit in die Niederlande gegangen. Das war eine super Zeit in der er viel gelernt hat. Klar gab es auch die ein oder andere Besonderheit, aber das gehört hier eher nicht hin. Dann noch schnell Marzipankugeln zur Stärkung und los gehts. Mark ist schon eine ganze Zeit im Büro, da sein Betreuer möchte, dass er jetzt auf seinen Daten mit  $n = 110$  Beobachtungen von Fleischrindern einen  $\chi^2$ -Test rechnet. Das ginge, da er als Behandlung *Automatische Fütterung [ja/nein]* bestimmt und zum anderen die Variable *Gewichtszuwachs erreicht [ja/nein]* ermittelt hat. Wie genau, das ist jetzt eine andere Frage. Am Ende des Tages möchte er dann noch sein Hobby Geocaching genießen. Das muss auch mal sein!

			59
			51
	69	41	110

Leider kennt sich Mark mit der Berechnung eines  $\chi^2$ -Test für kategoriale Daten überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Ergänzen Sie die Felder innerhalb der 2x2 Kreuztabelle, so dass *kein* signifikanter Effekt zu erwarten wäre! **(2 Punkte)**
4. Begründen Sie Ihr Vorgehen an der Formel des Chi-Quadrat-Tests. Erklären Sie Ihr Vorgehen an einem Beispiel! **(2 Punkte)**
5. Visualisieren Sie den Zusammenhang zwischen den beiden kategorialen Variablen! **(2 Punkte)**
6. Was ist die Mindestanzahl an Beobachtungen je Zelle? Wenn in einer der Zellen weniger Beobachtungen auftreten, welchen Test können Sie anstatt des Standard Chi-Quadrat-Tests anwenden? **(2 Punkte)**

## 92. Aufgabe

(10 Punkte)

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



**Der Chi-Quadrat-Test in R** Paula hat sich ein Herz gefasst und war für ihrem Projektbericht in die Niederlande gegangen. Das war eine super Zeit in der sie viel gelernt hat. Klar gab es auch die ein oder andere Besonderheit, aber das gehört hier eher nicht hin. Dann noch schnell Jagd auf roter Oktober starten und los gehts mit der Kraft von Smarties. Paula ist schon eine ganze Zeit im Büro, da ihr Betreuer möchte, dass sie jetzt auf ihren Daten mit  $n = 110$  Beobachtungen von Zandern einen  $\chi^2$ -Test rechnet. Das ginge, da sie als Behandlung *Außenklimakontakt* [*ja/nein*] bestimmt und zum anderen die Variable *Fettgehalt erreicht* [*ja/nein*] ermittelt hat. Wie genau, das ist jetzt eine andere Frage. Nach ihrem Experiment erhielt sie folgende 2x2 Kreuztabelle aus ihren erhobenen Daten.

```
##               Außenklimakontakt
## Fettgehalt erreicht ja  nein
##                ja    17    5
##                nein   4   12
```

Dann rechnete Paula den Fisher-Exakt-Test auf der 2x2-Kreuztabelle in R und erhielt folgende R Ausgabe der Funktion `fisher.test()`.

```
##
## Fisher's Exact Test for Count Data
##
## data:  Fettgehalt erreicht
## p-value = 0.002568
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.85845 61.14631
## sample estimates:
## odds ratio
##  9.451509
```

Leider kennt sich Paula mit der Berechnung eines  $\chi^2$ -Test für kategoriale Daten überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

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



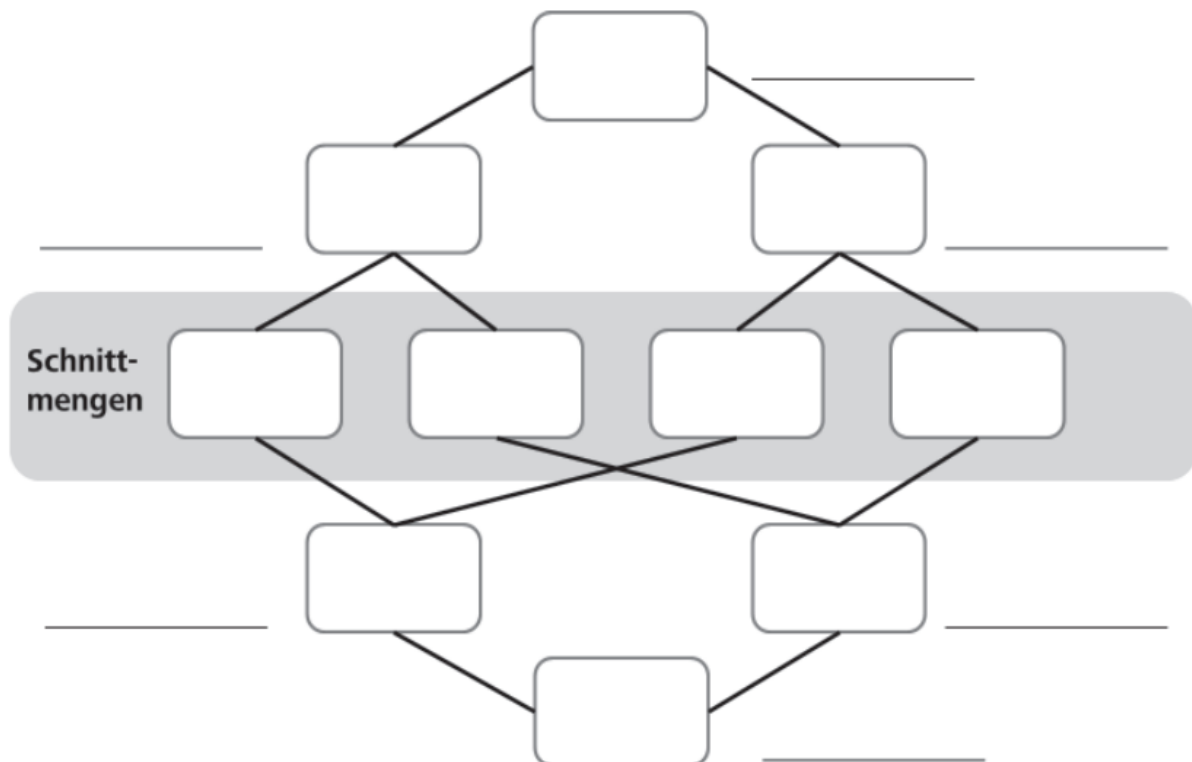
### 93. Aufgabe

(11 Punkte)

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



**Den diagnostische Test am Doppelbaum berechnen** Tina liest laut vor. 'Die Prävalenz von Klauenseuche bei Zandern wird mit 4% angenommen. In 80% der Fälle ist ein Test positiv, wenn das Tier erkrankt ist. In 8.5% der Fälle ist ein Test positiv, wenn das Tier nicht erkrankt ist und somit gesund ist. Wir führen einen Test auf Tuberkulose an 1000 Zandern mit einem diagnostischen Test durch.' Nilufar klappt die Kinnlade runter. In der Stille dudelt Tocotronic. Tina schaut kompetent und schmeißt sich mit offenen Mund Takis Blue Heat an den Kopf vorbei.



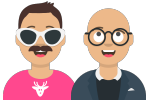
Leider kennen sich Tina und Nilufar mit dem diagnostischen Testen überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe!

1. Beschriften Sie die Äste des Doppelbaumes, mit denen Ihnen bekannten Informationen! (2 Punkte)
2. Beschriften Sie den Doppelbaum! (2 Punkte)
3. Füllen Sie freien Felder des Doppelbaums aus! (4 Punkte)
4. Berechnen Sie die Wahrscheinlichkeit  $Pr(K^+|T^+)$ ! (2 Punkte)
5. Was sagt Ihnen die Wahrscheinlichkeit  $Pr(K^+|T^+)$  aus? (1 Punkt)

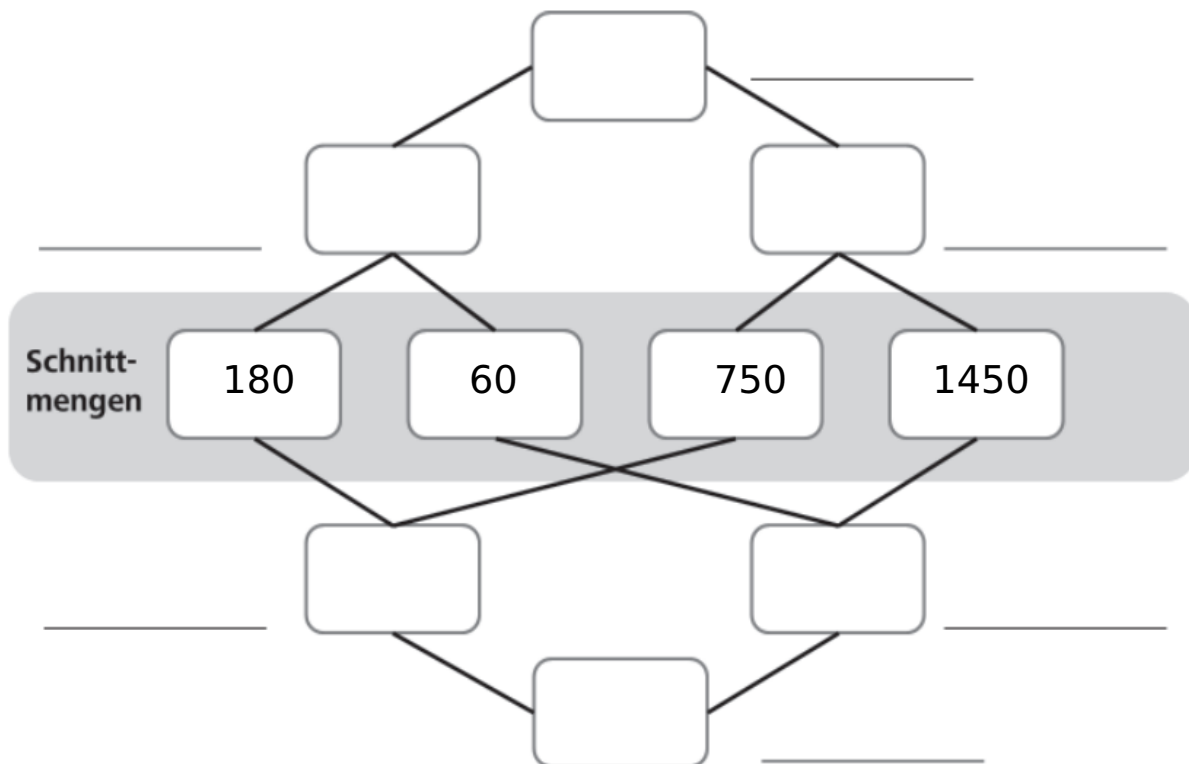
## 94. Aufgabe

(11 Punkte)

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



**Der diagnostische Test und statistische Maßzahlen** 'Was ist denn das?', entfährt es Alex. 'Hm... ich glaube es handelt sich um einen Doppelbaum, den wir beim diagnostischen Testen brauchen.', meint Yuki und dreht Alien auf dem Second Screen etwas leiser. Was jetzt beide von einem diagnostischen Test haben, ist ihnen auch nicht klar. Es ist ja schon alles komplex genug und die Gefälligkeit von Alex macht es heute auch nicht mehr einfacher. 'Es geht um Mehltau an Lauch.', stellt Yuki fest. Eigentlich wollte Yuki eher los um zu Boldern. Das wird aber wohl nichts mehr.



Leider kennen sich Alex und Yuki mit dem diagnostischen Testen überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe!

1. Beschriften Sie den Doppelbaum! **(2 Punkte)**
2. Füllen Sie freien Felder des Doppelbaums aus! **(4 Punkte)**
3. Berechnen Sie die Wahrscheinlichkeit  $Pr(K^+|T^+)$ ! **(2 Punkte)**
4. Berechnen Sie die Prävalenz für Klauenseuche! **(1 Punkt)**
5. Berechnen Sie die Sensifität und Spezifität des diagnostischen Tests! Erstellen Sie dafür zunächst eine 2x2 Kreuztabelle! **(2 Punkte)**

## Teil VII.

# Lineare Regression & Korrelation

### 95. Aufgabe

(10 Punkte)

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



**Visualisierung der linearen Regression** 'Hä? Hatten wir das als Aufgabe nicht schon mal, das wir aus kontinuierlichen Daten eine Abbildung bauen sollten?', fragt Jonas. Tina schaut fragend zurück. 'Kann mich wie immer an nichts erinnern. Können wir trotzdem jetzt erstmal die Daten auswerten? Columbo?', antwortet Tina leicht angespannt. Die beiden hatten ein Freilandversuch im Oldenburger Land mit Kartoffeln durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittlichen Anteil an Ton [%/l] und Proteingehalt [g/kg]. Jetzt haben die beiden eigentlich alles zusammen. *Eigentlich...*

Durchschnittlichen Anteil an Ton [%/l]	Proteingehalt [g/kg]
12.7	8.2
19.9	20.1
14.5	13.8
11.4	10.4
22.0	19.6
14.4	13.5
10.4	8.9
16.8	15.4
8.3	8.0
11.8	12.7
16.0	17.0
11.3	14.4

Leider kennen sich Jonas und Tina mit der linearen Regression für kontinuierliche Daten überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Erstellen Sie eine Visualisierung für die Datentabelle. Beschriften Sie die Achsen! **(2 Punkte)**
3. Schätzen Sie die Regressionsgleichung aus der obigen Abbildung ab! **(2 Punkte)**
4. Beschriften Sie die Grade mit den statistischen Maßzahlen der linearen Regressionsgleichung! **(2 Punkte)**
5. Liegt ein Zusammenhang zwischen x und y vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Wenn kein Zusammenhang zu beobachten wäre, wie würde die Grade aussehen? *Antworten Sie mit einer Skizze der Geraden!* **(1 Punkt)**

## 96. Aufgabe

(12 Punkte)

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



**Interpretation der Ergebnisse einer linearen Regression** 'Ich glaube du bringst da was durcheinander. Wir nutzen zwar auch für die ANOVA die Funktion `lm()` aber hier wollen wir, glaube ich, eine Gerade durch die Punkte zeichnen.', merkt Nilufar an. 'Ich sehe keine Punkte... ich sehe nur zwei Zeilen einer Tabelle und ich glaube du hast gerade was gelöscht.', antwortet Jonas sichtlich übernächtigt. 'Wir müssen die Koeffizienten der linearen Regression ja auch erst interpretieren!', spricht Nilufar sehr deutlich und langsam. Die beiden hatten einen Leistungssteigerungsversuch in der Uckermark mit Zandern durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Tagestemperatur [C/d] und Schlachtgewicht [kg]. Jetzt wollen sie erstmal schauen, ob es einen Zusammenhang gibt und das soll mit der `R` Ausgabe möglich sein.

term	estimate	std.error	t statistic	p-value
(Intercept)	6.66	2.34		
Durchschnittliche Tagestemperatur	1.53	0.23		

Leider kennen sich Nilufar und Jonas mit der linearen Regression für kontinuierliche Daten in `R` überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Erstellen Sie eine Visualisierung der `lm()`-Ausgabe. Beschriften Sie die Achsen! **(2 Punkte)**
3. Beschriften Sie die Visualisierung mit den statistischen Maßzahlen der `lm()`-Ausgabe! **(2 Punkte)**
4. Formulieren Sie die Regressionsgleichung! **(1 Punkt)**
5. Ergänzen Sie die t Statistik in der `lm()`-Ausgabe! **(2 Punkte)**
6. Ergänzen Sie den p-Wert in der `lm()`-Ausgabe mit  $T_{\alpha=5\%} = 1.96$ ! **(2 Punkte)**
7. Interpretieren Sie den p-Wert im Kontext der wissenschaftlichen Fragestellung! **(1 Punkt)**
8. Wie groß ist der Effekt im Kontext der wissenschaftlichen Fragestellung? **(1 Punkt)**

## 97. Aufgabe

(11 Punkte)

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



**Interpretation der Ergebnisse einer linearen Regression in R** 'Wichtig ist es, dass wir jetzt eine Gerade durch die Punkte zeichnen!', ruft Jessica. 'Ich sehe nur Kauderwelsch und keine Punkte. Wie soll ich da denn jetzt eine Gerade durchzeichnen? Und warum überhaupt? War das unsere Fragestellung?', fragt Mark. Jessica atmet schwer ein und starrt auf die R Ausgabe der Funktion `lm()`. Die beiden hatten einen Leistungssteigerungsversuch im Oldenburger Land mit Hühnern durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Tagestemperatur [C/d] und Fettgehalt [%/kg]. Jetzt will die Betreuung von den beiden die Interpretierung der Daten in Form einer linearen Regression gerechnet bekommen. Das haben beide in R gemacht, aber wie soll das jetzt gehen? Das mit der Interpretation?

```
##
## Call:
## Fettgehalt ~ Durchschnittliche_Tagestemperatur
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.86975 -0.44820  0.04785  0.78723  2.77377
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.4135     1.7992  -0.230    0.82
## Durchschnittliche_Tagestemperatur  1.3306     0.1783   7.463 9.75e-09
##
## Residual standard error: 1.279 on 35 degrees of freedom
## Multiple R-squared:  0.6141, Adjusted R-squared:  0.6031
## F-statistic: 55.69 on 1 and 35 DF, p-value: 9.753e-09
```

Leider kennen sich Jessica und Mark mit der linearen Regression für kontinuierliche Daten in R überhaupt nicht aus. Deshalb brauchen beide bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Wie groß ist der Effekt im Kontext der wissenschaftlichen Fragestellung? **(2 Punkte)**
3. Interpretieren Sie die  $p$ -Werte im Kontext der wissenschaftlichen Fragestellung! **(2 Punkte)**
4. Visualisieren Sie die Verteilung der Residuen! **(2 Punkte)**
5. Ist die Annahme der Normalverteilung erfüllt? Begründen Sie die Antwort! **(2 Punkte)**
6. Erklären Sie *kurz* den Begriff R-squared! Was sagt Ihnen der Wert aus? **(2 Punkte)**

## 98. Aufgabe

(10 Punkte)

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



**Interpretation der Ergebnisse einer Korrelationsanalyse in R** 'Hä? Was ist denn das? Das wird ja immer wilder! Hatten wir als Aufgabe eine Korrelation zu berechnen? Wir bauen aus kontinuierlichen Daten eine Abbildung und interpretieren diese dann?', fragt sich Yuki laut. 'Keine Ahnung... das ist jetzt jedenfalls keine Abbildung von irgendwas sondern eine R Ausgabe mit ganz wilden Bezeichnungen...', denkt er. Aber das hilft jetzt auch nicht. Irgendwie komisch, wenn er Matrix anmacht, dann ist das Minischwein eigentlich sofort vor dem Bildschirm und starrt hinein. Yuki hatte einen Versuch in einer Klimakammer in der Uckermark mit Spargel durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittlicher Niederschlag [ml/w] und Trockengewicht [kg/ha]. Jetzt hat er eigentlich alles zusammen. *Eigentlich...*, denn mit der R Ausgabe hat Yuki jetzt ein Problem. Yuki und die Faulheit, eine unendliche Geschichte mit kniffligen Wendungen.

```
##
## Pearson's correlation
##
## data: Durchschnittlicher Niederschlag and Trockengewicht
## t = 6.7147, df = 8, p-value = 0.0001504
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6955371 0.9816197
## sample estimates:
##      cor
## 0.9215781
```

Leider kennt sich Yuki mit der Korrelationsanalyse in R überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Erstellen Sie eine Visualisierung für den Korrelationskoeffizienten! Beschriften Sie die Abbildung! **(2 Punkte)**
4. Nennen Sie die zwei Eigenschaften des Korrelationskoeffizienten! **(2 Punkte)**
5. Interpretieren Sie den Korrelationskoeffizienten hinsichtlich des Effekts und der Signifikanz! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Visualisieren Sie das 95% Konfidenzintervall! Beschriften Sie die Abbildung! **(2 Punkte)**

## 99. Aufgabe

(12 Punkte)

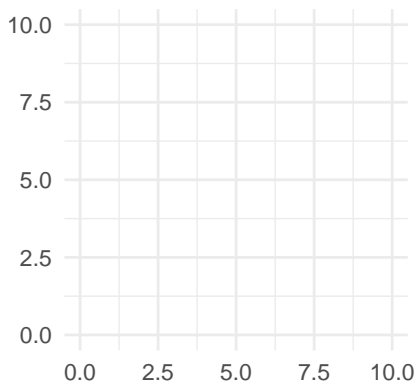
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



**Visualisierung der Korrelation und des Bestimmtheitsmaßes** 'Hm..., drei leere Abbildungen. Was soll ich da jetzt machen?', fragt sich Yuki. Yuki kennt sich nur begrenzt bis gar nicht mit der linearen Regression und Korrelation aus. Dafür mit etwas anderem. Yuki liebt Orchideen. Darin kann sie sich wirklich verlieren und immer wieder neu begeistern. Aber das hilft hier auch nur so halb, die Aufgabe zu lösen und mehr über den Korrelationskoeffizienten zu erfahren.

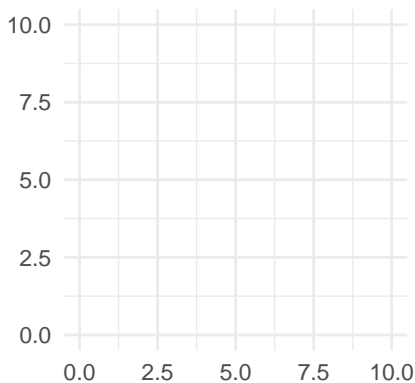
Pearsons  $\rho = 0.25$

$R^2 = 0.25$



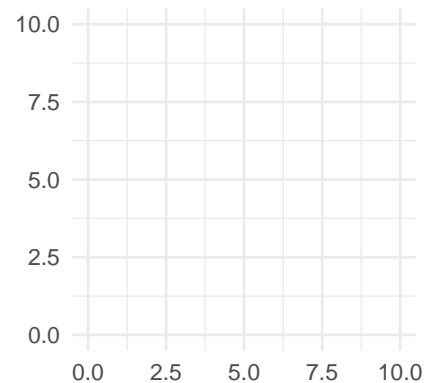
Pearsons  $\rho = 0.5$

$R^2 = 0.75$



Pearsons  $\rho = -0.75$

$R^2 = 1$



**Visualisierung der Korrelation und des Bestimmtheitsmaßes** Leider kennt sich Yuki mit der Korrelationsanalyse und der linearen Regression überhaupt nicht aus. Deshalb braucht sie bei der Auswertung Ihre Hilfe!

1. Zeichnen Sie für die  $\rho$ -Werte eine Gerade in die entsprechende Abbildung! **(3 Punkte)**
2. Zeichnen Sie für die  $R^2$ -Werte die entsprechende Punktwolke um die Gerade! **(3 Punkte)**
3. Nennen Sie die zwei Eigenschaften des Korrelationskoeffizienten! **(2 Punkte)**
4. Interpretieren Sie die  $R^2$ -Werte für die jeweilige Gerade! **(2 Punkte)**
5. Warum müssen Sie ein  $R^2$ -Wert berechnen, wenn Sie die einfachere Möglichkeit der visuellen Überprüfung haben? Begründen Sie Ihre Antwort! **(2 Punkte)**

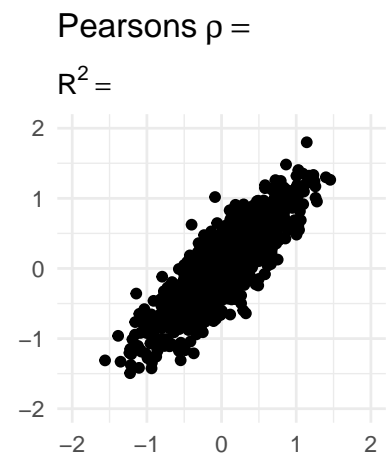
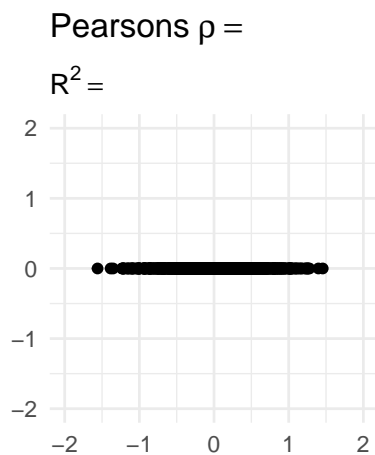
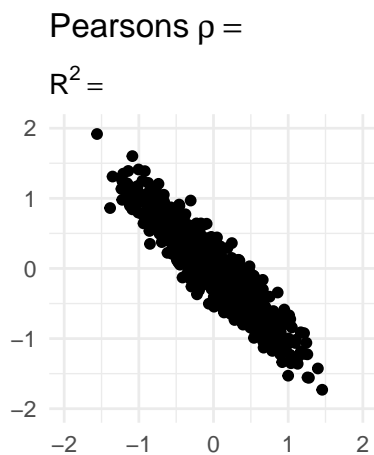
## 100. Aufgabe

(12 Punkte)

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



**Schätzen der Korrelation und des Bestimmtheitsmaßes** Der Bildschirm strahlt blau in das Gesicht von Jonas. Es ist schon spät. Und das hat einen Grund. Schon dutzende Male gesehen: Mission Impossible. Aber immer noch großartig zusammen mit Snickers. Jonas überlegt, aber seine Gedanken sind etwas zäh. 'Was soll das hier alles bedeuten?', fragt sich Jonas. Irgendwie ist ihm nicht klar wie er  $\rho$ -Werte oder  $R^2$ -Werte abschätzen soll. Alles nicht so einfach. Jonas und die Erschöpfung, eine unendliche Geschichte mit kniffligen Wendungen.



Leider kennt sich Jonas mit der Korrelationsanalyse und der linearen Regression überhaupt nicht aus. Deshalb braucht er bei der Auswertung Ihre Hilfe!

1. Schätzen Sie die  $\rho$ -Werte in den Abbildungen! **(2 Punkte)**
2. Schätzen Sie die  $R^2$ -Werte in den Abbildungen! **(2 Punkte)**
3. Interpretieren Sie die  $R^2$ -Werte für die jeweilige Gerade! **(2 Punkte)**
4. Was ist der optimale  $R^2$ -Wert im Kontext einer wissenschaftlichen Fragestellung? Begründen Sie Ihre Antwort an einem Beispiel! **(2 Punkte)**
5. Was ist der optimale  $\rho$ -Wert im Kontext einer wissenschaftlichen Fragestellung? Begründen Sie Ihre Antwort an einem Beispiel! **(2 Punkte)**
6. Erklären Sie die Aussage "Correlation does not imply causation!" an einem Beispiel! **(2 Punkte)**



## 101. Aufgabe

(11 Punkte)

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



**Modellgüte der linearen Regression** 'Oh! Residuen. Die waren wichtig um zu wissen, ob eine Modellierung funktioniert hat! Da schauen wir uns dann mit der Funktion `augment()` die Werte der einzelnen Residuen an. Oder gleich den Residuenplot...da sehen wir dann... ja was eigentlich?', verkündet Nilufar stolz. Leider hat Nilufar vergessen wie der `R` Code für den Residuenplot geht. Nilufar hatte anderes im Kopf. Hm, lecker Takis Blue Heat und dazu dann im Hintergrund Star Trek laufen lassen. Aber sowas hilft ihr natürlich hier nicht. Da schmeißt sich Nilufar noch ein paar Takis Blue Heat in den Mund und kaut los.

Gewichtszuwachs	Mittlere Anzahl	$\hat{y}$	$\epsilon$
26.3	11.0	26.9	
26.6	11.4	27.7	
17.1	6.4	19.3	
33.2	13.4	31.0	
11.4	1.3	10.7	
24.6	11.4	27.6	
14.8	3.3	14.2	
29.0	10.2	25.6	

Leider kennt sich Nilufar mit der linearen Regression überhaupt nicht aus. Deshalb braucht sie bei der Auswertung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Werte der Residuen  $\epsilon$  in der obigen Tabelle! **(2 Punkte)**
3. Zeichnen Sie den Boxplot der Residuen  $\epsilon$ . Beschriften Sie die Abbildung! **(2 Punkte)**
4. Zeichnen Sie den Residualplot. Beschriften Sie die Abbildung! **(2 Punkte)**
5. Gibt es auffällige Werte anhand des Residualplots? Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Erklären Sie die Eigenschaft eines statistischen Modells, welche mit dem Residualplot überprüft wird! Begründen Sie Ihre Antwort anhand einer Visualisierung! **(2 Punkte)**

## 102. Aufgabe


(12 Punkte)

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



**Visualisierung des Regressionskreuzes** Mark hat ein Gewächshausexperiment mit Brokkoli durchgeführt. Soweit so gut. Dann war er bei seinem Betreuer. Leider war der Schritt nicht so hilfreich. Mark und die Unsicherheit, eine unendliche Geschichte mit kniffligen Wendungen. Aber es muss ja weitergehen. Mark hatte dann in seiner Abschlusarbeit einfach zu viele Endpunkte gemessen und ist jetzt vollkommen durcheinander, welche Analyse er nun wie rechnen soll. Naja, dann heißt es jetzt eben Andrea Berg aufdrehen und darüber nachdenken, was hier eigentlich gemacht wurde. Mark fängt einfach an und nimmt den ersten Endpunkt Anzahl Läsionen auf den Blättern. Dann kann er sich voran arbeiten. Später dann noch raus um zu Reiten um mal zu entspannen und vielleicht ist Nilufar auch da. Wäre toll.

Leider kennt sich Mark mit dem Kontext der linearen Regression überhaupt nicht aus. Deshalb braucht er bei der Auswertung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Zeichnen Sie die Zeile des Regressionskreuzes für den Endpunkt mit drei Feldern! Beschriften Sie die Abbildung! **(4 Punkte)**
3. Ergänzen Sie die entsprechenden statistische Methoden zur Analyse in jedem Feld! **(2 Punkte)**
4. Formulieren Sie die Nullhypothese für die statistische Methode in jedem Feld! **(2 Punkte)**
5. Ergänzen Sie die entsprechenden Funktionen in  zur Analyse in jedem Feld! **(2 Punkte)**
6. Welchen Effekt erhalten Sie in jedem Feld? Geben Sie ein Beispiel! **(2 Punkte)**

## Teil VIII.

# Experimentelles Design

### 103. Aufgabe

(16 Punkte)



**Einfache experimentelle Designs** Neuer Versuch neues Glück! Yuki und Alex sind bei Mark um sich Hilfe für eine Versuchsplanung in **R** zu holen. Im Hintergrund läuft viel zu laut Columbo. Daher hat der Hamster schon lange reiaus genommen. In dem neuen Versuch geht es um den Zusammenhang zwischen der Behandlung Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) und dem Messwert Protein/Fettrate [%/kg] in Hnern. Der Versuch soll in einem Leistungssteigerungsversuch im Wendland durchgefhrt werden. Immerhin ist der Messwert normalverteilt, was einges einfacher macht. Was es nicht so einfacher macht ist, dass Yuki noch als zustzliche Herausforderung etwas anderes umtreibt: die Faulheit. Im ersten Schritt berlegt Mark ein einfaches experimentelles Design zu probieren. Daher entscheiden sich alle drei fr ein *Latin square design*. Ob es das jetzt einfacher macht?

Leider kennen sich Mark, Yuki und Alex mit dem *Latin square design* berhaupt nicht aus. Deshalb brauchen die Drei bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Skizzieren Sie das faktorielle Versuchsdesign! **(3 Punkte)**
4. Skizzieren Sie eine Datentabelle fr das faktorielle Versuchsdesign in **R**! **(2 Punkte)**
5. Erstellen Sie das statistische Modell in der in **R** blichen Schreibweise fr eine ANOVA! Skizzieren Sie die notwendige Funktionen in **R**! **(3 Punkte)**
6. Skizzieren Sie die weitere Datenanalyse hinsichtlich eines multiplen Gruppenvergleiches! **(2 Punkte)**
7. Skizzieren Sie eine mgliche Abbildung im Kontext der wissenschaftlichen Fragestellung! Beschriften Sie die Abbildung! **(2 Punkte)**
8. Ergnzen Sie zu der Abbildung ein mgliches Ergebnis des multiplen Gruppenvergleichs! Begrnden Sie Ihre Antwort! **(2 Punkte)**




## 104. Aufgabe

(20 Punkte)



**Fortgeschrittene experimentelle Designs** Der Hamster macht mal wieder Randalen in Marks Zimmer und rennt davon! Tina und Jessica sind bei Mark in im Teuteburgerwald wo der neue, bessere Versuch stattfinden soll. Dabei soll in einem Stallexperiment im Teuteburgerwald mit Schweinen durchgeführt werden. Auf dem Tisch stapeln sich Marzipankugeln aus Vollkorndinkelmehl. Eine Spezialität der Komune hier. Tina hasst Vollkorn wie Gummibärchen geliebt werden. In dem neuen Versuch geht es um den Zusammenhang zwischen der Behandlung Lüftungssystem (*keins, storm, tornado* und *thunder*) sowie Genotypen (*AA* und *BB*) sowie vier Blöcken und dem Messwert Gewichtszuwachs in der 1LW. Immerhin ist der Messwert normalverteilt, was einiges einfacher macht. Was es nicht so einfacher macht ist, dass Jessica als zusätzliche Herausforderung noch die Gefälligkeit mitgebracht hat. Daher entscheiden sich alle drei für ein *Strip plot design* oder auch *Streifenanlage*. 'Naja, so viel einfacher ist es dann doch nicht...', merkt Jessica an und sucht der Hamster.

Leider kennen sich Mark, Tina und Jessica mit dem *Strip plot design* oder auch *Streifenanlage* überhaupt nicht aus. Deshalb brauchen die Drei bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistische Hypothesenpaare! **(2 Punkte)**
3. Skizzieren Sie das faktorielle Versuchsdesign! **(3 Punkte)**
4. Skizzieren Sie eine Datentabelle für das faktorielle Versuchsdesign in ! **(2 Punkte)**
5. Erstellen Sie das statistische Modell in der in  üblichen Schreibweise für eine ANOVA! Skizzieren Sie die notwendige Funktionen in ! **(4 Punkte)**
6. Skizzieren Sie die weitere Datenanalyse hinsichtlich eines multiplen Gruppenvergleiches! **(2 Punkte)**
7. Skizzieren Sie eine mögliche Abbildung im Kontext der wissenschaftlichen Fragestellung! Beschriften Sie die Abbildung! **(3 Punkte)**
8. Ergänzen Sie zu der Abbildung ein mögliches Ergebnis des multiplen Gruppenvergleichs! Welche Annahme hinsichtlich der Modellierung haben Sie getroffen? Begründen Sie Ihre Antwort! **(3 Punkte)**

## Teil IX.

# Programmieren in R

### 105. Aufgabe

(9 Punkte)



**Grundlegende Kenntnisse der Programmierung in R** Steffen muss seiner Hausarbeit mit R arbeiten. Deshalb sitzt er jetzt mit Ihnen zusammen und hat einige Fragen zu den Grundlagen in R an Sie! Na dann wollen Sie mal helfen. Immerhin will seine Betreuerin, dass R genutzt wird.

Steffen: *Was ist der Unterschied zwischen `library()` und `Packages` und warum brauche ich sowas in R?* **(1 Punkt)**

Sie antworten:

Steffen: *Ich verstehe den Pipe-Operator nicht. Wie sieht der aus und was macht der? Gebe mal ein Beispiel!* **(1 Punkt)**

Sie antworten:

Steffen: *Was ist eigentlich ein Faktor in R?* **(1 Punkt)**

Sie antworten:

Steffen: *In R gibt es Objekte, Wörter und Funktionen. Wie unterscheiden sich diese voneinander?* **(1 Punkt)**

Sie antworten:

Steffen: *Ich habe gehört, dass es Vorteile gibt R zu nutzen. Nenne mir mal einen Vorteil!* **(1 Punkt)**

Sie antworten:

Steffen: *Wie sieht der Zuweisungs-Operator aus und was ist seine Funktion? Gerne mit Beispiel!* **(1 Punkt)**

Sie antworten:

Steffen: *Wie speichern wir in R intern Daten ab? Ich brauche nochmal den Namen der Funktion. Was sind den die Vorteile von dieser Art der Speicherung?* **(1 Punkt)**

Sie antworten:

Steffen: *Was ist der Unterschied zwischen dem RStudio und R?* **(1 Punkt)**

Sie antworten:

Steffen: *Ich habe die Namen der beiden R Pakete vergessen, die wir eigentlich immer laden. Wie heißen die noch gleich?* **(1 Punkt)**

Sie antworten:

## 106. Aufgabe

(9 Punkte)



**Fortgeschrittene Kenntnisse der Programmierung in R** 'Unter den Blinden ist der Einäuge König!', ruft Ihnen Paula entgegen. Das können Sie schon nicht mehr hören. Nur weil Sie einmal gesagt haben, dass Sie sich schonmal mit R beschäftigt haben, stehen hier alle Schlange. Aber gut, das hat Sie dann doch vorangebracht. Leider kennt sich Paula auch wieder überhaupt nicht mit R aus aber ihr Betreuer möchte gerne, dass die Auswertung in R gemacht wird. Da müssen Sie dann wohl mal nochmal ran und helfen.

Paula fragt: *Ich möchte ein CLD erstellen. Welche Funktionen muss ich in welcher Reihenfolge nutzen?* **(2 Punkte)**

Sie antworten:

Paula fragt: *Man kann doch die Funktion `emmeans()` von Varianzhomogenität auf Varianzheterogenität umstellen. Wie ging das noch gleich?* **(1 Punkt)**

Sie antworten:

Paula fragt: *Ich habe den Faktor  $f_1$  und  $f_2$  und möchte den Faktor  $f_1$  getrennt in jedem Level des Faktors  $f_2$  auszuwerten. Wie geht das in der Funktion `emmeans()`?* **(1 Punkt)**

Sie antworten:

Paula fragt: *Wozu nutze ich die Funktion `mutate()` hauptsächlich?* **(1 Punkt)**

Sie antworten:

Paula fragt: *Warum wurden jeweils die R Pakete `{emmeans}`, `{ggplot}` und `{readxl}` geladen?* **(2 Punkte)**

Sie antworten:

Paula fragt: *Viele Abbildungen gehen ja auch in Excel, aber `{ggplot}` ist schon besser. Wie verbindet `{ggplot}` eigentlich nochmal die einzelnen Ebenen?* **(1 Punkt)**

Sie antworten:

Paula fragt: *Welche Funktionen brauche ich in welcher Reihenfolge um eine ANOVA zu rechnen?* **(1 Punkt)**

Sie antworten:

## Teil X.

# Forschendes Lernen

### Zerforschen einer wissenschaftlichen Veröffentlichung

Das forschende Lernen basiert zum einen auf den folgenden wissenschaftlichen Veröffentlichungen. Für die Prüfung wird die vertiefende Kenntnis der folgenden Veröffentlichungen vorausgesetzt.

- Sánchez, M., et al. (2022). Hoverfly pollination enhances yield and fruit quality in mango under protected cultivation. *Scientia Horticulturae*, 304, 111320. [\[Link\]](#)
- Petersen, F., et al. (2022). Influence of light intensity and spectrum on duckweed growth and proteins in a small-scale, re-circulating indoor vertical farm. *Plants*, 11(8), 1010. [\[Link\]](#)
- Selle, P. H., et al. (2010). Implications of sorghum in broiler chicken nutrition. *Animal Feed Science and Technology*, 156(3-4), 57-74. [\[Link\]](#)
- Wu, G., et al. (2004). Arginine nutrition in neonatal pigs. *The Journal of Nutrition*, 134(10), 2783S-2790S. [\[Link\]](#)
- Graham, J., E., et al. (2024) Stock assessment models overstate sustainability of the world's fisheries. *Science* 385, 860-865. [\[Link\]](#)

*In der Prüfung erhalten Sie keinen Auszug der wissenschaftlichen Veröffentlichung! Die Veröffentlichungen werden als bekannt in der Prüfung vorgelegt. Sie haben sich vorab Notizen und Anmerkungen auf Ihrem Spickzettel gemacht.*

### Zerforschen eines wissenschaftlichen Datensatzes

Das forschende Lernen basiert zum anderen auf den folgenden wissenschaftlichen Datensätzen und deren vertiefende Analyse werden als bekannt vorausgesetzt. Die Datensätze werden über ILIAS bereitgestellt.

- salad\_fert\_weight.xlsx
- weight\_gain\_pig.xlsx
- flowercolor\_data.xlsx
- chickentype\_class.xlsx

*In der Prüfung erhalten Sie keinen Auszug aus den wissenschaftlichen Daten. Die Datensätze werden als bekannt in der Prüfung vorgelegt. Sie haben sich vorab Notizen und Anmerkungen auf Ihrem Spickzettel gemacht.*

## 107. Aufgabe



(20 Punkte)

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



**Zerforschen einer wissenschaftlichen Veröffentlichung** Jessica hält die wissenschaftliche Veröffentlichung *Selle, P. H., et al. (2010). Implications of sorghum in broiler chicken nutrition* unter einem Schnaufen in die Luft. 'Worum geht es denn eigentlich in dieser Arbeit?', fragt sie stirnrunzelnd und wirft die Arme in die Luft, da hilft dann auch nicht mehr die beruhigende Wirkung von David Bowie. Jessica soll die Veröffentlichung nutzen um das eigene Experiment zu planen. Als eine Vorlage sozusagen. Daher möchte ihr Betreuer, dass sie einmal die Veröffentlichung sinnvoll zusammenfasst. Das sollte dann doch etwas aufwendiger werden. Das wird dann vermutlich heute Abend nichts mehr mit ihrem Hobby Warhammer. Die Hündin schaut mitleidig.

Leider kennt sich Jessica mit dem Lesen einer wissenschaftlichen Veröffentlichung mit Fokus auf die Statistik überhaupt nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe! Glücklicherweise kennen Sie die wissenschaftliche Veröffentlichung schon im Detail und können sofort helfen.

1. Erläutern Sie die wissenschaftliche Fragestellung der wissenschaftlichen Veröffentlichung anhand des OCAR Prinzips nach Schimel (2012)<sup>3</sup> **(4 Punkte)**
2. Nennen Sie die untersuchten Endpunkte in der wissenschaftlichen Veröffentlichung! Wie lautet der primäre Endpunkt? **(2 Punkte)**
3. Erstellen Sie das statistische Modell in der in  üblichen Schreibweise! **(2 Punkte)**
4. Nennen Sie eine Auswahl an bedeutenden statistischen Maßzahlen in der wissenschaftlichen Veröffentlichung! **(1 Punkt)**
5. Interpretieren Sie die Hauptaussage der wissenschaftlichen Veröffentlichung hinsichtlich der Signifikanz für den primären Endpunkt! **(2 Punkte)**
6. Interpretieren Sie die Hauptaussage der wissenschaftlichen Veröffentlichung hinsichtlich der Effektstärke für den primären Endpunkt! **(2 Punkte)**
7. Diskutieren Sie die ökonomische Relevanz der Hauptaussage der wissenschaftlichen Veröffentlichung im Bezug auf Signifikanz und Effektstärke für den primären Endpunkt! **(1 Punkt)**
8. Skizzieren Sie für den primären Endpunkt den sich ergebenden Datensatz in  für eine ausgewählte Abbildung! **(2 Punkte)**
9. Skizzieren Sie einen möglichen Versuchsplan für den primären Endpunkt! **(2 Punkte)**
10. Schätzen Sie die benötigte Fallzahl für ein zukünftiges Experiment anhand der Ergebnisse in der wissenschaftlichen Veröffentlichung für den primären Endpunkt! **(2 Punkte)**

<sup>3</sup>Schimel, J. (2012). Writing science: how to write papers that get cited and proposals that get funded. OUP USA.



## 108. Aufgabe


(20 Punkte)

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



**Zerforschen eines wissenschaftlichen Datensatzes** Vor dem Start der eigenen Arbeit möchte ihr Betreuer, dass Tina einmal die wissenschaftlichen Daten *data2* sinnvoll zusammenfasst. Dann würde die eigene Arbeit auch leichter von der Hand gehen und Tina hätte dann schon eine Vorlage um die eigenen erhobenen Daten in eine Tabelle eintragen zu können. 'Das ist jetzt aber umfangreicher als gedacht!', schnauft sie und runzelt die Stirn als sie in ihren Laptop starrt. Dabei isst sie noch ein paar Katjes. Das wird dann vermutlich heute Abend nichts mehr mit Indiana Jones

Leider kennt sich Tina mit der Analyse eines wissenschaftlichen Datensatzes überhaupt nicht aus. Deshalb braucht sie bei der Auswertung Ihre Hilfe! Glücklicherweise kennen Sie den wissenschaftlichen Datensatz aus Ihren eigenen Analysen schon im Detail und können sofort helfen.

1. Formulieren Sie die wissenschaftliche Fragestellung des Datensatzes in Form einer PowerPoint Folie! **(2 Punkte)**
2. Nennen Sie zwei Besonderheiten des Datensatzes! Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Nennen Sie die untersuchten Endpunkte in dem Datensatz! Wie lautet der primäre Endpunkt für die Auswertung? **(2 Punkte)**
4. Skizzieren Sie die großen Analysebereiche der Statistik! Beschriften Sie die Abbildungen! **(2 Punkte)**
5. In welchen der großen Analysebereiche der Statistik fällt die Auswertung des primären Endpunktes? Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Skizzieren Sie eine ikonische Abbildung für den primären Endpunkt im Kontext der wissenschaftlichen Fragestellung! **(2 Punkte)**
7. Erstellen Sie das statistische Modell in der in  üblichen Schreibweise! **(2 Punkte)**
8. Skizzieren Sie die Datenanalyse hinsichtlich der Signifikanz für den primären Endpunkt! **(2 Punkte)**
9. Skizzieren Sie die Berechnung der Effektstärke für den primären Endpunkt! **(2 Punkte)**
10. Skizzieren Sie einen möglichen Versuchsplan für den primären Endpunkt! **(2 Punkte)**

# Teil XI.

## Mathematik

### 109. Aufgabe

(10 Punkte)

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



**Herodot – der Schimmel aus Ivenack** Die Lerngruppe *Die Pantoffeltieren* bestehend aus Jessica, Steffen, Alex und Tina waren auf Exkursion in Brandenburg und haben dort Folgendes erarbeitet. 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 verriet sein Versteck durch lautes Wiehern, woraufhin die französische Armee den Schimmel beschlagnahmte und nach Frankreich führte<sup>4</sup>. Jetzt wollen die vier herausfinden: "Konnten die Ivenacker den Apfelschimmel Herodot vor dem Zugriff von Napoleon in der 1000-jährigen Eiche verstecken?"

Helfen Sie der Lerngruppe *Die Pantoffeltieren* bei der Beantwortung der Forschungsfrage! Gehen Sie von einem radialen Wachstum der 1000-jährigen Eiche von 1.2mm pro Jahr aus. Es ist bekannt, dass die Eiche im Jahr 2022 einen Umfang von 13m in Brusthöhe hatte.

1. Wie groß war der Durchmesser in  $m$  der Eiche im Jahr 1805 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 190cm, eine Breite von 75cm sowie eine Länge von 230cm.

3. Berechnen Sie das effektive Volumen von Herodot in  $m^3$ , welches Herodot in der 1000-jährigen Eiche einnehmen würde! **(2 Punkte)**

Es wurde berichtet, dass sich Herodot in der 1000-jährigen Eiche *mitblutigerNase* 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 30cm bricht die Eiche zusammen. Beantworten Sie die Forschungsfrage! Begründen Sie Ihre Antwort! **(2 Punkte)**

<sup>4</sup>Die 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](#)

## 110. Aufgabe

(10 Punkte)

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



**Von Töpfen auf Tischen** Die Projektgruppe *B* bestehend aus Alex, Tina, Yuki und Jonas hat sich zusammengefunden um den ersten Versuch zu planen. In einem Experiment wollen sie die Wuchshöhe von 160 Stockrosen bestimmen. Bevor die Vier 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 muss die Projektgruppe 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 310 EUR.

Helfen Sie der Projektgruppe *B* bei der Planung des Versuches!

1. Skizzieren Sie den Versuchsplan auf zwei Tischen im Gewächshaus! **(2 Punkte)**
2. Berechnen Sie die benötigte Anzahl an Pflanztöpfen, wenn Sie Randpflanzen mit berücksichtigen wollen! **(1 Punkt)**
3. Berechnen Sie die benötigte (a) *Pflanztopf*fläche in  $m^2$  sowie die (b) *Tisch*fläche in  $m^2$  gegeben der Anzahl an Pflanztöpfen inklusive Randpflanzen am Anfang der Keimungsphase! **(4 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! **(2 Punkte)**
5. Berechnen Sie die Kosten in EUR für Ihre Torfbestellung! **(1 Punkt)**

## 111. Aufgabe

(10 Punkte)

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



**Solar- & Biogasanlagen** Paula bringt ein neues, tolles Projekt mit in die Lerngruppe *Die Kühe auf dem Deich* bestehend aus ihr, Nilufar, Tina sowie Yuki. Um die Energiekosten ihres Betriebes zu senken, will sie eine Solaranlage auf den Schweinestall montieren lassen. Dafür hat sie ihren Stall ausgemessen und findet folgende Maße wieder. Die vordere Seite des Schweinestall hat eine Höhe  $h_v$  von  $5.5m$ . Die hintere Seite des Schweinestall hat eine Höhe  $h_b$  von  $8.5m$ . Der Schweinestall hat eine Tiefe  $t$  von  $13m$  und eine Breite  $b$  von  $70m$ . 'Sag mal Paula, ist das eine Matheaufgabe oder rechnen wir hier gerade für dich kostenlos als menschliche Computer Sachen für deinen Betrieb?', fragt Tina mit erhobenen Augenbrauen. Yuki und Nilufar nicken zustimmend.

Wenn die Lerngruppe nicht will, dann müssen Sie bei der Planung helfen!

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 plant Paula eine neue Biogasanlage für ihren Betrieb. Der neue Methantank hat einen Radius  $r$  von  $1.8m$ . Leider gibt es ein paar bauliche Beschränkungen auf dem Grundstück. Das Fundament des zylindrischen Methantanks kann nur ein Gewicht von maximal  $12t$  aushalten bevor der Tank wegbricht. Paula rechnen eine Sicherheitstoleranz von 10% ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei  $-80^\circ C$  eine Dichte von  $235kg/m^3$ . Bei  $-100^\circ C$  hat Methan eine Dichte von  $300kg/m^3$ . Paula betreibt ihre Anlage bei  $-92^\circ C$ .

3. Extrapolieren Sie die effektive Dichte des Methans in dem Methantank! Welche Annahme haben Sie getroffen? **(1 Punkt)**
4. Berechnen Sie wie viel Kubikmeter  $m^3$  in den Methantank gefüllt werden 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)**

## 112. Aufgabe

(10 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 puderröte Gesicht von Paula. “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 Netto über sich ergehen lassen. Da kommen Sie nicht mehr raus, also können Sie auch gleich mitmachen. Das Problem liegt in Chile<sup>5</sup>. Tja, die Deutschen und Südamerika.

Zuerst werden Ihre Fähigkeiten getestet, der Mathematik folgen zu können. Oder berechnen Sie gerade den Einkauf von Paula?

1. Wenn 7 Blaubeerschalen 13.23 Euro kosten, wie viel kosten 12 Schalen? **(2 Punkte)**
2. Wenn Sie die 12 Blaubeerschalen gekauft haben, wie viele Aligatorbirnen zu je 2.89 EUR können Sie sich dann noch für 200 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 Netto über Gemüse erfährt.

- Ein Kilo Strauchtomaten benötigt 170l Wasser. Eine Strauchtomate wiegt 100 - 125g.
- Ein Kilo Salat benötigt 120l Wasser. Ein Salatkopf wiegt 320 - 520g.
- Ein Kilo Avocado benötigt 950l Wasser. Eine Avocado wiegt 130 - 400g.
- Ein Kilo Blaubeeren benötigt 830l Wasser. Eine Blaubeere wiegt 3.2 - 3.6g.

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! **(3 Punkte)**

Chile exportiert im großem Ausmaß Blaubeeren und Avocados. In dem Exportjahr 2022 blieben die Erträge von Blaubeeren mit  $7.9 \times 10^4$ t in dem prognostizierten Rahmen. Die Menge steigerte sich um 5.1%. Die Exporte für Avocados stiegen in dem gleichen Zeitraum um 22.1% auf  $1.9 \times 10^5$ t.

4. Wie viele Tonnen Wasser hat Chile in dem Exportjahr 2021 exportiert? **(2 Punkte)**

Chile ist eines der wenigen Länder der Welt, die ihr Wasser komplett privatisiert haben. Derzeit sind nur ein 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 52 Liter Wasser pro Kopf für den täglichen Bedarf. In *Deutschland* liegt der Verbrauch bei 10 - 15 Liter pro Minute Duschen und 3 - 12 Liter pro Minute Händewaschen.

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

Das alles hätten Sie nicht von Paula 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)**

<sup>5</sup>Die 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.

## 113. Aufgabe

(10 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 auf Empfehlung von von Paula 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 67 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 rechnen<sup>6</sup>.

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  $9.65 \text{ m/s}^2$  an. Im Weiteren hat die Erde einen ungefähren Durchmesser von  $1.235 \times 10^4 \text{ km}$  und eine mittlere Dichte  $\rho$  von  $5.44 \text{ g/cm}^3$ . Das Gewicht von einem heute lebenden afrikanischen Elefanten liegt bei 5t bis 7t und das Gewicht von einem Brachiosaurus bei bis zu 30t.

1. Welchen Durchmesser müsste die Erde vor 67 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 67 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 67 Millionen Jahren! **(2 Punkte)**
2. Beantworten Sie die Eingangsfrage mit 1-2 Antwortsätzen! **(1 Punkt)**

Die Distanz zwischen Sonne und Erde entspricht 1.01 astronomische Einheiten (AE). Die Einheit 1 AE wird mit  $1.48 \times 10^8 \text{ km}$  angegeben. Der *massebehaftete* Sonnenwind besteht aus 81% Wasserstoffkernen mit einer molaren Masse von  $1.02 \text{ g/mol}$ , 11% Heliumkernen mit  $4.32 \text{ g/mol}$  sowie 8% weiteren Atomkernen mit  $89.32 \text{ g/mol}$ . Die Teilchendichte bei Eintritt in die Erdatmosphäre liegt zwischen  $0.4$  bis  $100$  Teilchen  $\text{cm}^{-3}$  pro Sekunde mit einer mittleren Teilchendichte von  $7 \text{ cm}^{-3}$  pro Sekunde.

Lösen Sie den folgenden Aufgabenteil mit einer aussagekräftigen Skizze!

3. Berechnen Sie die Anzahl an massebehafteten Teilchen des Sonnenwindes, die die gesamte Erde pro Sekunde treffen! **(2 Punkte)**
4. Berechnen Sie die Masse, die die Erde pro Jahr durch die *massebehafteten* Teilchen des Sonnenwind zunimmt! **(2 Punkte)**

<sup>6</sup>Die Quelle der Inspiration für die Aufgabe war folgender Artikel: "Skeptische Anmerkungen — Die Erde als Dampfnudel" in Der Humanistische Pressedienst

## 114. 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 Enten in Mastställen sagen. Das ist natürlich etwas ungünstig, denn 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 Mark, Yuki, Jonas und Alex aber als vorsorgliche Enten-Halter:innen nicht<sup>7</sup>. Gemeinsam sind die Vier in einer Projektgruppe gelandet. Betrachten wir also gemeinsam einmal das Platzangebot (eng. *space allowance*, abk. *SA*) der Enten 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 Ente plus dem benötigten Radius für das Verhalten  $i$ .
- $R_i$  dem notwendigen Abstand zu den Nachbarn für das Verhalten  $i$ .
- $i$  dem Verhalten: (1) sitting, (2) wingflapping, (3) foraging incl. scratching und (4) walking.

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)
sitting	37cm; 44cm; 50.1%	38cm; 29cm; 25.4%	26cm; 42cm; 25.4%
wingflapping	41cm; 11cm; 1.8%	44cm; 10cm; 5.2%	65cm; 38cm; 3.5%
foraging incl. scratching	30cm; 30cm; 0.4%	42cm; 14cm; 0.6%	37cm; 41cm; 0.9%
walking	38cm; 34cm; 0.8%	28cm; 49cm; 0.8%	31cm; 54cm; 0.1%

Leider kennen sich die Vier nicht so gut mit der Berechnung aus! Daher brauchen die Vier Ihre Hilfe!

1. Skizzieren Sie die Werte  $r_i$ ,  $R_i$  und  $A_i$  für zwei nebeneinander agierende Enten für ein Verhalten  $i$ . Nutzen Sie hierfür vereinfachte geometrische Formen! **(2 Punkte)**
2. 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)**
3. Ergänzen Sie eine Spalte mit dem benötigten Platz  $A$  für das jeweilige Verhalten, welches sich aus den mittleren Werten ergibt! **(1 Punkt)**
4. Berechnen Sie das benötigte Platzangebot  $SA$  für alle betrachteten Verhalten! **(1 Punkt)**
5. Sie entnehmen der Literatur folgende Aussage zur Verteilung der Enten 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  $\omega$  and is calculated with  $\omega = \frac{A}{0.9069}$ .“ Veranschaulichen Sie die Fläche  $\omega$  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)**

<sup>7</sup>Die 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.](#)

## 115. Aufgabe

(8 Punkte)

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



**Nelken von den Molukken** Mark und Jonas waren gemeinsam in Berlin und sitzen nun im IC nach Amsterdam um zurück nach Osnabrück zu fahren. 'Weißt du was ich mich frage?', entfährt es Mark ziemlich plötzlich, so dass Jonas die Snickers aus dem Mund fallen. 'Nein, und ehrlich gesagt bin ich auch ziemlich müde...'. Das ist jetzt aber Mark egal, denn er möchte folgende Sachlage diskutieren. Und Mark hat jetzt 3 Stunden Zeit. Plus Verspätung. 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 42 Tagen die ersten Symptome ein; die ersten Toten sind nach 70 Tagen zu beklagen; nach 105 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 222 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. Was ist die Besonderheit der Überlebenszeitkurve? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. 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 Skrobut bei der Überquerung des Pazifiks, da durch Erkundungsfahrten weniger Zeit blieb, um wilden Sellerie aufzunehmen. Wilder Sellerie enthält  $5000\mu\text{g}/100\text{g}$  Vitamin C. Der Bedarf liegt bei  $120\text{mg}$  pro Tag für Männer.

3. Berechnen Sie die notwendige Menge in  $t$  an aufzunehmenden wilden Sellerie auf die *Concepción* für die ununterbrochene Fahrt von drei Monate und 20 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)**



## 116. Aufgabe

(10 Punkte)

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



**Event Horizon – Am Rande des Universums** Yuki ist bei Paula um gemeinsam *Event Horizon – Am Rande des Universums* zu streamen. Das war jetzt nicht die beste Idee. Denn Yuki kann Horror überhaupt nicht ab. Deshalb flüchtet sie sich in Logik um ihre Emotionen zu bändigen. Paula mampft ungerührt Smarties. Folgenden Gedankengang nutzt Yuki um dem Film zu entkommen. Die Sonne hat eine aktuelle, angenommene Masse von  $2 \times 10^{27}$  kg. Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 2500m kollabiert, wird die Sonne 35% der aktuellen Masse verloren haben. Ein Lichtteilchen mit der Masse  $m_f$  und der Fluchtgeschwindigkeit  $v_f$  will dem schwarzen Loch entkommen. An folgende Formeln erinnert sich Yuki für die kinetische Energie des Lichtteilchens  $E_{kin}$  und der Gravitationsenergie des schwarzen Lochs  $E_{grav}$ <sup>8</sup>.

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

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 Yuki bei der Ablenkung helfen und 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! **(2 Punkte)**
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! Erstellen Sie dafür eine Datentabelle mit fünf Werten für den Radius  $r$ ! **(3 Punkte)**

<sup>8</sup>Die Quelle der Inspiration für die Aufgabe war ein Montagnachtfilm: [Event Horizon – Am Rande des Universums](#)

## 117. Aufgabe

(10 Punkte)

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



**Stichworte:** Great filter • SETI • WOW-Signal • 5-Sigma • Voyager 1 • Voyager 2

**Das Fermi Paradoxon** Jonas und Paula wandern durch den Teutoburgerwald um mal vom Studium runterzukommen. 'Kennst du eigentlich Enrico Fermi?', fragt Jonas und fährt ohne die Antwort abzuwarten fort, 'Er war ein berühmter 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 extra-terrestrischer Technik zu beobachten?'. Paula schaut ihn irritiert und interessiert an. Die beiden hat das Problem gepackt. Deshalb wollen Jonas und Paula das Paradoxon mal mathematisch untersuchen! Wie lange würde eine außerirdische Zivilisation benötigen um die gesamte Milchstraße zu besuchen, wenn das maximale Reisetempo die Geschwindigkeit der Voyager 1 Sonde wäre?<sup>9</sup>

Die beiden 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 250 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  $2 \times 10^{11}$  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 annährungsweise benötigt um die gesamten Sterne der Milchstraße mit Sonden zu besuchen! **(2 Punkte)**
4. Bei einem vermutet Alter der Erde von  $4.3 \times 10^9$  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  $9 \times 10^7$  Jahre vollständig im Erdinneren umgewandelt haben! **(2 Punkte)**

<sup>9</sup>Die Quelle der Inspiration für die Aufgabe war folgender Wikipediaeintrag: [Fermi-Paradoxon](#)

## 118. Aufgabe

(10 Punkte)

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



**Pyramiden bauen** Es stehen die mecklemburgischen Pyramidentage an! Sie und Tina 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 73 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 und Tina bei dem Befüllen der Pyramide mit Schutt zu Seite stehen. Leider haben Ihre Sklaven zu allem Überfluss auch noch chronische Schulterschmerzen entwickelt, als die Sklaven von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 75% 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  $11^\circ$ ! **(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 *Chefredaktuer*) 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  $6^\circ$  ändert! **(2 Punkte)**

## 119. Aufgabe

(10 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. Mark und Jessica schwingen sich auf ihre Cachermobile um mit 16km/h, geleitet von modernster Satellitentechnologie und einem Supercomputer aus dem Jahr 2000 in den Händen, Plastikdosen in der Natur und an sehenswerten Orten zu finden. Mark und Jessica wollen diesmal endlich die abwärts Terrainchallenge durchführen. Die Reihenfolge der Caches nach Terrainwertung gibt daher die von den beiden 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 Mark und Jessica für die Planung der Route zu Verfügung<sup>10</sup>.

Ort	Cache	Wertung (S T G)
A	GCDI7FS	3.0   2.0   Klein
B	GCDNI8V	4.0   2.5   Mikro
C	GCCPNHB	2.0   3.0   Mikro
D	GCWXA2F	1.0   5.0   Normal
E	GCAVVPR	1.5   1.5   Mikro

Im Weiteren sind den beiden folgende Informationen zu den Entfernungen der Orte zugänglich. Der Entfernungsvektor  $\overrightarrow{AC}$  ist 4km. Im Weiteren ist Ihnen der Entfernungsvektor  $\overrightarrow{CB}$  mit 5.5km bekannt. Der Entfernungsvektor  $\overrightarrow{BE}$  ist das 2.1-fache des Entfernungsvektor  $\overrightarrow{CB}$ . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr 20° südlich. Wenn Sie von dem Ort C den Ort B anpeilen, so liegt der Ort D ungefähr 45° 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 C Ihre Cachertour.

Leider sind die beiden sehr schlecht im Navigieren und Entfernungen ausrechnen. Die beiden brauchen Ihre Hilfe!

1. Lösen Sie diese Aufgabe mit Hilfe einer aussagekräftigen Skizze der Orte und Caches. Bezeichnen Sie die Strecken und die Winkel Ihrer Skizze entsprechend! **(5 Punkte)**
2. Welche Strecke in km legen Sie bei der Bewältigung der abwärts Terrainchallenge zurück? **(2 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

$$\text{Suchzeit} = 0.2 + 0.13 \cdot \text{Schwierigkeit}$$

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

<sup>10</sup>Die Quelle der Inspiration für die Aufgabe war folgende Tätigkeit: [Geocaching – Mach mit bei der weltweit größten Schatzsuche](#).

## 120. Aufgabe

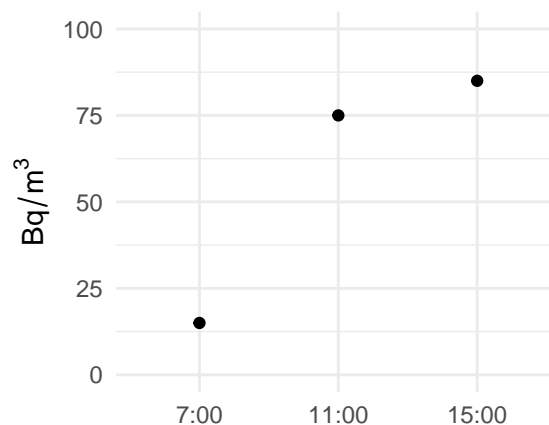
(10 Punkte)

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<sup>11</sup> vom Dorf will Nilufar das Ausmaß der Radonbelastung in ihrem Kellerzimmer bestimmen und lüften daher nicht. Passt schon. Spart dann auch Energie und lüften wird sowieso überschätzt. Während einer Messperiode von 7:00 Uhr bis 15:00 bestimmt sie dreimal automatisch die Radonbelastung in ihrem Kellerraum in  $Bq/m^3$ . Es ergibt sich folgende Abbildung<sup>12</sup>. Leider helfen die Messwerte Nilufar überhaupt nicht weiter. Sie müssen also helfen!



1. Wie lange dauert es in Stunden bis Sie eine kritische Belastung von  $320 Bq/m^3$  in Ihrem ungelüfteten Kellerraum erreicht haben? **(2 Punkte)**

Radon zerfällt mit einer Halbwertszeit von 3.7d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 180d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

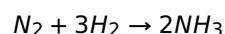
2. Wie lange dauert es in Stunden bis Ihre kritische Radonbelastung von  $320 Bq/m^3$  auf unter  $90 Bq/m^3$  gefallen ist? **(4 Punkte)**

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	77.1	27.9	
Sauerstoff	20.45	16.5	
Kohlenstoffdioxid	0.045	11.8	

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 Nilufar ihr etwas pappiges Toastbrot mampfen kommt Nilufar die Dokumentation über Brot aus Luft in den Sinn. Nilufar denkt darüber ein wenig nach. Für die Umwandlung von Stickstoff  $N_2$  mit Wasserstoff  $H_2$  zu Ammoniak  $NH_3$  gilt folgende Reaktionsgleichung<sup>13</sup>:



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. Wie viel Ammoniak in  $mol$  erhalten Sie aus einem Kubikmeter Luft? **(1 Punkt)**

<sup>11</sup>Tocotronic - Electric Guitar als passende Untermalung für diese Aufgabe.

<sup>12</sup>Die Quelle der Inspiration für die Aufgabe war folgender Artikel: [Atmende Wand](#)

<sup>13</sup>Die Quelle der Inspiration für die Aufgabe war folgender Artikel: [Haber-Bosch-Verfahren – Brot aus Luft](#)

## 121. Aufgabe

(10 Punkte)

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



**Armee der Finsternis** Der Studentenjob von Yuki war nach Ladenschluss bei IKEA die Regale einzuräumen. Dabei ist Yuki in der Auslage der Sonderangebote das Necronomicon<sup>14</sup> in die Hände gefallen. Nun ist sie eine Magierin der Zeichen geworden! Also eigentlich kann Yuki nur Mathe und das dämliche Necronomicon hat sie in die Vergangenheit geschleudert... aber gut, was tut man nicht alles im Jahr 292 n. Chr. für den neuen Lehnsherren Fürsten Arthur. Yuki baut natürlich einen Schrottkugelturm um sich den Horden der Finsternis mit genug Schrott erwehren zu können! Yuki stehen zwei mächtige magische Formeln zur Unterstützung zu Verfügung. Leider wird das nicht reichen, deshalb müssen Sie hier auch noch durch Zeit und Raum helfen!

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

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  $13m/s$  bilden.

1. Wie hoch müssen Sie den Schrottkugelturm bauen lassen, damit sich runde Bleikugeln durch die Fallgeschwindigkeit von  $13m/s$  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  $3.5mm$ . Die Löcher im Sieb erlauben ein Tropfendurchmesser von  $1.6mm$ . Welchen Durchmesser in mm haben Ihre produzierten Bleikugeln? **(3 Punkte)**

Sie haben jetzt die  $6.1 \times 10^4$  Bleikugeln zusammen. Blei hat eine Dichte von  $10.32g/cm^3$ .

4. Wie schwer in Kilogramm  $kg$  sind die  $6.1 \times 10^4$  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  $1.4cm$  Abstand haben müssen? **(1 Punkt)**

<sup>14</sup>Ein wirklich gefährliches Buch ist: *Du bist genug: Vom Mut, glücklich zu sein* von Fumitake Koga und Ichiro Kishimi

## 122. Aufgabe

(10 Punkte)

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



**Armee der Kaninchen** Leider hat es bei Yuki mit dem Krokodilreservat in Down Under nicht geklappt. War vielleicht auch nicht so die beste Idee... aber dafür hat Yuki eine neue Eingebung! Oder wie es Mike Tyson zugeschrieben wird: »Ich wurde nie niedergeschlagen, ich war immer am Aufstehen!«. Daher macht Yuki jetzt einen Großhandel mit Kaninchenfleisch und damit dem teuersten Fleisch in Australien auf. Moment, hopsen hier nicht, seit Thomas Austin im Jahr 1859 ungefähr 30 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Yuki wird stutzig und frag Sie, dem mal mathematisch nachzugehen!<sup>15</sup>

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

$$f(t) = 10^{10} - 10^9 \cdot 2.3^{-0.1 \cdot t + 2.4}$$

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 15 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 1.2 nach folgender Formel.

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

3. Wie viele Wochen benötigen die Viren um theoretisch die gesamte Kaninchenpopulation nach 13 Jahren Wachstum zu durchseuchen? **(1 Punkt)**

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

4. Wie lange in Jahren dauert es bis eine Kaninchenpopulation nach einer Viruspandemie wieder auf 30% der gesättigten Kaninchenpopulation angewachsen ist? **(2 Punkte)**

Thomas Austin entließ die Kaninchen im äußersten Osten von Australien. Australien hat eine West-Ost-Ausdehnung von 4100km und eine Nord-Süd-Ausdehnung von knapp 3400km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 11.5km 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 9\$ pro Tier und der durchführende Arzt verlangt ca. 35\$ pro Tier.

6. In Ihrem Stall leben 1100 Mastkaninchen. Mit welchen jährlichen Zusatzkosten für die Impfungen der Kaninchen müssen Sie daher kalkulieren? **(1 Punkt)**

<sup>15</sup>Die 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...](#)

## 123. Aufgabe

(12 Punkte)

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



**Ostfriesland. Unendliche Weiten.** Wir schreiben das Jahr 2024. Dies sind die Abenteuer des Esels Fridolin und Jonas. Grünes Gras unter Jonass Füßen und ein strammer Wind im Gesicht, egal wohin er schaut. Ein schmatzendes Geräusch ertönt unter Jonas. Jonas sinniert, sollte er seine weiten Graslandschaften jetzt schon düngen? Dafür benötigt Jonas die *Grünlandtemperatur*! Die Grünlandtemperatur (GLT) ist die Summe aller positiven Tagesmitteltemperaturen seit Jahresbeginn. Ab einer GLT von  $200^\circ$  kann mit der Stickstoffdüngung begonnen werden. Jonas sieht nicht ein, Geld für einen Agrarmetrologen zu bezahlen, wenn auch Sie mitrechnen können. Also rechnen Sie beide mit folgenden Informationen zu Monatsmultiplikatoren des GLT-Wertes: Januar mit  $0.4\times$ , Februar mit  $0.8\times$  und März mit  $1.05\times$ . Sie haben noch im letzten Jahr folgende Temperaturen gemessen.

Datum	C°
01. Jan 2023	0.4
01. Feb 2023	1.1
01. Mrz 2023	3.1
01. Apr 2023	5.8

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  $200^\circ\text{C}$  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 Jonass Boskoopplantage werden Sie beide auf dem Trecker von einer Gruppe elektrifizierter Renter abgedrängt. Der Trecker muss wieder aus dem Graben! Fridolin und die elektrifizierten Rentner ziehen an zwei, separaten Seilen. Dabei zieht Fridolin mit  $120\text{N}$ . Die elektrifizierten Renter bringen eine Kraft von  $190\text{N}$  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  $30^\circ$  Winkel mit ihrem Seil an dem Trecker. Welche Kraft wird aufgebracht? **(2 Punkte)**
7. Mit welcher Beschleunigung ziehen Sie den  $1.2\text{t}$  schweren Trecker *jeweils* aus dem Graben, wenn  $F = m \cdot a$  gilt? **(1 Punkt)**



## 124. Aufgabe

(10 Punkte)

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



**In der Kartonagenfabrik** Steffen, Yuki, Jessica und Jonas sitzen im Bus. Wenn man sich zu spät anmeldet, dann ist die Exkursion nicht so toll. Steffen hatte den Anderen in der Lerngruppe zu spät Bescheid gesagt. 'Was denn, bin ich eure Nanny oder was?!', entfährt es Steffen nachdem die vorwurfsvollen Blicke schon eine Weile auf ihm lasten. Also geht es eben mit Rektor Skinner und Mrs. Krabappel in die Kartonagenfabrik. Wie schon im vorherigen Semester... In der Kartonagenfabrik angekommen erfahren die Vier, dass die Kartons zum Versand von Nägeln nicht hier zusammengebaut werden sondern dass sich die Endfertigung in Flint, Michigan befindet. Unter anderem wird dort der berühmte *Doppelt gewellte, 8-mal-gefaltete, 0.8mm, 30-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 alle wenig Zuversicht zu erkennen.

Jetzt heißt es Kartons optimieren, wenn Sie auch nochmal nach Hause wollen. Warum jetzt *Sie* mit dabei sind, lassen wir mal weg. Der nun zu optimierende, flache Karton hat eine Länge von 30cm 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  $\text{cm}^2$ ? **(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 90m Zaun zu Verfügung. Auch hier sollen Sie mal helfen, sonst fährt der Bus Sie nicht nach Hause. Sie wollen nun die maximale Fläche des abgeschirmten Vorgartens in Abhängigkeit der Seitenlängen bei der Verwendung von 90m 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)**

## 125. Aufgabe

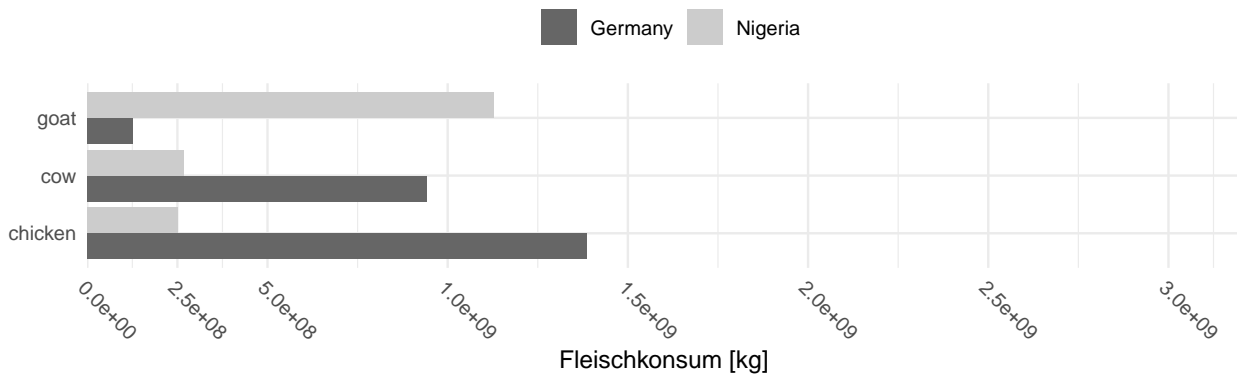
(10 Punkte)

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



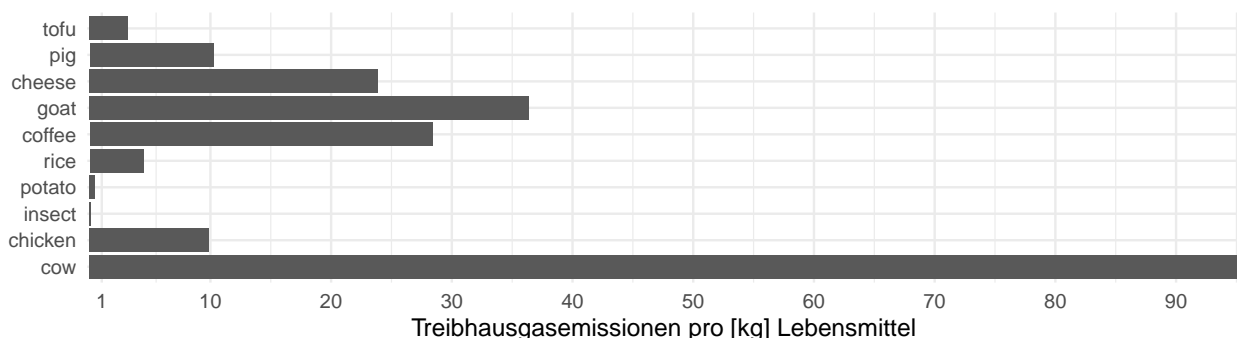
**Ein Pfund Insekten, bitte!** 'Das wird wohl häufiger gehört werden, wenn wir die Menschheit mit Proteinen ausreichend ernähren wollen<sup>16</sup>.', merkt Alex an. Die Lerngruppe um Mark, Tina und Jonas sind bei Alex um mal was außergewöhnliches zu essen. Um den Sinn der Nahrungsumstellung zu verdeutlichen, vergleicht Alex einmal Deutschland mit Nigeria. 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.  $8.4 \times 10^7$  Menschen in Deutschland und ca.  $1.79 \times 10^8$  Menschen in Nigeria. Mit den Informationen wollen Sie und Alex mit der Überzeugungsarbeit 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)**

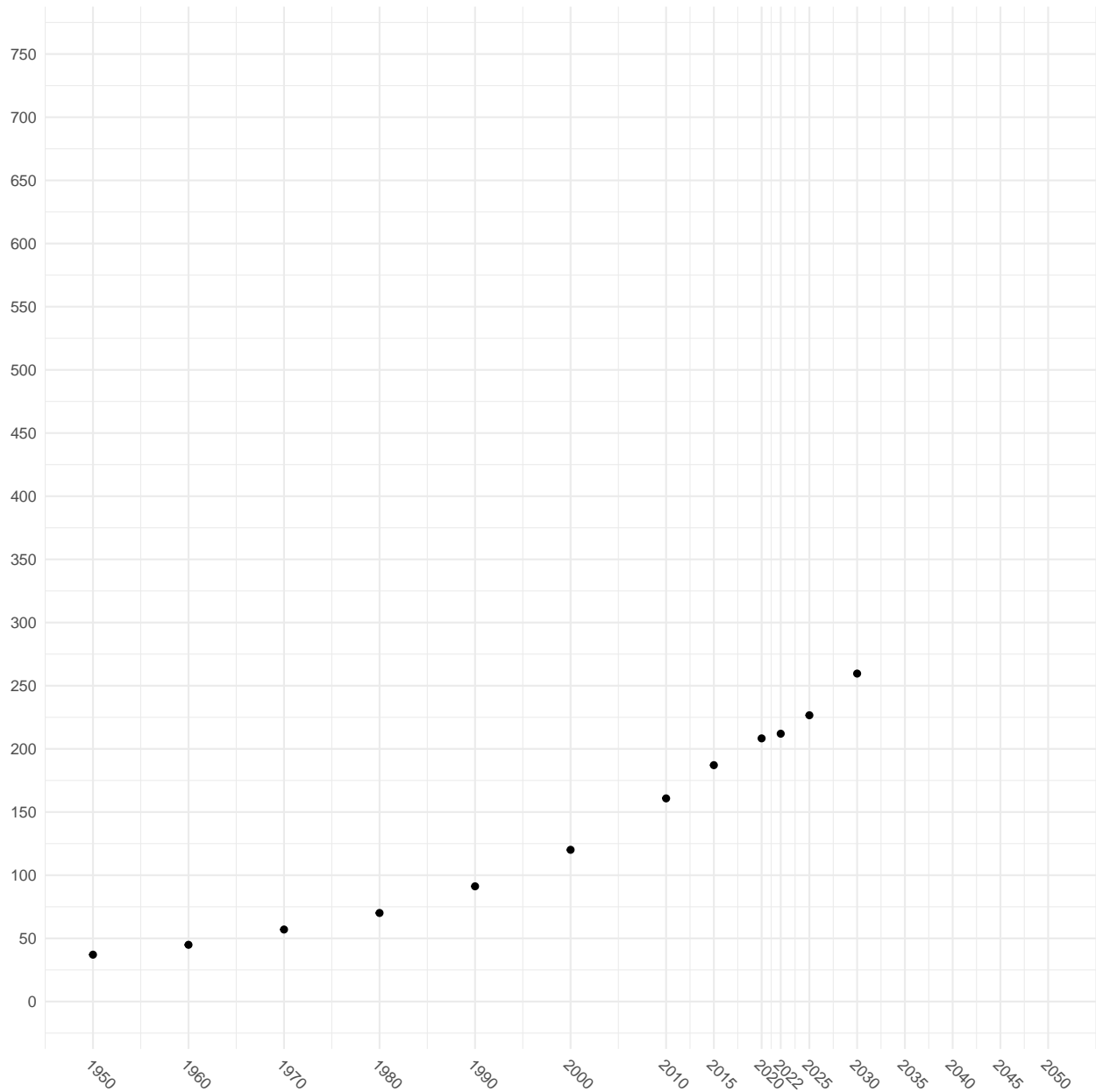
In der nächsten Abbildung finden Sie die CO<sub>2</sub>-Emission in [kg] nach Lebensmittel, die durch die Produktion entsteht, abgebildet.



3. Stellen Sie in einer Tabelle die Treibhausgasemissionen an CO<sub>2</sub> 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)**

<sup>16</sup>Die 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 80%-iger Angleichung der Lebensbedingungen zu Deutschland im Jahr 2021! **(1 Punkt)**
6. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an  $\text{CO}_2$  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  $\text{CO}_2$  in Nigeria, wenn die gesamte Proteinaufnahme durch Insekten ersetzt würde! **(1 Punkt)**

## 126. Aufgabe

(12 Punkte)

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



**Stichworte:** Immunsystem – Muskel vs. Interpol • Inzidenz • Prävalenz

**Tödliche Seuche AIDS – Die rätselhafte Krankheit** Irritiert legt Tina die historische Ausgabe des Spiegels aus den 80zigern beiseite. Tina und Mark sind bei ihrem Hausarzt und wollen einen AIDS-Test machen lassen. Woanders leider keinen Termin gekriegt... Immerhin denken die beiden über Nachwuchs nach und da geht es eben nur durch ungeschützten Sex. Was wissen Tina und Mark nun aber über AIDS und dem diagnostischen AIDS-Test, den die beiden nun machen werden? Leider zu wenig. Da brauchen dann Tina und Mark mal wieder Ihre Hilfe bei der Interpretation eines diagnostischen Tests!

Die Prävalenz von AIDS bei einem Menschen in Europa wird mit 0.5% angenommen. In 92% 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  $n = 10^4$  Patienten mit einem diagnostischen Test für AIDS aus. Sie nehmen sich also einen Kuli und fangen an auf der historischen Ausgabe des Spiegels zu rechnen<sup>17</sup>.

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 Tina und Mark, dass beim diagnostischen Testen *True Positives* ( $TP$ ), *True Negatives* ( $TN$ ), *False Positives* ( $FP$ ) und *False Negatives* ( $FN$ ) auftreten. Das verstehen beiden so noch nicht und deshalb stellen Sie für Tina und Mark 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)**

<sup>17</sup>Die 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

## 127. Aufgabe

(12 Punkte)

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



**Network-Marketing oder Schneeballschlacht!** Mark, Steffen und Alex sitzen bei Tina und hören sich etwas über Network-Marketing an. Tina ist jetzt im Network-Marketing tätig. 'Jetzt reicht es. Wir sind eine Lerngruppe und du versuchst uns hier abzuziehen!', poltert Mark und fährt fort, 'Ich erklär dir mal, wie falsch du liegst!'. 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. 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? Mark und Sie wollen hier einmal in die Untiefen des »passiven Einkommens« abtauchen und die Lerngruppe vor Schlimmeren bewahren<sup>18</sup>!

Das Jahr 2022 war das erfolgreichste Jahr in der Geschichte von Direct Finanzanlagen Left/Right (D-FL/R). Das Unternehmen steigerte den Umsatz um rund 27 Prozent von 290 Millionen Euro im Jahr 2021. Doch wie viel kommt bei den Partnern an? Laut D-FL/R habe das Unternehmen  $3.3 \times 10^5$  aktive Partner.

1. Berechnen Sie zuerst den Umsatz der Firma D-FL/R 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 35%? **(1 Punkt)**

Das von Tina zu vermarkende Produkt, hinter dem Tina voll steht, kostet 200EUR pro Einheit im Direktverkauf. Die direkte Provision für die erste Stufe beträgt 40%. Für die zweite, dritte und vierte Stufe betragen die indirekten Provisionen jeweils 1.75%, 0.75% und 0.25%. Jeder von Tina angeworbener »Partner« wirbt wiederum fünf Partner für sich selbst an. Pro Monat werden im Schnitt vier Einheiten vom Produkt verkauft. Tina will nun 1800EUR im Monat *passiv* – also durch indirekte Provisionen – erwirtschaften. Kann das klappen? Sie sind zusammen mit Mark skeptisch.

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)**

Tina musste zum Einstieg bei D-FL/R Einheiten des Produkts für 7000EUR kaufen. Diese Einheiten kann Tina nur direkt verkaufen. Das ganze Wohnzimmer ist voll davon. Leider musste Tina den Kauf über einen Kredit über 6% p.a. über 48 Monate finanzieren. Sie schütteln den Kopf und klären Tina über Zinsen auf.

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)**

<sup>18</sup>Die 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](#)

## 128. Aufgabe

(10 Punkte)

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



**Höhlen & Drachen** Nilufar, Paula und Steffen sitzen bei Yuki nachdem sich alle begeistert in der Serie *Stranger Thinks* verloren haben. Alle drei wollen jetzt einmal bei Yuki *Höhlen & Drachen* ausprobieren. Um Geld zu sparen, das Zeug kostet echt, wurde etwas an den Regeln gebastelt. Schnell stellen die Drei fest, dass hier ganz schön viele unterschiedliche Würfel durch die Gegend fliegen. Daher müssen Sie sich jetzt einiges an Fragen klären damit Yuki nicht so alleine ist.

In dem Spiel hat Nilufar nun auf einmal 7 zwölfseitige Würfel ( $7d_{12}$ ) zum würfeln in der Hand. Wenn Nilufar eine 12 würfelt, hat Nilufar einen Erfolg.

1. Berechnen Sie die Wahrscheinlichkeit *genau* 6 Erfolge zu erzielen! **(2 Punkte)**
2. Berechnen Sie die Wahrscheinlichkeit keinen Erfolg zu erzielen! **(1 Punkt)**

Paula betrachtet nun aufmerksam die ausufernden Ausrüstungstabellen. Paula wird aber geholfen und muss sich jetzt nur zwischen der Axt oder dem Schwert entscheiden.

3. Würden Sie die Axt mit zwei vierseitigen Würfeln ( $2d_4$ ) als Schaden oder das Schwert mit einem achtseitigen Würfel plus 2 ( $1d_8+2$ ) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

Jetzt wird es immer wilder, da Paula und Steffen sich jetzt überlegen müssen, wie wahrscheinlich es ist, dass der Rettungswurf gegen den zaubernden Hexer funktioniert. Paula und Steffen 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  $Pr(B) = 0.7$ . Sie haben aber mitgezählt und festgestellt, dass in 40 von 100 Fällen der Rettungswurf bei einem erfolgreichen Zauber funktioniert hat.

4. Erstellen Sie eine  $2 \times 2$  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  $2 \times 2$  Kreuztabelle! **(2 Punkte)**
7. Bestimmen Sie Wahrscheinlichkeit  $Pr(A|B)$ , dass Ihr Rettungswurf gelingt, wenn der Hexer erfolgreich gezaubert hat! **(1 Punkt)**

## 129. Aufgabe

(12 Punkte)

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



**Retrocheck im TV** Das war zu viel für Nilufar gestern. Die Lerngruppe mit Yuki und Tina ging viel zu lang. Während sie wegdämmert, kommen in ihr seltsame Bilder hoch. '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 Nilufar fragt sich, ob sie nicht doch lieber bezahlter Gast bei Barbara Salesch hätten sein sollten. Aber Nilufar braucht 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 Yuki und Tina das Team der drei Kandidaten. Nilufar braucht dringend Ihre Hilfe in ihren Wahnträumen. Sie gehen wie in *Interception* rein!<sup>19</sup>

Name	$P(\text{win})$	$P(\text{outbid})$
Yuki	0.1	0.08
Tina	0.2	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 0.076 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 mit Nilufar auf der Kirmes und spielen mit einem einäugigen Piraten um das große Geld. Das Glücksrad hat 22 Felder. Sie beide drehen das Glücksrad zweimal. Auf 10 Feldern gewinnen Nilufar und Sie 3000EUR 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 4000EUR? **(1 Punkt)**

Im Fiebertraum von Nilufar reisen sie beide im Zug nach Köln um bei »Geh aufs Ganze!« mitzuspielen. Nilufar und Sie schaffen es tatsächlich ins Finale und können als Hauptgewinn ein Auto hinter einer der drei Türen gewinnen. Und was braucht man mehr als ein Auto in einem Fiebertraum?

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«! Berechne Sie dazu die Wahrscheinlichkeiten der einzelnen Pfade. Lohnt sich ein Wechsel der anfangs gewählte Tür? Begründen Sie Ihre Antwort mathematisch! **(2 Punkte)**

<sup>19</sup>South Park Inception Spoof – Wunderbare South Park Folge

## Teil XII.

# Angewandte Nutztier- und Pflanzenwissenschaften (M.Sc.)

### 130. Aufgabe

(6 Punkte)

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

### 131. 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
- $\mu$ : 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.

### 132. Aufgabe

(6 Punkte)

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