

Nicht bestanden: ☐

**Vorname:** \_\_\_\_\_

**Matrikelnummer:** \_\_\_\_\_

**Endnote:** \_\_\_\_\_

**Fakultät Agrarwissenschaften und Landschaftsarchitektur (AuL)**

# Klausurfragen Bio Data Science

**für Pflichtmodule**

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

**(Prüfungsleistung der Wahlpflichtmodule ist eine Portfolioprüfung)**

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

Wintersemester 2024/25



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



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, dass 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 78 Punkten sind aus den Rechen- und Textaufgaben erreicht.  
\_\_\_\_\_ von 98 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
93.5 - 98.0	1,0
88.5 - 93.0	1,3
84.0 - 88.0	1,7
79.0 - 83.5	2,0
74.0 - 78.5	2,3
69.0 - 73.5	2,7
64.0 - 68.5	3,0
59.5 - 63.5	3,3
54.5 - 59.0	3,7
49.0 - 54.0	4,0

Es ergibt sich eine Endnote von \_\_\_\_\_.

## Multiple Choice Aufgaben

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

	A	B	C	D	E	✓
<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>	8	9	10	20	9	10	12

- Es sind \_\_\_\_ von 78 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)

Aus einem Feldversuch ergibt sich die Notwendigkeit der Berechnung einer einfaktoriellen ANOVA. Es ergibt sich ein  $\eta^2 = 0.78$ . Welche Aussage ist richtig?

- A** ☐ Das  $\eta^2$  ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.
- B** ☐ 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.
- C** ☐ Die Berechnung von  $\eta^2$  ist ein Wert für die Interaktion.
- 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  $\eta^2 = 0.28$ . Welche Aussage ist richtig?

- A** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Umweltbedingungen erklärt wird. Daher werden 28% der Varianz durch die Umweltbedingungen erklärt. Der Anteil der Varianz durch die Behandlungsgruppen ist dann 72%.
- B** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird. Daher werden 28% der Varianz durch die Behandlungsgruppen erklärt.
- C** ☐ 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.
- D** ☐ Es werden 28% der Varianz durch den Versuch erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der durch Fehler in der Versuchsdurchführung entsteht.
- E** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der durch den Forschenden entsteht. Es gilt die Regel, dass ca. 70% der Varianz eines Versuches durch die Versuchsdurchführung entstehen sollen.

#### 3. Aufgabe

(2 Punkte)

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

- A** ☐ Die ANOVA berechnet die F-Statistik indem die MS der Behandlung durch die MS des Fehlers geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese nicht abgelehnt werden.
- B** ☐ 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.

- C** ☐ 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.
- D** ☐ 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.
- 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)

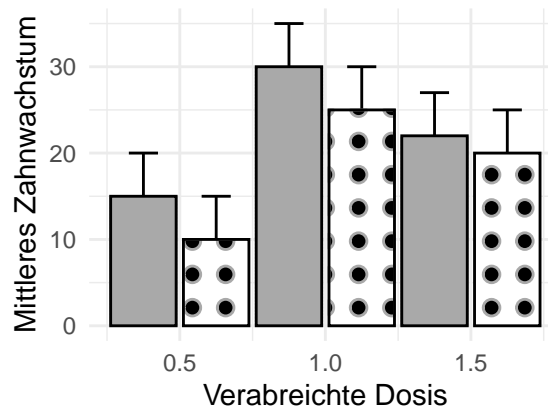
Viele statistische Verfahren nutzen eine Teststatistik um eine Aussage über den Zusammenhang zwischen der Grundgesamtheit und der Stichprobe abzubilden. Ein statistisches Testwerkzeug ist hierbei die ANOVA. Die ANOVA rechnet dabei...

- A** ☐ ... den Unterschied zwischen der Varianz über alle Behandlungsgruppen und der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, muss ein Posthoc-Test angeschlossen werden.
- B** ☐ ... den Unterschied zwischen der Mittelwerte und der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, ist bekannt welcher Vergleich konkret unterschiedlich ist.
- 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 über alle Behandlungsgruppen oder der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, muss sich zwischen einem der beiden Varianzquellen entschieden werden.
- E** ☐ ... den Unterschied zwischen der F-Statistik anhand der Varianz der Gruppen. Wenn die F-Statistik exakt 0 ist, kann die Nullhypothese abgelehnt werden.

#### 5. Aufgabe

(2 Punkte)

Ein Versuch wurde an 82 Tieren durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0.5, 1 und 1.5 mg/Tag) über eine von zwei Verabreichungsmethoden erhielt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung von Vitamin C auf das Zahnwachstum bei Hasen. Welche Aussage ist richtig, wenn Sie eine zweifaktorielle ANOVA rechnen?



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

## Deskriptive Statistik & Explorative Datenanalyse

### 6. Aufgabe

(2 Punkte)

Gegeben ist  $y$  mit 13, 9, 10, 7 und 9. Berechnen Sie den Mittelwert und Standardabweichung.

- A ☐ Es ergibt sich  $9.6 \pm 2.19$
- B ☐ Es ergibt sich  $10.6 \pm 1.095$
- C ☐ Es berechnet sich  $10.6 \pm 4.8$
- D ☐ Sie erhalten  $9.6 \pm 1.095$
- E ☐ Sie erhalten  $9.6 \pm 1.48$

### 7. Aufgabe

(2 Punkte)

Berechnen Sie den Median, das 1<sup>st</sup> Quartile sowie das 3<sup>rd</sup> Quartile von  $y$  mit 19, 21, 21, 21, 24, 31, 25, 21, 18, 14 und 42.

- A ☐ Es ergibt sich  $23 \pm 19$
- B ☐ Es ergibt sich  $21 \pm 19$
- C ☐ Es berechnet sich 23 [20; 26]
- D ☐ Sie erhalten 21 [19; 25]
- E ☐ Sie erhalten 21 [17; 23]

### 8. Aufgabe

(2 Punkte)

Sie überlegen Ihre Daten mit einem Histogramm zu visualisieren. Was ist die minimale Anzahl an Beobachtungen pro Gruppe ?

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

### 9. Aufgabe

(2 Punkte)

Die Standardabweichung ist eine bedeutende deskriptive Statistik für die Analyse von Daten. Wie müssen Sie vorgehen um die Standardabweichung zu berechnen?

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

## 10. Aufgabe

(2 Punkte)

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

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

## 11. Aufgabe

(2 Punkte)

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

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

## 12. Aufgabe

(2 Punkte)

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

- A** ☐ 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.
- B** ☐ In einer explorativen Datenanalyse nutzen wir den Violinplot. Dabei sollte der Bauch am Rand liegen. Dann können wir von einer Normalverteilung ausgehen.
- C** ☐ Nach dem Einlesen der Daten nutzen wir einen Barplot um zu schauen, ob alle Mittelwerte über alle Behandlungen in etwa gleich groß sind. Damit ist dann auch die Varianz in allen Behandlungen in etwa gleich.
- D** ☐ Wir erstellen uns für jede Behandlung einen Dotplot und schauen, ob die Dots und damit die Varianz für jede Behandlung gleich groß sind.
- E** ☐ Einen Boxplot. Das IQR muss über alle Behandlungen zusammen mit den Whiskers ungefähr gleich aussehen.



### 13. Aufgabe

(2 Punkte)

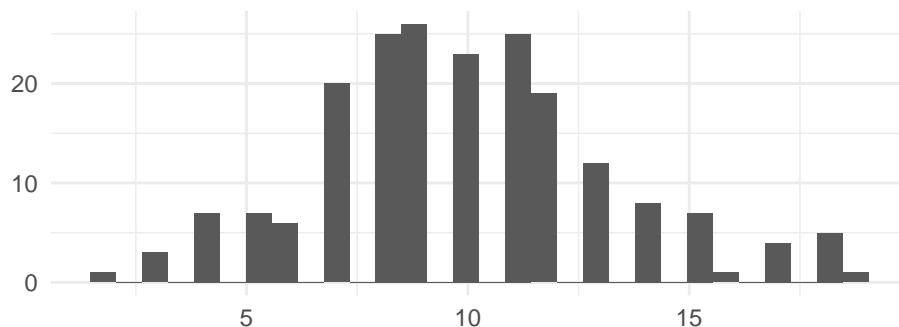
Sie wollen in Ihrer Abschlussarbeit über eine explorative Datenanalyse überprüfen, ob Ihr gemessener Endpunkt einer Normalverteilung folgt. Welche drei Abbildungen eignen sich insbesondere für die Überprüfung?

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

### 14. Aufgabe

(2 Punkte)

In dem folgenden Histogramm von  $n = 194$  Pflanzen ist welche Verteilung abgebildet?



- A** ☐ Es handelt sich um eine Poisson-Verteilung.
- B** ☐ Wir haben eine Gammaverteilung vorliegen.
- C** ☐ In dem Histogramm ist eine Normalverteilung dargestellt.
- D** ☐ In dem Histogramm ist eine Ordinalverteilung dargestellt.
- E** ☐ Eine Standardnormalverteilung.

## Lineare Regression & Korrelation

### 15. Aufgabe

(2 Punkte)

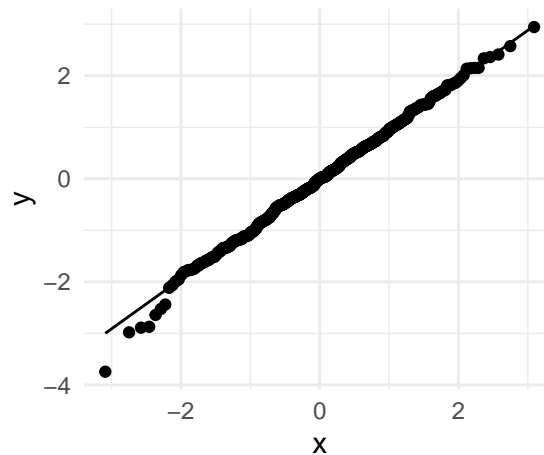
Im Allgemeinen gibt es zwei mögliche Ziele für ein Regressionsmodell. Wir können eine Vorhersagemodell oder ein kausales Modell rechnen. Welche Aussage ist für ein prädiktives Modell richtig?

- A** ☐ Es wird ein Trainingsdatensatz zum Modellieren des Trainingsmodells benötigt. Der Testdatensatz dient rein zur Visualisierung. Dies gilt vor allem für ein prädiktives Modell.
- B** ☐ Es wird ein Trainingsdatensatz zum Trainieren des Modells benötigt. Der Testdatensatz dient zur Validierung. Dies gilt insbesondere für ein prädiktives Modell.
- C** ☐ Wir modellieren den Zusammenhang zwischen  $X$  und  $Y$  wenn ein prädiktives Modell rerechnet wird. Dabei kann der gesamte Datensatz genutzt werden. Eine Aufteilung wie in einem prädiktiven Modell ist nicht notwendig.
- D** ☐ Ein prädiktives Modell wird auf einem Trainingsdatensatz trainiert und anschliessend über eine explorative Datenanalyse validiert. Signifikanzen über  $\beta_i$  können hier nicht festgestellt werden.
- E** ☐ Wenn ein prädiktives 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.

## 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?

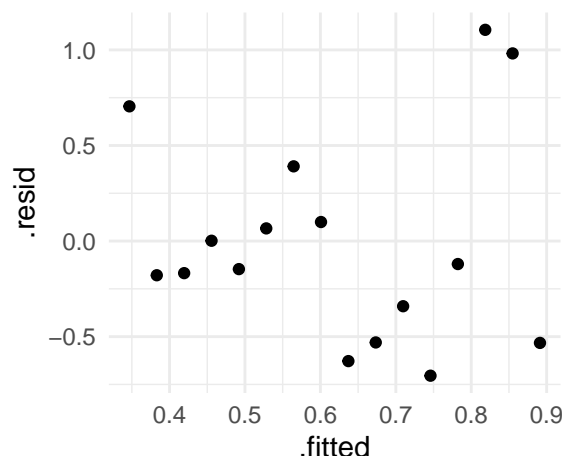


- A** ☐ Wir betrachten die Punkte. Wenn die Punkte einigermaßen gleichmäßig verteilt liegen, dann gehen wir von normalen Residuen aus.
- B** ☐ Wir betrachten die Gerade, die durch die einzelnen Punkte laufen sollte. Wenn die 95% der Punkte von der Geraden getroffen werden, dann gehen wir von normalverteilten Residuen aus.
- C** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden und Korrelation ist negativ.
- D** ☐ Wir betrachten die Punkte auf der Geraden. Wenn die Punkte einigermaßen auf der Geraden liegen, dann gehen wir von normalverteilten Residuen aus. Wir können hier von normalverteilten Residuen ausgehen.
- E** ☐ Wir betrachten die Gerade und dabei insbesondere die beiden Enden der Gerade. Hier sollten die Punkte auf der Geraden liegen, dann ist die Annahme an die Normalverteilung der Residuen erfüllt. Diese Annahme ist nicht erfüllt.

## 17. Aufgabe

(2 Punkte)

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



- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Es ist ein Muster zu erkennen und wir können damit auf die Signifkanz von  $x_1, \dots, x_p$  schließen.

- B** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Es ist kein Muster zu erkennen.
- C** ☐ Die Punkte müssen gleichmäßig in dem positiven Bereich liegen. Dies ist hier klar nicht der Fall. Einzelne Ausreißer können beobachtet werden. Die Analyse ist gescheitert.
- D** ☐ Die Punkte müssen gleichmäßig, mit ähnlichen Abständen, in dem positiven wie auch negativen Bereich liegen. Dies ist hier klar nicht der Fall. Einzelne Ausreißer können beobachtet werden. Wir können mit dem Model so nicht rechnen und müssen erst die auffälligen Werte gesondert betrachten.
- E** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Kein Muster ist zu erkennen und keine Outlier zu beobachten.

## 18. Aufgabe

(2 Punkte)

Welche Aussage über den Korrelationskoeffizienten  $\rho$  ist richtig?

- A** ☐ 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.
- B** ☐ Der Korrelationskoeffizienten  $\rho$  liegt zwischen -1 und 1. Darüber hinaus ist der Korrelationskoeffizienten  $\rho$  einheitslos und kann als standardisierte Steigung verstanden werden.
- C** ☐ 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.
- D** ☐ 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.
- E** ☐ Der Korrelationskoeffizienten  $\rho$  ist eine veraltete Darstellungsform von Effekten in der linearen Regression und wird wie das  $\eta^2$  aus der ANOVA interpretiert. Der Korrelationskoeffizienten  $\rho$  beschreibt den Anteil an erklärter Varianz durch die Regression.

## 19. Aufgabe

(2 Punkte)

Nach einer simplen linearen Regression zur Untersuchung vom Einfluss der  $PO_2$ -Konzentration in  $[\mu g]$  im Wasser auf das Wachstum von Spitzkohl in  $[kg]$  erhalten Sie einen  $\beta_{PO_2}$  Koeffizienten von  $6.9 \times 10^{-7}$  und einen hoch signifikanten  $p$ -Wert mit  $1e-04$ . Warum sehen Sie so einen kleinen Effekt bei einer so deutlichen Signifikanz?

- A** ☐ Manchmal ist die Einheit der Einflussvariable  $X$  zu klein gewählt, so dass der Anstieg von 1 Einheit in  $X$  zu einer zu kleinen Änderung in  $y$  führt. Daher kann der Effekt  $\beta_{PO_2}$  sehr klein wirken, aber auf einer anderen Einheit sehr viel größer sein. Der  $p$ -Wert wird auf einer einheitslosen Teststatistik bestimmt.
- B** ☐ 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.
- C** ☐ 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.
- D** ☐ Das Gewicht und die  $PO_2$ -Konzentration korrelieren sehr stark, deshalb wird der  $\beta_{PO_2}$  Koeffizient sehr klein. Mit einer ANOVA kann für die Korrelation korrigiert werden und der Effektschätzer passt dann zum  $p$ -Wert.
- E** ☐ 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.

## 20. Aufgabe

(2 Punkte)

Sie wollen nach der explorativen Datenanalyse (EDA) Ihre Daten in der Abschlussarbeit auswerten. Nach einiger Recherche finden Sie heraus, dass Sie zuerst die Daten mit der Funktion `lm()` in R modellieren müssen. Welche Anwendung folgt drauf?

- A** ☐ Die Funktion `lm()` in R ist der letzte Schritt für einen Gruppenvergleich. Vorher kann eine ANOVA oder aber ein multipler Vergleich in `{emmeans}` gerechnet werden. In der Funktion `lm()` werden die Gruppenvarianzen bestimmt.
- B** ☐ Ist die Einflussvariable  $X$  ein Faktor so werden die Gruppenmittelwerte geschätzt und eine anschließende ANOVA sowie multipler Gruppenvergleich mit `{emmeans}` ist möglich. Die Funktion `lm()` kann dabei eigentlich weggelassen werden, wird aber traditionell gerechnet.
- C** ☐ Die Funktion `lm()` in R wird klassischerweise für die nicht-lineare Regression genutzt. Ist die Einflussvariable  $X$  numerisch so werden die Gruppenmittelwerte geschätzt.
- D** ☐ Die Funktion `lm()` berechnet die Varianzstruktur für eine ANOVA. Dannach kann dann über eine explorative Datenanalyse nochmal eine Signifikanz berechnet werden. Sollte vor der Verwendung der Funktion `lm()` schon eine EDA gerechnet worden sein, so ist die Analyse wertlos.
- E** ☐ Ist die Einflussvariable  $X$  ein Faktor so werden die Gruppenmittelwerte geschätzt und eine anschließende ANOVA sowie multipler Gruppenvergleich mit `{emmeans}` ist möglich. Dennoch muss zuerst ein lineares Modell mit der Funktion `lm()` in R gerechnet werden.

## 21. Aufgabe

(2 Punkte)

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

- A** ☐ Das GLM erlaubt auch nicht normalverteilte Residuen in der Schätzung der Regressionsgrade.
- B** ☐ Dank dem *generalisierten linearen Modell (GLM)* können auch andere Verteilungsfamilien als die Normalverteilung mit einer linearen Regression modelliert werden.
- C** ☐ Das GLM ist ein faktisch maschineller Lernalgorithmus, der selbstständig die Verteilungsfamilie für  $Y$  wählt.
- D** ☐ 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.
- E** ☐ Das GLM ist eine Vereinfachung des LM in R. Mit dem GLM lassen sich polygonale Regressionen rechnen. Somit stehen neben der Normalverteilung noch weitere Verteilungen zu Verfügung.

## 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** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- B** ☐ Randomisierung ist die direkte Folge von Strukturgleichheit. Die Strukturgleichheit erlaubt es erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- C** ☐ Durch eine Randomisierung können wir von Strukturgleichheit zwischen der Stichprobe und der Grundgesamtheit ausgehen.
- D** ☐ 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.
- E** ☐ Randomisierung war bis 1952 bedeutend, wurde dann aber in Folge besserer Rechnerleistung nicht mehr verwendet. Aktuelle Statistik nutzt keine Randomisierung mehr.

### 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** ☐ Die Spracherkennung von **R** ist nicht in der Lage Deutsch zu verstehen.
- B** ☐ Programmiersprachen können nur englische Begriffe verarbeiten. Zusätzliche Pakete können zwar geladen werden, aber meist funktionieren diese Pakete nicht richtig. Deutsch ist International nicht bedeutend genug.
- C** ☐ Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Die Nutzung von englischer Sprache umgeht dieses Problem in eleganter Art.
- D** ☐ Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- E** ☐ Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Daher ist die Nutzung in Deutsch in den AGBs von **R** untersagt.

### 24. Aufgabe

(2 Punkte)

Nachdem Sie Ihr Feldexperiment als Vorversuch für Ihre Abschlussarbeit abgeschlossen haben, wollen Sie in einer explorativen Datenanalyse (EDA) in **R** einmal schauen, ob Sie überhaupt Effekte der Behandlung vorliegen haben. Welche Reihenfolge von Schritten müssen Sie in **R** durchführen, damit Sie eine EDA rechnen können?

- A** ☐ Die Funktionsreihenfolge ist wie folgt: `read_excel()` -> `mutate()` -> `ggplot()`. Dabei ist bei der Transformation der Daten darauf zu achten, dass keine Faktoren erstellt werden.
- B** ☐ 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.
- C** ☐ 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.
- D** ☐ Für eine explorativen Datenanalyse (EDA) in **R** müssen wir als erstes die Daten über `read_excel()` einlesen. Danach müssen wir schauen, dass wir die Zeilen richtig über `mutate()` transformiert haben. Insbesondere müssen Variablen mit kontinuierlichen Werten in einen Faktor umgewandelt werden. Am Ende nutzen wir die Funktion `ggplot()` für die eigentlich EDA.
- E** ☐ 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.

### 25. Aufgabe

(2 Punkte)

Es sei  $s_1^2 = s_2^2$  in dem Modell  $Y \sim X$ . Welche Aussage ist richtig?

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

## 26. Aufgabe

(2 Punkte)

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

- A** ☐ Untereinander stark korreliert. Die Ferkel sind von einer Mutter und somit miteinander korreliert. Dies wird in der Statistik jedoch meist nicht modelliert.
- B** ☐ Untereinander unabhängig. Sollten die Mütter verwandt sein, so ist die Varianzstruktur ähnlich und muss modelliert werden.
- C** ☐ Untereinander unabhängig. Die Ferkel sind eigenständig und benötigen keine zusätzliche Behandlung.
- D** ☐ 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.
- E** ☐ Die Ferkel stammen von der gleichen Sau und sind somit untereinander unabhängig.

## 27. Aufgabe

(2 Punkte)

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Kühen durch. Bei 6 Tieren finden Sie eine Erkrankung der Klauen vor und 7 Tiere sind gesund. Welche Aussage über den Effektschätzer Risk ratio ist richtig?

- A** ☐ Es ergibt sich ein Risk ratio von 0.46, da es sich um eine Chancenverhältnis handelt.
- B** ☐ Das Verhältnis von Chancen Risk ratio ergibt ein Chancenverhältnis von 0.86.
- C** ☐ Das Verhältnis der Chancen Risk ratio ergibt ein Chancenverhältnis von 0.46. Wir sind an der Chance krank zu sein interessiert.
- D** ☐ Der Anteil der Kranken wird berechnet. Da es sich um ein Anteil handelt ergibt sich ein Risk ratio von 0.46.
- E** ☐ Da es sich um ein Chancenverhältnis handelt ergibt sich ein Risk ratio von 2.17.

## 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** ☐ Riesige Datensatz haben mehr Fallzahl was zur  $\alpha$ -Inflation führt. Durch eine Adjustierung kann dem Problem entgegengewirkt werden.
- 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** ☐ Eine große Fallzahl führt zu mehr signifikanten Ergebnissen auch bei kleinen Effekten. Daher werden fast alle Vergleich signifikant, wenn die Fallzahl nur groß genug wird.
- 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** ☐ Aktuell werden immer größere Datensätze erhoben. Dadurch wird auch die Varianz immer höher was automatisch zu mehr signifikanten Ergebnissen führt.

## Multiple Gruppenvergleiche

### 29. Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.01, 0.21, 0.34 und 0.02. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.04, 0.84, 1 und 0.08. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0025, 0.0525, 0.085 und 0.005. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0025, 0.0525, 0.085 und 0.005. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.04, 0.84, 1 und 0.08. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.04, 0.84, 1.36 und 0.08. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.

### 30. Aufgabe

(2 Punkte)

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

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

### 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** ☐ Sie nutzen das R Paket {emmeans} für die Berechnung des multiplen Gruppenvergleichs. Die Ausgabe der Funktion emmeans() erlaubt zügig über {ggplot} einen Barplot zu erstellen und dann auch das CLD zu berechnen. Sie haben alles sofort zusammen.
- 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** ☐ 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.
- D** ☐ Das R Paket {lm}. Das Paket {lm} erstellt selbstständig Konfidenzintervalle und entsprechende p-Werte. Da wir in dem Paket nicht adjustieren müssen, ist es bei Anwendern sehr beliebt.
- E** ☐ Das R Paket {hmisc} erlaubt die Durchführung eines multiplen Gruppenvergleichs aus verschiedenen Modellen heraus. Aus einem hmisc Objekt lässt sich recht einfach das CLD erstellen und so über Barplots eine schnelle Interpretation der statistischen Auswertung durchführen.

### 32. Aufgabe


(2 Punkte)

Bei einem Posthoc-Test kann es zu einer überraschenden Besonderheit beim statistischen Testen kommen. Wie lautet der Fachbegriff und wie kann mit der überraschenden Besonderheit umgegangen werden?

- A** ☐ 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.
- B** ☐ Das globale Signifikanzniveau explodiert und erreicht Werte größer als Eins. Es kommt zu einer  $\alpha$ -Inflation. Dagegen kann mit der Adjustierung der  $\alpha$ -Werte nach Bonferroni vorgegangen werden.
- C** ☐ Beim multiplen Testen kann es zu Varianzheterogenität kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5%. Daher müssen die p-Werte entsprechend adjustiert werden. Das Verfahren nach Welch, bekannt aus dem t-Test, ist hier häufig anzuwenden.
- D** ☐ Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern sehr viel höher. Es kommt zu einer  $\alpha$ -Inflation. Dagegen kann mit der Adjustierung der p-Werte nach Bonferroni vorgegangen werden.
- E** ☐ 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.

### 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** ☐ 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.
- B** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  nicht adjustiert werden im Gegensatz zu den p-Werten.
- C** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  adjustiert werden im Gegensatz zu den p-Werten.
- D** ☐ 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.
- E** ☐ Beim multiplen Testen werden die Effekte der paarweisen Vergleiche ignoriert. Der Nachteil des multiplen Testens ist ja auch, dass wir am Ende keine Effekte mehr vorliegen haben. Eine ANOVA liefert hier bessere Informationen.

## Statistische Testtheorie

### 34. Aufgabe

(2 Punkte)

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

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



### 35. Aufgabe

(2 Punkte)

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

- A** ☐ ... dass ein schlechtes Modell durch ein schlechteres Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.
- B** ☐ ... dass ein minderwertes Modell durch ein weniger minderwertiges Modell ersetzt wird. Es gilt das Falsifikationsprinzip nach Karl Popper.
- C** ☐ ... dass Modelle meist falsch sind und selten richtig.
- D** ☐ ... dass Fehlerterme in statistischen Modellen nicht verifiziert werden können.
- E** ☐ ... dass ein minderwertes Modell durch ein minderwertiges Modell ersetzt wird. Es gilt das Verifikationsprinzip nach Karl Popper.

### 36. Aufgabe

(2 Punkte)

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

- A** ☐ Da Wissenschaftler eine Schwelle für die statistische Testentscheidung benötigen wurde  $\alpha$  in einer großen Konferenz 1945 gewählt. Damit ist  $\alpha = 5\%$  eine Kulturkonstante mit einem Rang einer Naturkonstante.
- B** ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.
- 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** ☐ Als Kulturkonstante hat  $\alpha = 5\%$  den Rang einer Naturkonstante und wurde nach langer Diskussion in der UN im Jahre 1983 festgesetzt. Damals auch schon mit der Zustimmung der UdSSR.
- E** ☐ Da Wissenschaftler eine Schwelle für die statistische Testentscheidung benötigen wurde  $\alpha$  historisch gewählt. Damit ist  $\alpha = 5\%$  eine Kulturkonstante.

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

### 38. Aufgabe

(2 Punkte)

Eine Analogie kann helfen einen Sachverhalt besser zu verstehen. Wie kann folgende Aussage richtig in die Analogie der statistischen Testtheorie gesetzt werden?

*$H_0$  beibehalten obwohl die  $H_0$  falsch ist*

- A ☐ *Alarm without fire*, dem  $\alpha$ -Fehler in der Analogie eines Rauchmelders.
- B ☐ *Alarm with fire*, dem  $\alpha$ -Fehler in der Analogie von Feuer.
- C ☐ In die Analogie eines brennenden Hauses ohne Rauchmelder: *House without noise*.
- D ☐ In die Analogie eines Rauchmelders: *Alarm without fire police*, dem  $\alpha$ -Fehler.
- E ☐ Dem  $\beta$ -Fehler mit der Analogie eines Rauchmelders: *Fire without alarm*.

### 39. Aufgabe

(2 Punkte)

Sie lesen eine wissenschaftliche Arbeit, die damit wirbt, dass Effekte und Signifikanz nicht separat dargestellt sind, sondern in einer statistischen Maßzahl zusammen. Welche Aussage ist richtig?

- A ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- B ☐ Der p-Wert. Durch den Vergleich mit  $\alpha$  lässt sich über die Signifikanz entscheiden und der  $\beta$ -Fehler erlaubt über die Power eine Einschätzung der Relevanz.
- C ☐ 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.
- D ☐ Das Konfidenzintervall. Durch die Visualisierung des Konfidenzintervalls kann eine Relevanzschwelle vom Anwender definiert werden. Zusätzlich erlaubt das Konfidenzintervall auch eine Entscheidung über die Signifikanz.
- E ☐ 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.

### 40. Aufgabe

(2 Punkte)

Sie haben ein Signifikanzniveau  $\alpha$  gleich 5% vorliegen. Welche Aussage zusammen mit dem p-Wert ist richtig?

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

#### 41. Aufgabe

(2 Punkte)

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

- A** ☐ In der Analogie der Regenwahrscheinlichkeit in einem bestimmten Gebiet: ein statistischer Test gibt die Wahrscheinlichkeit für ein Ereignis in einem Experiment mit den Daten  $D$  wieder und lässt sich kaum verallgemeinern.
- B** ☐ In der Analogie der Sonnenscheindauer: Wie lange kann mit einem entsprechenden Effekt gerechnet werden? Die Wahrscheinlichkeit für den Effekt gibt der statistische Test wieder.
- C** ☐ 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.
- D** ☐ In der Analogie der Durchschnittstemperatur: Wie oft tritt ein Effekt durchschnittlich ein? Wir erhalten eine Wahrscheinlichkeit für die Effekte. Zum Beispiel, wie hoch ist die Wahrscheinlichkeit für einen Mittelwert als Durchschnitt.
- E** ☐ Die Analogie der Regenwahrscheinlichkeit: der statistische Test erlaubt es die Wahrscheinlichkeit für Regen abzuschätzen jedoch nicht die Menge und somit den Effekt.

#### 42. Aufgabe

(2 Punkte)

Sie wollen eine Aussage über ein untersuchtes Individuum treffen. Dazu nutzen Sie einen statistischen Test. Können Sie eine valide Aussage aus einem statistischen Test erhalten?

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

#### 43. Aufgabe

(2 Punkte)

Welche Aussage über die *Power* ist richtig?

- A** ☐ Alle statistischen Tests sind so konstruiert, dass die  $H_A$  mit 20% *bewiesen wird*. Die Power ist  $1 - \beta$  mit  $\beta$  gleich 80% gesetzt.
- B** ☐ Es gilt  $\alpha + \beta = 1$  und somit liegt  $\beta$  meist bei 95%.
- C** ☐ Die Power beschreibt die Wahrscheinlichkeit die  $H_A$  abzulehnen. Wir testen die Power jedoch nicht.
- D** ☐ Die Power  $1 - \beta$  wird auf 80% gesetzt. Alle statistischen Tests sind so konstruiert, dass die  $H_A$  mit 80% *bewiesen wird*.
- E** ☐ Die Power  $1 - \beta$  wird auf 80% gesetzt. Damit liegt die Wahrscheinlichkeit für die  $H_0$  bei 20%.

#### 44. Aufgabe

(2 Punkte)

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

- A** ☐ 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.
- 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 den Output oder die Wiedergabe eines Tests in einem Computer.
- D** ☐ Der Forschende muss am Anfang wissen, ob das Ergebnis eines Experiments relevant für seine Forschung ist. Dafür kann der Effekt eines statistischen Tests genutzt werden oder auch der Prähoc-Test. Damit beschreibt der Effekt den biologischen interpretierbaren Teil eines Experiments vor der Durchführung. Zum Beispiel der Unterschied zwischen zwei Mittelwerten.
- E** ☐ 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.

#### 45. Aufgabe

(2 Punkte)

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

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

#### 46. Aufgabe

(2 Punkte)

Ein statistischer Test benötigt für die richtige Durchführung Hypothesen  $H$ , sonst ist der Test nicht zu interpretieren. Welche Aussage ist richtig?

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

## Statistische Tests für Gruppenvergleiche

### 47. Aufgabe

(2 Punkte)

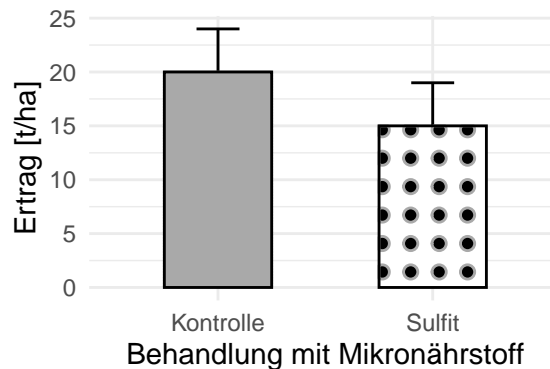
In Ihrer Abschlussarbeit rechnen Sie einen Student t-Test. Welche Aussage ist auch für den Welch t-Test richtig?

- A** ☐ Der t-Test ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern
- B** ☐ Der t-Test vergleicht zwei Gruppen indem die Mittelwerte miteinander verglichen werden.
- 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 testet generell zu einem erhöhten  $\alpha$ -Niveau von 20%.
- E** ☐ Der t-Test vergleicht zwei oder mehr Gruppen indem die Mittelwerte miteinander verglichen werden.

### 48. Aufgabe

(2 Punkte)

In einer Studie zur Bewertung der Wirkung des Mikronährstoff Sulfit auf den Ertrag in t/ha von Mais im Vergleich zu einer Kontrolle entstand folgende Abbildung. Der Versuch wurde in 9 Parzellen pro Gruppe durchgeführt. Welche Aussage ist im Bezug auf einen t-Test ist richtig?



- A** ☐ Der Test deutet auf einen signifikanten Unterschied hin. Der Effekt liegt vermutlich bei -5.
- B** ☐ Die Barplots deuten auf keinen signifikanten Unterschied. Der Effekt liegt vermutlich bei -5 unter einer groben Abschätzung.
- C** ☐ Nach Betrachtung des Barplots liegt kein signifikanter Unterschied vor. Der Effekt kann nicht bei einem t-Test aus Barplots bestimmt werden.
- D** ☐ Die Barplots deuten auf keinen signifikanten Unterschied. Der Effekt liegt vermutlich bei -5 unter einer groben Abschätzung. Wir müssen aber eine ANOVA rechnen um den Effekt wirklich bestimmen zu können.
- E** ☐ Es liegt ein signifikanter Unterschied vor. Der Effekt liegt bei -0.5.

### 49. Aufgabe

(2 Punkte)

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

- A** ☐ 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.
- B** ☐ 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.

- 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 nutzt die Varianz der Beobachtungen jeweils paarweise und bildet dafür eine verbundene Stichprobe. Dieser Datensatz  $d$  dient dann zur Differenzbildung.
- E** ☐ 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.

## 50. Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit passen die Ergebnisse einer ANOVA und eines multiplen Vergleiches nicht zusammen. Nach einem Experiment mit fünf Maissorten ergibt eine ANOVA ( $p = 0.048$ ). 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  $p_{3-2} = 0.051$ . Welche Aussage ist richtig?

- A** ☐ 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 Unterschied nachweisen.
- B** ☐ Es gibt einen Fehler in der Varianzstruktur. Daher kann die ANOVA nicht richtig sein und paarweise t-Tests liefern das richtige Ergebnis.
- C** ☐ Der Fehler liegt in den t-Tests. Wenn eine ANOVA signifikant ist, dann muss zwangsweise auch ein t-Test signifikant sein.
- D** ☐ Das ist kein Wunder. Die ANOVA testet nicht auf der gesamten Fallzahl und die paarweisen t-Tests gewinnen immer eine oder mehr Gruppen als Fallzahl dazu. Mit steigender Fallzahl sind mehr signifikante Unterschiede zu erwarten. Die p-Werte unterscheiden sich numerisch auch kaum.
- E** ☐ 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.

## Teil I.

# Deskriptive Statistik & Explorative Datenanalyse

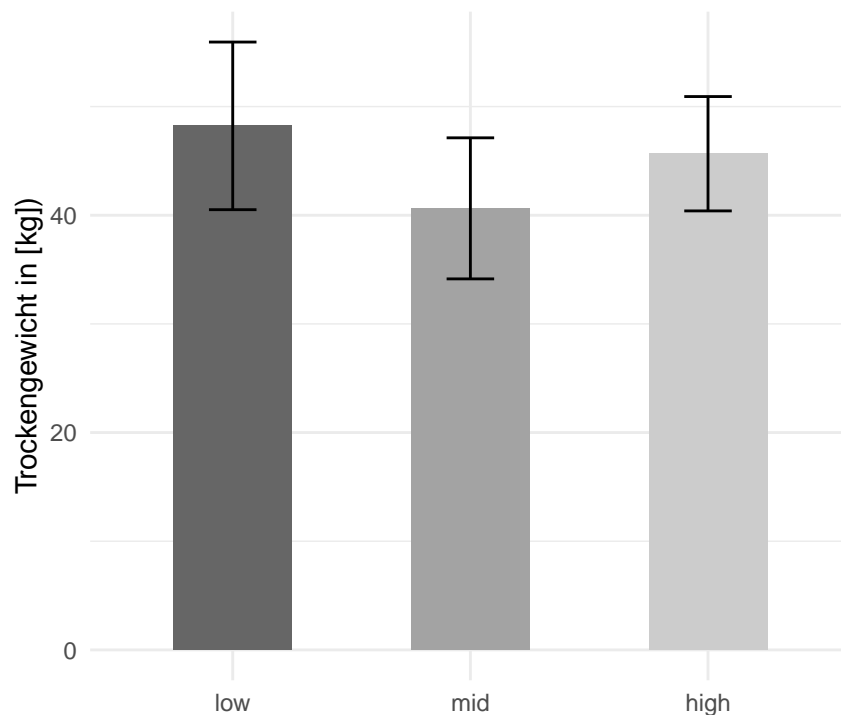
### 51. Aufgabe

(8 Punkte)


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



**Zerforschen des Barplots** Wenn die Gefälligkeit nicht wäre, ja dann wäre wohl vieles möglich für Alex! Aber so.. Deshalb gilt anschauen, was andere vor einem gemacht haben. Für Alex ist es eine Möglichkeit schneller ans Ziel zu gelangen. Alex soll in seiner Abschlussarbeit Maiss untersuchen. Die Behandlung in seiner Abschlussarbeit werden verschiedene Bewässerungstypen (*low*, *mid* und *high*) sein. Erheben wird Alex als Endpunkt (Y) *Trockengewicht* benannt als *drymatter* in seiner Exceldatei. Von seiner Betreuerin erhält er nun folgende Abbildung von Barplots, die er erstmal zur Übung nachbauen soll, bevor er mit dem eigentlichen Versuch beginnt.



Leider kennt sich Alex mit der Erstellung von Barplots in  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  üblichen Format, aus dem die drei Barplots *möglicherweise* erstellt wurden! **(2 Punkte)**
4. Kann Alex 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** Barplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Alex nicht gemerkt, welche statistischen Maßzahlen für einen Barplot erhoben werden müssen. Besser wäre was anderes gewesen. Am Ende dann doch besser Starcraft. Wunderbar. Eine echte Ablenkung für Alex. Das ist in soweit doof, da nach seiner Betreuerin nun Barplots aus seinen Daten gebaut werden sollen, bevor es mit dem statistischen Testen weitergeht. Na dann mal los. Alex schafft sich die nötige Stimmung. Alex streichelt liebevoll die Katze. Der Kopf ist in seinem Schloß vergraben um den Klang von Abba zu dämpfen. Die Behandlung für Kartoffeln waren verschiedene Genotypen (AA, AB und BB). Erfasst wurde von Alex als Messwert (Y) *Ertrag*. Alex hat dann *yield* in seiner Exceldatei eintragen. Aber auch irgendwie egal. Alex will später nochmal raus um zu Laufen. Druck ablassen, dass muss er auch.

treatment	yield
AB	33.7
AA	16.9
AA	23.1
BB	42.4
AB	26.6
AA	14.5
BB	38.4
AB	42.8
BB	39.7
AB	51.2
BB	42.4
AA	23.6

Leider kennt sich Alex mit der Erstellung von Barplots nicht aus. Deshalb braucht er 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 Kartoffeln! Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
3. Beschriften Sie *einen* Barplot mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
4. Wenn Alex *keinen Effekt* zwischen den Behandlungen von Kartoffeln erwarten würde, wie sehen dann die Barplots aus? Antworten Sie mit einer Skizze der Barplots! **(1 Punkt)**



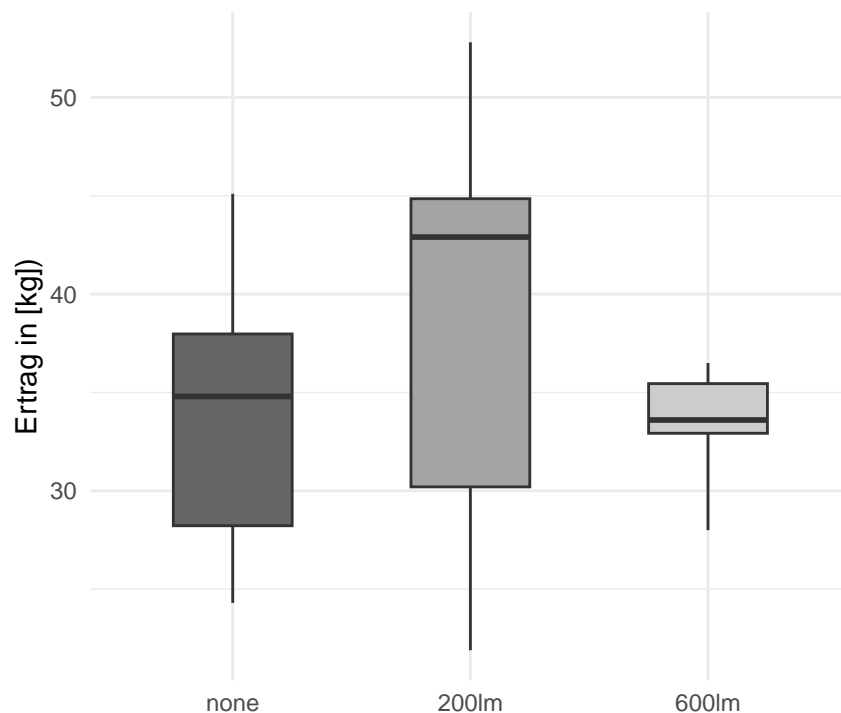
### 53. Aufgabe

(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 Paula nicht gemerkt, welche statistischen Maßzahlen für einen Barplot erhoben werden müssen. Besser wäre was anderes gewesen. Am Ende dann doch besser Harry Potter. Wunderbar. Eine echte Ablenkung für Paula. Das ist in soweit doof, da nach ihrer Betreuerin erstmal ein Barplot nachgebaut werden soll, bevor es mit ihrer Abschlussarbeit losgeht. Dann hat sie schonmal den Code vorliegen und nachher geht dann alles schneller. Na dann mal los. Paula schafft sich die nötige Stimmung. Paula nickt im Takt von White Lies und bemerkt dabei gar nicht was die Ratte schon wieder anstellt. In der Behandlung für Kartoffeln werden verschiedene Lichtstufen (*none*, *200lm* und *600lm*) sein. Erfasst wird als Outcome (Y) *Ertrag*. Paula soll dann *yield* in ihrer Exceldatei eintragen. Aber nur in passender Atmosphäre! Auf seinem Second Screen läuft Jagd auf roter Oktober und Paula schaufelt Smarties. Nicht effizient, aber gut.



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

1. Erstellen Sie eine Tabelle mit den statistischen Maßzahlen aus der obigen Abbildung der drei Boxplots! Beachten Sie die korrekte Darstellungsform der statistischen Maßzahlen! (3 Punkte)
2. Beschriften Sie *einen* der Boxplots mit den gängigen statistischen Maßzahlen! (2 Punkte)
3. Erstellen Sie einen beispielhaften Datensatz, aus dem die drei Boxplots *möglicherweise* erstellt wurden, im üblichen Format! (2 Punkte)
4. Kann Paula 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** Steffen und die Romantik, eine unendliche Geschichte mit kniffligen Wendungen. Deshalb gilt anschauen, was andere vor einem gemacht haben. Für Steffen ist es eine Möglichkeit schneller ans Ziel zu gelangen. Deshalb hat sich Steffen viele Poster in der Fakultät angeschaut und ist zum Schluß gekommen, dass Boxplots eine häufig genutzte Abbildung sind. Steffen soll nun in seiner Abschlussarbeit Brokoli untersuchen. Die Behandlung in seiner Abschlussarbeit sind verschiedene Lüftungssystemen und Folientunneln (*ctrl* und *tornado*). Erhoben wurden von Steffen als Messwert (Y) *Proteingehalt* benannt als *protein* in seiner Exceldatei. Erwartungsgemäß erhält er von seinem Betreuer den Auftrag die erhobenen Daten als Boxplots darzustellen. Dann kann Steffen auch schonmal abschätzen, was bei einem statistischen Test rauskommen könnte. Darüber hinaus kann Steffen anhand Boxplots eine Aussage über die Varianzhomogenität über die Behandlungsgruppen treffen. Na dann mal los. Steffen schafft sich die nötige Stimmung. Wenn Taylor Swift ertönt, dann sucht die Schlange schleunigst Schutz unter dem Sofa. Steffen schüttelt den Kopf.

treatment	drymatter
ctrl	42.3
tornado	23.4
ctrl	44.7
tornado	32.5
tornado	36.2
ctrl	41.8
ctrl	47.7
tornado	35.7
ctrl	45.3
ctrl	45.9
ctrl	49.4
tornado	32.8
tornado	40.6
ctrl	44.4
ctrl	44.5
ctrl	44.1
tornado	43.8

Leider kennt sich Steffen 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 Brokoli! 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 Brokoli 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** Aus den Boxen wummert Taylor Swift und sein Mund ist verklebt von Oreos. 'Herrlich', denkt Steffen. Steffen betrachtet die folgenden Daten nach einem Kreuzungsexperiment mit Hühnern. In dem Experiment wurden die Anzahl an gedrehten Haaren/Federn pro  $cm^2$  gezählt. Nach der Meinung seinem Betreuer muss als erstes geschaut werden, wie diese verteilt sind. Also welcher statistischen Verteilung die Anzahl an gedrehten Haaren/Federn pro  $cm^2$  folgen. Dazu soll Steffen 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 Romantik nicht wäre, ja dann wäre wohl vieles möglich für Steffen! Aber so.. Wenn Taylor Swift ertönt, dann sucht die Schlange schleunigst Schutz unter dem Sofa. Steffen schüttelt den Kopf.

Die Anzahl an gedrehten Haaren/Federn pro  $cm^2$ : 4, 6, 5, 3, 5, 7, 1, 1, 2, 5, 3, 2, 5, 2, 1, 12, 3, 3, 7, 7, 3, 3, 2, 7, 3, 5, 3, 0, 2, 4, 9, 6, 2

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

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

## 56. Aufgabe

(8 Punkte)

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



**Visualisierung des Histogramm für kontinuierliche Daten** In ihrer Abschlussarbeit möchte Jessica gerne die Daten aus einem Kreuzungsexperiment mit Milchvieh in einem Histogramm darstellen. Das Histogramm erlaubt ihr dabei Rückschlüsse auf die Verteilung über das Outcome ( $Y$ ) zu treffen. Jessica schmeißt noch eine Handvoll Schokobons in ihren Rachen. Im Hintergrund klirrt leise der Spiegel zum Sound von David Bowie. In seinem Experiment hat Jessica die mittlere Anzahl an gedrehten Haaren pro  $cm^2$  gezählt. Es wäre einfacher, wenn da nicht noch was wäre. Wenn der Mangel nicht wäre, ja dann wäre wohl vieles möglich für Jessica! Aber so.. Jessica streichelt liebevoll die Hündin. Der Kopf ist in ihrem Schloß vergraben um den Klang von David Bowie zu dämpfen.

Die mittlere Anzahl an gedrehten Haaren pro  $cm^2$ : 9.8, 11.1, 10.7, 8.9, 9.6, 11.3, 9.5, 9.9, 13.5, 9, 8, 10.2, 11.8, 10.9, 8.2, 10.7, 7.1, 10.8, 10.8, 12, 9.4, 12.1

Leider kennt sich Jessica 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** Yuki liest laut: 'Wenn zwei kontinuierliche Variablen vorliegen, können diese in einer explorativen Datenanalyse...'. Yuki stoppt. Aus den Boxen wummert London Grammar und ihr Mund ist verklebt von Reese's Peanut Butter Cups. 'Herrlich', denkt Yuki. Was waren noch gleich kontinuierliche Variablen? In ihrer Abschlussarbeit hatte sie ein Gewächshausexperiment im Teutoburgerwald durchgeführt. Dabei ging es um den Zusammenhang zwischen Frischgewicht [kg/ha] und durchschnittlichen Niederschlag [ml/w] im groben Kontext von Brokkoli. Nun stellt sich die Frage für sie, ob es überhaupt einen Zusammenhang zwischen den gemessenen Variablen gibt. Dafür war eine explorative Datenanalyse gut! Yuki und die Faulheit, eine unendliche Geschichte mit kniffligen Wendungen. Dann was anderes. Wenn Matrix läuft, dann ist das Minischwein nicht mehr da. Aber jetzt braucht sie mal Entspannung!

Frischgewicht [kg/ha]	Durchschnittlichen Niederschlag [ml/w]
23.4	33.4
25.3	34.0
21.8	31.6
19.9	25.9
33.1	43.6
22.6	30.3
17.5	26.9
27.9	40.5
14.8	23.9
18.9	25.7

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

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

## 58. Aufgabe

(10 Punkte)

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



**Visualisierung des Mosaicplots** In ein Feldexperiment im Wendland hatte Jessica sich zum einen die Behandlung Pestizideinsatz [ja/nein] und zum anderen die Messung Chlorophyllgehalt unter Zielwert [ja/nein] im Kontext von Kartoffeln angeschaut. Nun steht Jessica vor dem Problem, dass sie zwei kategoriale Variablen in ihrer Hausarbeit gemessen hat. Dazu kommt dann noch was anderes. Wenn der Mangel nicht wäre, ja dann wäre wohl vieles möglich für Jessica! Aber so.. Da ihr Betreuer erstmal die langen Tabellen mit ja/nein in einer explorativen Datenanalyse zusammengefasst und präsentiert bekommen möchte bevor es überhaupt weitergeht, muss sie jetzt eine Lösung finden. Was alles auch nicht einfacher macht. Am liebsten würde sie ja was anderes machen. Jessica liebt Warhammer. Darin kann sie sich wirklich verlieren und immer wieder neu begeistern.

Chlorophyllgehalt unter Zielwert	Pestizideinsatz	Chlorophyllgehalt unter Zielwert	Pestizideinsatz
ja	ja	nein	nein
nein	nein	nein	nein
ja	ja	nein	nein
nein	nein	ja	ja
nein	ja	ja	ja
ja	nein	ja	ja
nein	nein	ja	nein
ja	nein	nein	ja
nein	nein	ja	ja
nein	ja	ja	ja
nein	nein	nein	nein
ja	ja	nein	nein
ja	nein	nein	nein
ja	ja	nein	nein
ja	ja	nein	ja

Leider kennt sich Jessica 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 *kein* Effekt von der Behandlung vorliegen würde, wie würde die Tabelle und die Visualisierung aussehen? **(2 Punkt)**

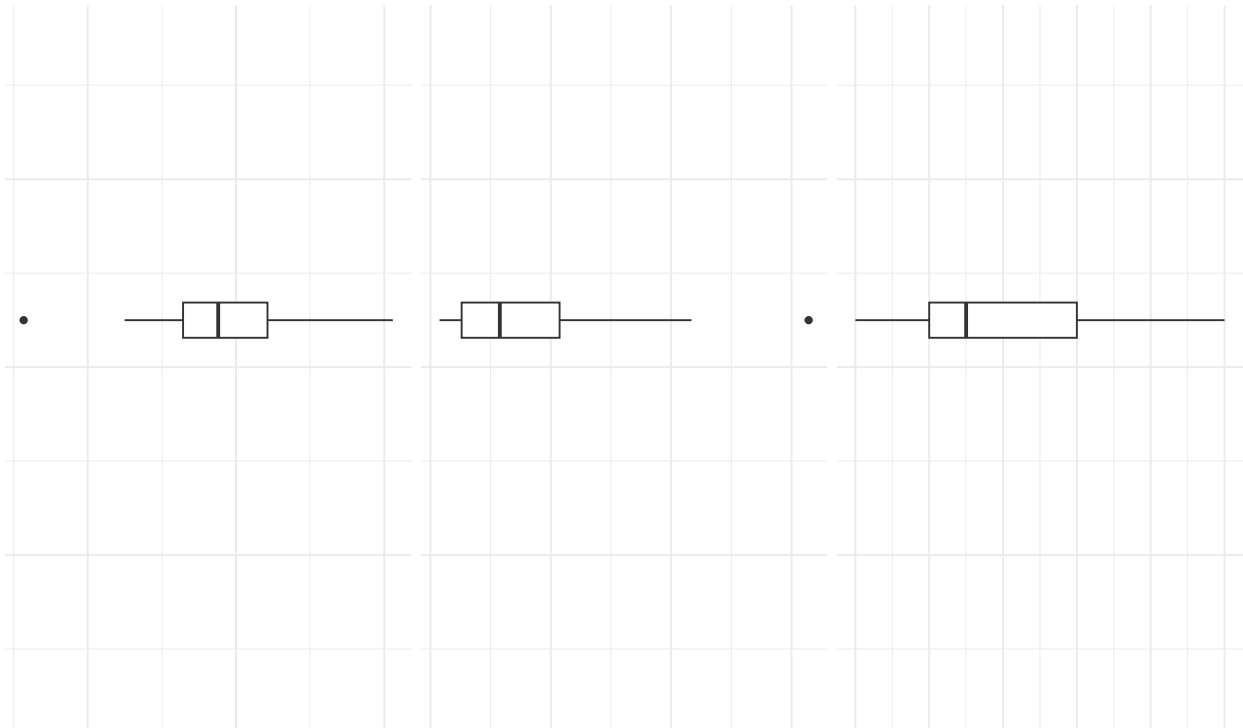
## 59. Aufgabe

(10 Punkte)

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



**Visualisierung von Verteilungen** 'Was hast du dir denn da hingeklebt? *Frei ist, wer missfallen kann.*<sup>1</sup>', liest Yuki vom Kühlschrank vor. Yuki und Mark sitzen zusammen in der Küche und versuchen zu verhindern, dass das Minischwein den Biomüll mampft. 'Können wir uns auf die etwas kryptische Aufgabe konzentrieren?', nöhlt Mark. Die beiden schauen angestrengt auf die drei Boxplots. Das Ziel ist es zu verstehen, wie eine Verteilung anhand eines Boxplots bewertet werden kann. Mark und die Unsicherheit machen die Sache nicht einfacher.



Jetzt brauchen Yuki und Mark 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** 'Hmpf, wir sollen Normalverteilungen visualisieren. Waren das nicht so komische Glockenkurven?', fragt sich Tina. Jonas hat Snickers mitgebracht und Iron Maiden aufgedreht. Immerhin etwas. 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. 'Wozu brauchen wir nochmal Normalverteilungen?', entfährt es Jonas. Durch das Mampfen von Katjes durch Tina versteht er kein Wort der Antwort.

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

1. Skizzieren Sie vier Normalverteilungen mit  $\bar{y}_1 \neq \bar{y}_2 \neq \bar{y}_3 \neq \bar{y}_4$  und  $s_1 = s_2 = s_3 = s_4$ ! **(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!

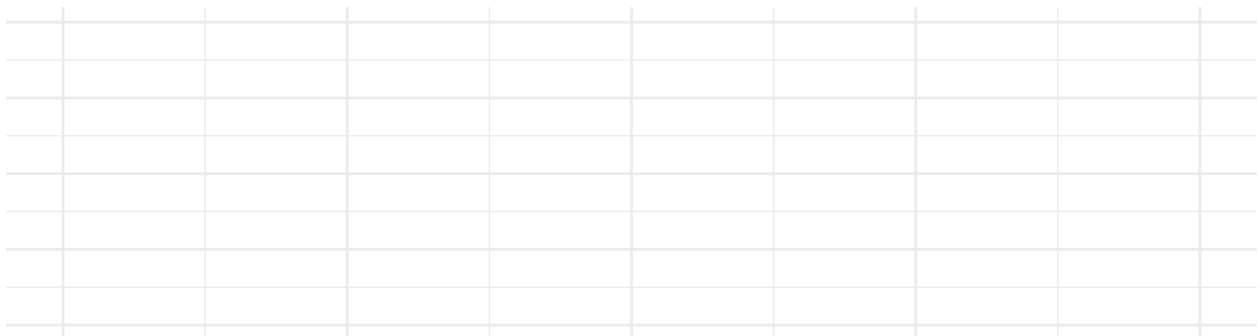


**Visualisierung der Normalverteilung und der Poissonverteilung** 'Wer hat sich denn sowas ausgedacht? Verteilungen?', entfährt es Jonas und schaut dabei Jessica an. In seiner Hand zerdrückt er Schokobons von Jessica. 'Wir sollen eine Normalverteilung mit einem Mittelwert von  $\bar{y}_1 = 4$  und einer Standardabweichung von  $s_1 = 9$  zeichnen. Sowie eine weitere Normalverteilung mit einem Mittelwert von  $\bar{y}_2 = 2$  und einer Standardabweichung von  $s_2 = 9$ . Keine Ahnung wie das geht. Darunter sollen dann noch eine Poissonverteilung mit einem Mittelwert von  $\lambda_1 = 3$  sowie einer weiteren Poissonverteilung mit einem Mittelwert von  $\lambda_2 = 15$  gezeichnet werden.', meint Jessica sichtlich eingeschüchtert und rettet noch ein paar Schokobons in seinem Mund. Jonas und die Erschöpfung machen die Suche nach der Lösung nicht einfacher. Im Hintergrund spielt viel zu laut David Bowie, die diesmal Jonas ausgewählt hat und das Meerschweinchen schon in die Küche vertrieben hat, wo es den Biomüll mampft.

### Zwei Normalverteilungen



### Zwei Poissonverteilungen



Jetzt brauchen Jonas 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** An einem schwülen Sommernachmittag sitzen Nilufar und Jonas in einem Eiskaffee und wollen sich auf die Klausur vorbereiten. In fast allen Fragen geht es ja um die Interpretation eines statistischen Tests. Daher wollen die beiden jetzt nochmal nacharbeiten, was die Grundlagen der Stichprobe (eng. *sample*) und der Grundgesamtheit (eng. *population* oder *ground truth*) sind. Nilufar hat sich Takis Blue Heat Eisbecher bestellt und Jonas bleibt lieber bei einem Snickers Eis. 'Irre, was die Lebensmittelindustrie alles auf die Beine kriegt', merkt Jonas an und Nilufar schüttelt anerkennend den Kopf.

Leider kennen sich Nilufar und Jonas 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** 'Das Känguruh hat mir gerade zugezwinkert. Das macht mir Angst', bemerkt Tina. Es reicht ja schon ein Problem. Wenn die Wut nicht wäre, ja dann wäre wohl vieles möglich für Tina! Aber so.. Da braucht es nicht noch Wahnvorstellungen. 'Ich glaube nicht, das Känguruhs zwinkern. Aber was Wichtigeres. Wo hast du eigentlich meine Schnapspralinen hingetan? Wir haben nur noch Katjes von dir. Bäh!', antwortet Alex. Beide sind im Zoo und wollen sich von der statistische Testtheorie ablenken lassen. Eigentlich wollte ja Tina stoppen wie lange Tiere pinkeln<sup>2</sup>, scheiterte aber an einer Oma mit Stock, die die beiden beschimpfte.

Leider kennen sich Tina und Alex 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)**

$H_0$  abgelehnt    Richtige Entscheidung    20%     $\beta$ -Fehler

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!

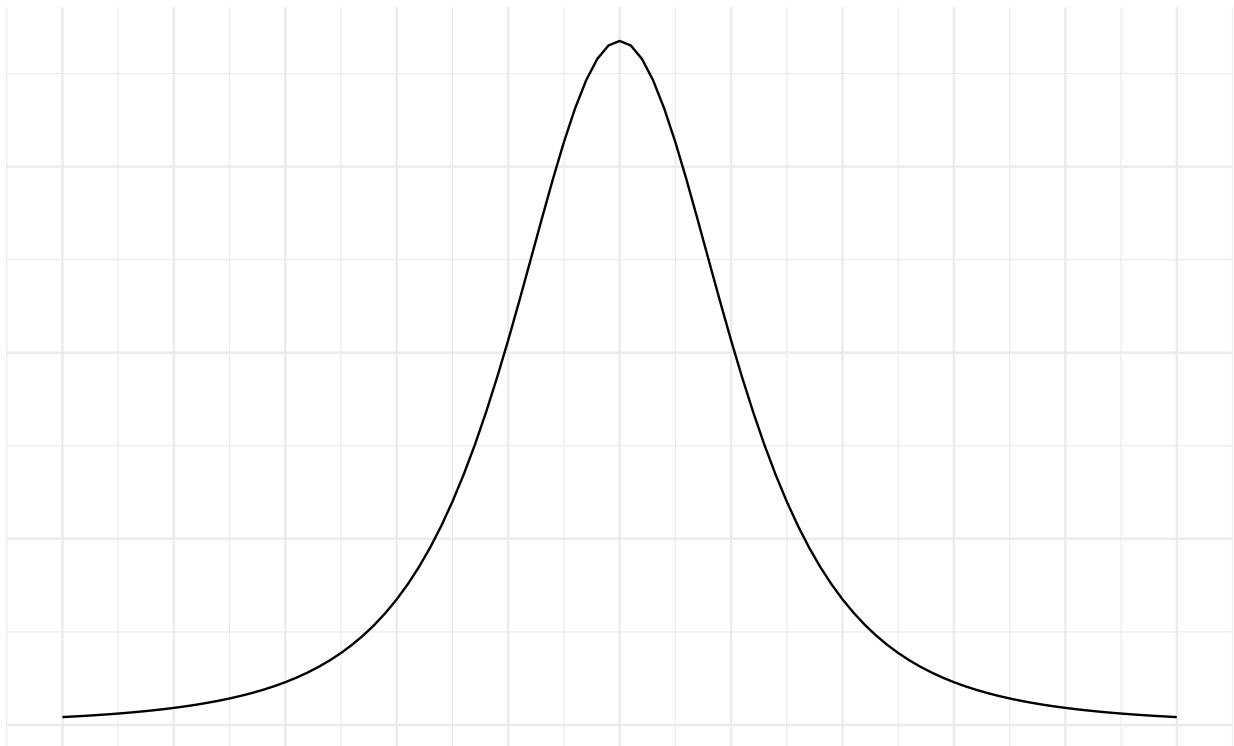


**Visualisierung der Teststatistik  $T_D$  und dem p-Wert** 'Wir sollen die Teststatistik  $T_D$  und dem p-Wert visualisieren, da mit einer Visualisierung vieles verständlicher wird!', ruft Paula um White Lies zu übertönen. 'Ich weiß nicht, was das jetzt helfen soll. Können wir nicht einfach schauen, ob der p-Wert kleiner als das Signifikanzniveau  $\alpha$  gleich 5% ist? Und gut ist?', merkt Tina an, was aber im Refrain von White Lies untergeht. Paula nickt im Beat. 'Wir haben hier eine t-Verteilung unter der Annahme der Nullhypothese!', singt sie.

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

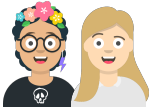
Beachten Sie, dass im Folgenden keine numerisch korrekte Darstellung verlangt wird! Es gilt Erkennbarkeit vor Genauigkeit!

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 signifikanten 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 Tina. Jessica schaut fragend. 'Hatten wir nicht alles zu einer Kontrolle verglichen? Das war doch so!', ruft Jessica laut aus. 'Wir haben doch Messwert *Pilzbefall nach Fungizid* erhoben.', stellt Tina fest. Jetzt haben beide das Problem, die möglichen 95% Konfidenzintervalle zu interpretieren.

Leider kennen sich Tina und Jessica 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, nicht relevantes 95% Konfidenzintervall
  - (e) Ein signifikantes, relevantes 90% Konfidenzintervall.
  - (f) Ein 95% Konfidenzintervall mit niedriger Varianz  $s_p$  in der Stichprobe als der Rest 95% der Konfidenzintervalle



## 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 kalten Dezembertag haben sich Yuki und Tina zum Lernen verabredet. Eine große Kanne Tee und eine große Packung Reese's Peanut Butter Cups warten darauf gegessen zu werden. Yuki 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  $KI_{1-\alpha}$ ?*

Tina hebt die Augenbraue. 'Irgendwie sagt mir die Aufgabe jetzt mal so gar nichts. Was soll da gemacht werden?', merkt Tina an und ergänzt: 'Schauen wir doch erstmal zur Entspannung Indiana Jones, den Film habe ich extra nochmal mitgebracht.' Yuki ist der Idee nicht abgeneigt und auch das Minischwein kommt unter dem Sofa hervor um mitzuschauen.

Leider kennen sich Yuki und Tina 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 90%-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** Yuki ist im Wendland für einen Pilotexperiment mit sehr geringer Fallzahl ( $n_1 = n_2 = 3$ ) mit Lamas. Allein diese Tatsache ist für ihn eine Erzählung wert. Eine echte Herausforderung für ihn war schon immer die Faulheit gewesen. Ein leidiges Lied. Für seiner Hausarbeit musste er ein Kreuzungsexperiment mit Lamas durchführen und das sollte laut seinem 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 seine Behandlung Genotypen (AA und BB) und der Messwert Protein/Fettrate [%/kg] sollen mit einem t-Test ausgewertet werden. Immerhin weiß er, dass sein Messwert einer Normalverteilung folgt. Hm..., was spannendes wäre gut. Yuki will später nochmal raus um zu Boldern. Druck ablassen, dass muss er auch.

treatment	weight
ctrl	17.4
ctrl	17.2
dose	12.5
dose	21.8
dose	11.2
ctrl	17.4

Leider kennt sich Yuki 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. Bestimmen Sie die Teststatistik  $T_D$  eines Student t-Tests! **(3 Punkte)**
3. Treffen Sie mit  $T_{\alpha=5\%} = 1.84$  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 Yuki 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 Tina laut. Das hilft jetzt auch nur bedingt weiter. Wenn die Wut nicht wäre, ja dann wäre wohl vieles möglich für Tina! Aber so.. Laut ihrer Betreuerin ist zwar ihr Messwert Gewichtszuwachs in der 1LW normalverteilt, aber wie rechnet sie jetzt einen t-Test? Für ihrer Hausarbeit musste sie ein Stallexperiment mit Schweinen im Teutoburgerwald durchführen. Als wäre das nicht schon anstrengend genug gewesen. Jetzt soll sie auch noch testen, ob die Behandlung Genotypen (AA und BB) ein signifikantes Ergebnis liefert. Hm..., was entspannendes wäre gut. Hm, lecker Katjes und dazu dann im Hintergrund Indiana Jones laufen lassen.

Genotypen	Gewichtszuwachs
BB	40.5
AA	24.0
AA	37.0
AA	46.5
AA	40.3
AA	48.6
BB	34.7
AA	50.8
BB	33.3
BB	33.6
BB	37.5
AA	33.9
BB	35.2
AA	31.8
BB	33.2
BB	31.7
AA	43.1
BB	38.2
BB	31.8

Leider kennt sich Tina 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  $T_{\alpha=5\%} = 1.84$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effekt des Student t-Tests! **(1 Punkt)**
6. Wenn Sie *einen* Unterschied zwischen den Behandlungsgruppen erwarten würden, wie groß wäre dann die Teststatistik  $T_D$ ? Begründen Sie Ihre Antwort! **(2 Punkte)**
7. Formulieren Sie eine Antwort an Tina 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** '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 ihrem Betreuer ist zwar ihr Messwert Frischegewicht [kg/ha] normalverteilt, aber wie rechnet sie jetzt einen t-Test? Für ihrer Hausarbeit musste sie ein Gewächshausexperiment mit Erbsen in der Uckermark durchführen. Als wäre das nicht schon anstrengend genug gewesen. Jetzt soll sie auch noch testen, ob die Behandlung Lichtstufen (*none* und *600lm*) 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.

Lichtstufen	Frischegewicht
600lm	44.1
none	33.5
600lm	33.2
600lm	49.3
600lm	39.3
600lm	47.4
none	32.3
600lm	28.2
none	28.0
none	33.4
none	24.2
none	31.3
600lm	47.4
600lm	40.7
600lm	39.3
none	24.2
none	33.4
600lm	33.4
none	30.8
none	33.6

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 Welch t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 1.84$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie das 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 Paula 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** Es gibt ja immer die Möglichkeit sich Hilfe zu holen. Das geht natürlich auch immer in einem Projektbericht. Deshalb arbeiten Jonas und Paula gemeinsam an einem Projektbericht. 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. In dem Hauptversuch wurde Folgendes von den beiden gemacht. Jonas und Paula haben sich Schweinen angeschaut. Dabei geht um Zusammenhang zwischen Flüssignahrung (1l/d und 5l/d) und Fettgehalt [%/kg]. Jetzt sollen beide einen gepaarten t-Test rechnen. Es würde auch besser funktionieren, wenn Jonas nicht die Erschöpfung im Weg stehen würde und Paula nicht das Problem hätte der Perfektionismus zu händeln. Gott sei Dank haben beide genug Snickers und Smarties auf dem Tisch aufgetürmt.

ID	treatment	freshmatter
2	5l/d	20.9
2	1l/d	30.1
4	1l/d	22.5
3	1l/d	35.7
8	5l/d	28.5
9	1l/d	22.7
8	1l/d	10.2
6	5l/d	20.6
1	1l/d	18.6
7	5l/d	25.6
5	5l/d	31.2
6	1l/d	13.8
5	1l/d	29.5
10	1l/d	2.6
3	5l/d	18.4
1	5l/d	22.8
7	1l/d	12.3
4	5l/d	25.6

Leider kennen sich Jonas 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  $T_{\alpha=5\%} = 2.68$  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 Jonas 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** 'Mit dem R Paket {emmeans} können wir gleich die Gruppenvergleiche rechnen und uns das *compact letter displac*' ausgeben lassen!', verkündet Yuki sichtlich stolz. Ein paar Mal hat sie schon die Faulheit gehindert weiterzumachen. 'Nach Meinung der Betreuerin soll es aber nur erstmal ein t-Test sein. Und die Ausgabe ist schon wirr genug.', merkt Alex an. Alex und Jonas sind bei Yuki um sich in R helfen zu lassen. Im Hintergrund wummert London Grammar. Jonas streichelt zur Beruhigung das Minischwein von Yuki. Die beiden waren 2 Monate im Emsland um einen Versuch mit Milchvieh in einem Leistungssteigerungsversuch durchzuführen. Ziel war es das Outcome (Y) Fettgehalt [%/kg] zu bestimmen. Yuki überlegt, ob sie die beiden nicht noch auf den Film *Matrix* einlädt oder dann doch lieber raus geht um zu Boldern? Vielleicht will ja Jonas mit. Besser als der Film.

```
##
## Two Sample t-test
##
## data: Fettgehalt by Elterlinie
## t = 3.1869, df = 14, p-value = 0.006589
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
##  4.148909 21.226091
## sample estimates:
## mean in group Standard      mean in group Xray
##                46.3000                33.6125
```

Helfen Sie Yuki bei der Interpretation des t-Tests! Sonst geht es auch für Alex und Jonas 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  $T_{\alpha=5\%} = |2.14|$  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** 'Wir waren in der Uckermark um Lauch in einem Feldexperiment zu messen.', Jonas legt das Dokument auf den Tisch und schaut Paula 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 Jonas gekommen, da sie sich nicht mit R auskennen und daher Hilfe bei der Interpretation des t-Tests brauchen. Im Hintergrund wummert Iron Maiden und leere Snickers Packungen stapeln sich auf dem Boden. 'Kein Problem', sagt Jonas und streichelt langsam das Meerschweinchen. 'Aber worum es in dem Versuch geht, lässt sich nur aus dem Text in seiner Hand errahnen.' merkt er an. Vielleicht hilft da ja die Ausgabe des t-Tests in R weiter. Draußen geht blutrot die Sonne unter.

```
##  
## Two Sample t-test  
##  
## data: Trockengewicht by Genotypen  
## t = -0.46089, df = 17, p-value = 0.6507  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -6.401709 4.106254  
## sample estimates:  
## mean in group AA mean in group BB  
## 32.52500 33.67273
```

Helfen Sie Jonas bei der Interpretation des t-Tests! Sonst geht es auch für Paula 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 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** Jessica und Tina sind bei Paula um sich Hilfe in R zu holen. Im Hintergrund wummert White Lies. Die beiden hatten zwar schon erste Kontakte mit R sind sich aber unsicher bei der Interpretierung der Ausgabe eines t-Tests für ihren gemeinsamen Versuch. Es würde auch besser funktionieren, wenn Paula nicht der Perfektionismus im Weg stehen würde und Tina nicht das Problem hätte der Mangel zu händeln. In einer Abschlussarbeit haben beide zusammen Erdbeeren untersucht. Dabei ging es um den Zusammenhang zwischen der Behandlung Bewässerungstypen (*low* und *high*) und dem Messwert Proteingehalt [g/kg]. Der Versuch wurde in einem Freilandversuch im Emsland durchgeführt. Nach des Betreuers ist der Messwert Proteingehalt [g/kg] normalverteilt und ein t-Test passt daher. Das wird jetzt nicht mehr angezweifelt...Paula überlegt, ob sie die beiden nicht noch auf den Film *Jagd auf roter Oktober* einlädt.

```
##
## Two Sample t-test
##
## data: Proteingehalt by Bewässerungstypen
## t = -7.3344, df = 12, p-value = 9.044e-06
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -12.062734 -6.537266
## sample estimates:
## mean in group low mean in group high
## 33.22857 42.52857
```

Helfen Sie Paula bei der Interpretation des t-Tests! Sonst geht es auch für Jessica 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** Tina und Mark 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 gut Idee soweit herausgestellt. Tina und die Wut, eine unendliche Geschichte mit kniffligen Wendungen. In einer Hausarbeit sollen beide herausfinden, ob es einen Zusammenhang zwischen Düngung (*vorher* und *nachher*) und Trockengewicht [kg/ha] gibt. Die Besonderheit ist hierbei, dass die Messungen an der gleichen Beobachtung stattfinden. Beide messen also zweimal an den gleichen Erbsen. Hier muss dann wohl auf einen normalverteilten Endpunkt ( $Y$ ) ein gepaarter t-Test gerechnet werden. Leider kennen sich beide nicht sehr gut in R aus. Aber wenigstens haben beide eine Menge an Katjes.

```
##
## Paired t-test
##
## data: Trockengewicht by Düngung
## t = 0.6084, df = 8, p-value = 0.5598
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -4.309445  7.398334
## sample estimates:
## mean difference
## 1.544444
```

Jetzt brauchen Tina und Mark 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** 'Uff... die einfaktorielle ANOVA. Und wir können jetzt anhand der Visualisierung sehen, ob da schon was signifikant ist?', Yuki hebt die Augenbraue. 'Ja, können wir. Dafür müssen wir aber erstmal in {ggplot} uns die Daten anschauen. Oder wir zeichnen es flott mit der Hand. Geht auch.', meint Jonas dazu. Die beiden hatten sich auf einem Konzert von Iron Maiden kennengelernt. Yuki hatte sich in ein Stallexperiment verschiedene Hühnern angeschaut. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Elterlinie (*Standard*, *Yray* und *Xray*) und dem Messwert Protein/Fettrate [%/kg] gibt. Später wird noch Mission Impossible geguckt. Jonas befürwortet das!

Elterlinie	Protein/Fettrate
Yray	26
Yray	25
Xray	30
Standard	46
Standard	45
Standard	44
Xray	30
Yray	25
Xray	30
Yray	25
Standard	47
Xray	29
Yray	26
Xray	30
Standard	48
Yray	26
Standard	44
Yray	25

Leider kennen sich Yuki und Jonas 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** 'Als erstes bauen wir uns aus unsere Daten die ANOVA Tabelle dann sehen wir schon, ob unser Gruppenvergleich in der ANOVA signifikant ist.', Yuki schaut Jonas fragend an und hofft auf eine positive Regung im Gesicht. Wird aber enttäuscht. Da hilft das Meerschweinchen von Jonas auch nur bedingt. Jonas tut sich auch sehr schwer mit der einfaktoriellen ANOVA. Beide waren im Wendland um einen Leistungssteigerungsversuch mit Milchvieh durchzuführen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Flüssignahrung (*ctrl*, *superIn* und *fLOW*) und dem Messwert Fettgehalt [%/kg] gibt. Nachher wollen sich beide noch mit dem Hobby Stricken von Jonas beschäftigen. Kennt Yuki noch nicht, klingt aber interessant.

Leider kennen sich Yuki 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)
Flüssignahrung	2				
error	15	478.13			
Total	17	762.44			

4. Schätzen Sie den p-Wert der Tabelle mit  $F_{\alpha=5\%} = 3.68$  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** 'Als erstes bauen wir uns aus unsere Daten die ANOVA Tabelle dann sehen wir schon, ob unser Gruppenvergleich in der ANOVA signifikant ist.', Mark schaut Paula fragend an und hofft auf eine positive Regung im Gesicht. Wird aber enttäuscht. Paula schmeißt sich noch ein paar Smarties in den Rachen. Beide tun sich sehr schwer mit der einfaktoriellen ANOVA. Nun möchte erstmal ihre Betreuung der Arbeit eine ANOVA Tabelle sehen. Was immer da auch drin zu erkennen sein mag. Beide waren im Wendland um ein Feldexperiment mit Kartoffeln durchzuführen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Lüftungssystemen und Folientunneln (*ctrl*, *storm*, *thunder* und *tornado*) und dem Messwert Trockengewicht [kg/ha] gibt. Später wollen die beiden dann noch raus um zu Fechten.

Leider kennen sich Mark und Paula 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)
<b>Lüftungssystemen</b>	3	4778.34			
<b>Error</b>	24	539.76			

4. Schätzen Sie den p-Wert der Tabelle mit  $F_{\alpha=5\%} = 3.01$  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)**

Lüftungssystemen	Fallzahl (n)	Mittelwert	Standardabweichung
ctrl	9	6.22	2.33
storm	6	2.33	5.16
thunder	8	3.12	3.68
tornado	5	38.00	8.19

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 sie kennt und sich mit R auskennt. Wird aber enttäuscht. Yuki war im Wendland um einen Versuch in einer Klimakammer mit Erbsen durchzuführen. Nun möchte ihre Betreuerin ihrem Projektbericht erstmal eine ANOVA sehen und die Ergebnisse präsentiert bekommen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Bewässerung (*ctrl*, *low*, *mid* und *high*) und dem Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] gibt. Und eigentlich wollte Yuki doch noch zum Sport! Yuki will später nochmal raus um zu Boldern. Druck ablassen, dass muss sie auch.

```
## Analysis of Variance Table
##
## Response: Chlorophyllgehalt
##           Df Sum Sq Mean Sq F value Pr(>F)
## Bewässerung  3  204.21   68.069   3.2813 0.04102
## Residuals    21  435.63   20.744
```

Leider kennen sich Yuki mit Berechnung einer einfaktoriellen 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. 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** Tina steht im Wendland. Und das ist noch langweiliger als es sich anhört. Wäre es nur so spannend wie bei ihren Kommilitonen, die in Almería waren. Ödnis wohin man nur blickt. Oder eben Milchvieh. Die Spinne durchbohrt sie mit leeren Blick. 'Woher zum Teufel!', entfährt es ihr. Aber da ist es schon weg. Ja, darum geht es in ihrer Hausarbeit. Und wäre das nicht noch alles genug, ist ihr Experiment auch noch als einen Leistungssteigerungsversuch komplex geraten. Es wurde der Messwert Schlachtgewicht [kg] mit dem Behandlung Ernährungszusatz (*ctrl*, *fedX* und *getIt*) sowie der Behandlung Genotypen (*AA* und *BB*) untersucht. 'Hmpf', denkt Tina und ruft 'Und jetzt!?' in die Leere. Und eigentlich wollte Tina doch noch ihrem Hobby nachgehen! Am Ende dann doch besser Astronomie. Wunderbar. Eine echte Ablenkung für Tina.

Leider kennen sich Tina 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>Ernährungszusatz</b>	3	54.97			
<b>Genotypen</b>	1	4.7			
<b>Ernährungszusatz:Genotypen</b>	3	438.94			
<b>Error</b>	18	360.84			

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

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

5. Was bedeutet ein signifikantes Ergebnis in einer zweifaktoriellen ANOVA? (2 Punkte)
6. Was sagt der Term *Ernährungszusatz:Genotypen* 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** Es ist schon kurz nach fünf und Alex wird langsam nervös. Alex wollte heute Abend noch seine E-Sport Qualifikation schauen. Stattdessen versucht sein Betreuer die Ausgabe der zweifaktoriellen ANOVA zu visualisieren und zu überprüfen, ob es mit der Visualisierung der Daten als Boxplots zusammenpasst. Alex hatte im Wendland einen Leistungssteigerungsversuch mit Zandern durchgeführt. Es gab dabei zwei Behandlungen. Einmal Ernährungszusatz (*ctrl*, *fedX* und *getIt*) sowie als zweite Behandlung Lüftungssystem (*keins* und *thunder*). Gemessen wurde der Messwert (Y) Fettgehalt [%/kg]. So kompliziert kann das jetzt doch nicht sein! Eigentlich wollte Alex nachher noch zum Sport. Um zu Laufen geht Alex dann später nochmal raus. Echte Entspannung.

```
## Analysis of Variance Table
##
## Response: Fettgehalt
##
##               Df Sum Sq Mean Sq F value Pr(>F)
## Ernährungszusatz      2  55.32   27.661    1.2064 0.32235
## Lüftungssystem         1 105.88  105.881    4.6180 0.04549
## Ernährungszusatz:Lüftungssystem  2 291.27 145.633    6.3517 0.00818
## Residuals             18 412.71   22.928
```

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** Das Meerschweinchen dreht durch und verwüstet Jonass Palme zu kleinen Schnetzeln. Aber dafür hat er jetzt keine Zeit. Jonas muss verstehen wie die Formeln der ANOVA und des t-Tests miteinander zusammen hängen und was das verbindende Konzept ist. Jonas dreht Iron Maiden lauter, damit das Meerschweinchen sie nicht mehr stört. Die Palme leidet still. Was hat Jonas eigentlich gemacht? In ein Stallexperiment wurden Puten mit der Behandlung Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) sowie der Behandlung Bestandsdichte (*standard* und *kontakt*) untersucht. Das hilft der Palme auch nicht mehr. Aber das ist nicht das einzige Problem von Jonas. Eine echte Herausforderung für ihn war schon immer die Erschöpfung gewesen. Ein leidiges Lied.

### 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 Jonas mit dem Zusammenhang zwischen der ANOVA und dem t-Test nicht aus. Deshalb braucht er 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 Stallexperiment wurden Schweinen mit der Behandlung Genotypen (*AA*, *AB* und *BB*) sowie der Behandlung Ernährungszusatz (*ctrl* und *getIt*) untersucht. Es wurde als Messwert Fettgehalt [%/kg] bestimmt. Jetzt starrt Yuki mit auf die R-Ausgabe einer zweifaktoriellen ANOVA. Leider starrt ihre Betreuerin in der gleichen Art Yuki 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. Yuki hätte doch nichts mit Schweinen machen sollen. Schweinen – was soll das auch bedeutendes sein? Eigentlich wollte Yuki nachher noch einen Film schauen. Irgendwie komisch, wenn sie Matrix anmacht, dann ist das Minischwein eigentlich sofort vor dem Bildschirm und starrt hinein.

Leider kennen sich Yuki und ihre Betreuerin 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 Orchideen 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!', ihre Betreuerin scheint die einfaktoriellen ANOVA zu verstehen. Warum jetzt sie jetzt nochmal alles wiederkauen muss, wird Tina echt nicht so klar. Wenn es doch so klar ist? 'Wir haben jetzt bei der ANOVA einen p-Wert mit 0.058 raus sowie eine F-Statistik  $F_D$  mit 1.51 berechnet. Nach den Boxplots müsste sich eigentlich ein Unterschied zwischen *fedX* und *ctrl* ergeben. Der Unterschied ist in {emmeans} auch signifikant mit einem p-Wert von 0.021. Wie kann das sein?', fragt Tina etwas provokant und dreht Tocotronic leiser. Tina war im Wendland und hatte dort einen Leistungssteigerungsversuch mit Milchvieh durchgeführt. Die Komune wo sie untergekommen war, war cool gewesen. Dort gab es selbstgemachte Katjes aus Vollkorn! Nur jetzt muss eben das Experiment fertig ausgewertet werden. Tina hatte eine Behandlungen Ernährungszusatz (*ctrl*, *fedX* und *getIt*) auf Milchvieh angewendet. Gemessen wurde der Messwert ( $Y$ ) Fettgehalt [%/kg]. Dabei wurden die Daten  $D$  erhoben. Jetzt muss das hier zu einem Ende kommen! Tina hat schon genug Probleme. Wenn die Wut 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 Tina und ihre Betreuerin mit der Interpretation einer ANOVA überhaupt nicht aus. Deshalb braucht sie 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** 'Moment, die haben ja das Gleiche gemacht wie wir!', ruft Alex laut aus. Yuki schaut etwas verwundert. 'Das glaube ich eher nicht. Lass uns mal unsere Daten mit den Ergebnissen von Totsdorf et al. (2018) vergleichen.', antwortet Yuki. In ein Freilandversuch mit Kartoffeln wurde die Behandlung Genotypen (00, AA, AB und BB) auf den Messwert Trockengewicht [kg/ha] untersucht. Jetzt müssen die beiden mal schauen, ob sie wirklich was Neues gefunden haben oder ob die Ergebnisse alle die gleichen sind wie schon bei Totsdorf et al. (2018). Es ergab sich dann die folgende Tabelle der rohen  $p$ -Werte für die Vergleiche zu Totsdorf et al. (2018).

Rohen $p$ -Werte	Adjustierte $p$ -Werte	Nullhypothese ablehnen?
0.06		
0.03		
0.08		
0.76		

Leider kennen sich Alex 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)** Alex sitzt schon etwas länger bei seinem Betreuer. So langsam macht Alex sich Gedanken, ob er nicht doch mal anmerken sollte, dass er von CLD noch nie was gehört hat. Aber noch kann gelauscht werden, ein Ende ist erstmal nicht in Sicht! Alex hatte in seinem Projektbericht ein Gewächshausexperiment durchgeführt. Deshalb sitzt er hier. Also eigentlich nein, deshalb nicht. Alex will fertig werden. Hat er sich doch mit Lüftungssysteme (*ctrl*, *storm*, *thunder* und *tornado*) und Chlorophyllgehalt (SPAD-502Plus) [SPAD] schon eine Menge angeschaut. Alex beugt sich leicht nach vorne. Nein, doch keine Pause. Weiter warten auf eine Lücke im Fluss...

Behandlung	Compact letter display
ctrl	a
storm	a
thunder	a
tornado	a

Leider kennen sich Alex 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 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** Jessica sitzt schon etwas länger bei ihre 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 die Projektbericht einen Leistungssteigerungsversuch durchgeführt. Deshalb sitzt sie hier. Also eigentlich nein, deshalb nicht. Jessica will fertig werden. Hat sie sich doch mit Genotypen (00, AA, AB und BB) und Gewichtszuwachs in der 1LW schon eine Menge angeschaut. Jessica beugt sich leicht nach vorne. Nein, doch keine Pause. Weiter warten auf eine Lücke im Fluss... 'Wir müssen als erstes die Gruppen nach absteigender Effektstärke sortieren!', hört Jessica noch aus der Ferne bevor sie einnickt.

Genotypen	Fallzahl (n)	Mittelwert	Standardabweichung
00	8	7.28	2.75
AA	8	16.04	1.68
AB	7	10.69	3.95
BB	8	5.25	3.40

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 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 Jessica 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** 'Das Problem ist, dass DataTab eben keine CLD kann. Die bräuchten wir dann schon dringend für unser Poster!', merkt ihre Betreuerin mit Nachdruck an. Jessica neigt den Kopf. 'Das wussten wir nicht vorher?', entfährt es ihr leicht entnervt. Da schaut ihre Betreuerin seltsam betroffen. Hilft jetzt auch so gar nicht. Jessica hatte sich zwei Variablen mit Genotypen (00, AA, AB und BB) und Trockengewicht [kg/ha] in ein Feldexperiment mit Brokkoli angeschaut. Jetzt möchte sie eigentlich fertig werden und nicht nochmal alles neu in R und {emmeans} machen. Dabei hatte sie schon echt ne Menge in im Wendland gemacht. Dann eben per Hand aus der Matrix der p-Wert. Jessica muss sich echt zusammenreißen.

	00	AA	AB	BB
00	1.0000000	0.5898325	0.7885156	0.0122095
AA	0.5898325	1.0000000	0.7751157	0.0310174
AB	0.7885156	0.7751157	1.0000000	0.0168766
BB	0.0122095	0.0310174	0.0168766	1.0000000

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 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 Jessica und Yuki! **(2 Punkte)**

## Teil VI.

# Der Chi-Quadrat-Test & Der diagnostische Test

### 88. Aufgabe

(12 Punkte)

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



**Den Chi-Quadrat-Test berechnen** Alex 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 Abba auf das Ohr und los gehts. Alex ist schon eine ganze Zeit im Büro, da seine Betreuerin möchte, dass er jetzt auf seinen Daten mit  $n = 124$  Beobachtungen von Zandern einen  $\chi^2$ -Test rechnet. Das ginge, da er als Behandlung *Außenklimakontakt [ja/nein]* bestimmt und zum anderen die Variable *Gewichtszuwachs erreicht [ja/nein]* ermittelt hat. Wie genau, das ist jetzt eine andere Frage. Eigentlich wollte Alex nachher noch einen Film schauen. Wenn Alien läuft, dann ist die Katze nicht mehr da. Aber jetzt braucht er mal Entspannung!

	38	19	
	23	44	

Leider kennt sich Alex 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. 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  $\chi^2_{\alpha=5\%} = 3.841$ ! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Skizzieren Sie die  $\chi^2$ -Verteilung, wenn die  $H_0$  wahr ist! Ergänzen Sie  $\chi^2_{\alpha=5\%}$  und  $\chi^2_D$  in der Abbildung! **(2 Punkte)**
7. Berechnen Sie den Effektschätzer *Cramers V*! Interpretieren Sie den Effektschätzer! **(2 Punkte)**

## 89. Aufgabe

(10 Punkte)

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



**Der Chi-Quadrat-Test konzeptionell verstehen** Am Ende hätte Jessica dann doch einen normalverteilten Endpunkt in ihrer Hausarbeit nehmen sollen. Dann noch schnell Schokobons zur Stärkung und los gehts. Vor ihr liegen jetzt die Daten von zwei Variablen als Kategorien oder wie es in  so schön heißt, als Faktoren. Aber immerhin, hofft sie das was bei den Daten rausgekommen ist. Gezählt hat Jessica einiges mit  $n = 147$  Beobachtungen von Lamas. Zum einen hat sie als Behandlung *Automatische Fütterung* [ja/nein] bestimmt und zum anderen die Variable *Protein/Fettrate im Zielbereich* [ja/nein] ermittelt. Nun möchte ihr Betreuer gerne einen  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Am Ende des Tages möchte sie dann noch ihr Hobby Warhammer genießen. Das muss auch mal sein!

			74
			73
	67	80	147

Leider kennt sich Jessica 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. 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)**

## 90. Aufgabe

(10 Punkte)

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



**Der Chi-Quadrat-Test in R** Am Ende hätte Nilufar dann doch einen normalverteilten Endpunkt in ihrem Projektbericht nehmen sollen. Vor ihr liegen jetzt die Daten von zwei Variablen als Kategorien oder wie es in R so schön heißt, als Faktoren. Aber immerhin, hofft sie das was bei den Daten rausgekommen ist. Dann noch schnell Star Trek starten und los gehts mit der Kraft von Takis Blue Heat. Gezählt hat Nilufar einiges mit  $n = 147$  Beobachtungen von Kartoffeln. Zum einen hat sie als Behandlung *Pestizideinsatz* [ja/nein] bestimmt und zum anderen die Variable *Trockengewicht über Zielwert* [ja/nein] ermittelt. Nun möchte ihr Betreuer gerne einen  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Nach ihrem Experiment erhielt sie folgende 2x2 Kreuztabelle aus ihren erhobenen Daten.

```
##                               Pestizideinsatz
## Trockengewicht über Zielwert ja  nein
##                               ja    6   12
##                               nein 10    7
```

Dann rechnete Nilufar 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: Trockengewicht über Zielwert
## p-value = 0.1811
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.07075797 1.67575062
## sample estimates:
## odds ratio
##  0.3611682
```

Leider kennt sich Nilufar 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)**

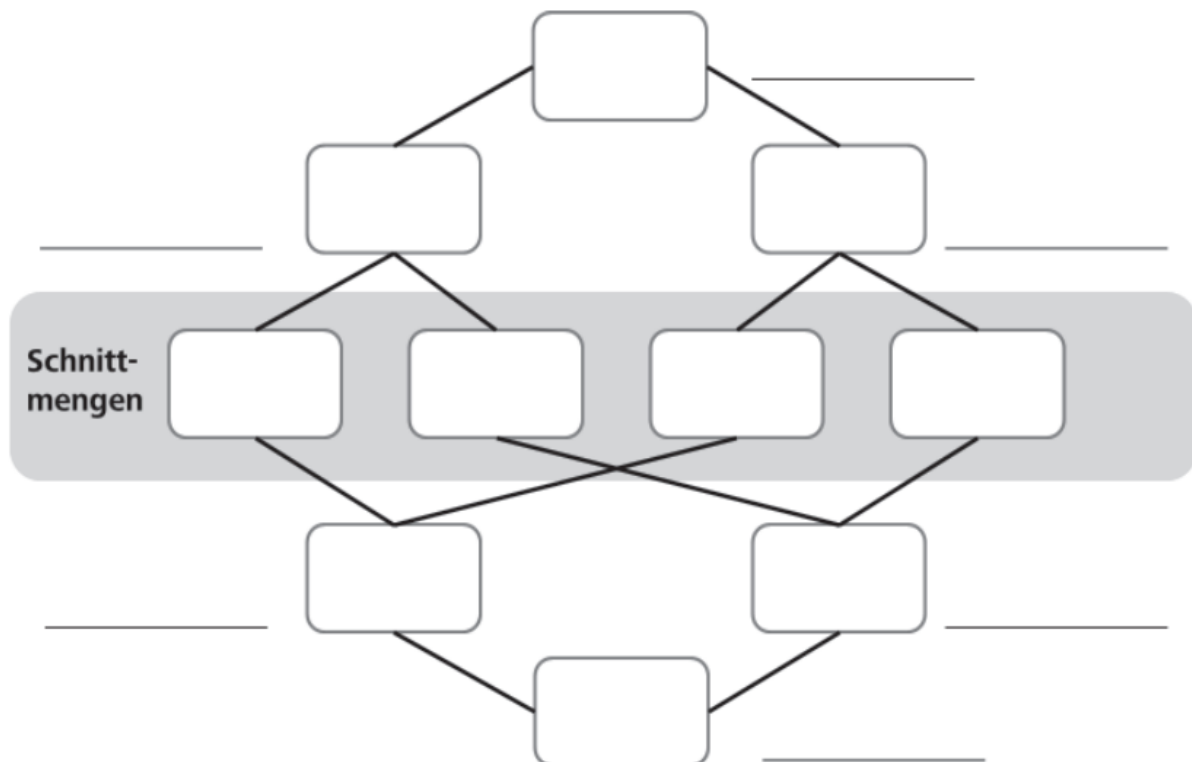
## 91. Aufgabe

(11 Punkte)

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



**Den diagnostische Test am Doppelbaum berechnen** Jessica liest laut vor. 'Die Prävalenz von Klauen-seuche bei Erbsen wird mit 3% angenommen. In 75% der Fälle ist ein Test positiv, wenn das Pflanze erkrankt ist. In 8.5% der Fälle ist ein Test positiv, wenn das Pflanze nicht erkrankt ist und somit gesund ist. Wir führen einen Test auf Schorf an 1000 Erbsen mit einem diagnostischen Test durch.' Paula klappt die Kinnlade runter. In der Stille dudelt David Bowie. Jessica schaut kompetent und schmeißt sich mit offenen Mund Smarties an den Kopf vorbei.



Leider kennen sich Jessica und Paula 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)**

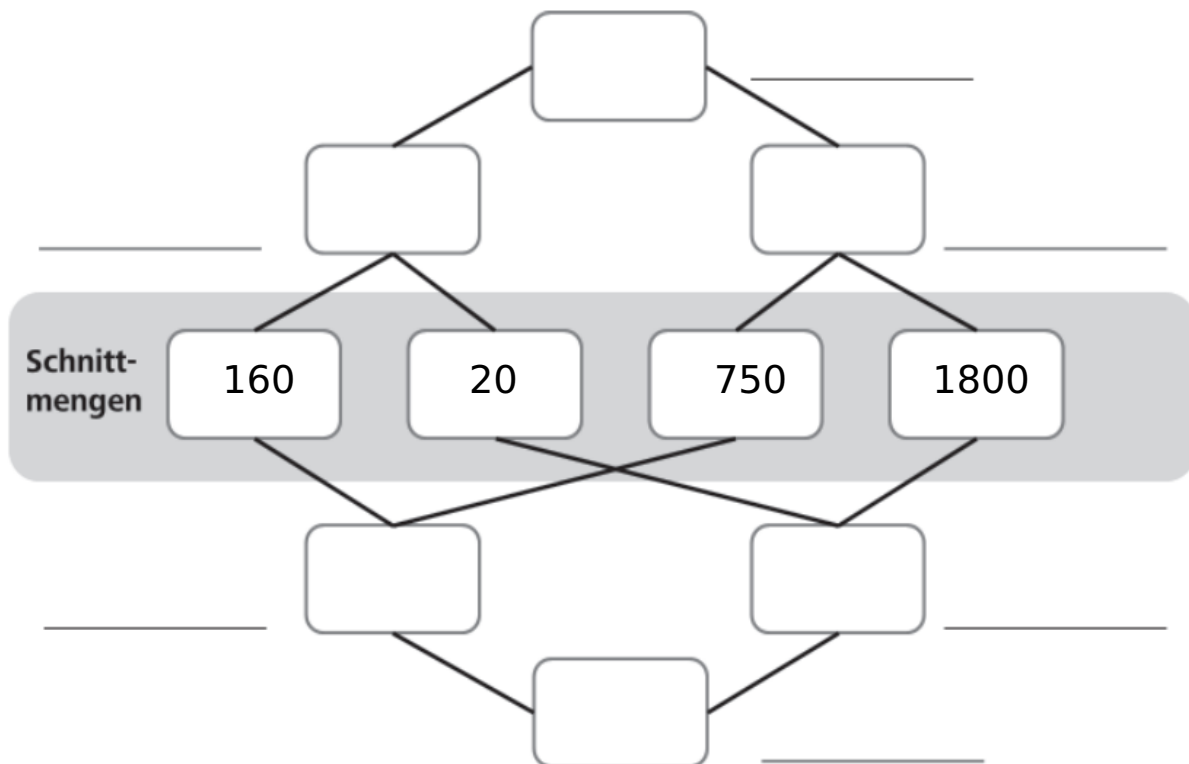
## 92. Aufgabe

(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 Steffen. 'Hm... ich glaube es handelt sich um einen Doppelbaum, den wir beim diagnostischen Testen brauchen.', meint Tina und dreht Harry Potter 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 Romantik von Steffen macht es heute auch nicht mehr einfacher. 'Es geht um Schorf an Brokkoli.', stellt Tina fest. Eigentlich wollte Tina eher los um zu Boxen. Das wird aber wohl nichts mehr.



Leider kennen sich Steffen und Tina 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 Sensitivität und Spezifität des diagnostischen Tests! Erstellen Sie dafür zunächst eine 2x2 Kreuztabelle! **(2 Punkte)**



## Teil VII.

# Lineare Regression & Korrelation

### 93. 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 Steffen. Yuki schaut fragend zurück. 'Kann mich wie immer an nichts erinnern. Können wir trotzdem jetzt erstmal die Daten auswerten? Columbo?', antwortet Yuki leicht angespannt. Die beiden hatten ein Stallexperiment im Oldenburger Land mit Hühnern durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittlichen Bewegungsscore [Movement/h] und Fettgehalt [%/kg]. Jetzt haben die beiden eigentlich alles zusammen. *Eigentlich...*

Fettgehalt [%/kg]	Durchschnittlichen Bewegungsscore [Movement/h]
25.7	20.7
25.4	15.6
24.0	16.0
32.5	25.2
33.4	22.2
25.9	19.8
27.3	18.8
25.4	14.5
28.2	16.3
29.7	16.1

Leider kennen sich Steffen und Yuki 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)**

## 94. 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 Jonas an. 'Ich sehe keine Punkte... ich sehe nur zwei Zeilen einer Tabelle und ich glaube du hast gerade was gelöscht.', antwortet Jessica sichtlich übernächtigt. 'Wir müssen die Koeffizienten der linearen Regression ja auch erst interpretieren!', spricht Jonas sehr deutlich und langsam. Die beiden hatten ein Stallexperiment im Teuteburgerwald mit Milchvieh durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: mittlere Eisenkonzentration [Fe/ml] und Gewichtszuwachs in der 1LW. 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)	6.02	1.88		
Mittlere Eisenkonzentration	1.59	0.18		

Leider kennen sich Jonas und Jessica 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)**

## 95. Aufgabe

(11 Punkte)

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



**Interpretation der Ergebnisse einer linearen Regression in R** 'Ich glaube du bringst da was durch-einander. Wir nutzen zwar auch für die ANOVA die Funktion `lm()` aber hier wollen wir, glaube ich, eine Gerade durch die Punkte zeichnen.', merkt Mark an. 'Ich sehe keine Punkte... das ist doch eine Ausgabe in R. Überhaupt, darum geht es doch gar nicht in unserem Versuch. Wir wollen doch keine Gerade zeichnen?.', antwortet Jessica sichtlich übernächtigt. 'Doch wir müssen nur die Koeffizienten der linearen Regression erst richtig interpretieren und vor unserem geistigen Auge erscheint eine Gerade!', spricht Mark sehr deutlich und langsam. Die beiden hatten ein Kreuzungsexperiment im Teutoburgerwald mit Milchvieh durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Tagestemperatur [C/d] und Gewichtszuwachs in der 1LW. Jetzt wollen sie erstmal schauen, ob es einen Zusammenhang gibt und das soll mit der R Ausgabe möglich sein.

```
##
## Call:
## Gewichtszuwachs ~ Durchschnittliche_Tagestemperatur
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0931 -0.9970 -0.1458  0.8779  2.6356
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.0442     1.5798   1.927   0.0624
## Durchschnittliche_Tagestemperatur  1.4383     0.1608   8.943 1.88e-10
##
## Residual standard error: 1.27 on 34 degrees of freedom
## Multiple R-squared:  0.7017, Adjusted R-squared:  0.6929
## F-statistic: 79.98 on 1 and 34 DF,  p-value: 1.878e-10
```

Leider kennen sich Mark und Jessica 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)**

## 96. 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 Mark 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. Das Verrückte ist, dass der Hamster Columbo wirklich liebt. Das ist Mark sehr recht, denn er braucht Entspannung. Mark hatte ein Stallexperiment im Teuteburgerwald mit Fleischrindern durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Tagestemperatur [C/d] und Schlachtgewicht [kg]. Jetzt hat er eigentlich alles zusammen. *Eigentlich...*, denn mit der R Ausgabe hat Mark jetzt ein Problem. Wenn die Unsicherheit nicht wäre, ja dann wäre wohl vieles möglich für Mark! Aber so..

```
##  
## Pearson's correlation  
##  
## data: Durchschnittliche Tagestemperatur and Schlachtgewicht  
## t = 2.2342, df = 8, p-value = 0.05593  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.01601285 0.89873080  
## sample estimates:  
## cor  
## 0.619863
```

Leider kennt sich Mark 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)**

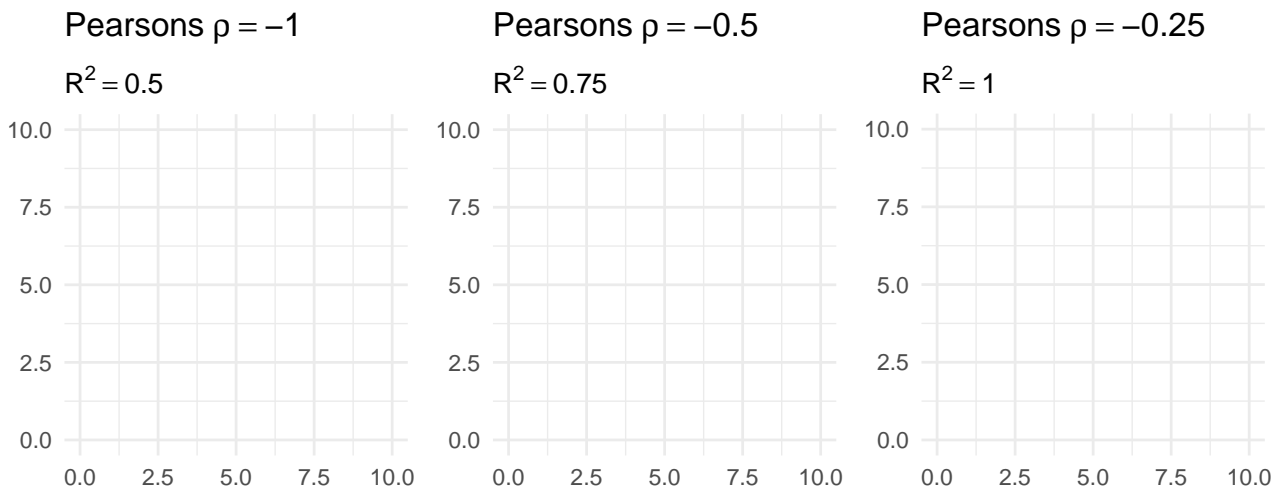
## 97. Aufgabe

(12 Punkte)

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



**Visualisierung der Korrelation und des Bestimmtheitsmaßes** Wenn Star Trek läuft, dann ist das Huhn nicht mehr da. Aber jetzt braucht sie mal Entspannung! Da hilft dann die Aufgabe auch nur bedingt. 'Hm..., drei leere Abbildungen. Was soll ich da jetzt machen?', fragt sich Nilufar und mampft noch ein paar Takis Blue Heat in sich hinein. Nilufar kennt sich nur begrenzt bis gar nicht mit der linearen Regression und Korrelation aus.



**Visualisierung der Korrelation und des Bestimmtheitsmaßes** Leider kennt sich Nilufar 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)**

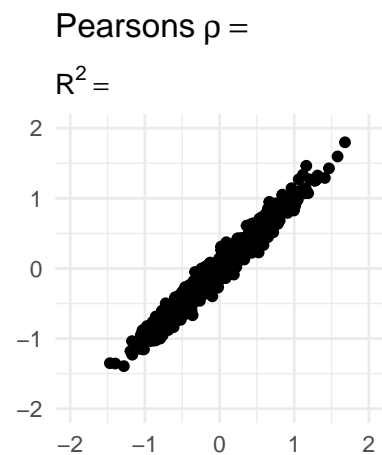
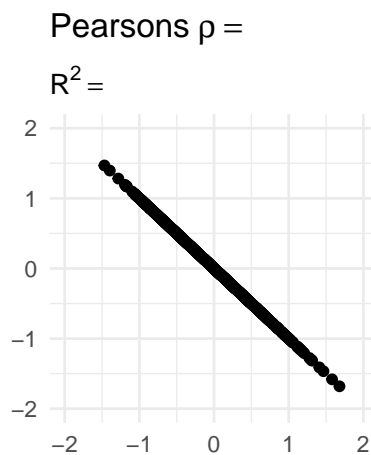
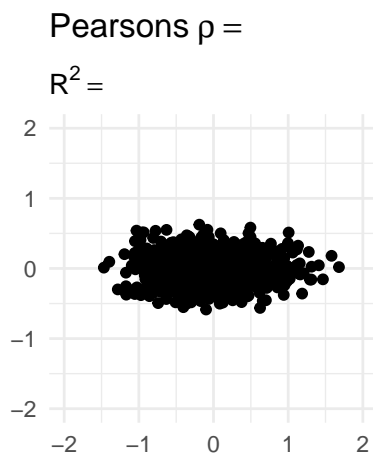
## 98. Aufgabe

(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 Paula. Es ist schon spät. Und das hat einen Grund. Auf seinem Second Screen läuft Jagd auf roter Oktober und Paula schaufelt Smarties. Nicht effizient, aber gut. Paula überlegt, aber ihre Gedanken sind etwas zäh. 'Was soll das hier alles bedeuten?', fragt sich Paula. Irgendwie ist ihr nicht klar wie sie  $\rho$ -Werte oder  $R^2$ -Werte abschätzen soll. Alles nicht so einfach. Wenn der Perfektionismus nicht wäre, ja dann wäre wohl vieles möglich für Paula! Aber so..



Leider kennt sich Paula mit der Korrelationsanalyse und der linearen Regression überhaupt nicht aus. Deshalb braucht sie 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)

## 99. 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 Jessica stolz. Leider hat Jessica vergessen wie der `R` Code für den Residuenplot geht. Jessica hatte anderes im Kopf. Hm, lecker Schokobons und dazu dann im Hintergrund Herr der Ringe laufen lassen. Aber sowas hilft ihr natürlich hier nicht. Da schmeißt sich Jessica noch ein paar Schokobons in den Mund und kaut los.

Mittlere Anzahl	Fettgehalt	$\hat{y}$	$\epsilon$
26.2	10.0	24.3	
32.7	13.9	31.1	
33.1	13.7	30.8	
25.7	10.2	24.7	
26.7	11.3	26.5	
31.7	12.8	29.3	
29.4	13.8	31.0	
27.1	13.3	30.0	
23.4	10.1	24.5	
23.5	11.7	27.2	

Leider kennt sich Jessica 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)**

## 100. Aufgabe


(12 Punkte)

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



**Visualisierung des Regressionskreuzes** Paula hat ein Stallexperiment mit Zandern durchgeführt. Soweit so gut. Dann war sie bei ihrer Betreuerin. Leider war der Schritt nicht so hilfreich. Paula und der Perfektionismus, eine unendliche Geschichte mit kniffligen Wendungen. Aber es muss ja weitergehen. Paula hatte dann in ihrer Abschlusarbeit einfach zu viele Endpunkte gemessen und ist jetzt vollkommen durcheinander, welche Analyse sie nun wie rechnen soll. Naja, dann heißt es jetzt eben White Lies aufdrehen und darüber nachdenken, was hier eigentlich gemacht wurde. Paula fängt einfach an und nimmt den ersten Endpunkt Anzahl dunkle Pigmentstörungen. Dann kann sie sich voran arbeiten. Später dann noch raus um zu Fechten um mal zu entspannen und vielleicht ist Steffen auch da. Wäre toll.

Leider kennt sich Paula mit dem Kontext der linearen Regression überhaupt nicht aus. Deshalb braucht sie 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

### 101. Aufgabe

(16 Punkte)



**Einfache experimentelle Designs** Steffen und Alex sind bei Mark um sich Hilfe für eine Versuchsplanung in **R** zu holen. Im Hintergrund läuft viel zu laut Andrea Berg. Dabei geht es um den Zusammenhang zwischen der Behandlung Lüftungssysteme (*ctrl*, *storm*, *thunder* und *tornado*) und dem Messwert Proteingehalt [g/kg] in Spargel. Der Versuch soll in einem Versuch in einer Klimakammer in der Uckermark durchgeführt werden. Nach der Dozentin ist der Messwert Proteingehalt [g/kg] normalverteilt. Die beiden entschieden sich für ein faktorielles Versuchsdesign. Im ersten Schritt überlegt Mark ein einfaches experimentelles Design zu probieren. Daher entscheiden sich alle drei für ein *Randomized complete block design (RCBD) mit drei Blöcken*. Das sollte für den Anfang erstmal reichen. 'Und jetzt, was machen wir jetzt?', Alex schaut die anderen beiden mit großen Augen an. Die zucken mit der Schulter. Alle mampfen Marzipankugeln.

Leider kennen sich Mark, Steffen und Alex mit dem *Randomized complete block design (RCBD) mit drei Blöcken* überhaupt 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 für das faktorielle Versuchsdesign in **R**! **(2 Punkte)**
5. Erstellen Sie das statistische Modell in der in **R** üblichen Schreibweise für 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 mögliche Abbildung im Kontext der wissenschaftlichen Fragestellung! Beschriften Sie die Abbildung! **(2 Punkte)**
8. Ergänzen Sie zu der Abbildung ein mögliches Ergebnis des multiplen Gruppenvergleichs! Begründen Sie Ihre Antwort! **(2 Punkte)**




## 102. Aufgabe

(20 Punkte)



**Fortgeschrittene experimentelle Designs** Der Hamster macht mal wieder Randalen in Tinas Zimmer und rennt davon! Nilufar und Yuki sind bei Tina in in der Uckermark wo der neue, bessere Versuch stattfinden soll. Dabei soll in einem Leistungssteigerungsversuch in der Uckermark mit Hühnern durchgeführt werden. Auf dem Tisch stapeln sich Marzipankugeln aus Vollkorndinkelmehl. Eine Spezialität der Komune hier. Nilufar hasst Vollkorn wie Gummibärchen geliebt werden. In dem neuen Versuch geht es um den Zusammenhang zwischen der Behandlung Bestandsdichte (*standard, eng, weit und kontakt*) sowie Ernährungszusatz (*ctrl* und *getIt*) sowie drei 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 Yuki als zusätzliche Herausforderung noch die Gefälligkeit mitgebracht hat. Daher entscheiden sich alle drei für ein *Split plot design oder auch Spaltanlage*. 'Naja, so viel einfacher ist es dann doch nicht...', merkt Yuki an und sucht der Hamster.

Leider kennen sich Tina, Nilufar und Yuki mit dem *Split plot design oder auch Spaltanlage* 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

### 103. Aufgabe

(9 Punkte)



**Grundlegende Kenntnisse der Programmierung in R** 'Unter den Blinden ist der Einäuge König!', ruft Ihnen Mark entgegen. Leider kennt sich Mark überhaupt nicht mit den Grundlagen in R aus aber sein Betreuer möchte gerne, dass die Auswertung in R gemacht wird. Da müssen Sie dann wohl mal ran und helfen.

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

Sie antworten:

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

Sie antworten:

Mark: *Was war eigentlich nochmal ein Vorteil von der Nutzung von R?* **(1 Punkt)**

Sie antworten:

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

Sie antworten:

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

Sie antworten:

Mark: *Ich sehe überall dieses c(). Was ist denn deren Nutzen?* **(1 Punkt)**

Sie antworten:

Mark: *Wie heißen nochmal die beiden R Pakete, die wir fast immer laden, wenn wir R nutzen wollen?* **(1 Punkt)**

Sie antworten:

Mark: *Ich habe doch die Spalte mutiert und geändert. Warum sehe ich das in R aber mein Datensatz ändert sich nicht?* **(1 Punkt)**

Sie antworten:

Mark: *Ich verstehe den Unterschied zwischen library() und Packages nicht. Warum gibt es die?* **(1 Punkt)**

Sie antworten:

## 104. Aufgabe

(9 Punkte)



**Fortgeschrittene Kenntnisse der Programmierung in R** Jessica muss ihrer Abschlussarbeit mit R arbeiten. Leider ist die Analyse etwas komplexer, so dass es eben in Excel dann nicht mehr geht. Deshalb also gleich alles in R. Das ist auch der Grund warum sie jetzt mit Ihnen in der Küche sitzt und einige vertiefende Fragen zu R an Sie hat! Na dann wollen Sie mal helfen. Immerhin will ihre Betreuerin, dass R genutzt wird und die Abgabe ist dann auch schon in gut einem Monat.

Jessica fragt: *Ich hatte mir eine Analogie für das R Paket {ggplot} gemerkt. Wie war noch gleich die Analogie und das damit verbundene Prinzip von {ggplot}?* **(2 Punkte)**

Sie antworten:

Jessica 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:

Jessica fragt: *Warum wurde jetzt nach dem Laden der Daten die Funktion `mutate()` genutzt?* **(1 Punkt)**

Sie antworten:

Jessica fragt: *Nach der EDA zu urteilen liegt eine Interaktion vor, wie spezifiziere ich diese im Modell, so dass ich die Interaktion zwischen zwei Faktoren  $f_1$  und  $f_2$  testen kann?* **(1 Punkt)**

Sie antworten:

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

Sie antworten:

Jessica fragt: *Wie ist noch gleich die Reihenfolge der Funktionen für die Erstellung eines CLD in R. Welche Funktionen brauche ich da?* **(2 Punkte)**

Sie antworten:

Jessica fragt: *Ich baue mir ja meinen Datensatz in Excel. Was muss ich da im Bezug auf die Namen der Spalten beachten?* **(1 Punkt)**

Sie antworten:

## Teil X.

# Forschendes Lernen

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

*In der Prüfung erhalten Sie einen Auszug der wissenschaftlichen Veröffentlichung. Für die Einarbeitung in die Veröffentlichung ist in der Prüfung ausdrücklich keine Zeit vorgesehen.*

- Sánchez, M., Velásquez, Y., González, M., & Cuevas, J. (2022). Hoverfly pollination enhances yield and fruit quality in mango under protected cultivation. *Scientia Horticulturae*, 304, 111320. [[Link](#)]
- Petersen, F., Demann, J., Restemeyer, D., Olf, H. W., Westendarp, H., Appenroth, K. J., & Ulbrich, A. (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., Cadogan, D. J., Li, X., & Bryden, W. L. (2010). Implications of sorghum in broiler chicken nutrition. *Animal Feed Science and Technology*, 156(3-4), 57-74. [[Link](#)]
- Wu, G., Knabe, D. A., & Kim, S. W. (2004). Arginine nutrition in neonatal pigs. *The Journal of Nutrition*, 134(10), 2783S-2790S. [[Link](#)]

Das forschende Lernen basiert zum anderen auf den folgenden wissenschaftlichen Datensätzen und deren vertiefende Analyse werden als bekannt vorausgesetzt. Die Teilaufgaben der Aufgaben stellen nur eine zufällige Auswahl an möglichen Fragen dar. Die Datensätze werden über ILIAS bereitgestellt.

*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 gemacht.*

- bar

## 105. Aufgabe



(20 Punkte)

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



**Zerforschen einer wissenschaftlichen Veröffentlichung** Jonas hält die wissenschaftliche Veröffentlichung 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* unter einem Schnaufen in die Luft. 'Worum geht es denn eigentlich in dieser Arbeit?', fragt er stirnrunzelnd und wirft die Arme in die Luft, da hilft dann auch nicht mehr die beruhigende Wirkung von Iron Maiden. Jonas soll die Veröffentlichung nutzen um das eigene Experiment zu planen. Als eine Vorlage sozusagen. Daher möchte seine Betreuerin, dass er einmal die Veröffentlichung sinnvoll zusammenfasst. Das sollte dann doch etwas aufwendiger werden. Das wird dann vermutlich heute Abend nichts mehr mit seinem Hobby Stricken. Das Meerschweinchen schaut mitleidig.

Leider kennt sich Jonas mit dem Lesen einer wissenschaftlichen Veröffentlichung mit Fokus auf die Statistik überhaupt nicht aus. Deshalb braucht er 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.

## 106. 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 sein Betreuer, dass Steffen einmal die wissenschaftlichen Daten *data4* sinnvoll zusammenfasst. Dann würde die eigene Arbeit auch leichter von der Hand gehen und Steffen hätte dann schon eine Vorlage um die eigenen erhobenen Daten in eine Tabelle eintragen zu können. 'Das ist jetzt aber umfangreicher als gedacht!', schnauft er und runzelt die Stirn als er in seinen Laptop starrt. Dabei isst er noch ein paar Oreos. Das wird dann vermutlich heute Abend nichts mehr mit Harry Potter

Leider kennt sich Steffen mit der Analyse eines wissenschaftlichen Datensatzes überhaupt nicht aus. Deshalb braucht er 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

### 107. Aufgabe

(10 Punkte)

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



**Herodot – der Schimmel aus Ivenack** Die Lerngruppe *Die Kühe auf dem Deich* bestehend aus Yuki, Alex, Nilufar und Steffen waren auf Exkursion in Mecklenburg-Vorpommern 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 Kühe auf dem Deich* 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 1820 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 180cm, eine Breite von 85cm sowie eine Länge von 250cm.

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 *bequem* 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 15cm 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](#)



## 108. Aufgabe

(10 Punkte)

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



**Von Töpfen auf Tischen** Die Projektgruppe I bestehend aus Jonas, Jessica, Steffen und Tina hat sich zusammengefunden um den ersten Versuch zu planen. In einem Experiment wollen sie die Wuchshöhe von 240 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 8.5cm und eine Höhe von 10cm. Der Kubikmeterpreis für Torf liegt bei 260 EUR.

Helfen Sie der Projektgruppe I 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. Welche *Tischfläche* in  $m^2$  gegeben der Anzahl an Pflanztöpfen inklusive Randpflanzen benötigen Sie im Gewächshaus am Anfang der Keimungsphase? **(3 Punkte)**
4. Berechnen Sie die benötigte Menge an Torf in Liter  $l$ , die Sie für das Befüllen der Pflanztöpfe benötigen! Gehen Sie von *einem Zylinder* für die Pflanztöpfe aus! **(3 Punkte)**
5. Berechnen Sie die Kosten in EUR für Ihre Torfbestellung! **(1 Punkt)**

## 109. Aufgabe

(10 Punkte)

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



**Solar- & Biogasanlagen** Yuki bringt ein neues, tolles Projekt mit in die Lerngruppe *Die Pantoffeltieren* bestehend aus ihr, Paula, Jessica sowie Mark. Um die Energiekosten ihres Betriebes zu senken, will sie eine Solaranlage auf den Rinderstall montieren lassen. Dafür hat sie ihren Stall ausgemessen und findet folgende Maße wieder. Die vordere Seite des Rinderstall hat eine Höhe  $h_v$  von  $6.5m$ . Die hintere Seite des Rinderstall hat eine Höhe  $h_b$  von  $9.5m$ . Der Rinderstall hat eine Tiefe  $t$  von  $16m$  und eine Breite  $b$  von  $50m$ . 'Sag mal Yuki, ist das eine Matheaufgabe oder rechnen wir hier gerade für dich kostenlos als menschliche Computer Sachen für deinen Betrieb?', fragt Jessica mit erhobenen Augenbrauen. Mark und Paula nicken zustimmend.

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

1. Skizzieren Sie den Rinderstall 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 Rinderstall! **(3 Punkte)**

Ebenfalls plant Yuki 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. Ihr Fundament des zylindrischen Methantanks kann nur ein Gewicht von maximal  $10t$  aushalten bevor der Tank wegbricht. Yuki rechnen eine Sicherheitstoleranz von 25% ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei  $-80^\circ C$  eine Dichte von  $220kg/m^3$ . Bei  $-100^\circ C$  hat Methan eine Dichte von  $290kg/m^3$ . Yuki betreibt ihre Anlage bei  $-85^\circ C$ .

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

## 110. 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 Steffen. “Wieso?“, entfährt es Ihnen und Sie bereuen sogleich die Frage. Sofort werden Sie zu einem Whiteboard voller roter Schnüre geschliffen und müssen folgenden mathematischen untermauerten Argumenten im Rewe über sich ergehen lassen. Da kommen Sie nicht mehr raus, also können Sie auch gleich mitmachen. Das Problem liegt in 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 Steffen?

1. Wenn 4 Blaubeerschalen 7.56 Euro kosten, wie viel kosten 7 Schalen? **(2 Punkte)**
2. Wenn Sie die 7 Blaubeerschalen gekauft haben, wie viele Aligatorbirnen zu je 2.89 EUR können Sie sich dann noch für 50 EUR leisten? **(1 Punkt)**

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

- Ein Kilo Strauchtomaten benötigt 180l Wasser. Eine Strauchtomate wiegt 110 - 120g.
- Ein Kilo Salat benötigt 130l Wasser. Ein Salatkopf wiegt 300 - 520g.
- Ein Kilo Avocado benötigt 1100l Wasser. Eine Avocado wiegt 130 - 380g.
- Ein Kilo Blaubeeren benötigt 830l Wasser. Eine Blaubeere wiegt 3 - 3.7g.

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

Chile exportiert im großem Ausmaß Blaubeeren und Avocados. In dem Exportjahr 2023 blieben die Erträge von Blaubeeren mit  $8 \times 10^4$ t in dem prognostizierten Rahmen. Die Menge reduzierte sich um 7.2%. Die Exporte für Avocados stiegen in dem gleichen Zeitraum um 23.8% auf  $2.3 \times 10^5$ t.

4. Wie viele Kubikmeter Wasser hat Chile in dem Exportjahr 2022 exportiert? **(2 Punkte)**

Chile ist eines der wenigen Länder der Welt, die ihr Wasser komplett privatisiert haben. Derzeit sind nur drei Prozent des Wassers des Landes für den häuslichen Verbrauch vorgesehen. In den Dörfern der Anbauregionen versorgen Tankwagen die Bevölkerung jede Woche mit Wasser, es gibt etwa 50 Liter Wasser pro Kopf für den täglichen Bedarf. In *Deutschland* liegt der Verbrauch bei 35 - 115 Liter pro Waschgang einer Waschmaschine und 10 - 15 Liter pro Minute Duschen.

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 Steffen 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.

## 111. Aufgabe

(12 Punkte)

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



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

**Die Dampfnudelerde** “Was für einen Unsinn!“, rufen Sie. Jetzt haben Sie auf Empfehlung von von Alex 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  $1.2742 \times 10^4 \text{ km}$  und eine mittlere Dichte  $\rho$  von  $5.51 \text{ g/cm}^3$ . Das Gewicht von einem heute lebenden asiatischen Elefanten liegt bei 3t bis 5t und das Gewicht von einem Tyrannosaurus rex (T. rex) bei 4.5t bis 8t.

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 1.05 astronomische Einheiten (AE). Die Einheit 1 AE wird mit  $1.52 \times 10^8 \text{ km}$  angegeben. Der *massebehaftete* Sonnenwind besteht aus 85% Wasserstoffkernen mit einer molaren Masse von  $1.02 \text{ g/mol}$ , 9% Heliumkernen mit  $4.01 \text{ g/mol}$  sowie 6% weiteren Atomkernen mit  $69.18 \text{ g/mol}$ . Die Teilchendichte bei Eintritt in die Erdatmosphäre liegt zwischen  $0.4$  bis  $100$  Teilchen  $\text{cm}^{-3}$  pro Sekunde mit einer mittleren Teilchendichte von  $5 \text{ cm}^{-3}$  pro Sekunde.

Lösen Sie den folgenden Aufgabenteil mit einer aussagekräftigen Skizze!

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

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

## 112. 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, den dann kommt es zu Picken und Kannibalismus. Denn wenn der Nachbar nervt, dann muss zu Maßnahmen gegriffen werden. Kennt jeder aus einer mittelmäßigen Wohngemeinschaft. Das wollen Tina, Jessica, Paula und Steffen 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) walking, (2) drinking/eating, (3) sitting und (4) wingflapping.

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

	Aldridge et al. (2021)	Baxter et al. (2022)	Jabcobs et al. (2019)
walking	44cm; 27cm; 25.4%	28cm; 22cm; 54.1%	35cm; 26cm; 25.4%
drinking/eating	31cm; 19cm; 2.1%	33cm; 28cm; 4.1%	43cm; 27cm; 1.8%
sitting	25cm; 22cm; 0.8%	35cm; 23cm; 1%	40cm; 23cm; 1%
wingflapping	38cm; 17cm; 16.4%	45cm; 23cm; 14.2%	34cm; 22cm; 12.8%

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

1. Erstellen Sie eine zusammenfassende Tabelle mit den mittleren Werten für  $r$ ,  $R$  und  $PB$  aus der obigen Tabelle 1 für die jeweiligen Verhalten! **(3 Punkte)**
2. Ergänzen Sie eine Spalte mit dem benötigten Platz  $A$  für das jeweilige Verhalten, welches sich aus den mittleren Werten ergibt! **(1 Punkt)**
3. Berechnen Sie das benötigte Platzangebot  $SA$  für alle betrachteten Verhalten! **(1 Punkt)**
4. Skizzieren Sie die Werte  $r_i$ ,  $R_i$  und  $A_i$  für zwei nebeneinander agierende Enten für ein Verhalten  $i$ . Nutzen Sie hierfür vereinfachte geometrische Formen! **(2 Punkte)**
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.](#)

## 113. Aufgabe

(8 Punkte)

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



**Nelken von den Molukken** Nilufar und Paula 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 Nilufar ziemlich plötzlich, so dass Paula die Smarties aus dem Mund fallen. 'Nein, und ehrlich gesagt bin ich auch ziemlich müde...'. Das ist jetzt aber Nilufar egal, denn sie möchte folgende Sachlage diskutieren. Und Nilufar 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 40 Tagen die ersten Symptome ein; die ersten Toten sind nach 65 Tagen zu beklagen; nach 110 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 Skorbut bei der Überquerung des Pazifiks, da durch Erkundungsfahrten weniger Zeit blieb, um wilden Sellerie aufzunehmen. Wilder Sellerie enthält  $8000\mu\text{g}/150\text{mg}$  Vitamin C. Der Bedarf liegt bei  $115\text{mg}$  pro Tag für Männer.

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

## 114. Aufgabe

(10 Punkte)

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



**Event Horizon – Am Rande des Universums** Steffen ist bei Paula um gemeinsam *Event Horizon – Am Rande des Universums* zu streamen. Das war jetzt nicht die beste Idee. Denn Steffen kann Horror überhaupt nicht ab. Deshalb flüchtet er sich in Logik um seine Emotionen zu bändigen. Paula mampft ungerührt Smarties. Folgenden Gedankengang nutzt Steffen um dem Film zu entkommen. Die Sonne hat eine aktuelle, angenommene Masse von  $2 \times 10^{29} \text{ kg}$ . Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 4000m kollabiert, wird die Sonne 20% 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 Steffen 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  $6.165 \cdot 10^{-11} \text{ m}^3 (\text{kg} \cdot \text{s}^2)^{-1}$

Im Folgenden wollen wir Steffen 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! **(1 Punkt)**
2. Überprüfen Sie Ihre umgestellte Formel nach  $v_f$  anhand der Einheiten! **(1 Punkt)**
3. Berechnen Sie die notwendige Fluchtgeschwindigkeit  $v_f$  des Lichtteilchens mit den angegebenen Informationen! **(2 Punkte)**
4. Gehen Sie von einer Lichtgeschwindigkeit von  $2.7 \times 10^8 \text{ m/s}$  aus. Kann das Lichtteilchen der Gravitation des schwarzen Lochs entkommen? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Stellen Sie den Zusammenhang zwischen dem sich verringernden Radius  $r$  des schwarzen Lochs bei gleichbleibender Masse  $m_s$  und der notwendigen Fluchtgeschwindigkeit  $v_f$  in einer Abbildung dar! **(2 Punkte)**
6. Ein Auto und eine Feder stürzen aus großer und gleicher Höhe in ein schwarzes Loch. Welches der beiden Objekte überschreitet zuerst den Ereignishorizont des schwarzen Loches? Begründen Sie Ihre Antwort mathematisch! **(2 Punkte)**

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

## 115. 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** Nilufar und Jessica wandern durch den Teuteburgerwald um mal vom Studium runterzukommen. 'Kennst du eigentlich Enrico Fermi?', fragt Nilufar 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 extraterrestrischer Technik zu beobachten?'. Jessica schaut sie irritiert und interessiert an. Die beiden hat das Problem gepackt. Deshalb wollen Nilufar und Jessica 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  $6.1198 \times 10^4 \text{ km/h}$  los um sich auf den erreichten Planeten selbst zu replizieren. Nach 500 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  $1.5 \times 10^{11}$  Sterne in der Milchstraße. Nehmen Sie eine Lichtgeschwindigkeit von  $2.8 \times 10^8 \text{ m/s}$  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 vermuteten Alter der Erde von  $4.1 \times 10^9$  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  $1.2 \times 10^8$  Jahre vollständig im Erdinneren umgewandelt haben! **(2 Punkte)**

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



## 116. Aufgabe

(10 Punkte)

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



**Pyramiden bauen** Es stehen die mecklemburgischen Pyramidentage! Sie und Paula 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 74 Grad im Vergleich zu den ägyptischen Pyramiden mit 55 Grad auf. Die durchschnittliche Seitenlänge der Grundfläche einer Pyramide beträgt 38 Königsellen. Eine Königselle misst 52.6cm.

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 38 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 7cm 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 5 Sklaven, die Ihnen und Paula bei dem Befüllen der Pyramide mit Schutt zu Seite stehen. Leider haben Ihre Sklaven zu allem Überfluss auch noch chronische Rückenschmerzen entwickelt, als die Sklaven von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 70% aus. In eine Schubkarre passen 90 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  $8^\circ$  ändert! **(2 Punkte)**

## 117. Aufgabe

(12 Punkte)

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



**Geocaching – Von Satelliten und Plastikdosen** Es ist Wochenende und das Wetter ist sweet. Nilufar und Mark schwingen sich auf ihre Cachermobile um mit 15km/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. Nilufar und Mark wollen diesmal endlich die aufwärts 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 Nilufar und Mark für die Planung der Route zu Verfügung<sup>10</sup>.

Ort	Cache	Wertung (S T G)
A	GCRLNPZ	5.0   2.5   Normal
B	GCC6BP2	3.0   4.5   Klein
C	GCYHYXR	2.5   3.0   Normal
D	GCSWCFV	4.0   2.0   Klein
E	GC8AHSF	1.0   4.0   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 6.5km bekannt. Der Entfernungsvektor  $\overrightarrow{BE}$  ist das 1.5-fache des Entfernungsvektor  $\overrightarrow{CB}$ . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr 30° südlich. Wenn Sie von dem Ort C den Ort B anpeilen, so liegt der Ort D ungefähr 55° 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! **(2 Punkte)**
2. Welche Strecke in km legen Sie bei der Bewältigung der aufwärts Terrainchallenge zurück? **(5 Punkte)**
3. Gehen Sie von einer zusätzlichen Suchzeit in Stunden für die Caches an den jeweiligen Orten zur reinen Reisezeit mit Ihrem Cachermobil aus. Die Suchzeit in Stunden für *jeden einzelnen* Cache wird durch die Funktion

$$\text{Suchzeit} = 0.05 + 0.13 \cdot \text{Schwierigkeit}$$

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

4. An der höchsten Schwierigkeit müssen Sie angeln. Ihre Angel ist ausgefahren 7m lang. Erreichen Sie einen Cache in der Höhe von 9.1m? Berechnen Sie dazu Ihre maximale mögliche Angelhöhe! Welche Annahmen mussten Sie treffen um die Aufgabe zu lösen? **(2 Punkte)**

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

## 118. Aufgabe

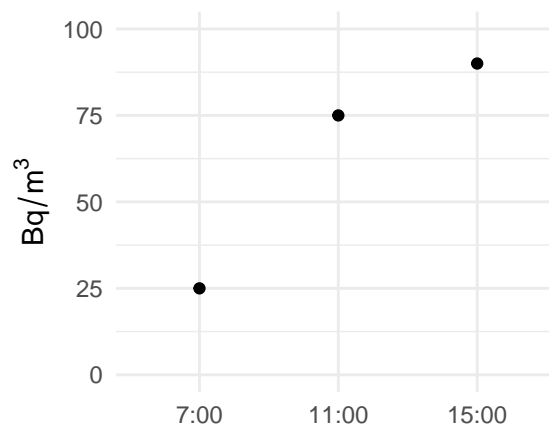
(10 Punkte)

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 Jessica 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 Jessica überhaupt nicht weiter. Sie müssen also helfen!



1. Wie lange dauert es in Stunden bis Sie eine kritische Belastung von  $300 Bq/m^3$  in Ihrem ungelüfteten Kellerraum erreicht haben? **(2 Punkte)**

Radon zerfällt mit einer Halbwertszeit von 1.8d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 140d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

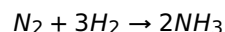
2. Wie lange dauert es in Stunden bis Ihre kritische Radonbelastung von  $300 Bq/m^3$  auf unter  $90 Bq/m^3$  gefallen ist? **(4 Punkte)**

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	77.1	27.9	
Sauerstoff	19.5	15.8	
Kohlenstoffdioxid	0.035	12.5	

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

Während Jessica ihr etwas pappiges Toastbrot mampfen kommt Jessica die Dokumentation über Brot aus Luft in den Sinn. Jessica 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. Wieviel 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](#)

## 119. Aufgabe

(10 Punkte)

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



**Armee der Finsternis** Der Studentenjob von Jonas war nach Ladenschluss bei Aldi die Regale einzuräumen. Dabei ist Jonas in der Auslage der Sonderangebote das Necronomicon<sup>14</sup> in die Hände gefallen. Nun ist er eine Magierin der Zeichen geworden! Also eigentlich kann Jonas nur Mathe und das dämliche Necronomicon hat ihn in die Vergangenheit geschleudert... aber gut, was tut man nicht alles im Jahr 691 n. Chr. für den neuen Lehnsherren Fürsten Arthur. Jonas baut natürlich einen Schrottkugelturm um sich den Horden der Finsternis mit genug Schrott erwehren zu können! Jonas 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  $10mg$  zu gleichförmigen Bleikugeln bei einer Geschwindigkeit von  $14m/s$  bilden.

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

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

2. Nennen Sie die beiden geometrischen Formen aus denen sich näherungsweise ein Tropfen zusammensetzt! Erstellen Sie eine beschriftete Skizze des Tropfens! **(2 Punkte)**
3. Sie messen eine Länge des Tropfens von  $3.5mm$ . Die Löcher im Sieb erlauben ein Tropfendurchmesser von  $1.5mm$ . 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  $11.23g/cm^3$ .

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. 700 Bleikugeln produzieren wollen und die Bleikugel im Fall  $1.2cm$  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

## 120. Aufgabe

(10 Punkte)

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



**Armee der Kaninchen** Leider hat es bei Alex mit der Surfschule in Down Under nicht geklappt. War vielleicht auch nicht so die beste Idee... aber dafür hat Alex eine neue Eingebung! Oder wie es Mike Tyson zugeschrieben wird: »Ich wurde nie niedergeschlagen, ich war immer am Aufstehen!«. Daher macht Alex jetzt einen Großhandel mit Kaninchenfleisch und damit dem teuersten Fleisch in Australien auf. Moment, hopsen hier nicht, seit Thomas Austin im Jahr 1860 ungefähr 32 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Alex 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.

$$f(t) = 10^{10} - 1.2 \times 10^9 \cdot 2.2^{-0.2 \cdot t + 4.1}$$

1. Skizzieren Sie die Sättigungsfunktion *annäherungsweise* in einer Abbildung! **(1 Punkt)**
2. Wie viele Kaninchen können nach der Sättigungsfunktion maximal im westlichen Australien leben? Ergänzen Sie den Wert in Ihrer Abbildung! **(2 Punkte)**
3. Wie viele Millionen Kaninchen leben nach der Sättigungsfunktion nach 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 20 Kaninchen ausgebracht. Da die Kaninchen keine Maßnahmen gegen die Ausbreitung vornehmen können, verläuft die Ausbreitung mit einem wöchentlichen Wachstumsfaktor von 1.8 nach folgender Formel.

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

3. Wie viele Wochen benötigen die Viren um theoretisch die gesamte Kaninchenpopulation nach 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 50% der gesättigten Kaninchenpopulation angewachsen ist? **(2 Punkte)**

Thomas Austin entließ die Kaninchen im äußersten Süden von Australien. Australien hat eine West-Ost-Ausdehnung von 4300km und eine Nord-Süd-Ausdehnung von knapp 3400km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 7.8km pro Jahr aus.

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

Eine jährliche Impfung gegen das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV) kosten 9\$ pro Tier und der durchführende Arzt verlangt ca. 35\$ pro Tier.

6. In Ihrem Stall leben 1200 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...](#)

## 121. Aufgabe

(12 Punkte)

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



**Uckermark. Unendliche Weiten.** Wir schreiben das Jahr 2024. Dies sind die Abenteuer des Schafs Frida und Tina. Grünes Gras unter Tinas Füßen und ein strammer Wind im Gesicht, egal wohin sie schaut. Ein schmatzendes Geräusch ertönt unter Tina. Tina sinniert, sollte sie ihre weiten Graslandschaften jetzt schon düngen? Dafür benötigt Tina 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. Tina 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  $0.6\times$ , 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	0.1
01. Feb 2023	1.5
01. Mrz 2023	3.5
01. Apr 2023	6.1

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  $190^\circ\text{C}$  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 Tinas Jonagoldplantage werden Sie beide auf dem Trecker von einer Gruppe elektrifizierter Renter abgedrängt. Der Trecker muss wieder aus dem Graben! Frida und die elektrifizierten Rentner ziehen an zwei, separaten Seilen. Dabei zieht Frida mit  $180\text{N}$ . Die elektrifizierten Renter bringen eine Kraft von  $210\text{N}$  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 Frida 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 Frida und die Rentner mit einem  $40^\circ$  Winkel mit ihrem Seil an dem Trecker. Welche Kraft wird aufgebracht? **(2 Punkte)**
7. Mit welcher Beschleunigung ziehen Sie den  $1.3\text{t}$  schweren Trecker *jeweils* aus dem Graben, wenn  $F = m \cdot a$  gilt? **(1 Punkt)**

## 122. Aufgabe

(10 Punkte)

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



**In der Kartonagenfabrik** Nilufar, Paula, Jessica und Alex sitzen im Bus. Wenn man sich zu spät anmeldet, dann ist die Exkursion nicht so toll. Nilufar hatte den Anderen in der Lerngruppe zu spät Bescheid gesagt. 'Was denn, bin ich eure Nanny oder was?!', entfährt es Nilufar nachdem die vorwurfsvollen Blicke schon eine Weile auf ihr 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, 6-mal-gefaltete, 0.8mm, 60-cm-Karton* durch Falzung hergestellt. Beim letzten Mal war Rektor Skinner die Stimmung zu schlecht und deshalb geht es erst nach Hause, wenn ein paar Aufgaben gelöst sind. Martin gefällt das. An dem Vorrat an Zigaretten von Mrs. Krabappel meinen 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 60cm und eine Breite von 21cm. Die Kartonagenmaschine in Flint soll dann einen quadratischen Eckenausschnitt der Länge  $x$  falzen.

1. Erstellen Sie eine Skizze des Kartonblattrohlings! 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 Kartondeckelblattrohlings 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 120m 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 120m 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)**

## 123. Aufgabe

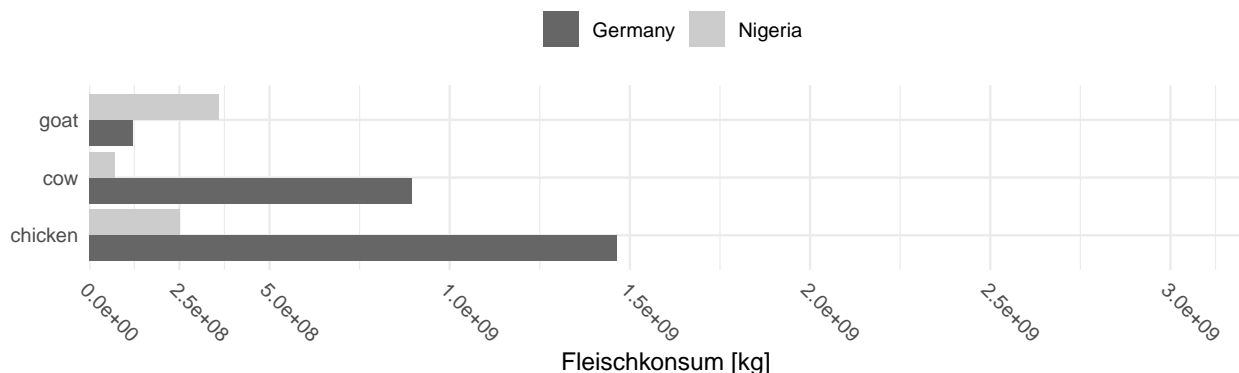
(10 Punkte)

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



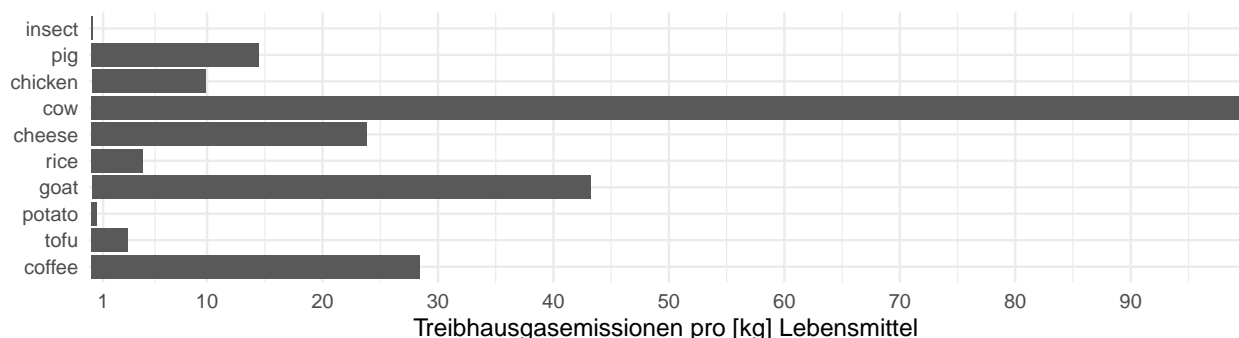
**Ein Pfund Insekten, bitte!** 'Das wird wohl häufiger gehört werden, wenn wir die Menschheit mit Proteinen ausreichend ernähren wollen<sup>16</sup>.', merkt Mark an. Die Lerngruppe um Nilufar, Jonas und Tina sind bei Mark um mal was außergewöhnliches zu essen. Um den Sinn der Nahrungsumstellung zu verdeutlichen,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

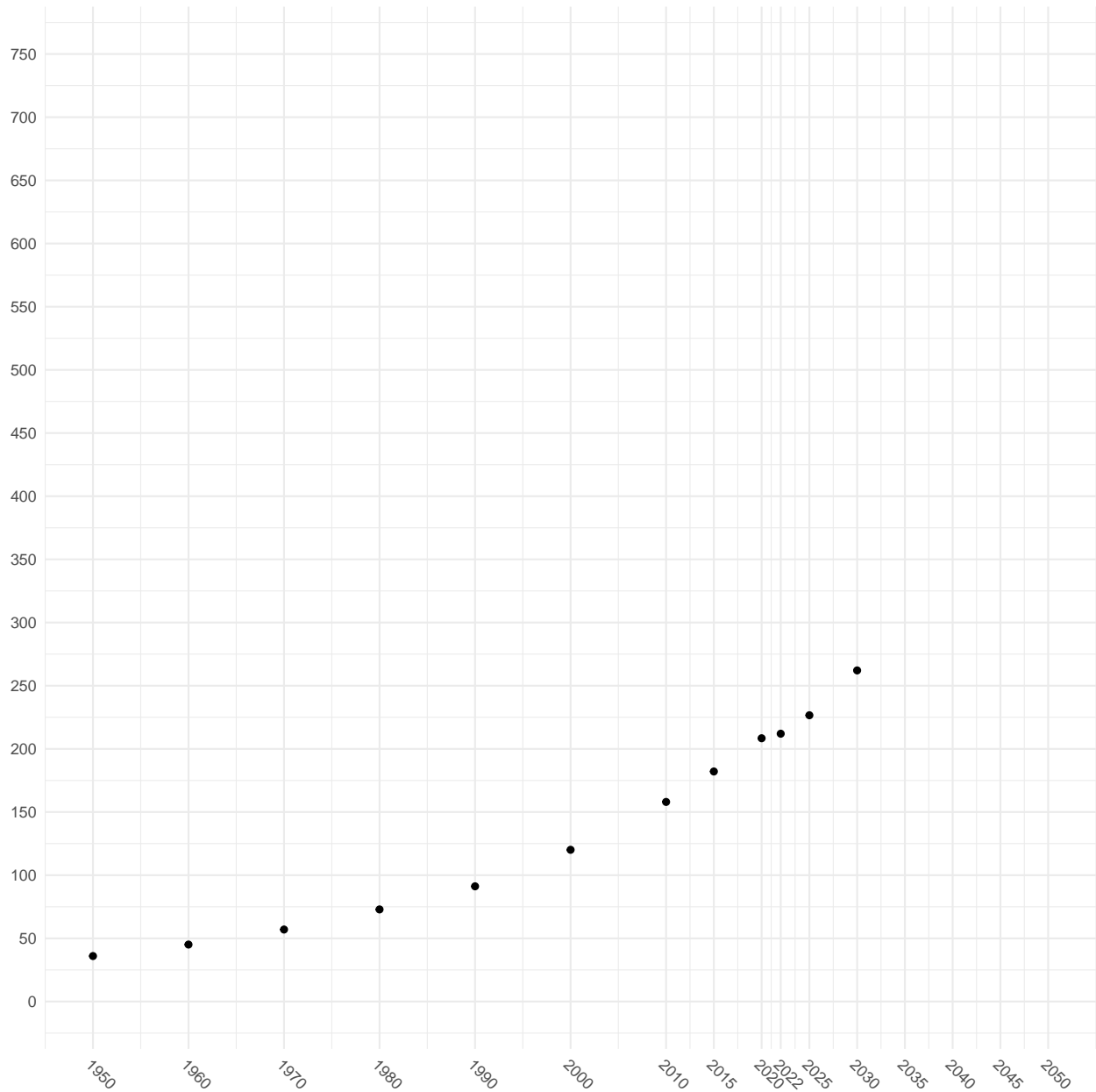


3. Stellen Sie in einer Tabelle die Treibhausgasemissionen an CO<sub>2</sub> pro Kopf, die durch den Fleischkonsum in Deutschland und Nigeria im Jahr 2021 entstehen, dar! Ergänzen Sie auch hier das Verhältnis Nigeria zu Deutschland! **(2 Punkte)**

<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)**

## 124. 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 Mark die historische Ausgabe des Spiegels aus den 80zigern beiseite. Mark und Nilufar sind bei ihrem Hautarzt 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 Mark und Nilufar nun aber über AIDS und dem diagnostischen AIDS-Test, den die beiden nun machen werden? Leider zu wenig. Da brauchen dann Mark und Nilufar mal wieder Ihre Hilfe bei der Interpretation eines diagnostischen Tests!

Die Prävalenz von AIDS bei einem Menschen in Europa wird mit 0.75% angenommen. In 96% der Fälle ist ein HIV-Test positiv, wenn der Patient erkrankt ist. In 1% der Fälle ist ein HIV-Test positiv, wenn der Patient *nicht* erkrankt ist und somit gesund ist. Sie stutzen. Wie wahrscheinlich ist es denn eigentlich an AIDS erkrankt zu sein ( $K^+$ ), wenn Sie einen positiven AIDS-Test vorliegen haben ( $T^+$ )? Gehen Sie für die folgenden Berechnungen von  $n = 4 \times 10^4$  Patienten mit einem diagnostischen Test für AIDS aus. Sie nehmen sich also einen Kuli und fangen an auf der historischen Ausgabe des Spiegels zu 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 Mark und Nilufar, 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 Mark und Nilufar 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

## 125. Aufgabe

(12 Punkte)

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



**Network-Marketing oder Schneeballschlacht!** Tina, Alex und Jessica sitzen bei Yuki und hören sich etwas über Network-Marketing an. Yuki ist jetzt im Network-Marketing tätig. 'Jetzt reicht es. Wir sind eine Lerngruppe und du versuchst uns hier abzuziehen!', poltert Tina 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? Tina 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 KH Gesund und Schön Components (KH-GSC). Das Unternehmen steigerte den Umsatz um rund 10 Prozent von 290 Millionen Euro im Jahr 2021. Doch wie viel kommt bei den Partnern an? Laut KH-GSC habe das Unternehmen  $3.3 \times 10^5$  aktive Partner.

1. Berechnen Sie zuerst den Umsatz der Firma KH-GSC im Jahr 2022! **(1 Punkt)**
2. Wie viel von dem Umsatz im Jahr 2022 wird im Durchschnitt von jedem aktiven Partner erwirtschaftet? **(1 Punkt)**
3. Welche *monatlicher* Umsatz ergibt sich dadurch im Durchschnitt für jeden aktiven Partner bei einer direkten Provision von 30%? **(1 Punkt)**

Das von Yuki zu vermarkende Produkt, hinter dem Yuki voll steht, kostet 100EUR 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 2.5%, 1.5% und 1%. Jeder von Yuki angeworbener »Partner« wirbt wiederum vier Partner für sich selbst an. Pro Monat werden im Schnitt vier Einheiten vom Produkt verkauft. Yuki will nun 5000EUR im Monat *passiv* – also durch indirekte Provisionen – erwirtschaften. Kann das klappen? Sie sind zusammen mit Tina 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)**

Yuki musste zum Einstieg bei KH-GSC Einheiten des Produkts für 4500EUR kaufen. Diese Einheiten kann Yuki nur direkt verkaufen. Das ganze Wohnzimmer ist voll davon. Leider musste Yuki den Kauf über einen Kredit über 5% p.a. über 48 Monate finanzieren. Sie schütteln den Kopf und klären Yuki 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](#)

## 126. Aufgabe

(10 Punkte)

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



**Höhlen & Drachen** Steffen, Mark und Jessica sitzen bei Tina nachdem sich alle begeistert in der Serie *Stranger Thinks* verloren haben. Alle drei wollen jetzt einmal bei Tina *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 Tina nicht so alleine ist.

In dem Spiel hat Steffen nun auf einmal 5 zwölfseitige Würfel ( $5d12$ ) zum würfeln in der Hand. Wenn Steffen eine 12 würfelt, hat Steffen einen Erfolg.

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

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

3. Würden Sie die Axt mit zwei zwölfseitigen Würfeln ( $2d12$ ) als Schaden oder das Schwert mit einem zwölfseitigen Würfel plus 3 ( $1d12+3$ ) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

Jetzt wird es immer wilder, da Mark und Jessica sich jetzt überlegen müssen, wie wahrscheinlich es ist, dass der Rettungswurf gegen den zaubernden Hexer funktioniert. Mark und Jessica haben folgende Wahrscheinlichkeiten gegeben. Die Wahrscheinlichkeit für das Ereignis  $A$ , der Rettungswurf ist erfolgreich, ist  $Pr(A) = 0.7$ , die Wahrscheinlichkeit für das Ereignis  $B$ , der Zauberwurf des Hexers ist erfolgreich, ist  $Pr(B) = 0.9$ . 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)**

## 127. Aufgabe

(12 Punkte)

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



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

Name	$P(\text{win})$	$P(\text{outbid})$
Tina	0.3	0.05
Steffen	0.4	0.076

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

Glücksrad für Arme auf der Kirmes! Leider hat es für Maren Gilzer nicht gereicht. Deshalb sind Sie jetzt mit Paula auf der Kirmes und spielen mit einem Typen in einem Tentakelkostüm um das große Geld. Das Glücksrad hat 24 Felder. Sie beide drehen das Glücksrad zweimal. Auf 10 Feldern gewinnen Paula und Sie 5000EUR sonst 1500EUR. 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 6500EUR? **(1 Punkt)**

Im Fiebertraum von Paula reisen sie beide im Zug nach Köln um bei »Geh aufs Ganze!« mitzuspielen. Paula 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«! Berechnen 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.)

### 128. Aufgabe

(6 Punkte)

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

### 129. 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.

### 130. Aufgabe

(6 Punkte)

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