

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 die Portfolioprüfung)**

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

»Sommersemester 2025«



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



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

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



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

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



Nach der Ausbildung wollte Nilufar eigentlich gleich anfangen zu arbeiten, aber nach einem Praktikum und der Probezeit stellte sie fest, dass es ohne einen Hochschulabschluss schwer wird Führungsverantwortung zu übernehmen. Mit Menschen kann sie schon immer und dann auch eigene Projekte mit anderen verwirklichen, 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 74 Punkten sind aus den Rechen- und Textaufgaben erreicht.  
\_\_\_\_\_ von 94 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
90.0 - 94.0	1,0
85.0 - 89.5	1,3
80.5 - 84.5	1,7
75.5 - 80.0	2,0
71.0 - 75.0	2,3
66.5 - 70.5	2,7
61.5 - 66.0	3,0
57.0 - 61.0	3,3
52.0 - 56.5	3,7
47.0 - 51.5	4,0

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

## Multiple Choice Aufgaben

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

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

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

## Rechen- und Textaufgaben

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

- Es sind \_\_\_\_ von 74 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.

### Programmieren in R

#### 1. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

- A ☐ **R** Pakete sind nur in englischer Sprache verfasst. Es macht keinen Sinn **R** daher in Deutsch zu bedienen.
- B ☐ Programmiersprachen 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 ☐ Alle Funktionen und auch Anwendungen sind in **R** in englischer Sprache. Die Nutzung von deutschen Wörtern ist nicht schick und das ist zu vermeiden.
- D ☐ Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Daher ist die Nutzung in Deutsch in den AGBs von **R** untersagt.
- E ☐ Im Allgemeinen haben Programmiersprachen Probleme mit Umlauten und Sonderzeichen, die in der deutschen Sprache vorkommen. Eine Nutzung der englischen Sprache umgeht dieses Problem auf einfache Art.

#### 2. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Die Ablage von Daten erfolgt normalerweise in Excel. Damit die Daten dann in **R** ausgewertet werden können, müssen die Daten in Excel im Long-Format angelegt werden. Welche Aussage ist zum Long-Format richtig?

- A ☐ In den Spalten finden sich die experimentellen Faktoren (X) sowie die Messwerte (Y). In den Zeilen finden sich dann die einzelnen Beobachtungen.
- B ☐ In den Spalten sind die Beobachtungen in den Zeilen die Variablen, wie die Messwerte und experimentellen Faktoren.
- C ☐ Wichtig ist, dass sich in den Zeilen die Variablen finden, wie die Messwerte (Y) sowie die experimentellen Faktoren (X). In den Spalten stehen die Variablen.
- D ☐ In den Spalten finden sich die einzelnen Beobachtungen und in den Zeilen die gemessenen Variablen (Y) sowie die experimentellen Faktoren (X).
- E ☐ Das Long-Format beschreibt in den Spalten die Beobachtungen sowie in den Zeilen die *unabhängigen* Beobachtungen.

#### 3. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

In Ihrer Abschlussarbeit wollen Sie zu Beginn eine explorativen Datenanalyse (EDA) in **R** rechnen. Dafür gibt es eine generelle Abfolge von Prozessschritten. Welche ist hierbei die richtige Reihenfolge?

- A ☐ Wir lesen als erstes die Daten über `read_excel()` ein, transformieren die Spalten über `mutate()` in die richtige Form und können dann über `ggplot()` uns die Abbildungen erstellen lassen. Wichtig ist, dass wir keine Faktoren sondern nur numerische Variablen vorliegen haben.
- B ☐ Wir lesen die Daten ein und mutieren die Daten. Dabei ist wichtig, dass wir nicht das Paket `tidyverse` nutzen, da dieses Paket veraltet ist. über die Funktion `library(tidyverse)` entfernen wir das Paket von der Analyse.
- 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 ☐ Wir lesen als erstes die Daten über `read_excel()` ein, transformieren die Spalten über `mutate()` in die richtige Form und können dann über `ggplot()` uns die Abbildungen erstellen lassen.
- E ☐ Wir lesen die Daten über eine generische Funktion `read()` ein und müssen dann die Funktion `ggplot()` nur noch installieren. Dann haben wir die Abbildungen als `*.png` vorliegen.

## Deskriptive Statistik & Explorative Datenanalyse

### 4. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik

Wie lautet der Mittelwert und Standardabweichung von  $y$  mit 9, 8, 11, 10 und 7.

- A ☐ Sie erhalten 9 +/- 1.26
- B ☐ Es ergibt sich 8 +/- 1.25
- C ☐ Es berechnet sich 9 +/- 2.5
- D ☐ Es berechnet sich 9 +/- 1.58
- E ☐ Es ergibt sich 10 +/- 0.79

### 5. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik

Gegeben ist  $y$  mit 15, 23, 13, 11, 31, 10, 14, 13, 14, 10 und 51. Berechnen Sie den Median, das 1<sup>st</sup> Quartile sowie das 3<sup>rd</sup> Quartile.

- A ☐ Es ergibt sich 14 +/- 11
- B ☐ Es berechnet sich 15 [12; 22]
- C ☐ Sie erhalten 14 +/- 23
- D ☐ Sie erhalten 14 [9; 21]
- E ☐ Es berechnet sich 14 [11; 23]

### 6. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik

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. Am Ende durch die Fallzahl ( $n - 1$ ) teilen
- B ☐ 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 ( $n - 1$ ).
- 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 ( $n - 1$ ).
- D ☐ Den Mittelwert berechnen und die Abstände quadrieren. Die Summe mit der Fallzahl ( $n - 1$ ) multiplizieren.
- E ☐ 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 ( $n - 1$ ). Als letzten Schritt ziehen wir die quadratische Wurzel.

### 7. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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 ☐ Durch die Abbildung des Barplot erhalten wir die Informationen über die Mittelwerte und die Varianz.
- B ☐ Durch die Abbildung des Barplot erhalten wir die Informationen über den Median und die Standardabweichung.
- C ☐ Der Barplot stellt die Mittelwerte und die Varianz dar.
- D ☐ Der Barplot stellt die Mittelwerte und die Standardabweichung dar.
- E ☐ Der Barplot stellt den Median und die Quartile dar.

## 8. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

In Ihrer Abschlussarbeit zu Lauch finden Sie auf einmal seltsame Daten. Jedenfalls kommt Ihnen das so vor. Daher berechnen Sie den Mittelwert mit 18.1 t/ha und den Median mit 13.8 t/ha. Der Mittelwert wird mit  $\bar{y}$  und der Median mit  $\tilde{y}$  abgekürzt. Welche Aussage ist richtig?

- A ☐ Wenn sich der Mittelwert und der Median nicht unterscheiden, liegen vermutlich Outlier in den Daten vor.
- B ☐ Da sich der Mittelwert und der Median unterscheiden, liegen vermutlich Outlier in den Daten vor. Wir untersuchen den Datensatz nach auffälligen Beobachtungen.
- C ☐ Nach Ihrer Betrachtung der Werte unterscheiden sich Mittelwert und Median nicht. Die Daten können also so nicht verwendet werden. Es gibt mindestens Outlier  $n - 1$  Ausreißer in den Daten.
- D ☐ Da sich der Mittelwert und der Median unterscheiden, liegen vermutlich keine Outlier in den Daten vor. Wir verwenden den Datensatz so wie er ist.
- E ☐ Nach der Betrachtung der Werte unterscheiden sich Mittelwert und Median nicht. Die Daten können so verwendet werden wie sie vorliegen. Es gibt keinen Outlier.

## 9. Aufgabe

(2 Punkte)

Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Im Folgenden sehen Sie eine Normalverteilung dargestellt. Welche Aussage zu der Normalverteilung und der Standardabweichung  $\sigma$  ist richtig?



- A ☐ Die Fläche rechts von  $2\sigma$  ist der p-Wert mit  $Pr(D|H_0)$  in der obigen Abbildung.
- B ☐ Dargestellt ist eine Standardnormalverteilung.
- C ☐ Die Fläche links von  $-2\sigma$  ist der p-Wert mit  $Pr(D|H_0)$  in der obigen Abbildung.
- D ☐ Die Fläche unter der Kurve ist 1, wenn die Nullhypothese falsch ist. Wenn die Nullhypothese gilt, dann ist die Fläche  $1 - \alpha$ . Somit ergibt sich auch eine Standardabweichung  $\sigma$  mit  $\alpha$  gleich 0.05 in beiden Fällen.
- E ☐ Es liegen 68% der Beobachtungen zwischen  $\bar{y} \pm 1\sigma$ . Angezeigt durch die Fläche zwischen  $-1\sigma$  und  $1\sigma$  in der obigen Abbildung.

## 10. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A ☐ Wir brauchen fünf oder mehr Beobachtungen.
- B ☐ Mindestens 20 Beobachtungen.
- C ☐ 1 Beobachtung.
- D ☐ Wir sollten eine Beobachtung mindestens pro Gruppe vorliegen haben.
- E ☐ Die untere Grenze liegt bei einer Beobachtung.

## 11. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

## 12. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A ☐ Nach dem Einlesen der Daten nutzen wir einen Barplot um zu schauen, ob alle Mittelwerte über alle Behandlungen in etwa gleich groß sind. Damit ist dann auch die Varianz in allen Behandlungen in etwa gleich.
- B ☐ 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 Varianzhomogenität ausgehen.
- C ☐ Einen Barplot. Die Mittelwerte müssen alle auf einer Höhe liegen. Die Fehlerbalken haben hier keine Informationen.
- D ☐ In einer explorativen Datenanalyse nutzen wir den Violinplot. Dabei sollte der Bauch am Rand liegen. Dann können wir von einer Varianzhomogenität ausgehen.
- E ☐ Nach dem Einlesen der Daten nutzen wir einen Boxplot um zu schauen, ob alle Boxen über alle Behandlungen in etwa gleich groß sind. Damit ist dann auch das IQR in allen Behandlungen in etwa gleich.

## Statistische Testtheorie

## 13. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

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



## 14. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Der Fehler 1. Art oder auch Signifikanzniveau  $\alpha$  genannt, liegt bei 5%. Welcher der folgenden Gründe für diese Festlegung auf 5% als Signifikanzschwelle ist richtig?

- ☐ A In der Wissenschaft gibt es neben der Naturkonstante, die sich aus der Beobachtung der Welt ergibt, noch die Kulturkonstante, die von einer Gruppe Menschen selbstgewählt wird. Dabei ist  $\alpha = 5\%$  eine Kulturkonstante und wurde somit eher zufällig gewählt.
- ☐ B 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.
- ☐ C Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.
- ☐ D Auf einer Statistikkonferenz in Genf im Jahre 1942 wurde dieser Cut-Off nach langen Diskussionen festgelegt. Bis heute ist der Cut Off aber umstritten, da wegen dem 2. Weltkrieg viele Wissenschaftler nicht teilnehmen konnten.
- ☐ E Der Wert ergab sich aus einer Auswertung von 1042 wissenschaftlichen Veröffentlichungen zwischen 1914 und 1948. Der Wert 5% wurde in 28% der Veröffentlichungen genutzt. Daher legte man sich auf diese Zahl fest.

## 15. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Das statistische Testen basiert auf dem Falsifikationsprinzip. Es besagt,

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

## 16. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Betrachten wir die Teststatistik aus einem abstrakteren Blickwinkel. Beim statistischen Testen wird das Testsignal mit dem Testnoise aus den Daten  $D$  zu einer Teststatistik  $T_D$  verrechnet. Welche der Formel berechnet korrekt die Teststatistik  $T_D$ ?

- ☐ A Es gilt  $T_D = (\text{signal} \cdot \text{noise})^2$ . Der Effekt *signal* wird mit der Varianz *noise* gewichtet.
- ☐ B Bei der Berechnung der Teststatistik  $T_D$  gewichten wir den Effekt *signal* mit der Varianz *noise*. Wir können verallgemeinert  $T_D = \text{signal}/\text{noise}^2$  schreiben.
- ☐ C Es gilt  $T_D = \text{noise}/\text{signal}$ . Der Effekt *noise* wird mit der Varianz *signal* gewichtet.
- ☐ D Bei der Berechnung der Teststatistik  $T_D$  gewichten wir den Effekt *signal* mit der Varianz *noise*. Wir können verallgemeinert  $T_D = \text{signal}/\text{noise}$  schreiben.
- ☐ E Wir gewichten den Effekt *noise* mit der Varianz *signal* und erhalten  $\text{signal}/\text{noise}^2$ .

## 17. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

- ☐ A Wir vergleichen mit dem  $p$ -Wert und dem Signifikanzniveau  $\alpha$  absolute Werte auf einem Zahlenstrahl und damit den Unterschied der Teststatistiken, wenn die  $H_0$  gilt.
- ☐ B 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.

- C** ☐ Wir vergleichen mit dem  $p$ -Wert und dem Signifikanzniveau  $\alpha$  Wahrscheinlichkeiten und damit die absoluten Werte auf einem Zahlenstrahl, wenn die  $H_0$  gilt.
- D** ☐ 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.
- E** ☐ Wir machen eine Aussage über die Flächen und der Kurve der Teststatistik, wenn die  $H_0$  gilt. Dabei werden Wahrscheinlichkeiten verglichen, die durch die Flächen unter der Kurve repräsentiert werden.

## 18. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Um die Testtheorie besser zu verstehen, mag es manchmal sinnvoll sein ein Beispiel aus dem Alltag zu wählen. Die Ergebnisse der Analyse durch einen statistischen Test können auch in grobe Analogie zur Wettervorhersage gebracht werden. Welche Aussage trifft am ehesten zu?

- A** ☐ In der Analogie der 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.
- B** ☐ In der Analogie der Wahrscheinlichkeit für Regen: ein statistischer Test erlaubt die Wahrscheinlichkeit für ein Ereignis abzuschätzen. Die Stärke des Effektes können wir nicht bestimmen.
- C** ☐ 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.
- D** ☐ In der Analogie der Sonnenscheindauer: Wie lange kann mit einem entsprechenden Effekt gerechnet werden? Die Wahrscheinlichkeit für den Effekt gibt der statistische Test wieder.
- E** ☐ In der Analogie des Niederschlags oder Regenmenge: ein statistischer Test gibt die Stärke eines Effektes wieder. Zum Beispiel, wie hoch ist der Mittelwertsunterschied.

## 19. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Sie wollen eine Aussage über die untersuchte Population treffen. Dazu nutzen Sie einen statistischen Test. Können Sie eine valide Aussage aus einem statistischen Test erhalten?

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

## 20. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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 Forschende muss am Ende wissen, ob das Ergebnis eines Experiments relevant für seine Forschung ist. Dafür kann der Effekt eines statistischen Tests genutzt werden. Damit beschreibt der Effekt den biologischen interpretierbaren Teil einer Ausgabe eines Tests. Zum Beispiel der Unterschied zwischen zwei Anteilen.

- C** ☐ Der Effekt eines statistischen Tests beschreibt den Output oder die Wiedergabe eines Tests in einem Computer.
- D** ☐ Der Effekt eines statistischen Tests beschreibt die biologisch interpretierbare Ausgabe eines Tests. Damit ist der Effekt direkt mit dem Begriff der Signifikanz verbunden. Die Entscheidung über die Signifikanz trifft der Forschende unabhängig von der Relevanz eines statistischen Tests.
- E** ☐ Der Effekt eines statistischen Tests beschreibt die biologisch interpretierbare Ausgabe eines Tests. Modernen Algorithmen liefern keine Effekte mehr sondern nur noch bedingte Wahrscheinlichkeiten. Der Effekt spielt in der modernen Statistik keine Rollen mehr.

## 21. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Sie führen einen agrarwissenschaftlichen Versuch durch. Dafür müssen Sie als ersten Schritt Ihre Versuchseinheiten zufällig den Behandlungen zuordnen. Dieses zufällige Zuordnen nennt man in der Statistik eine Randomisierung. Warum ist die Randomisierung so wichtig für das Gelingen eines Versuchs?

- A** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- B** ☐ Randomisierung war bis 1952 bedeutend, wurde dann aber in Folge besserer Rechnerleistung nicht mehr verwendet. Aktuelle Statistik nutzt keine Randomisierung mehr.
- C** ☐ Randomisierung sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- D** ☐ Strukturgleichheit ist durch Randomisierung gegeben. Leider hilft die Randomisierung noch nicht um von der Stichprobe auf die Grundgesamtheit zu schließen. Deshalb wurde das Falsifikationsprinzip entwickelt.
- E** ☐ Durch eine Randomisierung können wir nicht von Strukturgleichheit zwischen der Stichprobe und der Grundgesamtheit ausgehen.

## 22. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ Mit der Nullhypothese  $H_A$  und der Alternativhypothese  $H_0$  gibt es zwei Hypothesen, die aber selten genutzt werden.
- B** ☐ Es gibt ein statistisches Hypothesenpaar mit der Hypothese für und gegen die wissenschaftliche Fragestellung. Die Hypothesen werden  $H_{pro}$  und  $H_{contra}$  bezeichnet.
- C** ☐ Ein statistisches Hypothesenpaar gibt es. Zum einen die Nullhypothese und zum anderen die Alternativhypothese. Es ist aber nur notwendig die Alternative anzugeben, da die Nullhypothese nicht beim Testen benötigt wird.
- D** ☐ Ein statistisches Hypothesenpaar gibt es. Zum einen die Nullhypothese  $H_0$  und zum anderen die Alternativhypothese  $H_A$  oder  $H_1$ .
- E** ☐ Die Hypothesen  $H_0$  und  $H_A$  sind rein prosarischer Natur und bilden keinen mathematischen Hintergrund ab. In der Statistik wird die wissenschaftliche Fragestellung getestet. Daher stehen auch die verständlichen Hypothesen im Mittelpunkt der biologischen Interpretation.

## 23. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

$H_0$  beibehalten obwohl die  $H_0$  falsch ist

- A** ☐ Dem  $\beta$ -Fehler mit der Analogie eines brennenden Hauses: *Fire without alarm*.
- B** ☐ *Alarm without fire*, dem  $\alpha$ -Fehler in der Analogie eines Rauchmelders.
- C** ☐ *Fire without alarm*, dem  $\beta$ -Fehler als Analogie von Rauch im Haus.
- D** ☐ In die Analogie eines Rauchmelders: *Fire without alarm*, dem  $\beta$ -Fehler.
- E** ☐ In die Analogie eines Rauchmelders: *Alarm without fire police*, dem  $\alpha$ -Fehler.

## 24. Aufgabe

(2 Punkte)

Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ Einem Konfidenzintervall. Das Konfidenzintervall bringt durch eine Visualisierung und zwei Intervallgrenzen die Möglichkeit mit, eine Relevanzschwelle neben der definierten Signifikanzschwelle zu definieren.
- B** ☐ 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 OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- D** ☐ Die Teststatistik. Durch den Vergleich von  $T_c$  zu  $T_k$  ist es möglich die  $H_0$  abzulehnen. Die Relevanz ergibt sich aus der Fläche rechts vom dem  $T_c$ -Wert.
- E** ☐ 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.

## 25. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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** ☐ Mehr Fallzahl in Datensätzen bedeutet mehr signifikante Ergebnisse, da in mehr Daten auch mehr Informationen beinhaltet sind. Deshalb lohnen sich riesige Datensätze, die durch die vielen signifikanten Ergebnisse auch eine Menge an relevanten Erkenntnissen liefern.
- B** ☐ Aktuell werden zu grosse Datensätze für die gängige Statistik gemessen. Daher wendet man maschinelle Lernverfahren für kausale Modelle an. Hier ist die Relevanz gleich Signifikanz.
- C** ☐ Eine erhöhte Fallzahl führt automatisch zu mehr signifikanten Ergebnissen auch wenn der Effekt klein ist und damit nicht relevant. Dadurch sind die Informationen zur Signifikanz in riesigen Datensätzen schwer zu verwerten, da fast alle Vergleiche signifikant sind.
- D** ☐ Riesige Datensatz haben mehr Fallzahl was zur  $\alpha$ -Inflation führt. Durch eine Adjustierung kann dem Problem entgegengewirkt werden.
- E** ☐ Big Data ist ein Problem der parametrischen Statistik. Parameter lassen sich nur auf kleinen Datensätzen berechnen, da es sich sonst nicht mehr um eine Stichprobe im engen Sinne der Statistik handelt.

## 26. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Im Rahmen Ihrer Abschlussarbeit werten Sie ein Experiment mit Ferkel aus. Es geht um die Leistungssteigerung der Ferkelproduktion. Sie messen jeweils die Gewichtszunahme der Ferkel. Die Ferkel einer Muttersau sind dabei...

- A** ☐ Abhängig von der Stallanlage und des Experiments können die Ferkel abhängig oder unabhängig sein. Allgemein gilt, dass Ferkel von unterschiedlichen Sauen näher miteinander verwandt sind als Ferkel von gleichen Sauen. Das Fisher-Axiom.
- B** ☐ Untereinander unabhängig. Die Ferkel sind eigenständig und benötigen keine zusätzliche Behandlung.
- C** ☐ Untereinander stark korreliert. Die Ferkel sind von einer Mutter und somit miteinander korreliert. Dies wird in der Statistik jedoch meist nicht modelliert.
- D** ☐ Untereinander abhängig, wenn die Mütter ebenfalls miteinander verwandt sind. Erst die Abhängigkeit 2. Grades wird in der Statistik modelliert.
- E** ☐ 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.

## 27. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Sie rechnen einen statistischen Test und wollen anhand des 95%-Konfidenzintervalls eine Entscheidung gegen die Nullhypothese treffen. Welche Aussage ist richtig?

- A ☐ Wir betrachten das Ganze Intervall des 95%-Konfidenzintervalls. Liegt die Null mit in dem Intervall, dann kann die Nullhypothese abgelehnt werden.
- B ☐ Ist  $Pr(D|H_0)$  kleiner als das Signifikanzniveau  $\alpha$  gleich 5% dann wird die Nullhypothese  $H_0$  abgelehnt.
- C ☐ Liegt das berechnete Intervall über dem kritische Wert  $T_{\alpha=5\%}$ , dann kann die Nullhypothese nicht abgelehnt werden.
- D ☐ 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.
- E ☐ Der kritische Wert  $T_{\alpha=5\%}$  ist tabelarisch festgelegt und gegeben. Ist  $T_D$  größer als der kritische Wert, kann die Nullhypothese abgelehnt werden

## 28. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Welche Aussage über die Power ist richtig?

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

## Statistische Tests für Gruppenvergleiche

## 29. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik

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

- A ☐ Der t-Test vergleicht die Mittelwerte von zwei Gruppen unter der strikten Annahme von Varianzhomogenität. Sollte keine Varianzhomogenität vorliegen, so gibt es keine Möglichkeit den t-Test in einer Variante anzuwenden.
- B ☐ Der t-Test testet generell zu einem erhöhten  $\alpha$ -Niveau von 20%.
- C ☐ Der t-Test ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern
- D ☐ Der t-Test vergleicht die Varianzen von mindestens zwei oder mehr Gruppen
- E ☐ Der t-Test vergleicht zwei Gruppen indem die Mittelwerte miteinander verglichen werden.

## 30. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A ☐ Der gepaarte t-Test wird genutzt, wenn die Differenzen der Beobachtungen verbunden sind und wir dadurch die Unabhängigkeit nicht mehr vorliegen haben.
- 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.

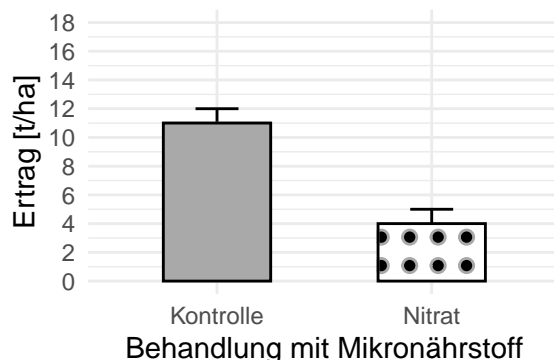
- C** ☐ Wenn die Beobachtungen unabhängig voneinander sind, rechnen wir einen gepaarten t-Test. Messen wir wiederholt an dem gleichen Tier oder Pflanze dann bilden wir das Produkt zwischen den zwei Messpunkten.
- D** ☐ Der gepaarte t-Test wird gerechnet, wenn die Beobachtungen nicht unabhängig voneinander sind. Wir messen wiederholt an dem gleichen Probanden oder Tier oder Pflanze. Wir bilden die Differenzen um den gepaarten t-Test rechnen zu können.
- E** ☐ Der gepaarte t-Test wird gerechnet, wenn die Beobachtungen abhängig voneinander sind. Wir messen jede Beobachtung nur einmal und berechnen dann die Differenz zu dem Mittel der anderen Beobachtungen.

### 31. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

In einer Studie zur Bewertung der Wirkung des Mikronährstoff Nitrat auf den Ertrag in t/ha von Mais im Vergleich zu einer Kontrolle entstand folgende Abbildung. Der Versuch wurde in 12 Parzellen pro Gruppe durchgeführt. Welche Aussage ist im Bezug auf einen t-Test ist richtig?



- A** ☐ Der Effekt und die Signifikanz lassen sich nicht aus Barplots abschätzen. Höchstens der Effekt als relativer Unterschied zwischen der Höhe der Barplots. Standard ist der mediane Unterschied aus Boxplots.
- B** ☐ Es liegt kein signifikanter Unterschied vor. Der Effekt liegt bei -7.
- C** ☐ Die Barplots deuten auf einen signifikanten Unterschied. Der Effekt liegt vermutlich bei -7 unter einer groben Abschätzung. Wir müssen aber eine ANOVA rechnen um den Effekt wirklich bestimmen zu können.
- D** ☐ Die Barplots deuten auf keinen signifikanten Unterschied. Der Effekt liegt vermutlich bei -7. Wir müssen aber einen Posthoc-Test rechnen um den Effekt wirklich bestimmen zu können.
- E** ☐ Der Test deutet auf einen signifikanten Unterschied hin. Der Effekt liegt vermutlich bei -7.

### 32. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Sie führen paarweise t-Tests für alle Vergleiche der verschiedenen Rapssorten in Ihrem Experiment durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der  $\alpha$ -Schwelle. Ihr Experiment beinhaltet drei Rapssorten und eine ANOVA ergibt  $p = 0.049$  für den Ertrag. 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** ☐ 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.
- B** ☐ Hier kommt der Effekt der steigenden Fallzahl auf die Anzahl an signifikante Ergebnisse zu tragen. Da die ANOVA auf mehr Fallzahl testet als die einzelnen paarweisen t-Tests, kann die ANOVA leichter einen signifikanten Unterscheid nachweisen. Die p-Werte sind immer etwas kleiner als bei den t-Tests.
- C** ☐ Die ANOVA testet auf der gesamten Fallzahl. Es wäre besser die ANOVA auf der gleichen Fallzahl wie die einzelnen t-Tests zu rechnen.
- D** ☐ Die adjustierten p-Werte deuten in die richtige Richtung. Zusammen mit den nicht signifikanten rohen p-Werten ist von einem Fehler in der ANOVA auszugehen.
- E** ☐ Das Beispiel kann so nicht auftreten, da die ANOVA und die t-Tests algorithmisch miteinander verschränkt sind.

## ANOVA

### 33. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Sie führen einen Versuch mit einer Behandlung und drei Faktorleveln durch. Danach rechnen Sie eine einfaktorielle ANOVA und es ergibt sich ein  $\eta^2 = 0.31$ . Welche Aussage ist richtig?

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

### 34. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

- A** ☐ 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.
- B** ☐ Es werden 28% der Varianz durch den Versuch erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der durch Fehler in der Versuchsdurchführung entsteht.
- C** ☐ Es werden 72% der Varianz durch die Behandlung erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den unterschiedlichen Behandlungsbedingungen nicht erklärt wird.
- D** ☐ 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%.
- E** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird. Daher werden 28% der Varianz erklärt.

### 35. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ 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.
- B** ☐ Die ANOVA berechnet die F-Statistik aus den SS Behandlung geteilt durch die SS Fehler.
- C** ☐ Wenn die F-Statistik kleiner als der kritische Wert ist kann die Nullhypothese nicht abgelehnt werden. Die F-Statistik ist der Quotient der MS der Behandlung durch die MS des Fehlers.
- D** ☐ Wenn die F-Statistik höher ist als der kritische Wert kann die Nullhypothese nicht abgelehnt werden. Die F-Statistik ist die Differenz der MS der Behandlung durch die MS des Fehlers.
- E** ☐ Die ANOVA berechnet die 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.

### 36. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

Die ANOVA ist ein statistisches Verfahren welches häufig in den Auswertungen von Experimenten in den Agrarwissenschaften angewendet wird. Dabei wird die ANOVA als ein erstes statistischen Werkzeug für die Übersicht über die Daten benutzt. Eine ANOVA testet dabei...

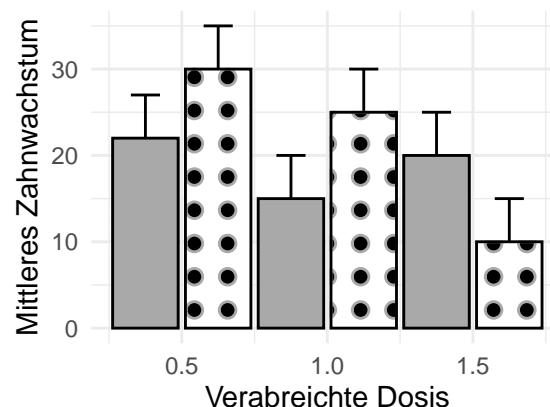
- ☐ A ... den Unterschied zwischen der Varianz in den verschiedenen Behandlungsgruppen und der Varianz in einer der Behandlungsgruppen. Wenn die ANOVA signifikant ist, muss über einen Posthoc-Test nachgedacht werden um den signifikanten Unterschied in einer der Gruppen exakt zu bestimmen.
- ☐ B ... den Unterschied zwischen der Varianz über alle Behandlungsgruppen und der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, muss ein Posthoc-Test angeschlossen werden.
- ☐ C ... den Unterschied zwischen der F-Statistik anhand der Varianz der Gruppen. Wenn die F-Statistik exakt 0 ist, kann die Nullhypothese abgelehnt werden.
- ☐ D ... den Unterschied zwischen der Varianz ausgelöst durch alle Behandlungsgruppen und der Varianz aus globalen Behandlungsgruppen der Kontrollen. Wenn die ANOVA nicht signifikant ist, muss ein Posthoc-Test ausgeschlossen werden.
- ☐ E ... den Unterschied zwischen der Mittelwerte und der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, ist bekannt welcher Vergleich konkret unterschiedlich ist.

### 37. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Die folgende Abbildung enthält die Daten aus einer Studie zur Bewertung der Wirkung von Vitamin E auf das Zahnwachstum bei Schweine. Der Versuch wurde an 56 Tieren durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0,5, 1 und 1,5 mg/Tag) über eine von zwei Verabreichungsmethoden erhielt. Welche Aussage ist richtig im Bezug auf eine zweifaktorielle ANOVA?



- ☐ A Eine Korrelation liegt vor ( $p \leq 0.05$ ).
- ☐ B Keine Interaktion liegt vor ( $p \leq 0.05$ ).
- ☐ C Das Bestimmtheitsmaß  $R^2$  ist klein.
- ☐ D Eine positive Interaktion liegt vor ( $p \leq -0.5$ )
- ☐ E Eine mittlere bis starke Interaktion liegt vor ( $p \leq 0.05$ )

### Multiple Gruppenvergleiche

### 38. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Sie haben folgende unadjustierten p-Werte gegeben: 0.21, 0.34, 0.42 und 0.89. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- ☐ A Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.84, 1.36, 1.68 und 3.56. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.



- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0525, 0.085, 0.105 und 0.2225. 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 0.84, 1, 1 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0525, 0.085, 0.105 und 0.2225. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.84, 1, 1 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.

### 39. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Auf wissenschaftlichen Postern finden Sie unter Abbildungen häufig die Abkürzung *CLD*. Für welchen statistischen Fachbegriff steht die Abkürzung und wie interpretieren Sie ein *CLD*?

- A** ☐ 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.
- B** ☐ Compact letter detection. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt.
- C** ☐ 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.
- D** ☐ Compact letter display. Gleiche Buchstaben bedeuten, dass sich die Behandlungen unterscheiden. Daher ist das CLD sehr unintuitiv. Es wäre besser, wenn gleiche Buchstaben Gleichheit anzeigen würden. Dies ist aber leider in der statistischen Testtheorie nicht möglich.
- E** ☐ Contrast letter display. Unterschiede in den Behandlungen werden durch den gleichen Buchstaben oder Symbol dargestellt. Die Interpretation des CLD führt häufig in die Irre.

### 40. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ Das R Paket {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.
- B** ☐ Das R Paket {ggplot}. Wir erhalten hier sofort eine Visualisierung der Daten. Anhand der Visualisierung lässt sich eine explorative Datenanalyse durchführen, die gleichwertig zu einem Posthoc-Test ist.
- C** ☐ 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.
- D** ☐ 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.
- E** ☐ Das R Paket {emmeans} erlaubt die Durchführung eines multiplen Gruppenvergleichs. Aus einem {emmeans} Objekt lässt sich recht einfach das CLD erstellen und so über Barplots eine schnelle Interpretation der statistischen Auswertung durchführen.

## 41. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

In den Humanwissenschaften werden multiple Vergleiche häufig anders behandelt als in den Agrarwissenschaften. In beiden Bereichen tritt jedoch das gleiche Phänomen bei multiplen Testen auf. Wie muss mit dem Phänomen umgegangen werden und wie ist es benannt?

- ☐ A Beim multiplen Testen kann es zu einer  $\alpha$ -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern weit darunter. Daher müssen die p-Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der p-Werte nach Welch das bekannteste Verfahren ist.
- ☐ B Die Adjustierung der p-Werte nach Bonferroni erlaubt es gegen die  $\beta$ -Inflation vorzugehen, die häufig beim multiplen Testen auftritt. Das globale Powerniveau liegt nicht mehr bei 80% sondern sehr viel niedriger.
- ☐ C 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.
- ☐ D 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.
- ☐ E 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.

## 42. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- ☐ A Beim multiplen Testen kann es zu einer  $\Delta$ -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 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.
- ☐ C Beim multiplen Testen kann es zu einer Effektüberschätzung ( $\Delta$ -Inflation) kommen. Daher müssen die Effekte angepasst werden. Dies geschieht nicht händisch sondern intern in den angewendeten Algorithmen.
- ☐ D Beim multiplen Testen muss der Effekt, wie der Mittelwertsunterschied  $\Delta$  aus einem t-Test, nicht adjustiert werden.
- ☐ E Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  nicht adjustiert werden. Bei einem Effekt im multiplen Testen handelt es sich um eine Wahrscheinlichkeit für das Auftreten der Nullhypothese.

## Lineare Regression & Korrelation

## 43. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

In Ihrer Abschlussarbeit wollen Sie ein prädiktives Modell rechnen. Jetzt stellt sich die Frage, was diese Entscheidung für Ihre Auswertung bedeutet. Welche Aussage ist 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 Wir modellieren den Zusammenhang zwischen X und Y wenn ein prädiktives Modell berechnet wird. Dabei kann der gesamte Datensatz genutzt werden. Eine Aufteilung wie in einem prädiktiven Modell ist nicht notwendig.
- ☐ C Wenn ein prädiktives Modell gerechnet werden soll, dann muss zum einen ein Trainingsdatensatz sowie ein Testdatensatz definiert werden. Dabei ist der Trainingsdatensatz meist 2/3 und der Testdatensatz 1/3 der Fallzahl groß. Der Testdatensatz dient zur Validierung.

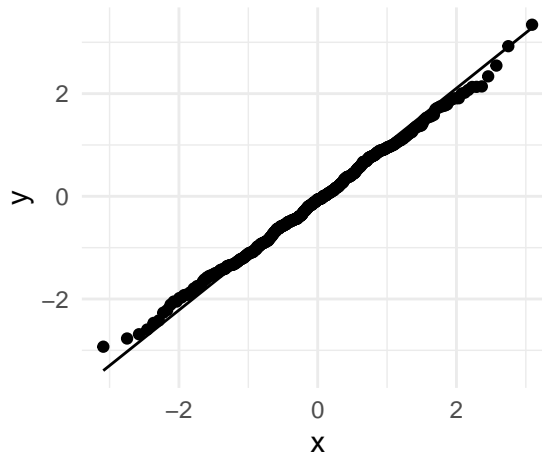
- D** ☐ Ein prädiktives Modell schliesst grundsätzlich lineare Modell aus. Es muss ein Graph gefunden werden, der alle Punkte beinhaltet. Erst dann kann das  $R^2$  berechnet werden.
- E** ☐ Ein prädiktives Modell benötigt mindestens eine Fallzahl von über 100 Beobachtungen und darf keine fehlenden Werte beinhalten. Die Varianzkomponenten müssen homogen sein.

#### 44. Aufgabe

(2 Punkte)

Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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



- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden und Korrelation ist negativ.
- B** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- C** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden.
- D** ☐ Wir betrachten insbesondere die beiden Enden der Gerade. Der Rest ist mehr oder minder egal, dann ist die Annahme an die Normalverteilung der Residuen erfüllt.
- E** ☐ Wir betrachten die Gerade. Wenn die Punkte einigermaßen gleichmäßig um die Gerade verteilt liegen, dann gehen wir von normalverteilten Residuen aus. Dies ist hier nicht der Fall. Wir haben keine normalverteilten Residuen vorliegen.

#### 45. Aufgabe

(2 Punkte)

Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

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

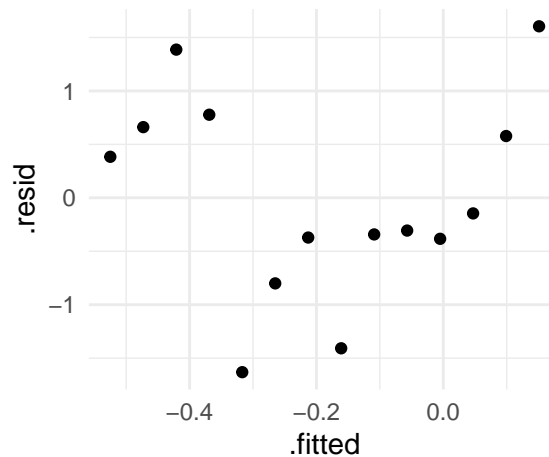
- A** ☐ Der Korrelationskoeffizienten  $\rho$  ist eine veraltete Darstellungsform von Effekten in der linearen Regression und wird wie das  $\eta^2$  aus der ANOVA interpretiert. Der Korrelationskoeffizienten  $\rho$  beschreibt den Anteil an erklärter Varianz durch die Regression.
- B** ☐ Korrelationskoeffizienten  $\rho$  liegt zwischen 0 und 1. Darüber hinaus ist der Korrelationskoeffizienten  $\rho$  einheitslos und kann als Standardisierung 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  $\rho$  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.
- E** ☐ Der Korrelationskoeffizienten  $\rho$  zeigt keinen Zusammenhang zwischen zwei Variablen  $x$  und  $y$  bei einem Wert von 0. Einen maximalen negativen Zusammenhang bei -1 und somit auch einen maximalen positiven Zusammenhang bei 1. Korrelationskoeffizienten  $\rho$  ist einheitslos.

## 46. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Sie rechnen eine lineare Regression und erhalten folgende Abbildung der Residuen (.resid). Welche Aussage ist richtig?



- ☐ A Wenn wir die Nulllinie betrachten so müssen die Punkte gleichmäßig unter der Nulllinie liegen. Unser Modell erfüllt somit nicht die Annahme von normalverteilten Residuen mit einem Mittelwert von  $> 0$  und einer Streuung von  $s$ .
- ☐ B Die Annahme der normalverteilten Residuen ist nicht erfüllt. Vereinzelte Punkte liegen oberhalb bzw. unterhalb der Geraden um die 0 Linie weiter entfernt. Ein klares Muster ist zu erkennen.
- ☐ C Wir betrachten die Nulllinie und alle Punkte sollten ohne Muster gleichmäßig um die Nulllinie liegen. Da dies der Fall ist, gehen wir von keinen Ausreißern aus.
- ☐ D Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Diagonalen. Damit ist das Modell erfolgreich geschätzt worden.
- ☐ E Die Annahme der normalverteilten Residuen ist nicht erfüllt. Ein klares Muster ist zu erkennen und/oder einige Outlier sind zu beobachten.

## 47. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Nach einer simplen linearen Regression zur Untersuchung vom Einfluss der  $PO_2$ -Konzentration in  $[\mu g]$  im Wasser auf das Trockengewicht von Erbsen 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 Die Einheit der  $PO_2$ -Konzentration ist zu klein gewählt. Dadurch sehen wir den sehr kleinen  $p$ -Wert. Der  $p$ -Wert und die Einheit von der  $PO_2$ -Konzentration hängen antiproportional zusammen.
- ☐ B Wenn der Effekt  $\beta_{PO_2}$  sehr klein 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  $y$ . Daher ist hier mit einer anderen Einheit in den Daten zu rechnen, so dass wir hier einen besser formatierten Effekt sehen. Der  $p$ -Wert stammt aus einer einheitslosen Teststatistik.
- ☐ C 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.
- ☐ 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 Fallzahl ist zu hoch angesetzt. Je höher die Fallzahl ist, desto kleiner ist die Teststatistik und damit ist dann auch der  $p$ -Wert sehr klein. Es sollte über eine Reduzierung der Fallzahl nachgedacht werden. Dann sollte der Effekt zum  $p$ -Wert passen.

## 48. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

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

## 49. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

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

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

## 50. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Schafe durch. Bei 3 Tieren finden Sie eine Erkrankung der Klauen vor und 12 Tiere sind gesund. Welche Aussage über den Effektschätzer Risk ratio ist richtig?

- ☐ A Es ergibt sich ein Risk ratio von 0.25, da es sich um ein Anteil handelt.
- ☐ B Der Anteil der Gesunden wird berechnet. Da es sich um ein Anteil handelt ergibt sich ein Risk ratio von 0.2.
- ☐ C Das Verhältnis von Chancen Risk ratio ergibt ein Chancenverhältnis von 0.25.
- ☐ D Da es sich um ein Chancenverhältnis handelt ergibt sich ein Risk ratio von 5.
- ☐ E Das Verhältnis der Anteile Risk ratio ergibt ein Anteilsverhältnis von 0.2. Wir sind am Anteil der Kranken interessiert.

# Teil I.

## Programmieren in R

### 51. Aufgabe

(9 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik



**Grundlegende Kenntnisse der Programmierung in R** 'Hm. R ist eigentlich gar nicht so schwer, wenn man die Grundlagen kann.', meint Yuki ganz zuversichtlich. Nur leider kennt er sich überhaupt nicht mit R aus! Das heißt, Sie müssen hier einmal Rede und Antwort stehen und helfen.

Yuki: Jetzt sehe ich wieder diese Tilde (~) in R. Wo brauchen wir diese denn nochmal? (1 Punkt)

Sie antworten:

Yuki: Der Zuweisungs-Operator wird sehr häufig genutzt. Wie sieht der aus und wie funktioniert der an einem Beispiel? (1 Punkt)

Sie antworten:

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

Sie antworten:

Yuki: Warum gibt es eigentlich Objekte, Wörter und Funktionen in R? Wie unterscheiden sich diese überhaupt? (1 Punkt)

Sie antworten:

Yuki: Der Pipe-Operator wird sehr häufig genutzt. Wie sieht der aus und wie funktioniert der an einem Beispiel? (1 Punkt)

Sie antworten:

Yuki: Wie war nochmal der Name der Funktion in dem wir in R Daten intern abspeichern? Was waren da nochmal die Vorteile? (1 Punkt)

Sie antworten:

Yuki: Teilweise brauche ich das Konzept des Faktors in R. Was ist ein Faktor? (1 Punkt)

Sie antworten:

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

Sie antworten:

Yuki: Warum nutzen wir eigentlich dieses c() in R? (1 Punkt)

Sie antworten:

## 52. Aufgabe

(9 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Fortgeschrittene Kenntnisse der Programmierung in R** 'Hm...am Ende ist dann **R** eigentlich gar nicht so schwer, wenn ich Hilfe habe.', meint Jessica stolz und lacht Sie an. Nur leider kennt sie sich überhaupt nicht mit **R** aus! Das heißt, Sie müssen hier einmal Rede und Antwort stehen und helfen. Sonst wird es für Jessica dann in ihrer Abschlussarbeit nichts mit der Auswertung und Abgabe. Das kann auch keine Lösung für Jessica und Sie sein. Immerhin haben Sie schon so viel gelernt.

Jessica fragt: *Es gibt ja nur ein richtiges Format für die Eingabe eines Datums. Wie lautet das Format?* **(1 Punkt)**  
Sie antworten:

Jessica fragt: *Oh, wie baue ich mir nochmal die Daten in R? Wie heißt das Dateiformat? Gerne einmal mit Beispiel!* **(1 Punkt)**  
Sie antworten:

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

Jessica fragt: *Ich will das R Paket `{ggplot}` nutzen, da war so eine Analogie an die ich mich nicht erinnern kann. Was war noch gleich das Prinzip von `{ggplot}`? Wie funktioniert `{ggplot}` konzeptionell?* **(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:

Jessica fragt: *Hm... wenn ich in R Daten mit Gruppen eingelesen habe, was muss ich dann nochmal als erstes machen? Und warum?* **(2 Punkte)**  
Sie antworten:

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

## Teil II.

# Deskriptive Statistik & Explorative Datenanalyse

### 53. Aufgabe

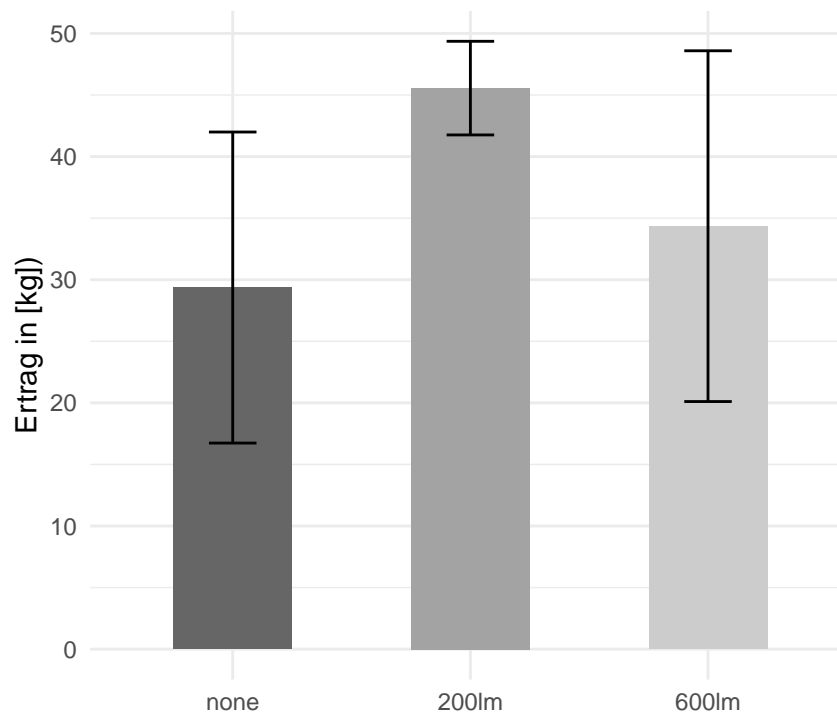
(8 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Zerforschen des Barplots** Nilufar steht vor einem ersten Problem, denn wenn es nach ihrer Betreuer geht, soll sie in einem Feldexperiment Erbsen auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Hip Hop. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Nilufar denkt gerne über Hip Hop nach. Das heißt erstmal überlegen für Nilufar. Aus den Boxen wummert Deichkind und ihr Mund ist verklebt von Takis Blue Heat. 'Herrlich', denkt Nilufar. Die Behandlung werden verschiedene Lichtstufen (*none*, *200lm* und *600lm*) sein. In ihrer Exceldatei wird sie den Outcome (Y) *Ertrag* als *yield* aufnehmen. Vorab soll Nilufar aber einmal die folgenden Barplots ihrer Betreuer nachbauen, damit sie den R Code schonmal für später vorliegen hat. Damit geht das Problem schon los. Wenn die Erwartung nicht wäre, ja dann wäre wohl vieles möglich für Nilufar! Aber so..



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

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



## 54. Aufgabe

(8 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Visualisierung des Barplots** Steffen steht vor einem ersten Problem, denn wenn es nach seiner Betreuerin geht, soll er in einem einem Gewächshausexperiment Maiss auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Am Ende dann doch besser Klemmbausteine. Wunderbar. Eine echte Ablenkung für Steffen. Die Behandlung waren verschiedene Lichtstufen (*none*, *200lm* und *600lm*). In seiner Exceldatei hat er den Messwert (Y) *Ertrag* als *yield* aufgenommen. Nun soll Steffen die Daten einmal als Barplots in einer Präsentation visualisieren, damit seiner Betreuerin wieder klar wird, was er eigentlich nochmal gemacht hat und was für ein Ergebnis in einem statistischen Test zu erwarten wäre. Wäre da nicht noch etwas. Steffen und die Romantik, eine unendliche Geschichte mit kniffligen Wendungen. Aber egal. Um zu Ringen geht Steffen dann später nochmal raus. Echte Entspannung.

treatment	yield
200lm	41.5
600lm	31.3
none	40.7
none	25.6
none	44.6
200lm	42.3
600lm	21.7
600lm	35.4
none	32.1
200lm	42.3
600lm	34.5
200lm	44.0
none	23.3
600lm	37.1
200lm	45.6

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

## 55. Aufgabe

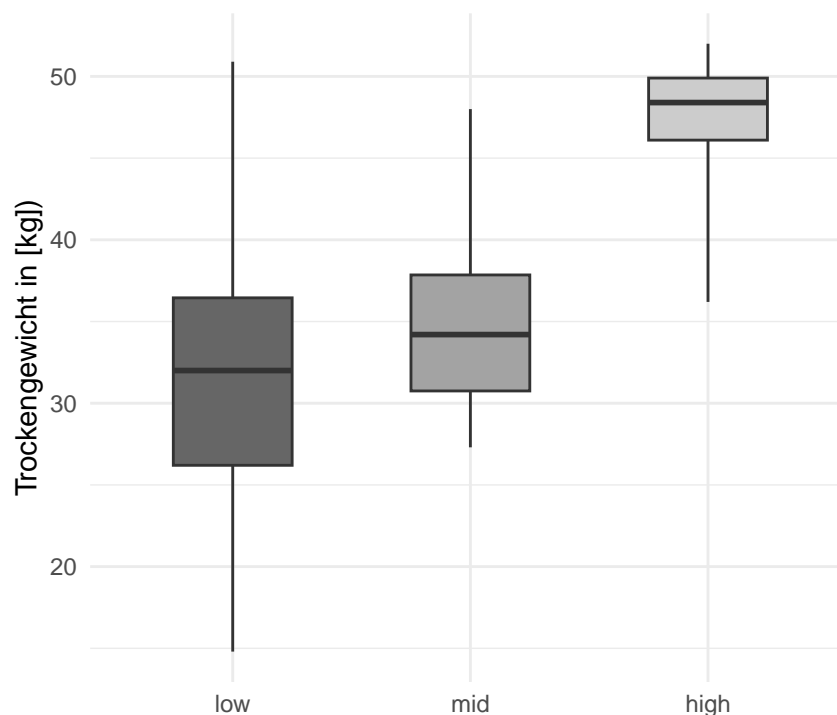
(9 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Zerforschen des Boxplots** Tina steht vor einem ersten Problem, denn wenn es nach ihrer Betreuerin geht, soll sie in einem Versuch in einer Klimakammer Maiss auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Am Ende dann doch besser Astronomie. Wunderbar. Eine echte Ablenkung für Tina. Das heißt erstmal überlegen für Tina. Tina schmeißt noch eine Handvoll Katjes in ihren Rachen. Im Hintergrund klirrt leise der Spiegel zum Sound von Tocotronic. Die Behandlung werden verschiedene Bewässerungstypen (*low*, *mid* und *high*) sein. In ihrer Exceldatei wird sie den Outcome (Y) *Trockengewicht* als *drymatter* aufnehmen. Vorab soll Tina aber einmal die folgenden Boxplots ihrer Betreuerin nachbauen, damit sie den R Code schonmal für später vorliegen hat. Damit geht das Problem schon los. Tina und die Wut, eine unendliche Geschichte mit kniffligen Wendungen.



Leider kennt sich Tina mit der Erstellung von Boxplots in R 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 R üblichen Format! **(2 Punkte)**
4. Kann Tina einen Unterschied zwischen den Behandlungen erwarten? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 56. Aufgabe

(9 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Visualisierung des Boxplots** Boxplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Steffen nicht gemerkt, welche statistischen Maßzahlen für einen Boxplot erhoben werden müssen. Besser wäre was anderes gewesen. Klemmbausteine. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Steffen denkt gerne über Klemmbausteine nach. Das ist in soweit doof, da nach seinem Betreuer nun Boxplots aus seinen Daten gebaut werden sollen, bevor es mit dem statistischen Testen weitergeht. Anhand von Boxplots lässt sich eine Aussage über die Normalverteilung von  $Y$  treffen. Die Behandlung für Erbsen waren verschiedene Genotypen ( $AA$  und  $BB$ ). Erfasst wurde von Steffen als Messwert ( $Y$ ) *Trockengewicht*. Steffen hat dann *drymatter* in seiner Exceldatei eingetragen. Aber nur in passender Atmosphäre! Schon dutzende Male gesehen: Harry Potter. Aber immer noch großartig zusammen mit Oreos.

treatment	drymatter
AA	34.9
AA	35.4
BB	40.1
BB	28.1
BB	34.1
AA	41.3
BB	18.4
BB	20.3
AA	25.3
AA	34.5
AA	20.1
AA	21.1
AA	39.6
BB	27.4
AA	25.7
BB	33.6
BB	39.5
BB	35.0

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

## 57. Aufgabe

(8 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Visualisierung des Histogramm für kategoriale Daten** Aus den Boxen wummert Tocotronic und ihr Mund ist verklebt von Katjes. 'Herrlich', denkt Tina. Tina betrachtet die folgenden Daten nach einem Stallexperiment mit Schweinen. In dem Experiment wurden die Anzahl an weißen Blutkörperchen gezählt. Nach der Meinung ihrem Betreuer muss als erstes geschaut werden, wie diese verteilt sind. Also welcher statistischen Verteilung die Anzahl an weißen Blutkörperchen folgen. Dazu soll Tina ein Histogramm verwenden. Dann hätte man auch einen guten Überblick über das Outcome (Y). Es wäre einfacher, wenn da nicht noch was wäre. Tina und die Wut, eine unendliche Geschichte mit kniffligen Wendungen. Tina nickt im Takt von Tocotronic und bemerkt dabei gar nicht was die Spinne schon wieder anstellt.

Die Anzahl an weißen Blutkörperchen: 3, 9, 4, 2, 2, 8, 3, 4, 4, 1, 6, 6, 3, 0, 5, 4, 1, 6, 5, 6, 5, 7, 2, 2, 5, 4, 2, 3, 1, 4, 4, 5, 3, 5, 5, 1

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

1. Zeichnen Sie ein Histogramm um die Verteilung der Daten zu visualisieren! **(3 Punkte)**
2. 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* gleich oder mehr als die Anzahl 6 zu beobachten! **(1 Punkt)**
5. Berechnen Sie aus den Daten die *Chance* gleich oder mehr als die Anzahl 6 zu beobachten! **(1 Punkt)**

## 58. Aufgabe

(8 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Visualisierung des Histogramm für kontinuierliche Daten** 'Hm...', Reese's Peanut Butter Cups und London Grammar. Das ist und bleibt die beste Kombination zum Nachdenken für Yuki. Yuki betrachtet die folgenden Daten nach einem Versuch in einer Klimakammer mit Spargel. In dem Experiment wurden die mittleren Knötchen gezählt. Nach der Meinung seiner Betreuerin muss als erstes geschaut werden, wie diese verteilt sind. Also welcher statistischen Verteilung die mittleren Knötchen folgen. Dazu soll Yuki ein Histogramm verwenden. Dann hätte man auch einen guten Überblick über das Outcome (Y). Es wäre einfacher, wenn da nicht noch was wäre. Eine echte Herausforderung für ihn war schon immer die Faulheit gewesen. Ein leidiges Lied. Wenn London Grammar ertönt, dann sucht das Minischwein schleunigst Schutz unter dem Sofa. Yuki schüttelt den Kopf.

Die mittleren Knötchen: 9.2, 9, 10.1, 10, 9, 7, 8.4, 10.4, 8.7, 12.8, 8.8, 15.1, 11.6, 11.5, 12.5, 11.9, 10.1, 10.1, 9.7, 6.9, 11.2, 8.7, 9.4, 7.2, 11.7, 10.7, 7.6, 11.3

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

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

## 59. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Visualisierung des Scatterplots** Wenn es nach Jonas ginge, wäre er schon längst fertig mit seiner Hausarbeit. Geht es aber nicht. Jonas schmeißt noch eine Handvoll Snickers in seinen Rachen. Im Hintergrund klirrt leise der Spiegel zum Sound von Iron Maiden. In seiner Hausarbeit hatte er ein Freilandversuch in der Uckermark durchgeführt. Nach der Meinung seiner Betreuerin sieht das jedoch etwas anders aus. Jetzt soll er doch noch eine explorative Datenanalyse für den Zusammenhang zwischen durchschnittlichen Niederschlag [ml/w] und Trockengewicht [kg/ha] in Lauch durchführen. Wie nervig! Jonas und die Erschöpfung, eine unendliche Geschichte mit kniffligen Wendungen. Da zwei kontinuierliche Variablen vorliegen, geht die explorative Datenanalyse leider nicht mit Boxplots oder Barplots. Dann was anderes. Wenn Mission Impossible läuft, dann ist das Meerschweinchen nicht mehr da. Aber jetzt braucht er mal Entspannung!

Durchschnittlichen Niederschlag [ml/w]	Trockengewicht [kg/ha]
24.7	25.6
25.2	27.4
22.9	24.1
23.3	28.3
22.2	21.6
24.5	28.6
26.7	31.5
28.7	27.3
25.1	27.1
24.2	26.9
27.3	29.5
24.9	29.0

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

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

## 60. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Visualisierung des Mosaicplots** Zwei kategoriale Variablen darzustellen ist nicht so einfach. Jessica hatte erst über einen Mittelwert nachgedacht, dann aber die Idee verworfen. Wäre da nicht noch was anderes. Jessica und der Mangel, eine unendliche Geschichte mit kniffligen Wendungen. Dabei hatte sie sich in einen Leistungssteigerungsversuch im Emsland zum einen die Behandlung Außenklimakontakt [ja/nein] und zum anderen die Messung Fettgehalt erreicht [ja/nein] im Kontext von Schweinen angeschaut. Jetzt möchte ihre Betreuerin erstmal die langen Tabellen mit ja/nein in einer explorativen Datenanalyse zusammengefasst bekommen. Sonst geht es bei ihrem Projektbericht nicht weiter. Was super nervig ist. Einfach mal raus um Rad zu fahren. Ohne Ziel und Uhr. Das ist was für Jessica.

Außenklimakontakt	Fettgehalt erreicht
nein	nein
ja	nein
ja	nein
nein	nein
nein	nein
nein	nein
nein	nein
ja	ja
ja	nein
nein	nein
ja	ja
nein	nein
nein	ja
nein	nein
nein	nein

Außenklimakontakt	Fettgehalt erreicht
ja	nein
ja	nein
nein	nein
ja	ja
ja	nein
nein	nein
ja	nein
ja	nein
nein	nein
nein	nein
nein	nein
ja	nein
ja	nein
ja	nein
ja	nein

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 *ein* Effekt von der Behandlung vorliegen würde, wie würde die Tabelle und die Visualisierung aussehen? **(2 Punkt)**

## 61. Aufgabe

(10 Punkte)

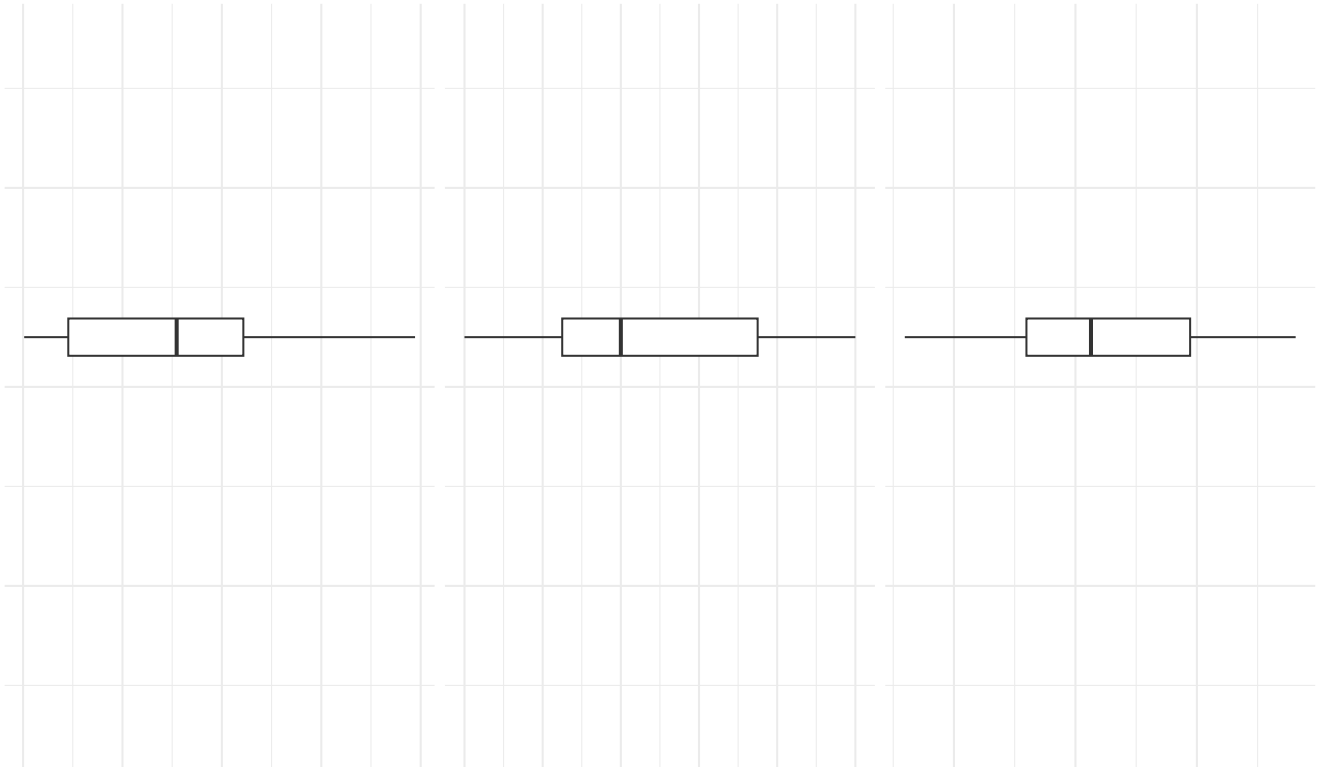
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



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



Jetzt brauchen Jessica und Jonas 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



## Teil III.

# Statistisches Testen & statistische Testtheorie

## 62. Aufgabe

(9 Punkte)



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Grundgesamtheit und experimentelle Stichprobe** 'Grundlage des statistischen Testen ist das Verständnis von der Grundgesamtheit (eng. *population* oder *ground truth*) und der experimentellen Stichprobe (eng. *sample*).', liest Jonas laut aus dem Skript vor. Paula war kurz eingenickt und wird mit einem Stoß geweckt. 'Reiz dich zusammen und iss noch ein paar Snickers das hilft mir immer. Alleine komme ich hier nicht weiter.', tadelt Jonas Paula etwas zu forsch. 'War ne lange Nacht', mault Paula. Beide sollen in ihrer Abschlussarbeit einen statistischen Test interpretieren und versuchen die Grundlagen zu wiederholen. Paula war auf einem Konzert von White Lies.

Leider kennen sich Jonas und Paula 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 ein aussagekräftiges Beispiel!* **(3 Punkte)**
3. Erweitern Sie das Schaubild um die Entstehung von  $Pr(D|H_0)$ ! *Nutzen Sie hierfür als Veranschaulichung ein aussagekräftiges Beispiel!* **(3 Punkte)**

## 63. Aufgabe

(9 Punkte)



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Das Nullritual - Die statistische Testtheorie** 'Alien ist der beste Film, den es gibt.', meint Alex. Steffen entgegnet, 'Ich empfehle ja immer allen Harry Potter.' Die beiden sind im Zoo und diskutieren, ob Pinguine Knie haben. Eigentlich wollten beide nochmal die statistische Testtheorie durchgehen, sind dann aber auf abenteuerlichen Wege im Zoo gelandet. Alex starrt wie hypnotisiert auf einen strullenden Elefanten und stoppt die Zeit.<sup>2</sup> 'Du bist so peinlich.', entfährt es Steffen und schmeißt sich noch ein paar überteuerte Oreos rein.

Leider kennen sich Alex und Steffen 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    $H_0$  falsch    $H_0$  beibehalten    $H_0$  wahr

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

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

3. In der Analogie des Feuermelders, wie lautet der  $\alpha$ -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 Monat 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!



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

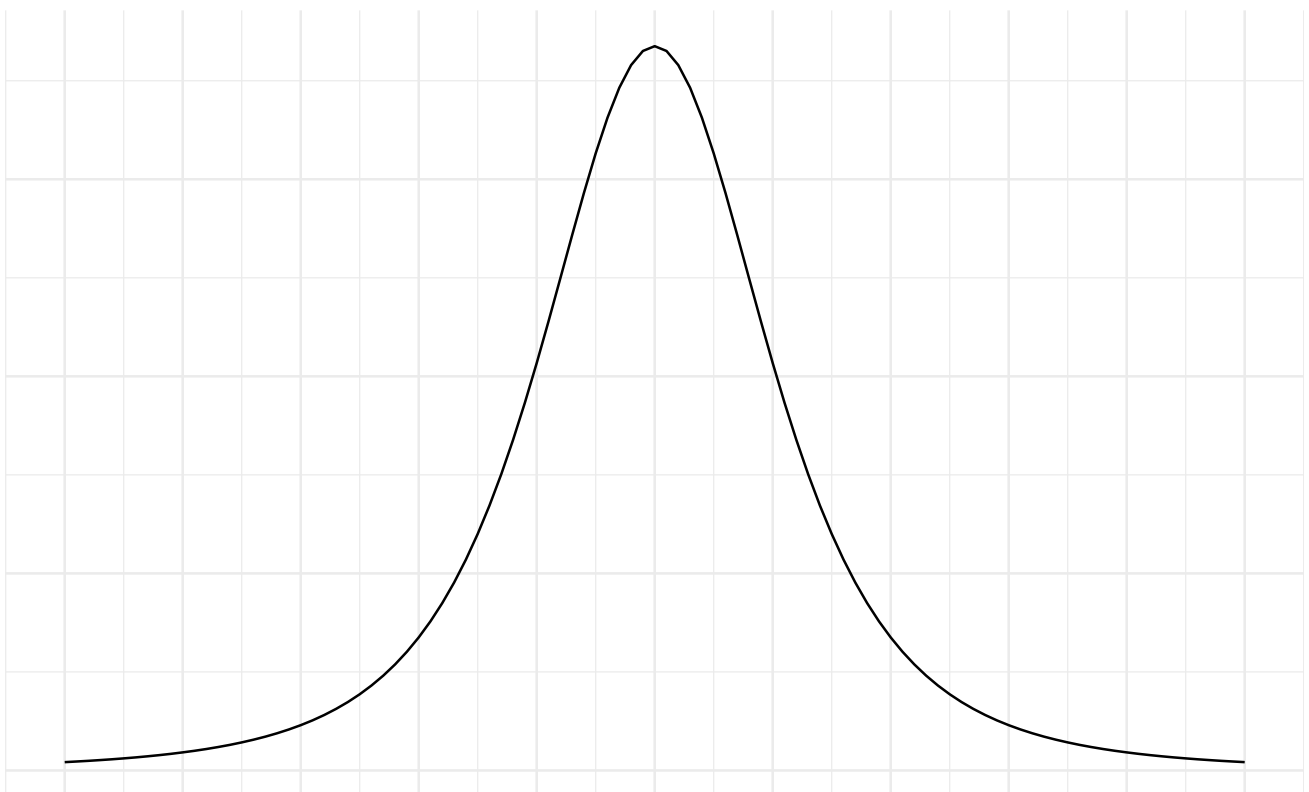


**Visualisierung der Teststatistik  $T_D$  und dem p-Wert** 'Wir sollen die Teststatistik  $T_D$  und dem p-Wert visualisieren, da mit einer Visualisierung vieles verständlicher wird!', ruft Mark um Andrea Berg zu übertönen. 'Ich weiß nicht, was das jetzt helfen soll. Können wir nicht einfach schauen,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

Leider kennen sich Mark und Paula 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 nicht signifikant p-Wert in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**



## 65. Aufgabe

(10 Punkte)



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



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

Leider kennen sich Steffen und Paula 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 Fallzahl  $n$  in der Stichprobe als der Rest der 95% Konfidenzintervalle
  - (b) Ein signifikantes, relevantes 90% Konfidenzintervall.
  - (c) Ein signifikantes, relevantes 95% Konfidenzintervall
  - (d) Ein nicht signifikantes, nicht relevantes 95% Konfidenzintervall
  - (e) Ein 95% Konfidenzintervall mit niedriger Fallzahl  $n$  in der Stichprobe als der Rest 95% der Konfidenzintervalle
  - (f) Ein signifikantes, nicht relevantes 95% Konfidenzintervall



## 66. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Zusammenhang zwischen dem Effekt, der Streuung sowie der Fallzahl** An einem heißen Juliabend haben sich Tina und Yuki zum Lernen verabredet. Eine große Kanne Eistee und Berge von Katjes schmelzen in der Sonne und warten darauf gegessen zu werden. Tina 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  $KI_{1-\alpha}$ ?*

Yuki hebt die Augenbraue. 'Ich bleibe dabei. Wir sollten erstmal Matrix schauen, bis es kühler ist. Den Film habe ich doch extra mitgebracht!' Tina ist der Idee nicht abgeneigt und auch die Spinne kommt auf die Terasse um dabei zu sein.

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

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

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

3. Visualisieren Sie ein 95%-iges Konfidenzintervall im Vergleich zu einem 99%-igen Konfidenzintervall! Begründen Sie Ihre Visualisierung anhand der Formel des Konfidenzintervalls des t-Tests mathematisch! **(3 Punkte)**

## Teil IV.

# 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!



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Berechnung des Student t-Test oder Welch t-Test** Der t-Test. Tina erschauert. Wenn die Wut nicht wäre, ja dann wäre wohl vieles möglich für Tina! Aber so.. Ein mächtiges Werkzeug ist der t-Test in den Händen desjenigen, der ein normalverteiltes Outcome ( $Y$ ) hat. Aber erstmal überhaupt den t-Test rechnen können. Wie sah das Experiment von Tina überhaupt aus? 'Hm...', Katjes und TocoTronic. Das ist und bleibt die beste Kombination zum Nachdenken für Tina. Tina hat ein Freilandversuch mit Kartoffeln durchgeführt um eine neue technische Versuchsanlage zu testen. Bei dem Pilotexperiment mit sehr geringer Fallzahl ( $n_1 = n_2 = 3$ ) wurde die Behandlung Lichtstufen (*none* und *600lm*) an den Kartoffeln getestet und dabei wurde geschaut, ob der Versuch überhaupt technisch klappen könnte. Gemessen hat Tina dann als Messwert Trockengewicht [kg/ha]. Warum der Versuch im Emsland für ihrer Hausarbeit stattfinden musste, ist ihr bis heute ein Rätsel. Egal. Gibt es jetzt einen Zusammenhang zwischen der Behandlung und Trockengewicht [kg/ha]?

treatment	weight
dose	21.8
ctrl	12.2
ctrl	9.9
dose	25.4
ctrl	16.7
dose	36.3

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

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

## 68. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Berechnung des Student t-Test** 'Der t-Test testet einen normalverteilten Endpunkt (Y).', liest Alex laut. Das hilft jetzt auch nur bedingt weiter. Eine echte Herausforderung für ihn war schon immer die Gefälligkeit gewesen. Ein leidiges Lied. Laut seiner Betreuerin ist zwar ihm Messwert Protein/Fettrate [%/kg] normalverteilt, aber wie rechnet er jetzt einen t-Test? Für seinen Projektbericht musste er einen Leistungssteigerungsversuch mit Hühnern in der Uckermark durchführen. Als wäre das nicht schon anstrengend genug gewesen. Jetzt soll er auch noch testen, ob die Behandlung Flüssignahrung (*ctrl* und *fIOW*) ein signifikantes Ergebnis liefert. Hm..., was entspannendes wäre gut. Hm, lecker Gummibärchen und dazu dann im Hintergrund Alien laufen lassen.

Flüssignahrung	Protein/Fettrate
ctrl	43.0
fIOW	23.9
fIOW	32.5
fIOW	28.7
fIOW	31.0
ctrl	29.2
fIOW	17.5
ctrl	43.8
fIOW	22.8
ctrl	47.7
ctrl	36.5
ctrl	50.8
fIOW	21.6
ctrl	49.5
ctrl	43.6
ctrl	50.9
ctrl	39.6
fIOW	22.3

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

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

## 69. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Berechnung des Welch t-Test** Der t-Test. Jessica erschauert. Eine echte Herausforderung für sie war schon immer der Mangel gewesen. Ein leidiges Lied. Ein mächtiges Werkzeug ist der t-Test in den Händen desjenigen, der ein normalverteiltes Outcome ( $Y$ ) hat. Aber erstmal überhaupt den t-Test rechnen können. Wie sah das Experiment von Jessica überhaupt aus? 'Hm...', Schokobons und David Bowie. Das ist und bleibt die beste Kombination zum Nachdenken für Jessica. Jessica hat ein Kreuzungsexperiment mit Zandern durchgeführt. Dabei wurde die Behandlung Bestandsdichte (*Verordnung* und *Gesteigert*) an den Zandern getestet. Gemessen hat Jessica dann als Messwert Schlachtgewicht [kg]. Warum der Versuch im Teutoburgerwald für ihren Projektbericht stattfinden musste, ist ihr bis heute ein Rätsel. Egal. Gibt es jetzt einen Zusammenhang zwischen der Behandlung und Schlachtgewicht [kg]?

Bestandsdichte	Schlachtgewicht
Verordnung	40.9
Verordnung	39.8
Gesteigert	37.3
Gesteigert	37.9
Verordnung	37.0
Gesteigert	39.7
Verordnung	47.6
Verordnung	41.4
Verordnung	42.1
Gesteigert	46.1
Gesteigert	41.2
Gesteigert	40.3
Gesteigert	37.0
Gesteigert	37.4
Gesteigert	30.2
Verordnung	39.3
Verordnung	39.6

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Bestimmen Sie die Teststatistik  $T_D$  eines Welch t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 1.96$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie das 95% Konfidenzintervall. Welche Annahmen haben Sie getroffen? **(2 Punkte)**
6. Nennen Sie den statistischen Grund, warum Sie sich zwischen einem Student t-Test und einem Welch t-Test entscheiden müssen! **(1 Punkt)**
7. Formulieren Sie eine Antwort an Jessica über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**



## 70. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Interpretation des t-Tests in R - die Teststatistik und der p-Wert** Programmieren ist nicht so einfach. R ist schon ein tolles Programm, wenn man mit dem Ding umgehen kann. Super umgehen kann damit Nilufar. Eine echte Herausforderung für Nilufar ist die Erwartung, aber das ist noch eine andere Sache. Deshalb sind aber Jonas und Alex nicht bei ihr. Sondern um sich bei einem gemeinsamen Projekt helfen zu lassen. Im Hintergrund wummert Deichkind. Beide arbeiten gemeinsam an einer Abschlussarbeit. In dem zu beschreibenden Versuch geht es im Teuteburgerwald um einem Kreuzungsexperiment mit Schweinen. Dabei ging darum herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Lüftungssystem (*keins* und *vorhanden*) und dem Messwert Gewichtszuwachs in der 1LW gibt. Da der Messwert Gewichtszuwachs in der 1LW normalverteilt ist kann ein t-Test gerechnet werden. Jonas möchte dann später noch mehr über Nilufars Hobby Hip Hop erfahren.

```
##  
## Two Sample t-test  
##  
## data: Gewichtszuwachs by Lüftungssystem  
## t = 3.653, df = 20, p-value = 0.001581  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## 3.170485 11.611333  
## sample estimates:  
## mean in group keins mean in group vorhanden  
## 31.00000 23.60909
```

Helfen Sie Nilufar bei der Interpretation des t-Tests! Sonst geht es auch für Jonas und Alex 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  $T_{\alpha=5\%} = |2.09|$  einzeichnen! **(4 Punkte)**
5. Beschriften Sie die Abbildung! **(1 Punkt)**
6. Berechnen Sie den Effekt des t-Tests! **(1 Punkt)**

## 71. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Interpretation des t-Tests in R - das 95% Konfidenzintervall** Jonas und Jessica sind bei Nilufar um sich Hilfe in R zu holen. Im Hintergrund wummert Deichkind. 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 Nilufar nicht die Erwartung im Weg stehen würde und Jessica nicht das Problem hätte die Erschöpfung zu händeln. In einer Hausarbeit haben beide zusammen Lauch untersucht. Dabei ging es um den Zusammenhang zwischen der Behandlung Bewässerungstypen (*low* und *high*) und dem Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD]. Der Versuch wurde in einem Freilandversuch im Wendland durchgeführt. Nach der Betreuerin ist der Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] normalverteilt und ein t-Test passt daher. Das wird jetzt nicht mehr angezweifelt...Nilufar überlegt, ob sie die beiden nicht noch auf den Film *Star Trek* einlädt.

```
##  
## Two Sample t-test  
##  
## data: Chlorophyllgehalt by Bewässerungstypen  
## t = 0.47105, df = 17, p-value = 0.6436  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -6.084247 9.582025  
## sample estimates:  
## mean in group low mean in group high  
## 29.18889 27.44000
```

Helfen Sie Nilufar bei der Interpretation des t-Tests! Sonst geht es auch für Jonas und Jessica 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)**

## 72. Aufgabe

(9 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Interpretation des t-Tests in R - die Visualisierung** 'Wir sind uns relativ sicher, dass unser Messwert Trockengewicht [kg/ha] ist!', ruft Alex wild gestikulierend. Alex wäre mehr präsent, wenn es die Gefälligkeit nicht gäbe. Als würde sowas die Ausgabe von R interessieren. Alex und Nilufar sind in einem Café mit Jessica um sich Hilfe von ihr in R zu holen. Während Jessica Kirschstreuselkuchen und Schokobons mampft, versuchen die Alex und Nilufar ihren Versuch im Teuteburgerwald mit Erdbeeren in einem Versuch in einer Klimakammer zu erklären. Jessica hofft insgeheim, dass die R Ausgabe des t-Tests ihr mehr Informationen liefert. Eigentlich würde sie dann doch lieber raus um Rad zu fahren vielleicht mit Nilufar?

```
##  
## Two Sample t-test  
##  
## data: Trockengewicht by Bewässerungstypen  
## t = -2.4336, df = 19, p-value = 0.02499  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -14.975070 -1.126748  
## sample estimates:  
## mean in group low mean in group high  
## 28.60909 36.66000
```

Helfen Sie Jessica bei der Interpretation des t-Tests! Sonst geht es auch für Alex 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 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)**

## Teil V.

# Die einfaktorielle & zweifaktorielle ANOVA

### 73. Aufgabe

(11 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Visualisierung der einfaktoriellen ANOVA** 'Als erstes visualisieren wir unsere Daten und dann können wir schon abschätzen, ob unser Gruppenvergleich in der ANOVA signifikant werden würde?', Tina schaut Mark fragend an und hofft auf eine positive Regung im Gesicht. Wird aber enttäuscht. Die beiden hatten sich auf einem Konzert von Andrea Berg kennengelernt. Mark tut sich auch sehr schwer mit der einfaktoriellen ANOVA. Beide waren im Teuteburgerwald um ein Kreuzungsexperiment mit Schweinen durchzuführen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Elterlinie (*Standard*, *Yray* und *Xray*) und dem Messwert Fettgehalt [%/kg] gibt. Später wird noch Columbo geguckt. Mark befürwortet das!

Elterlinie	Fettgehalt
Xray	25
Xray	26
Yray	28
Standard	25
Xray	24
Standard	24
Xray	26
Standard	25
Standard	25
Yray	30
Standard	25
Xray	25
Xray	24
Yray	21
Yray	25
Yray	28

Leider kennen sich Tina und Mark 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!
  - Den globalen Mittelwert  $\beta_0$  **(1 Punkt)**
  - Die Mittelwerte der einzelnen Behandlungsstufen **(1 Punkt)**
  - Die Mittelwertsdifferenz der einzelnen Behandlungsstufen mit  $\beta_{Standard}$ ,  $\beta_{Yray}$  und  $\beta_{Xray}$  **(1 Punkt)**
  - Die Residuen oder Fehler mit  $\epsilon$  **(1 Punkt)**
4. Liegt ein *vermutlicher* signifikanter Unterschied vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Schätzen Sie die Effekte der Behandlungsstufen! **(2 Punkte)**

## 74. Aufgabe

(9 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**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 Teuteburgerwald um ein Freilandversuch mit Spargel durchzuführen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Lichtstufen (*none*, *200lm*, *400lm* und *600lm*) und dem Messwert Proteingehalt [g/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)
<b>Lichtstufen</b>	3				
<b>error</b>	25	681.79			
<b>Total</b>	28	1243.31			

4. Schätzen Sie den p-Wert der Tabelle mit  $F_{\alpha=5\%} = 2.99$  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)**

## 75. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Die einfaktoriellen ANOVA und der Student t-Test** Mark und Tina schauen sich etwas entnervt an. Gemeinsam schreiben die beiden ihre Abschlussarbeit und sollen nun als erstes einmal die Daten mit einer einfaktoriellen ANOVA auswerten damit abgeschätzt werden kann, ob überhaupt signifikante Ergebnisse in den multiplen Gruppenvergleichen zu erwarten sind. Deshalb erstmal Katjes mampfen, die Tina mitgebracht hat. Nun möchte erstmal ihre Betreuung der Arbeit eine ANOVA Tabelle sehen. Was immer da auch drin zu erkennen sein mag. Mark schaut Tina sehen erstmal gar nichts. Die beiden waren im Teutoburgerwald um ein Stallexperiment mit Milchvieh durchzuführen. Dabei haben Mark und Tina den Messwert Fettgehalt [%/kg] unter der Behandlung Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) ermittelt. Später wollen die beiden dann noch raus um zu Boxen.

Leider kennen sich Mark und Tina 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>Elterlinie</b>	3	4885.62			
<b>Error</b>	30	1066.15			

4. Schätzen Sie den p-Wert der Tabelle mit  $F_{\alpha=5\%} = 2.92$  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)

Elterlinie	Fallzahl (n)	Mittelwert	Standardabweichung
ctrl	8	6.00	1.69
Standard	7	35.86	10.79
Yray	10	9.40	5.62
Xray	9	4.11	2.80

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

## 76. Aufgabe

(9 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Die einfaktorielle ANOVA in R** Jonas schaut entnervt auf und klappt den Laptop zu. Jonas dreht Iron Maiden auf, so dass sich die Nachbarn beschweren werden. Nun möchte seine Betreuerin seiner Hausarbeit erstmal eine ANOVA sehen und *dann* die Ergebnisse präsentiert bekommen bevor es überhaupt mit der Abschlussarbeit weitergeht. Dabei war er extra im Teuteburgerwald um ein Gewächshausexperiment mit Maiss durchzuführen. Und dort was es wirklich nicht schön geschweige denn spannend wie bei seinen Kommilitonen, die in Almería waren. Hätte er es vorher gewusst, dann hätte er die Abschlussarbeit bei wem anders geschrieben. Aber gut, jetzt als die ANOVA in R.

```
## Analysis of Variance Table
##
## Response: Chlorophyllgehalt
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Substrattypen  3 291.61   97.203   12.053 3.441e-05
## Residuals    27 217.75    8.065
```

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

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

## 77. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Ergebnistabelle der zweifaktoriellen ANOVA** In einen Leistungssteigerungsversuch wurden Puten mit dem Behandlung Flüssignahrung (*ctrl*, *superIn* und *fLOW*) sowie der Behandlung Lüftungssystem (*keins* und *thunder*) untersucht. Es wurde als Messwert Protein/Fettrate [%/kg] bestimmt. Alex ahnte schon, dass es komplexer wird, als er mit seiner Abschlussarbeit angefangen hat. Das es jetzt aber so kompliziert wird, hätte er jetzt aber auch nicht gedacht. Alex kratzt sich am Kopf. Alex mampft aus Frust noch eine Handvoll Gummibärchen. Eventuell muss er dann doch nochmal Hilfe in der statistischen Beratung holen. Jetzt versucht er es aber erstmal selber. Und eigentlich wollte Alex doch noch seinem Hobby nachgehen! Alex liebt Starcraft. Darin kann er sich wirklich verlieren und immer wieder neu begeistern.

Leider kennen 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. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus! (3 Punkte)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Flüssignahrung</b>	3	669.56			
<b>Lüftungssystem</b>	1	238.47			
<b>Flüssignahrung:Lüftungssystem</b>	3	18.42			
<b>Error</b>	18	195.7			

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

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

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



## 78. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Die zweifaktorielle ANOVA in R** In ein Gewächshausexperiment wurden Brokkoli mit der Behandlung Bewässerungstypen (*ctrl*, *low*, *mid* und *high*) sowie der Behandlung Lichtstufen (*none*, und *600lm*) untersucht. Es wurde als Messwert Proteingehalt [g/kg] bestimmt. Jetzt starrt Jessica mit auf die R Ausgabe einer zweifaktoriellen ANOVA. Leider starrt ihre Betreuerin in der gleichen Art Jessica zurück an. 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. Jessica hätte doch nichts mit Brokkoli machen sollen. Brokkoli – was soll das auch bedeutendes sein? Eigentlich wollte Jessica nachher noch einen Film schauen. Das Verrückte ist, dass die Hündin Herr der Ringe wirklich liebt. Das ist Jessica sehr recht, denn sie braucht Entspannung.

```
## Analysis of Variance Table
##
## Response: Proteingehalt
##
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Bewässerungstypen      2 346.68  173.338  13.5440 0.0002576
## Lichtstufen             1  61.79   61.785   4.8276 0.0413393
## Bewässerungstypen:Lichtstufen  2 225.44  112.722   8.8077 0.0021514
## Residuals             18 230.37   12.798
```

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. 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)**

## 79. Aufgabe


(11 Punkte)

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




Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Interaktion in der zweifaktoriellen ANOVA** In ein Kreuzungsexperiment wurden Zandern mit der Behandlung Ernährungszusatz (*ctrl*, *fedX* und *getIt*) sowie der Behandlung Lüftungssystem (*keins* und *thunder*) untersucht. Es wurde als Messwert Gewichtszuwachs in der 1LW bestimmt. Jetzt starrt Paula mit auf die  Ausgabe einer zweifaktoriellen ANOVA. Leider starrt ihre Betreuerin in der gleichen Art Paula 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 die Projektbericht. Paula hätte doch nichts mit Zandern machen sollen. Zandern – was soll das auch bedeutendes sein? Eigentlich wollte Paula nachher noch einen Film schauen. Irgendwie komisch, wenn sie Jagd auf roter Oktober anmacht, dann ist die Ratte eigentlich sofort vor dem Bildschirm und starrt hinein.

Leider kennen sich Paula 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 Harry Potter 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  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)**

## 80. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Biostatistik



**Zusammenhang zwischen der ANOVA und dem t-Test** In ein Stallexperiment wurden Zandern mit der Behandlung Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) sowie der Behandlung Lüftungssystem (*keins* und *thunder*) untersucht. Jessica schaut konzentriert auf die Formeln der ANOVA und des t-Tests. In ihrem Experiment wurde als Messwert Protein/Fettrate [%/kg] bestimmt. Wenn der Mangel nicht wäre, ja dann wäre wohl vieles möglich für Jessica! Aber so.. Dann wäre es nicht noch komplizierter. Was war da jetzt nochmal der Zusammenhang zwischen den beiden statistischen Verfahren? Beide Verfahren haben ja irgendwie etwas miteinander zu tun und ihre Betreuerin möchte das jetzt auch noch verstehen. Muss das nicht eigentlich klar sein? Immerhin ist Jessica nicht die erste Betreuung einer Hausarbeit. Immerhin hat sie die beiden Formeln vorliegen. Schon dutzende Male gesehen: Herr der Ringe. Aber immer noch großartig zusammen mit Schokobons.

### 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 Jessica mit dem Zusammenhang zwischen der ANOVA und dem t-Test nicht aus. Deshalb braucht sie bei der Erstellung Ihre Hilfe!

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

## 81. Aufgabe

(11 Punkte)

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

Inhalt folgender Module: Biostatistik



**Zusammenhang zwischen der ANOVA und dem Post-hoc-Test** Es ist schon kurz nach fünf und Nilufar wird langsam nervös. Nilufar wollte heute Abend noch ihre E-Sport Qualifikation schauen. Hoffentlich kommt sie noch rechtzeitig zum Streamen. Angestrengend krault sie das Huhn. Stattdessen versucht ihr Betreuer die Ausgabe der einfaktoriellen ANOVA zu visualisieren und zu überprüfen, ob es mit der Visualisierung der Daten als Boxplots zusammenpasst. Anscheinend gibt es ein Problem mit der Annahme der Normalverteilung und der Varianzhomogenität der ANOVA in den Daten. 'Wir haben jetzt bei der ANOVA einen p-Wert mit 0.071 raus sowie eine F-Statistik  $F_D$  mit 1.51 berechnet. Nach den Boxplots müsste sich eigentlich ein Unterschied zwischen *none* und *200lm* ergeben. Der Unterschied ist in {emmeans} auch signifikant mit einem p-Wert von 0.049. Wie kann das sein?', grummelt ihr Betreuer. Nilufar hatte im Teutoburgerwald ein Freilandversuch mit Brokkoli durchgeführt. Dabei wurden die Daten  $D$  erhoben. Es gab dabei eine Behandlungen Lichtstufen (*none*, *200lm*, *400lm* und *600lm*). Gemessen wurde der Messwert ( $Y$ ) Chlorophyllgehalt (SPAD-502Plus) [SPAD]. So kompliziert kann das jetzt doch nicht sein! Nilufar hat schon genug Probleme. Wenn die Erwartung 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 Nilufar und ihr Betreuer 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 VI.

## Multiple Gruppenvergleiche

### 82. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Adjustierung multipler Vergleiche** In ein Stallexperiment mit Hühnern wurde die Behandlung Bestandsdichte (*effizient, standard, eng, weit und kontakt*) gegen die Ergebnisse einer früheren Studie von Qui et al. (2017) verglichen. Im Rahmen des Experiments haben Jessica und Steffen verschiedene Student t-Tests für den Mittelwertsvergleich für den Messwert Fettgehalt [%/kg] gerechnet. Es ergab sich dann die folgende Tabelle der rohen  $p$ -Werte für die Vergleiche zu Qui et al. (2017). Jetzt sollen die beiden einmal schauen, was in den Daten so drin ist.

Rohen $p$ -Werte	Adjustierte $p$ -Werte	Nullhypothese ablehnen?
0.002		
0.340		
0.070		
0.230		
0.012		

Leider kennen sich Jessica und Steffen 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)**

### 83. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Visualisierung des Compact Letter Displays (CLD)** Mark betrachtet in sich gekehrt die Poster vor dem Büro von seiner Betreuerin. Viele der explorativen Abbildungen sagen ihm etwas. Die Barplots und die Boxplots könnte er dann schon nachbauen. Das macht ihn dann zuversichtlich die Abschlussarbeit auch hinzukriegen. Etwas komischer sind die seltsamen Buchstaben über den Barplots. Mark betrachtet ein Poster das sich mit Lamas beschäftigt. Genotypen (00, AA, AB und BB) und Protein/Fettrate [%/kg] wurden dort bestimmt. So richtig schlau, wird er daraus nicht.

Behandlung	Compact letter display
00	A
AA	B
AB	BC
BB	C

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Zeichnen Sie die sich anhand 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)**

## 84. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Berechnung des Compact Letter Displays (CLD) anhand von t-Tests** Mark hatte in seine Abschlussarbeit einen Leistungssteigerungsversuch durchgeführt. Soweit so gut. Dabei hat er sich mit Hühnern beschäftigt. Angeblich der neueste heiße Kram... aber das ist wiederum was anderes. So richtig mitgenommen hat Mark das Thema dann doch nicht. Hat er sich doch mit Bestandsdichte (*effizient*, *standard*, *eng* und *kontakt*) und Fettgehalt [%/kg] schon eine Menge an Daten angeschaut. Nach sein Betreuer soll er nun ein CLD bestimmen. Weder weiß er was ein CLD ist, noch war sein erster Gedanke mit Köln und die LGBTQ Community richtig... Als erstes solle er die Gruppen nach absteigender Effektstärke sortieren. Was immer das jetzt bringen soll.

Bestandsdichte	Fallzahl (n)	Mittelwert	Standardabweichung
effizient	7	6.03	2.70
standard	8	14.09	2.78
eng	9	14.39	1.43
kontakt	9	8.86	2.50

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Zeichnen Sie die sich ergebenden Barplots! **(1 Punkt)**
4. Berechnen Sie die Matrix der  $p$ -Werte anhand von Student t-Tests! *Nutzen Sie hierfür ein globales  $s_p$  sowie eine gemittelte Fallzahl  $n$  für die Berechnung der Teststatistik!* **(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 Mark! **(1 Punkt)**

## 85. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Biostatistik



**Berechnung des Compact Letter Displays (CLD) anhand der Matrix der p-Werte** 'Okay, dann nochmal für mich. Ich habe jetzt alles in DataTab gemacht, aber das Wichtigste, was gemacht werden soll, nämlich das CLD, das kann ich nicht in DataTab machen?', Paula muss sich echt beherrschen. Immerhin betreut ihr Betreuer ja erst nicht seit gestern Abschlussarbeiten und wusste ja was gemacht werden soll! Paula hatte sich zwei Variablen mit Lichtstufen (*none*, *200lm*, *400lm* und *800lm*) und Chlorophyllgehalt (SPAD-502Plus) [SPAD] in einen Versuch in einer Klimakammer mit Maiss angeschaut. Jetzt möchte sie eigentlich fertig werden und nicht nochmal alles neu in `R` und `{emmeans}` machen. Deshalb soll jetzt das CLD per Hand aus der Matrix der *p*-Wert abgeleitet werden. 'Ich glaube ich wechsel nochmal das Thema...', denkt Paula, verwirft dann aber den Gedanken.

	none	200lm	400lm	800lm
none	1.0000000	0.6516772	0.0190677	0.9747201
200lm	0.6516772	1.0000000	0.0033381	0.6554077
400lm	0.0190677	0.0033381	1.0000000	0.0123926
800lm	0.9747201	0.6554077	0.0123926	1.0000000

Leider kennen sich Paula 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 Paula und Steffen! **(2 Punkte)**



## Teil VII.

# Der Chi-Quadrat-Test & Der diagnostische Test

### 86. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Den Chi-Quadrat-Test berechnen** Yuki hat sich ein Herz gefasst und war für seiner Hausarbeit 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 London Grammar auf das Ohr und los gehts. Yuki ist schon eine ganze Zeit im Büro, da sein Betreuer möchte, dass er jetzt auf seinen Daten mit  $n = 140$  Beobachtungen von Maiss einen  $\chi^2$ -Test rechnet. Das ginge, da er als Behandlung *KI-gesteuert [ja/nein]* bestimmt und zum anderen die Variable *Proteingehalt im Zielbereich [ja/nein]* ermittelt hat. Wie genau, das ist jetzt eine andere Frage. Eigentlich wollte Yuki nachher noch einen Film schauen. Wenn Matrix läuft, dann ist das Minischwein nicht mehr da. Aber jetzt braucht er mal Entspannung!

	56	19	
	27	38	

Leider kennt sich Yuki 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  $\chi^2_{\alpha=5\%} = 5.61$ ! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Skizzieren Sie in einer Abbildung die  $\chi^2$ -Verteilung, wenn die  $H_0$  wahr ist! Ergänzen Sie  $\chi^2_{\alpha=5\%}$  und  $\chi^2_D$  in der Abbildung! Beachten Sie folgenden Informationen zur  $\chi^2$ -Verteilung. Die  $\chi^2$ -Verteilung hat ein Maxima bei  $\chi^2 = 5$  sowie ein Minima bei  $\chi^2 = 9$ . **(2 Punkte)**

## 87. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Den Chi-Quadrat-Test in einem Fragebogen berechnen** Jessica hatte sich gleich von Beginn an in ihrer Abschlussarbeit für eine Umfrage im Marketing interessiert. Jetzt geht es um den Haupt- und Nebenerwerb von Erlebnishöfen in Norddeutschland. Viele Höfe haben angefangen auch Großkatzen zu halten, damit mehr Kunden auf die Höfe kommen. Für den Verband der Großkatzenbesitzer e.V. möchte sie nun einen Fragebogen zur Zukunftsfähigkeit Schritt für Schritt auswerten. Dabei teilt sie zuerst die Antwortenden in die beiden Gruppen 'Höfe mit Großkatzen [ja]' und 'Höfe mit Großkatzen [nein]' ein. Daraufhin möchte sie für folgende Frage *f5verband* einmal auswerten, ob es einen Unterschied zwischen den beiden Höfen mit oder ohne Großkatzen gibt.

*Sehen Sie die Notwendigkeit von genetisch veränderten Großkatzen zur Steigerung der Attraktivität des Erlebnishofes?*

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

<i>f5verband</i>	trifft gar nicht zu	trifft nicht zu	weder noch	trifft zu	trifft voll zu	
ja	4	9	11	29	17	
nein	9	21	11	9	5	

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

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

## 88. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Den Chi-Quadrat-Test mit Effektmaß berechnen** Mark hat sich ein Herz gefasst und war für seiner Abschlussarbeit in die Niederlande gegangen. Das war eine super Zeit in der er viel gelernt hat. Klar gab es auch die ein oder andere Besonderheit, aber das gehört hier eher nicht hin. Dann noch schnell Andrea Berg auf das Ohr und los gehts. Mark ist schon eine ganze Zeit im Büro, da sein Betreuer möchte, dass er jetzt auf seinen Daten mit  $n = 164$  Beobachtungen von Spargel einen  $\chi^2$ -Test rechnet. Das ginge, da er als Behandlung *KI-gesteuert* [*ja/nein*] bestimmt und zum anderen die Variable *Frischegewicht über Zielwert* [*ja/nein*] ermittelt hat. Wie genau, das ist jetzt eine andere Frage. Eigentlich wollte Mark nachher noch einen Film schauen. Wenn Columbo läuft, dann ist der Hamster nicht mehr da. Aber jetzt braucht er mal Entspannung!

	56	41	
	23	44	

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

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

## 89. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik



**Den Chi-Quadrat-Test konzeptionell verstehen** Paula hat sich ein Herz gefasst und war für ihrer Abschlussarbeit in die Niederlande gegangen. Das war eine super Zeit in der sie viel gelernt hat. Klar gab es auch die ein oder andere Besonderheit, aber das gehört hier eher nicht hin. Dann noch schnell Smarties zur Stärkung und los gehts. Paula ist schon eine ganze Zeit im Büro, da ihr Betreuer möchte, dass sie jetzt auf ihren Daten mit  $n = 120$  Beobachtungen von Fleischrindern einen  $\chi^2$ -Test rechnet. Das ginge, da sie als Behandlung *Ökologisch* [ja/nein] bestimmt und zum anderen die Variable *Fettgehalt erreicht* [ja/nein] ermittelt hat. Wie genau, das ist jetzt eine andere Frage. Am Ende des Tages möchte sie dann noch ihr Hobby Harry Potter genießen. Das muss auch mal sein!

			53
			67
	66	54	120

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. 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)**

## Teil VIII.

# Lineare Regression & Korrelation

### 90. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen



**Visualisierung der linearen Regression** 'Ich glaube du bringst da was durcheinander. Wir nutzen zwar auch für die ANOVA die Funktion  $\text{lm}()$  aber hier wollen wir, glaube ich, eine Gerade durch die Punkte zeichnen.', merkt Tina an. 'Ich sehe keine Punkte...', antwortet Alex sichtlich übernächtigt. 'Wir müssen die Daten ja auch erst visualisieren!', spricht Tina sehr deutlich und langsam. Die beiden hatten ein Gewächshausexperiment im Oldenburger Land mit Erdbeeren durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche UV Einstrahlung [UV/d] und Proteingehalt [g/kg]. Jetzt wollen sie erstmal schauen, ob es einen Zusammenhang gibt.

Durchschnittliche UV Einstrahlung [UV/d]	Proteingehalt [g/kg]
14.4	22.3
13.4	20.0
14.2	19.5
21.9	21.6
17.1	20.9
15.9	20.0
15.4	21.7
11.3	18.2
16.6	22.5
13.5	22.4
15.0	21.8
18.4	23.9

Leider kennen sich Tina und Alex 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)**

## 91. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen



**Interpretation der Ergebnisse einer linearen Regression** 'Wichtig ist es, dass wir jetzt eine Gerade durch die Punkte zeichnen!', ruft Paula. 'Ich sehe nur zwei Zeilen und keine Punkte. Wie soll ich da denn jetzt eine Gerade durchzeichnen?', fragt Nilufar. Paula atmet schwer ein und starrt auf die Ausgabe der Funktion `lm()`. Die beiden hatten einen Leistungssteigerungsversuch im Oldenburger Land mit Schweinen durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Tagestemperatur [C/d] und Schlachtgewicht [kg]. Jetzt will die Betreuung von den beiden einmal die Visualisierung der Daten und auch gleich noch die lineare Regression gerechnet bekommen. Das haben beide in gemacht, aber wie soll das jetzt gehen?

term	estimate	std.error	t statistic	p-value
(Intercept)	1.57	2.93		
Durchschnittliche Tagestemperatur	1.02	0.28		

Leider kennen sich Paula und Nilufar mit der linearen Regression für kontinuierliche Daten in 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)**

## 92. Aufgabe

(11 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



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

```
##
## Call:
## Schlachtgewicht ~ Durchschnittliche_Tagestemperatur
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.201  -1.123  -0.195   1.061   3.202
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.219      1.943   1.142   0.261
## Durchschnittliche_Tagestemperatur  2.497      0.196  12.739 2.74e-15
##
## Residual standard error: 1.638 on 38 degrees of freedom
## Multiple R-squared:  0.8103, Adjusted R-squared:  0.8053
## F-statistic: 162.3 on 1 and 38 DF,  p-value: 2.736e-15
```

Leider kennen sich Nilufar und Steffen 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)**

### 93. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Interpretation der Ergebnisse einer Korrelationsanalyse in R** 'Ich glaube ich bringe da was durcheinander. Ich möchte eine Gerade durch die Punkte zeichnen oder doch eine Korrelation berechnen?', merkt Jonas laut an. 'Ich sehe keine Punkte... das ist doch eine Ausgabe in R. Überhaupt, darum geht es doch gar nicht in meinem Versuch. Ich wollte doch keine Gerade zeichnen?.', antwortet Jonas sich sichtlich übernächtigt selber. Wenn Mission Impossible läuft, dann ist das Meerschweinchen nicht mehr da. Aber jetzt braucht er mal Entspannung! Die Nacht war zu lang und überhaupt. Jonas und die Erschöpfung, eine unendliche Geschichte mit kniffligen Wendungen. Jonas hatte ein Gewächshausexperiment im Teuteburgerwald mit Lauch durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche UV-Einstrahlung [UV/d] und Frischegewicht [kg/ha]. Jetzt will er erstmal schauen, ob es einen Zusammenhang gibt und das soll mit der R Ausgabe möglich sein.

```
##  
## Pearson's correlation  
##  
## data: Durchschnittliche UV-Einstrahlung and Frischegewicht  
## t = 5.6334, df = 8, p-value = 0.0004908  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6038070 0.9748011  
## sample estimates:  
## cor  
## 0.8936804
```

Leider kennt sich Jonas 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)**



## 94. Aufgabe

(12 Punkte)

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

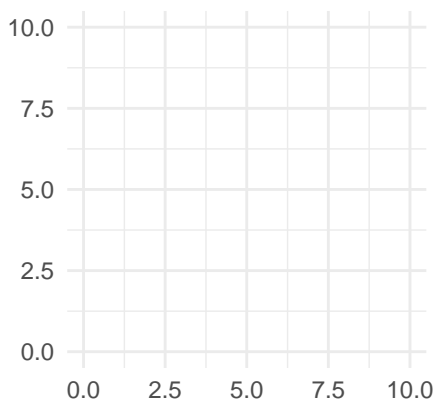
Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



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

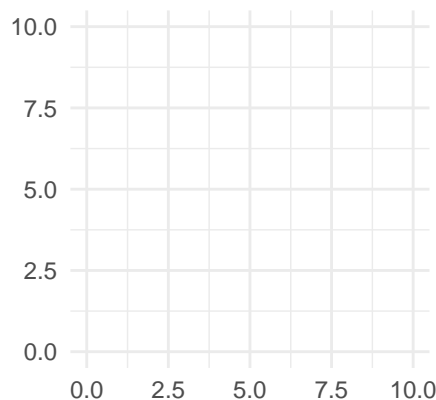
Pearsons  $\rho = 0.75$

$R^2 = 0.25$



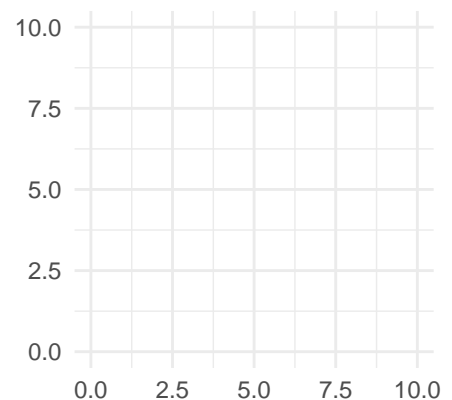
Pearsons  $\rho = 0.25$

$R^2 = 1$



Pearsons  $\rho = -0.5$

$R^2 = 0.75$



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)**

## 95. Aufgabe

(12 Punkte)

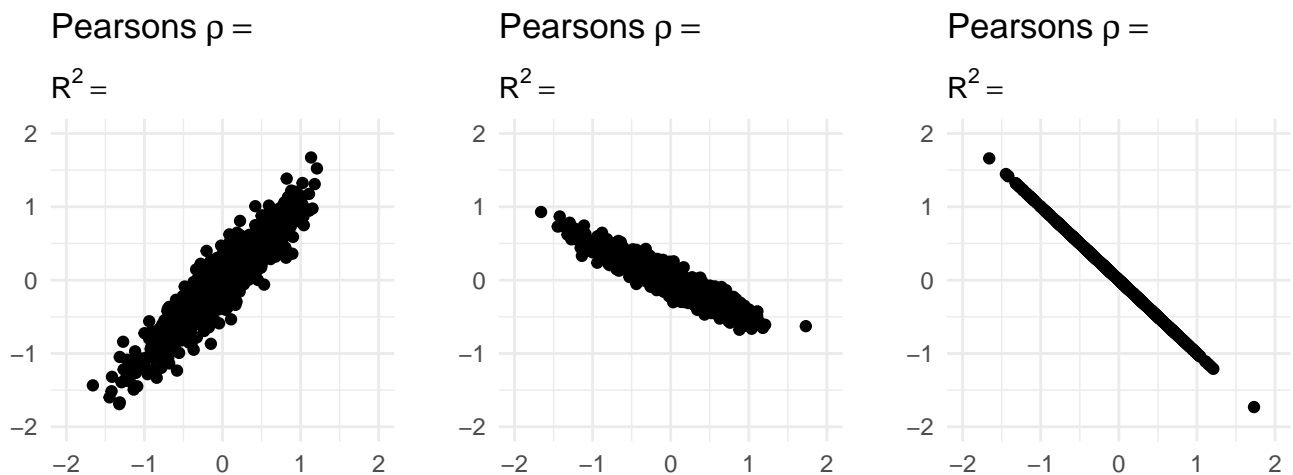
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Inhalt folgender Module: Statistik • Angewandte Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Schätzen der Korrelation und des Bestimmtheitsmaßes** Heute hat leider Mark Lerngruppe keine Zeit, so dass dann Mark alles hier alleine machen muss. Das ist schon recht nervig. Mark versteht nicht viel von  $p$ -Werten oder gar  $R^2$ -Werten. Geschweige denn hat er eine Idee, wie er die Werte abschätzen sollte. Müde dreht er Andrea Berg auf und lässt sich in Gedanken treiben. Gestern war er viel zu lange auf.



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

1. Schätzen Sie die  $p$ -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  $p$ -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)**

## 96. Aufgabe

(12 Punkte)


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

Inhalt folgender Module: Biostatistik



**Visualisierung des Regressionskreuzes** 'Ich sehe die Sinnhaftigkeit nicht. Eine Regression in einem Kreuz darzustellen ergibt für mich keinen Sinn.', meckert Alex so laut, dass die Cafébesucher an den anderen Tischen zu ihm rüberschauen. Alex hatte ein Kreuzungsexperiment mit Fleischrindern durchgeführt. Dabei hatte er dann Gewichtszuwachs erreicht [ja/nein] gemessen. Jetzt war es an ihm die Auswertung für seine Gruppe zu erledigen und dafür sollte Alex sich an dem Regressionskreuz orientieren.

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

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

# Teil IX.


## Experimentelles Design

### 97. Aufgabe

(16 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Einfache experimentelle Designs** Neuer Versuch neues Glück! Paula und Steffen sind bei Alex um sich Hilfe für eine Versuchsplanung in  zu holen. Im Hintergrund läuft viel zu laut Alien. Daher hat die Katze schon lange reißaus genommen. In dem neuen Versuch geht es um den Zusammenhang zwischen der Behandlung Genotypen (AA, AB und BB) und dem Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] in Lauch. Der Versuch soll in einem Versuch in einer Klimakammer in der Uckermark durchgeführt werden. Immerhin ist der Messwert normalverteilt, was einges einfacher macht. Was es nicht so einfacher macht ist, dass Paula noch als zusätzliche Herausforderung etwas anderes umtreibt: der Perfektionismus. Im ersten Schritt überlegt Alex ein einfaches experimentelles Design zu probieren. Daher entscheiden sich alle drei für ein *Randomized complete block design (RCBD)* mit zwei Blöcken. Ob es das jetzt einfacher macht?

Leider kennen sich Alex, Paula und Steffen mit dem *Randomized complete block design (RCBD)* mit zwei 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 ! **(2 Punkte)**
5. Erstellen Sie das statistische Modell in der in  üblichen Schreibweise für eine ANOVA! Skizzieren Sie die notwendigen Funktionen in ! **(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)**

## 98. Aufgabe




(20 Punkte)

Inhalt folgender Module: Biostatistik



**Fortgeschrittene experimentelle Designs** Nilufar und Yuki sind bei Steffen um sich Hilfe für eine Versuchsplanung in  zu holen. Im Hintergrund läuft viel zu laut Abba. Dabei geht es um den Zusammenhang zwischen der Behandlung Lichtstufen (*none*, *200lm*, *400lm* und *600lm*) sowie Bewässerungstypen (*ctrl*, und *high*) sowie drei Blöcken und dem Messwert Proteingehalt [g/kg] in Erbsen. Der Versuch soll in einem Feldexperiment in der Uckermark durchgeführt werden. Nach dem Dozenten ist der Messwert Proteingehalt [g/kg] normalverteilt. Die beiden entschieden sich für ein faktorielles Versuchsdesign. Im ersten Schritt überlegt Steffen ein komplexeres experimentelles Design zu probieren. Daher entscheiden sich alle drei für ein *Split plot design* oder auch *Spaltanlage* mit Berücksichtigung einer Interaktion. Das sollte für den anfang erstmal reichen. 'Und jetzt, was machen wir jetzt?', Yuki schaut die anderen beiden mit großen Augen an. Die zucken mit der Schulter. Alle mampfen Gummibärchen.

Leider kennen sich Steffen, 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 X.

# Forschendes Lernen

### Zerforschen einer wissenschaftlichen Veröffentlichung

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

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

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

## 99. Aufgabe

(20 Punkte)

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





Inhalt folgender Module: Biostatistik



**Zerforschen einer wissenschaftlichen Veröffentlichung** Vor dem Start der eigenen Arbeit möchte seine Betreuerin, dass Steffen einmal die wissenschaftliche Veröffentlichung *Edgar, G., J., et al. (2024) Stock assessment models overstate sustainability of the world's fisheries* sinnvoll zusammenfasst. Dann würde die eigene Arbeit auch leichter von der Hand gehen und Steffen hätte dann schon eine Vorlage. 'Das ist jetzt aber umfangreicher als gedacht!', schnauft er und runzelt die Stirn. Im Hintergrund spielt viel zu leise Taylor Swift. Das wird dann vermutlich heute Abend nichts mehr mit seinem Hobby Klemmbausteine. Die Schlange schaut mitleidig.

Leider kennt sich Steffen 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.

## 100. Aufgabe

(20 Punkte)

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

Inhalt folgender Module: Biostatistik



**Zerforschen eines wissenschaftlichen Datensatzes** Vor dem Start der eigenen Arbeit möchte seine Betreuerin, dass Mark einmal die wissenschaftlichen Daten *in der Tabelle 1* sinnvoll zusammenfasst. Insgesamt wurden  $n$  Beobachtungen erhoben. Dann würde die eigene Arbeit auch leichter von der Hand gehen und Mark 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 Marzipankugeln. Das wird dann vermutlich heute Abend nichts mehr mit Columbo.

$f_1$	$f_2$	$x_1$	$y_1$	$y_2$
< >	< >	< >	< >	< >
1	1	2.3	10.1	0
1	1	4.1	13.1	0
1	1	5.7	16.5	1
1	1	3.4	14.6	0
1	2	2.8	12.1	1
1	2	6.1	13.4	1
⋮	⋮	⋮	⋮	⋮
3	2	1.9	9.6	0

Mark fühlt sich mit der Analyse der Daten in der Tabelle 1 überfordert. 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.

Beantworten Sie die folgenden Fragen anhand eines selbst gewählten Beispiels!

### Allgemeiner Aufgabenteil

1. Ergänzen Sie die Eigenschaften der Spalten in der Form eines tibbles! **(2 Punkte)**
2. Skizzieren Sie die übergeordneten Analysebereiche der Statistik passend zur obigen Datentabelle! Beschriften Sie die Abbildungen! **(4 Punkte)**
3. Formulieren Sie zwei mögliche wissenschaftliche Fragestellungen in Form einer PowerPoint Folie aus der obigen Datentabelle! **(2 Punkte)**

### Spezieller Aufgabenteil für die Variablen $f_1$ , $f_2$ und $y_2$

4. In welchen der übergeordneten Analysebereiche der Statistik gehört die Auswertung Ihres Endpunktes? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Skizzieren Sie eine ikonische Abbildung für Ihren Endpunkt im Kontext der wissenschaftlichen Fragestellung! **(2 Punkte)**
6. Erstellen Sie das statistische Modell in der in R üblichen Schreibweise! **(1 Punkt)**
7. Skizzieren Sie die Datenanalyse für Ihren Endpunkt! **(4 Punkte)**
8. Auf welche Eigenschaft der Daten müssen Sie für Ihre statistische Analyse im Besonderen achten? Erklären Sie eine mögliche Lösung in der Modellierung! Begründen Sie Ihre Antwort! **(2 Punkte)**
9. Welche statistische Maßzahlen können Sie berichten? Begründen Sie Ihre Antwort! **(1 Punkt)**



# Teil XI.

## Mathematik

### 101. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**Herodot – der Schimmel aus Ivenack** Die Lerngruppe *Die Pantoffeltierenen* bestehend aus Alex, Yuki, Jonas und Tina waren auf Exkursion in Brandenburg und haben dort Folgendes erarbeitet. Während der Besetzung Mecklenburgs durch die Franzosen kamen Napoleon die Geschichten des berühmten Apfelschimmels Herodot aus Ivenack zu Gehör. Herodot lief zwar niemals Rennen, war aber eines der berühmtesten Pferde dieser Zeit. Napoleon selbst gab den Auftrag, diesen Schimmel durch die Armee nach Frankreich zu bringen. Der Legende nach sollen Arbeiter den Schimmel im hohlen Stamm einer 1000-jährigen Eiche aus Ivenack vor den Franzosen versteckt haben. Doch Herodot verriet sein Versteck durch lautes Wiehern, woraufhin die französische Armee den Schimmel beschlagnahmte und nach Frankreich führte<sup>4</sup>. Jetzt wollen die vier herausfinden: "Konnten die Ivenacker den Apfelschimmel Herodot vor dem Zugriff von Napoleon in der 1000-jährigen Eiche verstecken?"

Helfen Sie der Lerngruppe *Die Pantoffeltierenen* bei der Beantwortung der Forschungsfrage! Gehen Sie von einem radialen Wachstum der 1000-jährigen Eiche von 1mm pro Jahr aus. Es ist bekannt, dass die Eiche im Jahr 2022 einen Umfang von 12.5m in Brusthöhe hatte.

1. Wie groß war der Durchmesser in  $m$  der Eiche im Jahr 1815 als Herodot in der Eiche versteckt werden sollte? **(2 Punkte)**
2. Skizzieren Sie in einer Abbildung einen linearen Zusammenhang und einen exponentiellen Zusammenhang für das Wachstum der 1000-jährigen Eiche. Erklären Sie die Auswirkungen der Entscheidung für linear oder exponentiell auf Ihre Berechnungen! **(2 Punkte)**

Herodot hatte eine Schulterhöhe von 195cm, eine Breite von 75cm 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 *sehr bequem mit Abstand* um die eigene Achse drehen konnte.

4. Berechnen Sie die Dicke der Eichenwand in  $cm$ ! Verdeutlichen Sie Ihre Berechnungen an einer aussagekräftigen Skizze für Pferd und Eiche! **(2 Punkte)**
5. Unter einer Dicke der Eichenwand von 30cm bricht die Eiche zusammen. Beantworten Sie die Forschungsfrage! Begründen Sie Ihre Antwort! **(2 Punkte)**

<sup>4</sup>Die Quelle der Inspiration für die Aufgabe war eine Fahrt an die Ostsee und folgender Artikel: [Entdecke das erste Nationale Naturmonument Deutschlands - Ivenacker Eichen und Hutewald](#)

## 102. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**Von Töpfen auf Tischen** Die Projektgruppe / bestehend aus Jonas, Nilufar, Jessica und Paula hat sich zusammengefunden um den ersten Versuch zu planen. In einem Experiment wollen sie die Wuchshöhe von 160 Stockrosen bestimmen. Bevor die Vier überhaupt mit dem Experiment beginnen können, gibt es aber ein paar Abschätzungen über die Kosten und den Aufwand zu treffen. Zum einen müssen sie die Stockrosen einpflanzen und müssen dafür Substrat bestellen. Zum anderen muss die Projektgruppe die Stockrosen auch bewegen und in ein Gewächshaus platzieren. Die Töpfe für die Keimung haben einen Durchmesser von 8.5cm und eine Höhe von 10cm. Der Kubikmeterpreis für Torf liegt bei 290 EUR.

Helfen Sie der Projektgruppe / bei der Planung des Versuches!

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

## 103. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**Solar- & Biogasanlagen** Steffen bringt ein neues, tolles Projekt mit in die Lerngruppe *Die Pantoffeltieren* bestehend aus ihm, Nilufar, Paula sowie Yuki. Um die Energiekosten seines Betriebes zu senken, will er eine Solaranlage auf den Rinderstall montieren lassen. Dafür hat er seinen Stall ausgemessen und findet folgende Maße wieder. Die vordere Seite des Rinderstall hat eine Höhe  $h_v$  von  $5m$ . Die hintere Seite des Rinderstall hat eine Höhe  $h_b$  von  $8m$ . Der Rinderstall hat eine Tiefe  $t$  von  $13m$  und eine Breite  $b$  von  $60m$ . 'Sag mal Steffen, ist das eine Matheaufgabe oder rechnen wir hier gerade für dich kostenlos als menschliche Computer Sachen für deinen Betrieb?', fragt Paula mit erhobenen Augenbrauen. Yuki und Nilufar 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 Steffen eine neue Biogasanlage für seinen Betrieb. Der neue Methantank hat einen Radius  $r$  von  $1.5m$ . Leider gibt es ein paar bauliche Beschränkungen auf dem Grundstück. Das Fundament des zylindrischen Methantanks kann nur ein Gewicht von maximal  $15t$  aushalten bevor der Tank wegbricht. Steffen rechnen eine Sicherheitstoleranz von 15% ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei  $-80^\circ C$  eine Dichte von  $235kg/m^3$ . Bei  $-100^\circ C$  hat Methan eine Dichte von  $280kg/m^3$ . Steffen betreibt seine Anlage bei  $-92^\circ C$ .

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

## 104. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



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

**Aligatorenbirnen und Blaubeeren** “Sind Sie ein Riesenfaultier oder warum kaufen Sie so viele Aligatorenbirnen?”, spricht es hinter Ihnen. Irritiert drehen Sie sich um und blicken in das puderrote Gesicht von Mark. “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 Lidl 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 Mark?

1. Wenn 3 Blaubeerschalen 4.77 Euro kosten, wie viel kosten 7 Schalen? **(2 Punkte)**
2. Wenn Sie die 7 Blaubeerschalen gekauft haben, wie viele Aligatorbirnen zu je 0.99 EUR können Sie sich dann noch für 200 EUR leisten? **(1 Punkt)**

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

- Ein Kilo Strauchtomaten benötigt 190l Wasser. Eine Strauchtomate wiegt 120 - 125g.
  - Ein Kilo Salat benötigt 130l Wasser. Ein Salatkopf wiegt 320 - 520g.
  - Ein Kilo Avocado benötigt 1100l Wasser. Eine Avocado wiegt 140 - 420g.
  - Ein Kilo Blaubeeren benötigt 850l Wasser. Eine Blaubeere wiegt 3.2 - 3.5g.
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  $9 \times 10^4$ t in dem prognostizierten Rahmen. Die Menge steigerte sich um 8.6%. Die Exporte für Avocados stiegen in dem gleichen Zeitraum um 21.2% auf  $2.1 \times 10^5$ 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 48 Liter Wasser pro Kopf für den täglichen Bedarf. In *Deutschland* liegt der Verbrauch bei 8 - 17 Liter pro Spülmaschinenlauf und 3 - 12 Liter pro Minute Händewaschen.

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

Das alles hätten Sie nicht von Mark 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.

## 105. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



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 Nilufar 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  $9.78 \text{ m/s}^2$  an. Im Weiteren hat die Erde einen ungefähren Durchmesser von  $1.289 \times 10^4 \text{ km}$  und eine mittlere Dichte  $\rho$  von  $5.51 \text{ g/cm}^3$ . Das Gewicht von einem heute lebenden Waldelefanten mit 2.7t liegt bei 6t und das Gewicht von einem Tyrannosaurus rex (T. rex) bei 4.5t bis 8t.

1. Welchen Durchmesser müsste die Erde vor 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 0.99 astronomische Einheiten (AE). Die Einheit 1 AE wird mit  $1.48 \times 10^8 \text{ km}$  angegeben. Der *massebehaftete* Sonnenwind besteht aus 85% Wasserstoffkernen mit einer molaren Masse von  $1.02 \text{ g/mol}$ , 8% Heliumkernen mit  $4.01 \text{ g/mol}$  sowie 7% weiteren Atomkernen mit  $145.31 \text{ g/mol}$ . Die Teilchendichte bei Eintritt in die Erdatmosphäre liegt zwischen  $0.4$  bis  $100$  Teilchen  $\text{cm}^{-3}$  pro Sekunde mit einer mittleren Teilchendichte von  $6 \text{ cm}^{-3}$  pro Sekunde.

Lösen Sie den folgenden Aufgabenteil mit einer aussagekräftigen Skizze!

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

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

## 106. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**'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 Paula, Jessica, Nilufar und Tina 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) wingflapping, (2) sitting, (3) drinking/eating und (4) wing/leg stretching.

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)	Jabcoobs et al. (2019)
wingflapping	29cm; 35cm; 50.1%	35cm; 16cm; 25.4%	30cm; 14cm; 25.4%
sitting	24cm; 27cm; 5.1%	27cm; 37cm; 6.1%	43cm; 49cm; 3.2%
drinking/eating	23cm; 26cm; 0.8%	27cm; 41cm; 1%	30cm; 30cm; 1%
wing/leg stretching	26cm; 40cm; 13.2%	19cm; 25cm; 12.8%	24cm; 18cm; 12.3%

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

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

<sup>7</sup>Die Quelle der Inspiration für die Aufgabe war der folgende wissenschaftliche Artikel: [EFSA Panel on Animal Health and Welfare, et al. \(2023\) Welfare of broilers on farm. EFSA Journal 21.2.](#)

## 107. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**Nelken von den Molukken** Jonas und Yuki 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 Jonas ziemlich plötzlich, so dass Yuki die Reese's Peanut Butter Cups aus dem Mund fallen. 'Nein, und ehrlich gesagt bin ich auch ziemlich müde...'. Das ist jetzt aber Jonas egal, denn er möchte folgende Sachlage diskutieren. Und Jonas hat jetzt 3 Stunden Zeit. Plus Verspätung. In der Ausstellung *Europa und das Meer* im Deutschen Historischen Museum in Berlin gab es folgendes Zitat über die Probleme der frühen Hochseeschifffahrt.

»Ohne ausreichende Zufuhr von Vitamin C stellen sich nach 42 Tagen die ersten Symptome ein; die ersten Toten sind nach 70 Tagen zu beklagen; nach 105 Tagen rafft die Skorbut eine ganze Schiffsbesatzung dahin«

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

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

Der Chronist an Bord der *Trinidad*, Antonio Pigafetta, schrieb in seinem Bericht '[...] Um nicht Hungers zu sterben, aßen wir das Leder, mit dem die große Rahe zum Schutz der Taue umwunden war.' Insbesondere die Mannschaft der *Concepción* erlitt große Verluste durch die Skrobut bei der Überquerung des Pazifiks, da durch Erkundungsfahrten weniger Zeit blieb, um wilden Sellerie aufzunehmen. Wilder Sellerie enthält  $7000\mu\text{g}/100\text{g}$  Vitamin C. Der Bedarf liegt bei  $115\text{mg}$  pro Tag für Männer.

3. Berechnen Sie die notwendige Menge in  $t$  an aufzunehmenden wilden Sellerie auf die *Concepción* für die ununterbrochene Fahrt von drei Monate und 18 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)**

## 108. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**Event Horizon – Am Rande des Universums** Jonas ist bei Mark um gemeinsam *Event Horizon – Am Rande des Universums* zu streamen. Das war jetzt nicht die beste Idee. Denn Jonas kann Horror überhaupt nicht ab. Deshalb flüchtet er sich in Logik um seine Emotionen zu bändigen. Mark mampft ungerührt Marzipankugeln. Folgenden Gedankengang nutzt Jonas um dem Film zu entkommen. Die Sonne hat eine aktuelle, angenommene Masse von  $2 \times 10^{27} \text{ kg}$ . Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 3000m kollabiert, wird die Sonne 30% der aktuellen Masse verloren haben. Ein Lichtteilchen mit der Masse  $m_f$  und der Fluchtgeschwindigkeit  $v_f$  will dem schwarzen Loch entkommen. An folgende Formeln erinnert sich Jonas 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  $6.674 \cdot 10^{-11} \text{ m}^3 (\text{kg} \cdot \text{s}^2)^{-1}$

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

1. Geben Sie die Formel für die Fluchtgeschwindigkeit  $v_f$  an! **(2 Punkte)**
2. Überprüfen Sie Ihre umgestellte Formel nach  $v_f$  anhand der Einheiten! **(1 Punkt)**
3. Berechnen Sie die notwendige Fluchtgeschwindigkeit  $v_f$  des Lichtteilchens mit den angegebenen Informationen! **(2 Punkte)**
4. Gehen Sie von einer Lichtgeschwindigkeit von  $2.9 \times 10^8 \text{ m/s}$  aus. Kann das Lichtteilchen der Gravitation des schwarzen Lochs entkommen? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Stellen Sie den Zusammenhang zwischen dem sich verringernden Radius  $r$  des schwarzen Lochs bei gleichbleibender Masse  $m_s$  und der notwendigen Fluchtgeschwindigkeit  $v_f$  in einer Abbildung dar! Erstellen Sie dafür eine Datentabelle mit fünf Werten für den Radius  $r$ ! **(3 Punkte)**

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



## 109. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



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

**Das Fermi Paradoxon** Mark und Alex wandern durch den Teutoburgerwald um mal vom Studium runterzukommen. 'Kennst du eigentlich Enrico Fermi?', fragt Mark 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?'. Alex schaut ihn irritiert und interessiert an. Die beiden hat das Problem gepackt. Deshalb wollen Mark und Alex das Paradoxon mal mathematisch untersuchen! Wie lange würde eine außerirdische Zivilisation benötigen um die gesamte Milchstraße zu besuchen, wenn das maximale Reisetempo die Geschwindigkeit der Voyager 1 Sonde wäre?<sup>9</sup>

Die beiden treffen folgende Annahmen. Eine außerirdische Zivilisation schickt *vier* Voyager 1 ähnliche Sonden mit der Geschwindigkeit von  $6.0523 \times 10^4 \text{ km/h}$  los um sich auf den erreichten Planeten selbst zu replizieren. Nach 250 Jahren ist die Replikation abgeschlossen und wiederum *vier* Sonden werden ausgesendet. Gehen Sie von 4.24 Lichtjahren als mittlerer Abstand der Sterne in der Milchstraße aus. Es gibt  $10^{11}$  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 annäherungsweise 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  $10^8$  Jahre vollständig im Erdinneren umgewandelt haben! **(2 Punkte)**

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

## 110. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**Pyramiden bauen** Es stehen die oldenburgischen Pyramidentage an! 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 70 Grad im Vergleich zu den ägyptischen Pyramiden mit 51 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 6cm 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 Schulterschmerzen entwickelt, als die Sklaven von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 75% aus. In eine Schubkarre passen 110 Liter.

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

Bei der Besichtigung der Pyramide teilt Ihnen der leicht übergewichtige Pharao (Nebenberuf *Chefredakteur*) mit, das die Pyramide zu flach sei und somit nicht in die oldenburgische 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)**

## 111. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**Geocaching – Von Satelliten und Plastikdosen** Es ist Wochenende und das Wetter ist sweet. Tina und Mark schwingen sich auf ihre Cachermobile um mit 18km/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. Tina und Mark wollen diesmal endlich die abwärts Schwierigkeitschallenge durchführen. Die Reihenfolge der Caches nach Schwierigkeitswertung 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 Tina und Mark für die Planung der Route zu Verfügung<sup>10</sup>.

Ort	Cache	Wertung (S T G)
A	GCUQXIW	2.0   2.5   Klein
B	GCVAYHV	3.0   2.0   Normal
C	GCJEAUE	2.5   1.5   Normal
D	GCECERU	1.0   4.0   Mikro
E	GCFPVYD	5.0   4.5   Mikro

Im Weiteren sind den beiden folgende Informationen zu den Entfernungen der Orte zugänglich. Der Entfernungsvektor  $\vec{AB}$  ist 4km. Im Weiteren ist Ihnen der Entfernungsvektor  $\vec{CB}$  mit 6.5km bekannt. Der Entfernungsvektor  $\vec{BE}$  ist das 1.2-fache des Entfernungsvektor  $\vec{CB}$ . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr 40° 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! **(5 Punkte)**
2. Welche Strecke in km legen Sie bei der Bewältigung der abwärts Schwierigkeitschallenge zurück? **(2 Punkte)**
3. Gehen Sie von einer zusätzlichen Suchzeit in Stunden für die Caches an den jeweiligen Orten zur reinen Reisezeit mit Ihrem Cachermobil aus. Die Suchzeit in Stunden für *jeden einzelnen* Cache wird durch die Funktion

$$\text{Suchzeit} = 0.1 + 0.13 \cdot \text{Schwierigkeit}$$

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

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

## 112. Aufgabe

(10 Punkte)

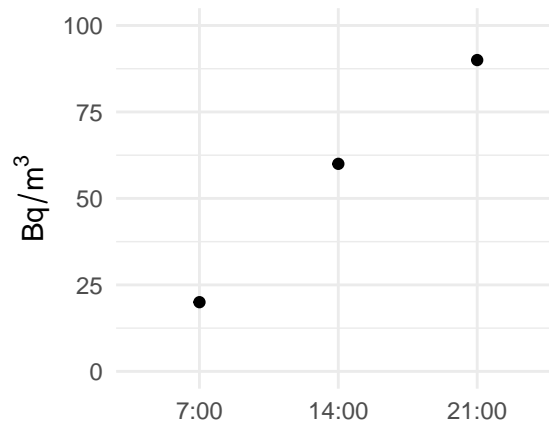
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!

Inhalt folgender Module: Mathematik & Statistik



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 Yuki 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 21:00 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 Yuki 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 Keller-raum erreicht haben? **(2 Punkte)**

Radon zerfällt mit einer Halbwertszeit von 3,7d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 135d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

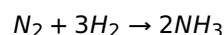
2. Wie lange dauert es in Stunden bis Ihre kritische Radonbelastung von  $300 Bq/m^3$  auf unter  $90 Bq/m^3$  gefallen ist? **(4 Punkte)**

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	79.7	28.1	
Sauerstoff	21.3	16.5	
Kohlenstoffdioxid	0.029	12.1	

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



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

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

<sup>11</sup>Tocotronic - Electric Guitar als passende Untermauerung 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](#)

## 113. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**Armee der Finsternis** Der Studentenjob von Paula war nach Ladenschluss bei Aldi die Regale einzuräumen. Dabei ist Paula in der Auslage der Sonderangebote das Necronomicon<sup>14</sup> in die Hände gefallen. Nun ist sie eine Magierin der Zeichen geworden! Also eigentlich kann Paula nur Mathe und das dämliche Necronomicon hat sie in die Vergangenheit geschleudert... aber gut, was tut man nicht alles im Jahr 481 n. Chr. für den neuen Lehnsherren Fürsten Arthur. Paula baut natürlich einen Schrottkugelturm um sich den Horden der Finsternis mit genug Schrott erwehren zu können! Paula 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  $30mg$  zu gleichförmigen *Bleikugeln* bei einer Geschwindigkeit von  $14m/s$  bilden.

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

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

2. Nennen Sie die beiden geometrischen Formen aus denen sich näherungsweise ein Tropfen zusammensetzt! Erstellen Sie eine beschriftete Skizze des Tropfens! **(2 Punkte)**
3. Sie messen eine Länge des Tropfens von  $3.2mm$ . Die Löcher im Sieb erlauben ein Tropfendurchmesser von  $1.5mm$ . 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  $11.34g/cm^3$ .

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

## 114. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**Armee der Kaninchen** Leider hat es bei Steffen mit dem Krokodilreservat in Down Under nicht geklappt. War vielleicht auch nicht so die beste Idee... aber dafür hat Steffen eine neue Eingebung! Oder wie es Mike Tyson zugeschrieben wird: »Ich wurde nie niedergeschlagen, ich war immer am Aufstehen!«. Daher macht Steffen jetzt einen Großhandel mit Kaninchenfleisch und damit dem teuersten Fleisch in Australien auf. Moment, hopsen hier nicht, seit Thomas Austin im Jahr 1859 ungefähr 32 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Steffen 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.

$$f(t) = 1.2 \times 10^{10} - 9 \times 10^8 \cdot 2^{-0.2 \cdot t + 2.4}$$

1. Skizzieren Sie die Sättigungsfunktion *annäherungsweise* in einer Abbildung! **(1 Punkt)**
2. Wie viele Kaninchen können nach der Sättigungsfunktion maximal im westlichen Australien leben? Ergänzen Sie den Wert in Ihrer Abbildung! **(2 Punkte)**
3. Wie viele Millionen Kaninchen leben nach der Sättigungsfunktion nach 9 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 14 Kaninchen ausgebracht. Da die Kaninchen keine Maßnahmen gegen die Ausbreitung vornehmen können, verläuft die Ausbreitung mit einem wöchentlichen Wachstumsfaktor von 1.4 nach folgender Formel.

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

3. Wie viele Wochen benötigen die Viren um theoretisch die gesamte Kaninchenpopulation nach 7 Jahren Wachstum zu durchseuchen? **(1 Punkt)**

Das Myxoma Virus und das RHDV töten 99.7% der Kaninchenpopulation innerhalb weniger Wochen.

4. Wie lange in Jahren dauert es bis eine Kaninchenpopulation nach einer Viruspandemie wieder auf 60% der gesättigten Kaninchenpopulation angewachsen ist? **(2 Punkte)**

Thomas Austin entließ die Kaninchen im äußersten Osten von Australien. Australien hat eine West-Ost-Ausdehnung von 4000km und eine Nord-Süd-Ausdehnung von knapp 3800km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 9.8km 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 12\$ pro Tier und der durchführende Arzt verlangt ca. 45\$ 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...](#)

## 115. Aufgabe

(12 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**Ostfriesland. Unendliche Weiten.** Wir schreiben das Jahr 2025. Dies sind die Abenteuer des Hängebauchschweins Frida und Nilufar. Grünes Gras unter Nilufars Füßen und ein strammer Wind im Gesicht, egal wohin sie schaut. Ein schmatzendes Geräusch ertönt unter Nilufar. Nilufar sinniert, sollte sie ihre weiten Graslandschaften jetzt schon düngen? Dafür benötigt Nilufar 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. Nilufar 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  $0.5\times$ , Februar mit  $0.7\times$  und März mit  $1.1\times$ . Sie haben noch im letzten Jahr folgende Temperaturen gemessen.

Datum	C°
01. Jan 2024	0.1
01. Feb 2024	1.1
01. Mrz 2024	3.5
01. Apr 2024	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  $200^\circ\text{C}$  zu blühen. An welchem Tag im 1. Quartal des Jahres 2024 war dies der Fall? *Ignorieren Sie ein eventuelles Schaltjahr in Ihrer Berechnung.* **(4 Punkte)**

Auf dem Weg zu Nilufars Pink Lady Plantage 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  $120\text{N}$ . Die elektrifizierten Renter bringen eine Kraft von  $140\text{N}$  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)**

## 116. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**In der Kartonagenfabrik** Paula, Nilufar, Tina und Alex sitzen im Bus. Wenn man sich zu spät anmeldet, dann ist die Exkursion nicht so toll. Paula hatte den Anderen in der Lerngruppe zu spät Bescheid gesagt. 'Was denn, bin ich eure Nanny oder was?!', entfährt es Paula 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, 8-mal-gefaltete, 0.6mm, 50-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 50cm und eine Breite von 22cm. Die Kartonagenmaschine in Flint soll dann einen quadratischen Eckenausschnitt der Länge  $x$  falzen.

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

Rektor Skinner möchte sich gerne wieder in seinem Vorgarten aufhalten und nicht die ganze Zeit von Bart mit Erdnüssen beworfen werden. Deshalb möchte er einen geräumigen Teil seines Vorgartens einzäunen. Ein Teil der Umzäunung bildet seine Vorderhauswand. Wegen Lieferschwierigkeiten stehen Rektor Skinner nur 90m Zaun zu Verfügung. Auch hier sollen Sie mal helfen, sonst fährt der Bus Sie nicht nach Hause. Sie wollen nun die maximale Fläche des abgeschirmten Vorgartens in Abhängigkeit der Seitenlängen bei der Verwendung von 90m Zaun bestimmen!

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



## 117. Aufgabe

(10 Punkte)

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

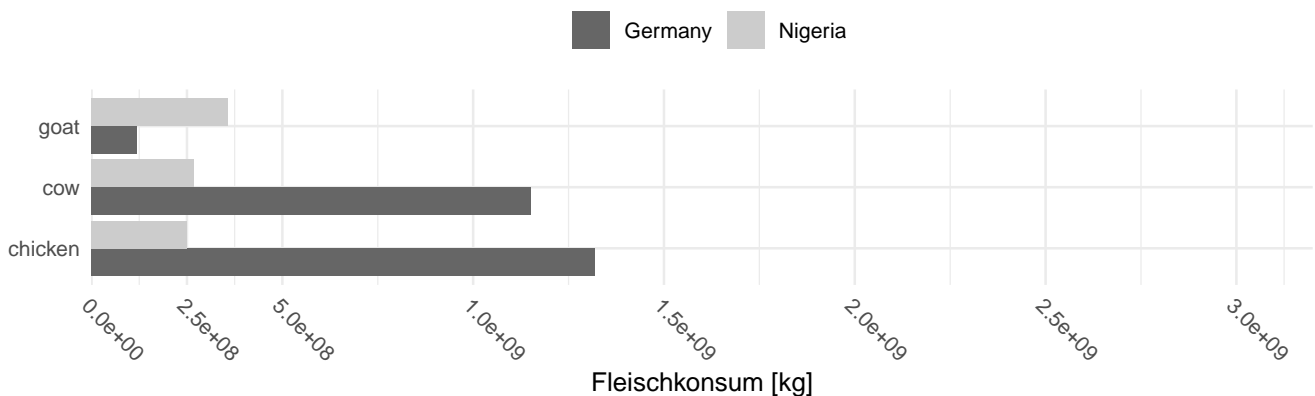


Inhalt folgender Module: Mathematik & Statistik



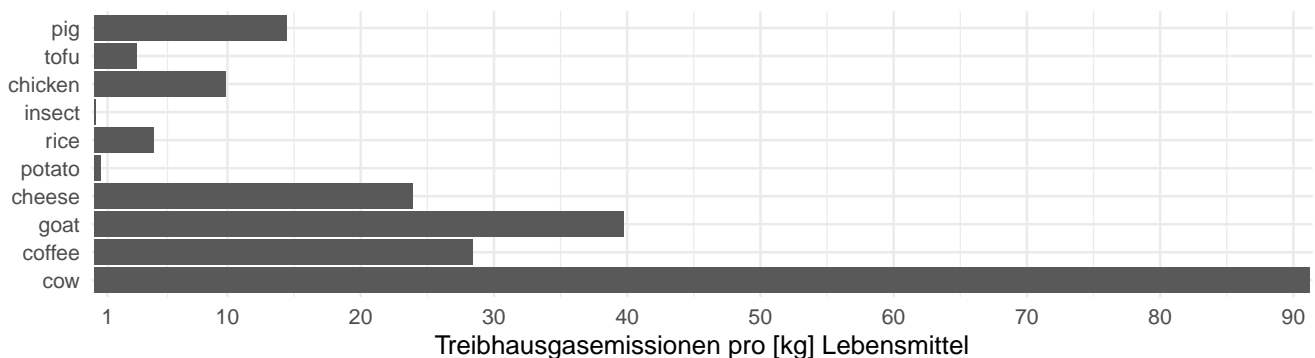
**Ein Pfund Insekten, bitte!** 'Das wird wohl häufiger gehört werden, wenn wir die Menschheit mit Proteinen ausreichend ernähren wollen<sup>16</sup>.', merkt Tina an. Die Lerngruppe um Yuki, Steffen und Mark sind bei Tina um mal was außergewöhnliches zu essen. Um den Sinn der Nahrungsumstellung zu verdeutlichen, vergleicht Tina 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 Tina 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)**

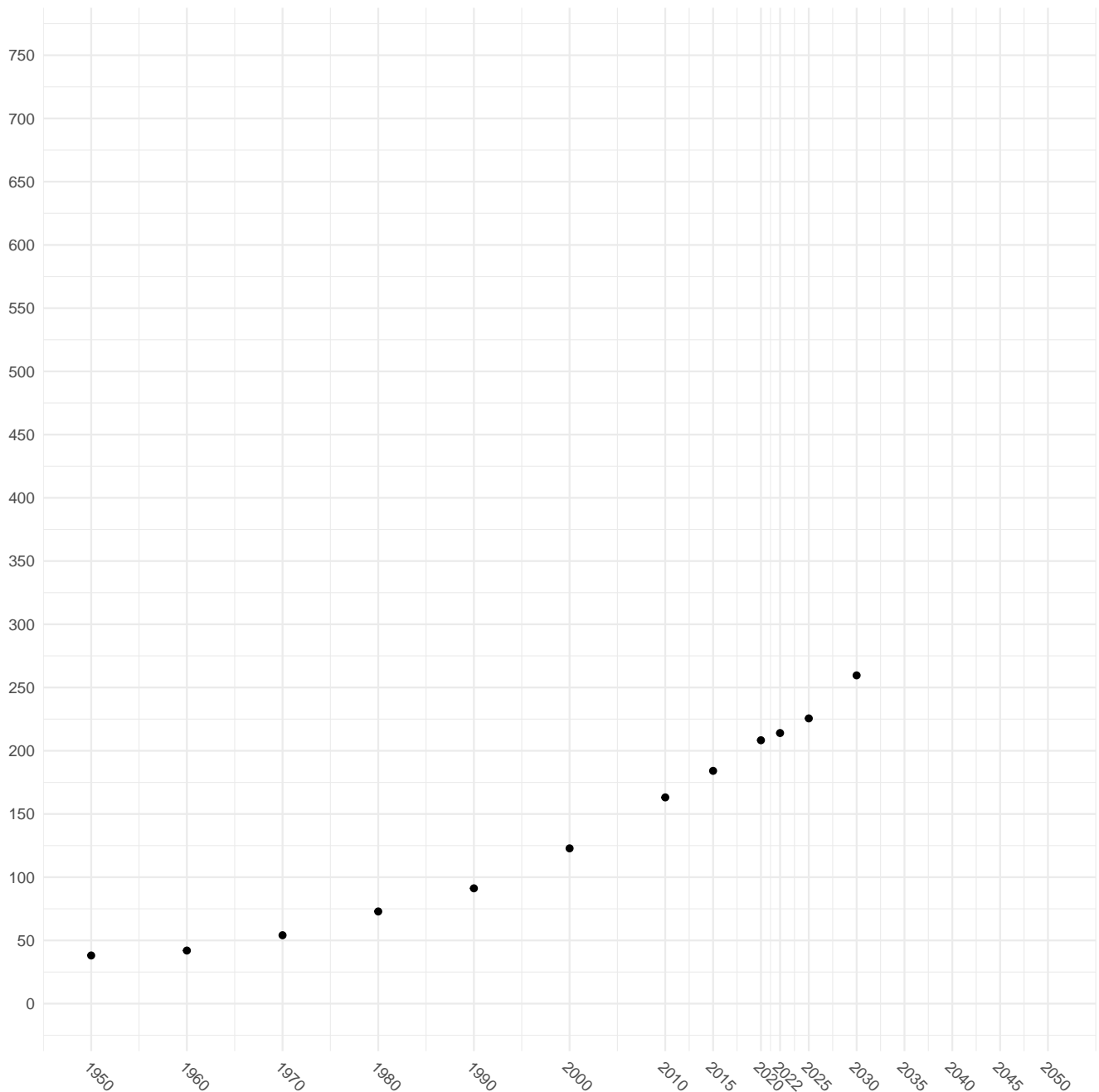
In der nächsten Abbildung finden Sie die CO<sub>2</sub>-Emission in [kg] nach Lebensmittel, die durch die Produktion entsteht, abgebildet.



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)**

## 118. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



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 Tina sind bei ihrem HNO-Arzt 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 Tina nun aber über AIDS und dem diagnostischen AIDS-Test, den die beiden nun machen werden? Leider zu wenig. Da brauchen dann Mark und Tina 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 90% der Fälle ist ein HIV-Test positiv, wenn der Patient erkrankt ist. In 1% der Fälle ist ein HIV-Test positiv, wenn der Patient *nicht* erkrankt ist und somit gesund ist. Sie stutzen. Wie wahrscheinlich ist es denn eigentlich an AIDS erkrankt zu sein ( $K^+$ ), wenn Sie einen positiven AIDS-Test vorliegen haben ( $T^+$ )? Gehen Sie für die folgenden Berechnungen von  $n = 2 \times 10^4$  Patienten mit einem diagnostischen Test für AIDS aus. Sie nehmen sich also einen Kuli und fangen an auf der historischen Ausgabe des Spiegels zu 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 Tina, 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 Tina 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

## 119. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



**Network-Marketing oder Schneeballschlacht!** Tina, Alex und Yuki sitzen bei Paula und hören sich etwas über Network-Marketing an. Paula 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 2023 war das erfolgreichste Jahr in der Geschichte von KH Gesund und Schön Components (KH-GSC). Das Unternehmen steigerte den Umsatz um rund 18 Prozent von 260 Millionen Euro im Jahr 2022. Doch wie viel kommt bei den Partnern an? Laut KH-GSC habe das Unternehmen  $2.8 \times 10^5$  aktive Partner.

1. Berechnen Sie zuerst den Umsatz der Firma KH-GSC im Jahr 2023! **(1 Punkt)**
2. Wie viel von dem Umsatz im Jahr 2023 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 20%? **(1 Punkt)**

Das von Paula zu vermarkende Produkt, hinter dem Paula voll steht, kostet 100EUR pro Einheit im Direktverkauf. Die direkte Provision für die erste Stufe beträgt 25%. Für die zweite, dritte und vierte Stufe betragen die indirekten Provisionen jeweils 3%, 2% und 1.5%. Jeder von Paula angeworbener »Partner« wirbt wiederum vier Partner für sich selbst an. Pro Monat werden im Schnitt fünf Einheiten vom Produkt verkauft. Paula will nun 3200EUR 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)**

Paula musste zum Einstieg bei KH-GSC Einheiten des Produkts für 4000EUR kaufen. Diese Einheiten kann Paula nur direkt verkaufen. Das ganze Wohnzimmer ist voll davon. Leider musste Paula den Kauf über einen Kredit über 4.7% p.a. über 36 Monate finanzieren. Sie schütteln den Kopf und klären Paula 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](#)

## 120. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Mathematik & Statistik



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

In dem Spiel hat Tina nun auf einmal 4 vierseitige Würfel ( $4d4$ ) zum würfeln in der Hand. Wenn Tina eine 4 würfelt, hat Tina einen Erfolg.

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

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

3. Würden Sie die Axt mit zwei achtseitigen Würfeln ( $2d8$ ) als Schaden oder das Schwert mit einem vierseitigen Würfel plus 7 ( $1d4+7$ ) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

Jetzt wird es immer wilder, da Paula und Jessica sich jetzt überlegen müssen, wie wahrscheinlich es ist, dass der Rettungswurf gegen den zaubernden Hexer funktioniert. Paula und Jessica haben folgende Wahrscheinlichkeiten gegeben. Die Wahrscheinlichkeit für das Ereignis  $A$ , der Rettungswurf ist erfolgreich, ist  $Pr(A) = 0.65$ , 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 gezauert hat! **(1 Punkt)**

## 121. Aufgabe

(12 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik



**Retrocheck im TV** Das war zu viel für Tina gestern. Die Lerngruppe mit Yuki und Alex 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 Tina fragt sich, ob sie nicht doch lieber bezahlter Gast bei Barbara Salesch hätten sein sollten. Aber Tina braucht das Geld und jetzt heißt es Spielschows farmen! Erstmal eine Kaffemaschine von Mitropa gewinnen. Ein Kandidat gewinnt die Kaffeemaschine von Mitropa, wenn nicht alle Kandidaten überbieten (eng. *outbid*). Mit Ihnen bilden Yuki und Alex das Team der drei Kandidaten. Tina braucht dringend Ihre Hilfe in ihren Wahnträumen. Sie gehen wie in *Interception* rein!<sup>19</sup>

Name	$P(\text{win})$	$P(\text{outbid})$
Yuki	0.2	0.043
Alex	0.3	0.11

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 0.02 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 Tina auf der Kirmes und spielen mit einem Typen in einem Tentakelkostüm um das große Geld. Das Glücksrad hat 20 Felder. Sie beide drehen das Glücksrad zweimal. Auf 8 Feldern gewinnen Tina und Sie 3000EUR 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 4500EUR? **(1 Punkt)**

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

#### 122. Aufgabe

(6 Punkte)

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

#### 123. 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.

#### 124. Aufgabe

(6 Punkte)

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