

Name: \_\_\_\_\_

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

**»Wintersemester 2025/26«**



„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 Studierte mehr verdienen, würde dann in Teilzeit auch mehr rauspringen. 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

Aufgabe	11	12	13	14	15	16	17
Punkte	9	12	10	11	12	10	10

- 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 und Versuchswesen • Biostatistik

Sie wollen Ihren Datensatz in R einlesen und stehen nun vor einem Problem. Sie stellen fest, dass die Hilfeseiten alle in englischer Sprache verfasst sind. Warum mag die Nutzung von Deutsch problematisch sein?

- A  Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Daher ist die Nutzung in Deutsch in den AGBs von R untersagt.
- B  Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- C  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.
- D  R Pakete sind nur in englischer Sprache verfasst. Es macht keinen Sinn R daher in Deutsch zu bedienen.
- E  Programmiersprachen haben Probleme mit Umlauten und Sonderzeichen der deutschen Sprache. Die Nutzung von englischer Sprache umgeht dieses Problem in eleganter Art.

### 2. Aufgabe

(2 Punkte)

In R müssen die Daten eines Experiments in einem bestimmten Format vorliegen. Wir sprechen auch von dem log-Format der Daten. Welche Aussage zu dem Long-Format ist richtig?

- A  In den Zeilen finden sich die experimentellen Faktoren (X) sowie die Messwerte (Y). In den Spalten 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  Das Long-Format beschreibt in den Spalten die Beobachtungen sowie in den Zeilen die *unabhängigen* Beobachtungen.
- D  In den Spalten finden sich die experimentellen Faktoren (X) sowie die Messwerte (Y). In den Zeilen finden sich dann die einzelnen Beobachtungen.
- E  In den Spalten finden sich die einzelnen Beobachtungen und in den Zeilen die gemessenen Variablen (Y) sowie die experimentellen Faktoren (X).

### 3. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • 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  Für eine explorativen Datenanalyse (EDA) in R müssen wir als erstes die Daten über `read_excel()` einlesen. Danach müssen wir schauen, dass wir die Zeilen richtig über `mutate()` transformiert haben. Insbesondere müssen Variablen mit kontinuierlichen Werten in einen Faktor umgewandelt werden. Am Ende nutzen wir die Funktion `ggplot()` für die eigentlich EDA.
- C  Für eine explorativen Datenanalyse (EDA) in R müssen wir als erstes die Daten über `read_excel()` einlesen. Danach müssen wir schauen, dass wir die Spalten richtig über `mutate()` transformiert haben. Insbesondere müssen Variablen mit Kategorien in einen Faktor umgewandelt werden. Am Ende nutzen wir die Funktion `ggplot()` für die eigentlich EDA.

- D**  Wir lesen die Daten über eine generische Funktion `read()` ein und müssen dann die Funktion `ggplot()` nur noch installieren. Dann haben wir die Abbildungen als `*.png` vorliegen.
- E**  Die Funktionsreihenfolge ist wie folgt: `read_excel() -> mutate() -> ggplot()`. Dabei ist bei der Transformation der Daten darauf zu achten, dass keine Faktoren erstellt werden.

## Deskriptive Statistik & Explorative Datenanalyse

### 4. Aufgabe

(2 Punkte)

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

Gegeben ist  $y$  mit 14, 20, 11, 15 und 14. Berechnen Sie den Mittelwert und Standardabweichung.

- A**  Es berechnet sich 15.8 +/- 10.7
- B**  Es ergibt sich 15.8 +/- 1.635
- C**  Es berechnet sich 14.8 +/- 10.7
- D**  Sie erhalten 14.8 +/- 1.81
- E**  Sie erhalten 14.8 +/- 3.27

### 5. Aufgabe

(2 Punkte)

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

Berechnen Sie den Median, das 1<sup>st</sup> Quartile sowie das 3<sup>rd</sup> Quartile von  $y$  mit 12, 22, 22, 17, 26, 28 und 63.

- A**  Sie erhalten 22 [15; 26]
- B**  Es ergibt sich 22 +/- 17
- C**  Sie erhalten 22 +/- 28
- D**  Es ergibt sich 22 [17; 28]
- E**  Es berechnet sich 23 [18; 27]

### 6. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Mathematik & 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**  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 ( $n - 1$ ).
- B**  Den Mittelwert berechnen, dann die absoluten Abstände zum Mittelwert aufsummieren. Die Fallzahl ( $n - 1$ ) entsprechend gewichten.
- C**  Den Mittelwert berechnen und die Abstände quadrieren. Die Summe mit der Fallzahl ( $n - 1$ ) multiplizieren.
- D**  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.
- E**  Den Mittelwert berechnen, dann die quadratischen Abstände zum Mittelwert aufsummieren und durch die Fallzahl ( $n - 1$ ) teilen.

### 7. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Mathematik & 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**  Der Barplot stellt die Mittelwerte und die Varianz dar.
- B**  Den Median und die Standardabweichung.
- C**  Der Barplot stellt die Mittelwerte und die Standardabweichung dar.
- D**  Durch die Abbildung des Barplot erhalten wir die Informationen über den Median und die Quartile.
- E**  Den Mittelwert und die Varianz.

## 8. Aufgabe

(2 Punkte)

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

In Ihrer Abschlussarbeit zu Erbsen finden Sie aufeinmal seltsame Daten. Jedenfalls kommt Ihnen das so vor. Daher berechnen Sie den Mittelwert mit 9.7 t/ha und den Median mit 10.1 t/ha. Der Mittelwert wird mit  $\bar{y}$  und der Median mit  $\tilde{y}$  abgekürzt. Welche Aussage ist richtig?

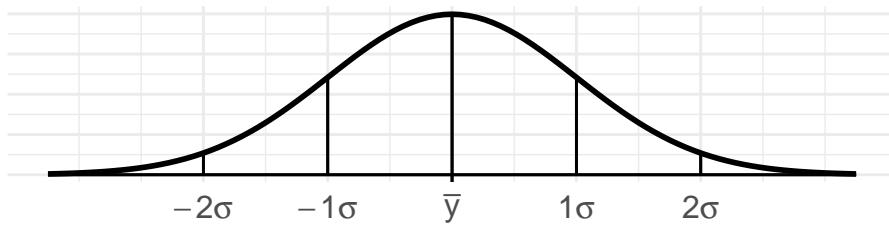
- A**  Wenn sich der Mittelwert und der Median unterscheiden, liegen vermutlich Outlier in den Daten vor.
- B**  Da sich der Mittelwert und der Median nicht unterscheiden, liegen vermutlich Outlier in den Daten vor. Wir untersuchen den Datensatz nach auffälligen Beobachtungen.
- C**  Da sich der Mittelwert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

## 9. Aufgabe

(2 Punkte)

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

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



- A**  Die Fläche links von  $-2\sigma$  ist der p-Wert mit  $Pr(D|H_0)$  in der obigen Abbildung.
- B**  Die Fläche unter der Kurve entspricht dem Signifikanzniveau  $\alpha$  von 5%. Damit ist die Standardabweichung  $\sigma$  gleich 1 in der obigen Abbildung.
- C**  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.
- D**  Die Fläche zwischen  $-2\sigma$  und  $2\sigma$  ist 0.68 und 68% der Beobachtungen liegen somit zwischen  $\bar{y} \pm \sigma$  in der obigen Verteilung.
- E**  Die Fläche zwischen  $-1\sigma$  und  $1\sigma$  ist 0.95 und 95% der Beobachtungen liegen somit zwischen  $\bar{y} \pm \sigma$  in der obigen Verteilung.

## 10. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A**  Die Mindestanzahl liegt bei fünf Beobachtungen.
- B**  2-5 Beobachtungen.
- C**  Wir sollten eine Beobachtung mindestens pro Gruppe vorliegen haben.
- D**  Damit wir hier sauber eine Abbildung von einem Histogramm erhalten, sollten wir mindestens zwanzig Beobachtungen haben.
- E**  Die optimale Anzahl ist größer als hundert Beobachtungen, wobei es gerne sehr viel mehr sein können.

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

## 12. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Um zu Überprüfen, ob die Daten die Annahme einer Varianzhomogenität genügen, können wir folgende Visualisierung nutzen. Dabei kommt dann auch die entsprechende Regel zur Abschätzung der Annahme einer Varianzhomogenität zur Anwendung.

- A**  In einer explorativen Datanalyse nutzen wir den Violinplot. Dabei sollte der Bauch am Rand liegen. Dann können wir von einer Varianzhomogenität ausgehen.
- B**  Nach der Erstellung eines Boxplots schauen wir, ob der Median in der Mitte der Box liegt. Dabei ist der Median als dicke Linie dargestellt und die Box ist das IQR.
- C**  Einen Boxplot. Das IQR muss über alle Behandlungen zusammen mit den Whiskers ungefähr gleich aussehen.
- D**  Wir erstellen uns für jede Behandlung einen Dotplot und schauen, ob die Dots und damit die Varianz für jede Behandlung gleich groß sind.
- E**  Einen Barplot. Die Mittelwerte müssen alle auf einer Höhe liegen. Die Fehlerbalken haben hier keine Informationen.

## Statistische Testtheorie

### 13. Aufgabe

(2 Punkte)

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

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

- A**   $Pr(D|H_0)$  ist die Wahrscheinlichkeit der Alternativehypothese und somit  $1 - Pr(H_A)$
- B**   $Pr(D|H_0)$  ist die Wahrscheinlichkeit nicht die Daten  $D$  zu beobachten sondern die Nullhypothese, wenn diese wahr ist.
- C**  Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativehypothese überdeckt.
- D**  Die Wahrscheinlichkeit der Daten unter der Nullhypothese in der Grundgesamtheit.
- 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 Mathematik & 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**  Im Rahmen eines langen Disputs zwischen Neyman und Fischer wurde  $\alpha = 5\%$  festgelegt. Leider werden die Randbedingungen und Voraussetzungen an statistische Modelle heute immer wieder ignoriert.
- B**  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.

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

## 15. Aufgabe

(2 Punkte)

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

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

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

## 16. Aufgabe

(2 Punkte)

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

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

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

## 17. Aufgabe

(2 Punkte)

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

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

- 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 schauen, ob der  $p$ -Wert kleiner ist als das Signifikanzniveau  $\alpha$ . Wir sprechen von einem signifikanten Ergebnis. Dabei vergleichen wir somit Wahrscheinlichkeiten. Die Wahrscheinlichkeiten werden als Flächen unter der Kurve der Teststaistik dargestellt, wenn die  $H_0$  gilt.
- C**  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.
- 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 Teststaistik dargestellt, wenn die  $H_A$  gilt.
- E**  Wir machen eine Aussage über die Flächen unter der Kurve der Teststatistik der Hypothese  $H_0$  und über die Flächen unter den Kurve der Teststatistik der Hypothese  $H_A$ . Dabei werden Wahrscheinlichkeiten verglichen, die durch die Flächen unter der Kurve der beiden Testverteilungen repräsentiert werden.

## 18. Aufgabe

(2 Punkte)

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

Die Ergebnisse der einer statistischen Analyse können in die Analogie einer Wettervorhersage gebracht werden. Welche Analogie für die Ergebnisse eines statistischen Tests trifft am besten zu?

- A**  In der Analogie des Niederschlags oder Regenmenge: ein statistischer Test gibt die Stärke eines Effektes wieder. Zum Beispiel, wie hoch ist der Mittelwertsunterschied.
- B**  In der Analogie der Durchschnittstemperatur: Wie oft tritt ein Effekt durchschnittlich ein? Wir erhalten eine Wahrscheinlichkeit für die Effekte. Zum Beispiel, wie hoch ist die Wahrscheinlichkeit für einen Mittelwert als Durchschnitt.
- C**  In der Analogie der 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 Wahrscheinlichkeit für Regen: ein statistischer Test erlaubt die Wahrscheinlichkeit für ein Ereignis abzuschätzen. Die Stärke des Effektes können wir nicht bestimmen.
- E**  In der Analogie der Maximaltemperatur: Was ist der maximale Unterschied zwischen zwei Gruppen. Wir erhalten hier eine Aussage über die Spannweite und den maximalen Effekt.

## 19. Aufgabe

(2 Punkte)

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

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

- A**  Weder eine Aussage über die Population noch über das Individuum ist mit einem statistischen Test möglich. Wir erhalten eine Aussage über ein Experiment.
- 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**  Nein, wir können die untersuchte Population nicht mit einer statistischen Analyse auswerten. Wir erhalten keine Aussage zur Population.
- D**  Ja, es ist möglich die untersuchte Population mit einem statistischen Test wie einem t-Test auszuwerten. Wir erhalten dann eine Aussage zur Population.
- E**  Nein, wir erhalten eine einfache Aussage. Wir müssen daher das Individuum im Kontext der Population adjustieren.

## 20. Aufgabe

(2 Punkte)

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

Welche Aussage über den Effekt eines statistischen Tests ist richtig?

- A**  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.
- 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**  Durch den Effekt erfahren wir die statistische interpretierbare Ausgabe eines statistischen Tests. Zum Beispiel das  $\eta^2$  aus einer ANOVA. Damit können wir die Signifikanz direkt mit dem Effekt verbinden. Am Ende muss der Forschende aber entscheiden, ob der Effekt entsprechend seinen Erwartungen als bedeutet zu bewerten ist.
- D**  Der Effekt eines statistischen Tests beschreibt die biologisch interpretierbare Ausgabe eines Tests. Moderne Algorithmen liefern keine Effekte mehr sondern nur noch bedingte Wahrscheinlichkeiten. Der Effekt spielt in der modernen Statistik keine Rollen mehr.
- E**  Der Forschende muss am Anfang wissen, ob das Ergebnis eines Experiments relevant für seine Forschung ist. Dafür kann der Effekt eines statistischen Tests genutzt werden oder auch der Prähoc-Test. Damit beschreibt der Effekt den biologischen interpretierbaren Teil eines Experiments vor der Durchführung. Zum Beispiel der Unterschied zwischen zwei Mittelwerten.

## 21. Aufgabe

(2 Punkte)

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

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

- A**  Randomisierung ist die direkte Folge von Strukturgleichheit. Die Strukturgleichheit erlaubt es erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- B**  Strukturgleichheit ist durch Randomisierung gegeben. Somit kann von der Stichprobe auf die Grundgesamtheit geschlossen werden
- C**  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.
- D**  Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte. Ohne Mittelwerte keine Varianz und somit auch kein statistischer Test.
- E**  Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.

## 22. Aufgabe

(2 Punkte)

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

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

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

## 23. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Mathematik & 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  $\alpha$ -Fehler in der Analogie eines Rauchmelder: *Alarm without fire*.
- B**  *Alarm with fire*, dem  $\alpha$ -Fehler in der Analogie von Feuer.
- C**  In die Analogie eines Rauchmelders: *Alarm with fire*.
- D**  In die Analogie eines brennenden Hauses ohne Rauchmelder: *House without noise*.
- E**  Dem  $\beta$ -Fehler mit der Analogie eines Rauchmelders: *Fire without alarm*.

## 24. Aufgabe

(2 Punkte)

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

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

- A**  Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da  $\Delta$  antiproportional zum p-Wert ist, bedeutet auch ein hohes  $\Delta$  ein sehr kleinen p-Wert.
- B**  Über das Konfidenzintervall. Das Konfidenzinterval bietet eine Entscheidung über die Signifikanz und zusätzlich kann über die Visualisierung des Konfidenzintervalls eine Relevanzschwelle definiert werden.
- C**  Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- D**  Über das Konfidenzintervall. Das Konfidenzinterval inkludiert eine Entscheidung über die Relevanz und zusätzlich kann über die Visualisierung des Konfidenzintervalls eine Signifikanzschwelle vom Forschenden definiert werden.
- E**  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.

## 25. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Sie werten in Ihrer Abschlussarbeit einen sehr großen Datensatz aus einer öffentlichen Datenbank aus. Nun stellen Sie fest, dass Sie ein Problem mit der Bewertung Ihrer Ergebnisse anhand der Signifikanz bekommen. Wie Sie herausfinden, scheint dies ein häufiges Problem in der Bio Data Science zu sein. Welche Aussage ist richtig?

- A**  Aktuell werden zu grosse Datensätze für die gänigige Statistik gemessen. Daher wendet man maschinelle Lernverfahren für kausale Modelle an. Hier ist die Relevanz gleich Signifikanz.
- B**  Relevanz und Signifikanz haben nichts miteinander zu tun. Daher gibt es auch keinen Zusammenhang zwischen hoher Fallzahl ( $n > 10000$ ) und einem signifikanten Test. Ein Effekt ist immer relevant und somit signifikant.
- C**  Eine große Fallzahl führt zu mehr signifikanten Ergebnissen auch bei kleinen Effekten. Daher werden fast alle Vergleich esignifikant, wenn die Fallzahl nur groß genug wird.
- D**  Riesige Datensätz haben mehr Fallzahl was zur  $\alpha$ -Inflation führt. Durch eine Adjustierung kann dem Problem entgegengewirkt werden.
- E**  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.

## 26. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Die Leistung von Sauen soll auf einem Zuchtbetrieb gesteigert werden. Dafür werden die Ferkel verschiedener Sauen gemessen. Die Ferkel einer Muttersaue sind daher...

- A**  Untereinander unabhängig. Sollten die Mütter verwandt sein, so ist die Varianzstruktur ähnlich und muss modelliert werden.
- B**  Untereinander stark korreliert. Die Ferkel sind von einer Mutter und sommit miteinander korreliert. Dies wird in der Statistik jedoch meist nicht modelliert.
- C**  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.
- 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 der berechneten Teststatistik eine Entscheidung gegen die Nullhypothese treffen. Welche Aussage ist richtig?

- A □ Das Signifikanzniveau  $\alpha$  ist gleich 5%.  $Pr(D|H_0)$  muss kleiner sein als das Signifikanzniveau.
- B □ Wir betrachten das ganze Intervall des 95%-Konfidenzintervalls. Liegt die Null mit in dem Intervall, dann kann die Nullhypothese nicht abgelehnt werden.
- C □ Der kritische Wert  $T_{\alpha=5\%}$  ist tabellarisch festgelegt und gegeben. Ist  $T_D$  größer als der kritische Wert, kann die Nullhypothese abgelehnt werden
- D □ Anhand der berechneten Teststatistik lässt sich wie folgt eine Entscheidung treffen. Liegt der Wert in dem Signifikanzniveaumittelwert  $\alpha$  dann kann die Nullhypothese abgelehnt werden.
- E □ Das Signifikanzniveaumittelwert  $\alpha$  ist gleich 5% und damit muss der berechnete Wert unter dem Signifikanzniveaumittelwert  $\alpha$  liegen, dann kann die Nullhypothese nicht abgelehnt werden.

## 28. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

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

- A □ Die Power wird nicht berechnet sondern ist eine Eigenschaft des Tests. Die Power wird auf 80% gesetzt und beschreibt mit welcher Wahrscheinlichkeit  $H_A$  bewiesen wird
- B □ Die Power beschreibt die Wahrscheinlichkeit die  $H_A$  abzulehnen. Wir testen die Power jedoch nicht.
- C □ Es gilt  $\alpha + \beta = 1$  und somit liegt  $\beta$  meist bei 95%.
- D □ Die Power ist nicht in der aktuellen Testtheorie mehr vertreten. Wir rechnen nur noch mit dem Fehler 1. Art.
- E □ Die Power wird berechnet und ist keine Eigenschaft des Tests. Die Power wird auf 80% gesetzt und beschreibt mit welcher Wahrscheinlichkeit  $H_0$  bewiesen wird

## Statistische Tests für Gruppenvergleiche

## 29. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Mathematik & 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 ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern
- B □ Der t-Test testet generell zu einem erhöhten  $\alpha$ -Niveau von 20%.
- C □ Der t-Test vergleicht zwei Gruppen indem die Mittelwerte miteinander verglichen werden.
- D □ 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.
- E □ Der t-Test vergleicht zwei oder mehr 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?

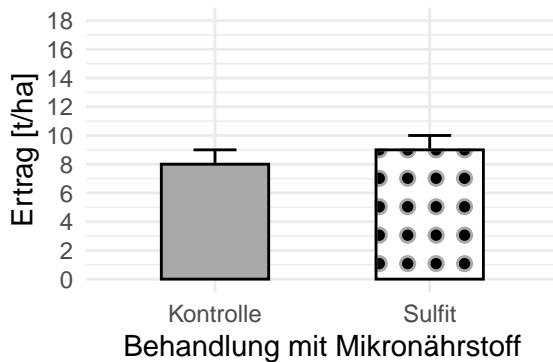
- A □ Der gepaarte t-Test nutzt die Varianz der Beobachtungen jeweils paarweise und bildet dafür eine verbundene Stichprobe. Dieser Datensatz  $d$  dient dann zur Differenzbildung.
- B □ Der gepaarte t-Test wird genutzt, wenn die Differenzen der Beobachtungen verbunden sind und wir dadurch die Unabhängigkeit nicht mehr vorliegen haben.
- C □ Wenn die Beobachtungen nicht unabhängig voneinander sind, rechnen wir einen gepaarten t-Test. Messen wir wiederholt an dem gleichen Tier oder Pflanze dann bilden wir die Differenz zwischen den zwei Messpunkten.
- D □ 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.
- E □ Beim gepaarten t-Test kombinieren wir die Vorteile des Student t-Test für Varianzhomogenität mit den Vorteilen des Welch t-Test für Varianzheterogenität. Wir bilden dafür die Differenz der Einzelbeobachtungen.

### 31. Aufgabe

(2 Punkte)

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

Ein Versuch wurde in 7 Parzellen pro Gruppe durchgeführt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung des Mikronährstoff Sulfit auf den Ertrag in t/ha von Mango im Vergleich zu einer Kontrolle. Welche Aussage ist richtig, wenn Sie einen t-Test rechnen?



- A**  Es liegt ein signifikanter Unterschied vor. Der Effekt liegt bei 0.1.
- B**  Der Test deutet auf keinen signifikanten Unterschied hin. Der Effekt liegt vermutlich bei 1.
- C**  Die Barplots deuten auf ein signifikanten Unterschied. Der Effekt liegt vermutlich bei 1.
- D**  Die Barplots deuten auf keinen signifikanten Unterschied. Der Effekt liegt vermutlich bei 1 unter einer groben Abschätzung. Wir müssen aber eine ANOVA rechnen um den Effekt wirklich bestimmen zu können.
- E**  Nach Betrachtung des Barplots liegt kein signifikanter Unterschied vor. Der Effekt kann nicht bei einem t-Test aus Barplots bestimmt werden.

### 32. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

In Ihrer Abschlussarbeit passen die Ergebnisse einer ANOVA und eines multiplen Vergleiches nicht zusammen. Nach einem Experiment mit drei Maissorten ergibt eine ANOVA ( $p = 0.045$ ). Sie führen anschließend die paarweisen t-Tests für alle Vergleiche durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der  $\alpha$ -Schwelle. Sie schauen sich auch die rohen, unadjustierten p-Werte an und finden hier als niedrigsten p-Wert  $p_{3-2} = 0.052$ . Welche Aussage ist richtig?

- A**  Hier kommt der Effekt der steigenden Fallzahl auf die Anzahl an signifikante Ergebnisse zu tragen. Da die ANOVA auf weniger Fallzahl testet als die paarweisen t-Tests, kann die ANOVA schwerer einen signifikanten Unterschied nachweisen.
- B**  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.
- C**  Das Beispiel kann so nicht auftreten, da die ANOVA und die t-Tests algorithmisch miteinander verschränkt sind.
- D**  Es gibt einen Fehler in der Varianzstruktur. Daher kann die ANOVA nicht richtig sein und paarweise t-Tests liefern das richtige Ergebnis.
- E**  Hier kommt der Effekt der steigenden Fallzahl auf die Anzahl an signifikante Ergebnisse zu tragen. Da die ANOVA auf mehr Fallzahl testet als die einzelnen paarweisen t-Tests, kann die ANOVA leichter einen signifikanten Unterschied nachweisen. Die p-Werte sind immer etwas kleiner als bei den t-Tests.

### ANOVA

### 33. Aufgabe

(2 Punkte)

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

Aus einem Feldversuch ergibt sich die Notwendigkeit der Berechnung einer einfaktoriellen ANOVA. Es ergibt sich ein  $\eta^2 = 0.31$ . Welche Aussage ist richtig?

- A**  Die Berechnung von  $\eta^2$  ist ein Wert für die Interaktion in der einfaktoriellen ANOVA.

- B**  Das  $\eta^2$  beschreibt den Anteil der globalen Varianz, der von den Umweltbedingungen erklärt wird.
- C**  Der Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird, wird durch das  $\eta^2$  beschrieben.
- D**  Der Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird, wird durch das  $1 - \eta^2$  beschrieben.
- E**  Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen nicht erklärt wird. Somit der Rest an nicht erklärbare Varianz.

### 34. Aufgabe

(2 Punkte)

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

In Ihrem Feldexperiment wollen Sie den Proteingehalt von Erbsen steigern. Die Pflanzen wachsen unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen. Um die Daten auszuwerten führen Sie eine einfaktoriellen ANOVA durch. Es ergibt sich ein  $\eta^2 = 0.28$ . Welche Aussage ist richtig?

- A**  Es werden 28% der Varianz durch den Versuch erklärt. Das  $\eta^2$  beschreibt den Anteil der Varianz, der durch Fehler in der Versuchsdurchführung entsteht.
- B**  Mit dem  $\eta^2$  lässt sich auf die Qualität der Randomisierung und damit der Strukturgleichheit zwischen der Grundgesamtheit und der Stichprobe schließen. Es gilt dabei die Regel, dass ein  $\eta^2$ -Wert von 1 zu bevorzugen ist.
- C**  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%.
- D**  Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird. Daher werden 28% der Varianz durch die Behandlungsgruppen erklärt.
- E**  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.

### 35. Aufgabe

(2 Punkte)

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

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

- A**  Die ANOVA berechnet die F-Statistik indem die MS des Fehlers durch die MS der Behandlung geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese abgelehnt werden.
- B**  Die ANOVA berechnet die T-Statistik aus der Multiplikation der MS Behandlung mit der MS der Fehler. Wenn die F-Statistik genau 0 ist, kann die Nullhypothese nicht abgelehnt werden.
- C**  Die ANOVA berechnet die F-Statistik indem die MS des Fehlers durch die MS der Behandlung geteilt werden. Wenn die F-Statistik sich der 1 annähert kann die Nullhypothese nicht abgelehnt werden.
- D**  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.
- E**  Die ANOVA berechnet die F-Statistik indem die MS der Behandlung durch die MS des Fehlers geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese nicht abgelehnt werden.

### 36. Aufgabe

(2 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Statistik • Angewandte Mathematik & 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.

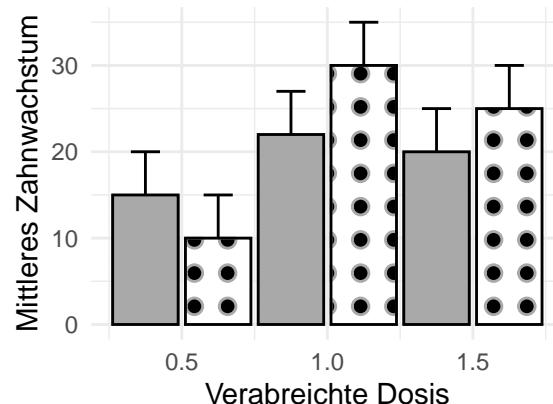
- C**  ... den Unterschied zwischen der Mittelwerte und der Varianz aus verschiedenen Behandlungsguppen. Wenn die ANOVA signifikant ist, ist bekannt welcher Vergleich konkret unterschiedlich ist.
- D**  ... den Unterschied zwischen mehreren Varianzen aus verschiedenen Behandlungsguppen. Wenn die ANOVA signifikant ist, ist bekannt welcher Vergleich konkret unterschiedlich ist.
- E**  ... den Unterschied zwischen der F-Statistik anhand der Varianz der Gruppen. Wenn die F-Statistik exakt 0 ist, kann die Nullhypothese abgelehnt werden.

## 37. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Ein Versuch wurde an 61 Tieren durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0.5, 1 und 1.5 mg/Tag) über eine von zwei Verabreichungsmethoden erhielt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung von Vitamin D auf das Zahnwachstum bei Hasen. Welche Aussage ist richtig, wenn Sie eine zweifaktorielle ANOVA rechnen?



- A**  Es liegt keine Interaktion vor ( $p > 0.05$ ).
- B**  Das Bestimmtheitsmaß  $R^2$  ist klein.
- C**  Eine Korrelation liegt vor ( $p \leq 0.05$ ).
- D**  Eine negative Interaktion liegt vor ( $\rho \geq 0.5$ ).
- 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: 0.02, 0.03, 0.42, 0.001 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 0.1, 0.15, 1, 0.005 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- B**  Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.1, 0.15, 1, 0.005 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1% verglichen.
- C**  Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.004, 0.006, 0.084, 2e-04 und 0.178. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1% verglichen.
- D**  Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.1, 0.15, 2.1, 0.005 und 4.45. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E**  Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.004, 0.006, 0.084, 2e-04 und 0.178. 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**  Contrast letter display. Unterschiede in den Behandlungen werden durch den gleichen Buchstaben oder Symbol dargestellt. Die Interpretation des CLD führt häufig in die Irre.
- B**  Compound letter display. Gleichheit in den 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.
- 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 line display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Früher wurden keine Buchstaben sondern eine durchgezogene Linie verwendet. Bei mehr als drei Gruppen funktioniert die Linie aber graphisch nicht mehr.
- E**  Compact letter detection. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt.

## 40. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A**  Das R Paket {emmeans} erlaubt die Durchführung eines multiplen Gruppenvergleichs. Aus einem emmeans Objekt lässt sich leider kein CLD erstellen. Dennoch ist das Paket einfach zu bedienen und wird deshalb genutzt. Die Interpretation der statistischen Auswertung wird über einen Barplot abgebildet.
- B**  Das R Paket {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.
- C**  Da Sie für Ihre Bachelorarbeit einen Barplot mit CLD brauchen nutzen Sie das R Paket {emmeans} welches Ihnen schnell die notwendigen Informationen liefert um einen Barplot zu erstellen. Die Berechnung eines CLD ist hierbei auch einfach.
- D**  Das R Paket {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.
- E**  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.

## 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 sehr viel höher. Daher müssen die p-Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der p-Werte nach Bonferroni das bekannteste Verfahren ist.
- 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**  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.

- 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

Sie rechnen mehrere t-Tests für einen multiplen Vergleich nachdem eine einfaktorielle ANOVA sich als signifikant herausgestellt hat. Welche Aussage im Bezug auf den Effekt ist richtig?

- A**  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.
- B**  Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  adjustiert werden im Gegensatz zu den p-Werten.
- C**  Beim multiplen Testen kann es zu einer  $\Delta$ -Inflation kommen. Das globale Effektniveau liegt nicht mehr bei 20%. Daher müssen die Effekte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der Effekte nach Bonferroni das bekannteste Verfahren ist.
- D**  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.
- E**  Wenn ein multipler Test gerechnet wird, dann muss der Effekt  $\Delta$  nicht adjustiert werden im Gegensatz zu den p-Werten.

## Lineare Regression & Korrelation

### 43. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

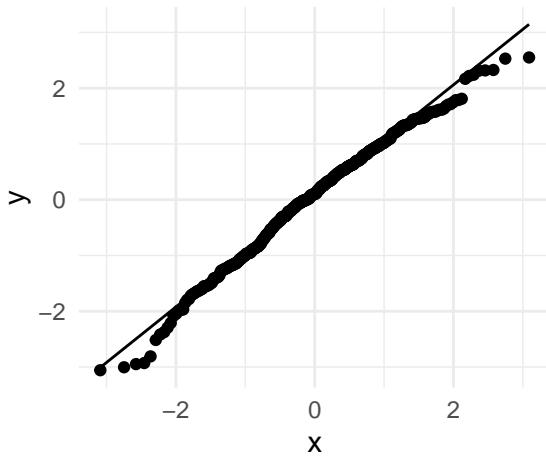
- A**  Wenn ein prädiktives Modell gerechnet werden soll, dann muss zum einen ein Trainingsdatensatz sowie ein Testdatensatz definiert werden. Dabei ist der Trainingsdatensatz meist 1/10 und der Testdatensatz 1/3 der Fallzahl groß. Der Testdatensatz dient zur Validierung.
- B**  Ein prädiktives Modell wird auf einem Trainingsdatensatz trainiert und anschliessend über eine explorative Datenanalyse validiert. Signifikanzen über  $\beta_i$  können hier nicht festgestellt werden.
- C**  Es wird ein Trainingsdatensatz zum Trainieren des Modells benötigt. Der Testdatensatz dient zur Validierung. Dies gilt insbesondere für ein prädiktives Modell.
- D**  Wir modellieren den Zusammenhang zwischen  $X$  und  $Y$  wenn ein prädiktives Modell gerechnet wird. Dabei kann nicht der gesamte Datensatz genutzt werden. Es wird ein Trainingsdatensatz zum Trainieren des Modells benötigt.
- E**  Wenn ein prädiktives Modell gerechnet werden soll dann kann dies auf dem gesamten Datensatz geschehen. Das Ziel ist es einen Zusammenhang von  $X$  auf  $Y$  zu modellieren. Wie wirken sich die Einflussvariablen  $X$  auf den gemessenen Endpunkt  $Y$  aus?

### 44. Aufgabe

(2 Punkte)

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

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



- A**  Wir betrachten die Gerade und dabei insbesondere die beiden Enden der Gerade in dem IQR, also dem ersten und dritten Quartile. Hier sollten die Punkte auf der Geraden liegen, dann ist die Annahme an die Normalverteilung der Residuen erfüllt.
- B**  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.
- C**  Die Annahme der normalverteilten Residuen ist nicht erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- D**  Die Annahme der normalverteilten Residuen ist nicht erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden.
- E**  Wir betrachten die Punkte auf der Geraden. Wenn die Punkte einigermaßen auf der Geraden liegen, dann gehen wir von normalverteilten Residuen aus. Wir können hier von normalverteilten Residuen ausgehen.

## 45. Aufgabe (2 Punkte)

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

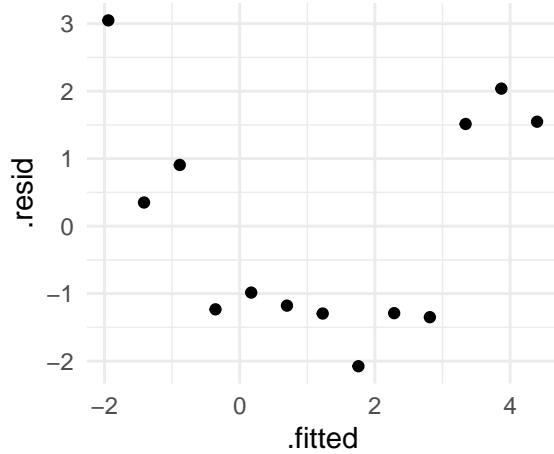
Sie berechnen in Ihrer Abschlussarbeit den Korrelationskoeffizienten  $\rho$ . Welche Aussage über den Korrelationskoeffizienten  $\rho$  ist richtig?

- A**  Der Korrelationskoeffizienten  $\rho$  ist eine standardisierte, statistische Maßzahl, die zwischen 0 und 1 liegt. Dabei ist Korrelationskoeffizienten  $\rho$  einheitslos. Eine Signifikanz kann nicht nachgewiesen werden.
- B**  Der Korrelationskoeffizienten  $\rho$  ist eine standardisierte, statistische Maßzahl, die zwischen -1 und 1 liegt. Dabei ist Korrelationskoeffizienten  $\rho$  einheitslos.
- C**  Korrelationskoeffizienten  $\rho$  liegt zwischen 0 und 1. Darüber hinaus ist der Korrelationskoeffizienten  $\rho$  einheitslos und kann als Standardisierung verstanden werden.
- D**  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.
- E**  Der Korrelationskoeffizienten  $\rho$  ist eine veraltete Darstellungsform von Effekten in der linearen Regression und wird wie das  $\eta^2$  aus der ANOVA interpretiert. Der Korrelationskoeffizienten  $\rho$  beschreibt den Anteil an erklärter Varianz durch die Regression.

## 46. Aufgabe (2 Punkte)

**Inhalt folgender Module:** Angewandte Statistik und Versuchswesen • Biostatistik

Nach der Modellierung einer Regression stellt sich die Frage, ob die Residuen (.resid) gleichmäßig um die gefitte Gerade liegen. Sie können folgende Abbildung für die visuelle Überprüfung der Residuen nutzen. Welche Aussage ist richtig?



- A**  Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Diagonalen. Damit ist das Modell erfolgreich geschätzt worden.
- B**  Wenn wir die Nulllinie betrachten so liegen die Punkte nicht gleichmäßig über und unter der Nulllinie. Unser Modell erfüllt nicht die Annahme von normalverteilten Residuen mit einem Mittelwert von 0 und einer Streuung von  $s^2$ .
- C**  Die Punkte müssen gleichmäßig in dem negativen Bereich liegen. Dies ist hier klar nicht der Fall. Einzelne Ausreißer können beobachtet werden. Die Analyse ist gescheitert.
- D**  Wir betrachten die Nulllinie und alle Punkte sollten ohne Muster gleichmäßig um die Nulllinie liegen. Da dies der Fal ist, gehen wir von keinen Ausreißern aus.
- E**  Die Annahme der normalverteilten Residuen ist erfüllt. Es ist ein Muster zu erkennen und wir können damit auf die Signifikanz von  $x_1, \dots, x_p$  schließen.

## 47. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Sie haben ein Feldexperiment mit Spitzkohl durchgeführt und wollen nun in einer simplen linearen Regression den Einfluss der  $PO_2$ -Konzentration in [ $\mu\text{g}$ ] im Wasser auf das Trockengewicht in [kg] untersuchen. Sie erhalten einen  $\beta_{PO_2}$  Koeffizienten von  $1.1 \times 10^{-5}$  und einen  $p$ -Wert mit  $1e-04$ . Welche Aussage zu der Signifikanz und dem Effekt ist richtig?

- A**  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.
- B**  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.
- 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**  Wenn der Effekt  $\beta_{PO_2}$  winzig ist, dann kann es an einer falsch gewählten Einheit liegen. Der Anstieg von einer Einheit in  $X$  führt ja zu einer Änderung von  $\beta_{PO_2}$  in  $x$ . Wir müssen daher die Einheit von  $y$  entsprechend anpassen.
- E**  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.

## 48. Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Neben der klassischen Regression kann die Funktion `lm()` in auch für welche andere Art von Anwendung genutzt werden?

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

## 49. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Wenn Ihr gemessener Endpunkt nicht einer Normalverteilung folgt, so können Sie dennoch Ihre Daten modellieren. Hierzu nutzen Sie dann das *generalisierte lineare Modell (GLM)*. Welche Aussage ist richtig?

- A**  Das GLM erlaubt auch nicht normalverteilte Residuen in der Schätzung der Regressionsgrade.
- B**  Dank dem *generalisierten linearen Modell (GLM)* können auch andere Verteilungsfamilien – außer die Normalverteilung – mit einer linearen Regression modelliert werden. Dafür werden alle Verteilungen in eine Normalverteilung überführt und anschließend standardisiert.
- C**  Das *generalisierte lineare Modell (GLM)* erlaubt auch weitere Verteilungsfamilien für das  $Y$  bzw. das Outcome in einer linearen Regression zu wählen.
- D**  Das GLM ist eine allgemeine Erweiterung der linearen Regression auf die Normalverteilung.
- E**  Das GLM ist eine Vereinfachung des LM in R. Mit dem GLM lassen sich polygonale Regressionen rechnen. Somit stehen neben der Normalverteilung noch weitere Verteilungen zu Verfügung.

## 50. Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Neben der Mittelwertsdifferenz als Effektschätzer bei normalverteilten Endpunkten wird auch häufig der Effektschätzer Odds ratio bei binären Endpunkten verwendet. Welche Aussage über den Effektschätzer Odds ratio ist im folgenden Beispiel zur Behandlung von Klaueninfektionen bei Schafe richtig? Dabei sind 3 Tiere krank und 7 Tiere sind gesund.

- A**  Der Anteil der Kranken wird berechnet. Da es sich um ein Anteil handelt ergibt sich ein Odds ratio von 0.3.
- B**  Da es sich um ein Chancenverhältnis handelt ergibt sich ein Odds ratio von 3.33.
- C**  Es ergibt sich ein Odds ratio von 0.3, da es sich um eine Chancenverhältnis handelt.
- D**  Das Verhältnis der Chancen Odds ratio ergibt ein Chancenverhältnis von 0.3. Wir sind an der Chance krank zu sein interessiert.
- E**  Es ergibt sich ein Odds ratio von 0.43, da es sich um eine Chancenverhältnis handelt

## Teil I.

# Programmieren in R

## 51. Aufgabe

(9 Punkte)

Inhalt folgender Module: Mathematik & Statistik • Modellierung landwirtschaftlicher Daten • Statistik



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

Nilufar: Jetzt lese ich hier von einem Faktor. Was ist ein Faktor in R? (1 Punkt)

Sie antworten:

Nilufar: Was war eigentlich nochmal ein Vorteil von der Nutzng von R? (1 Punkt)

Sie antworten:

Nilufar: Warum gibt es eigentlich das RStudio und R? Wie unterscheiden sich beide voneinander? (1 Punkt)

Sie antworten:

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

Sie antworten:

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

Sie antworten:

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

Sie antworten:

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

Sie antworten:

Nilufar: Was macht überhaupt dieses c() hier überall im R Code? (1 Punkt)

Sie antworten:

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

Sie antworten:

## 52. Aufgabe

(9 Punkte)



Inhalt folgender Module: Modellierung landwirtschaftlicher Daten • 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 Alex stolz und lacht Sie an. Nur leider kennt er sich überhaupt nicht mit R aus! Das heißt, Sie müssen hier einmal Rede und Antwort stehen und helfen. Sonst wird es für Alex dann in seiner Hausarbeit nichts mit der Auswertung und Abgabe. Das kann auch keine Lösung für Alex und Sie sein. Immerhin haben Sie schon so viel gelernt.

Alex fragt: *Wie nennt sich das Datenformat in R? Bitte mit kurzem Beispiel! (1 Punkt)*

Sie antworten:

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

Sie antworten:

Alex fragt: *Datumsangaben sind schwierig, da es nur ein gültiges Format gibt, was zwischen Programmen funktioniert. Wie lautet das Format? (1 Punkt)*

Sie antworten:

Alex 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:

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

Sie antworten:

Alex fragt: *Ich habe hier einen Gruppenvergleich als Experiment durchgeführt. Wenn ich jetzt die Daten in R reingeladen habe, was muss ich dann als erstes machen und warum nochmal? (2 Punkte)*

Sie antworten:

Alex fragt: *Ich habe den Faktor f<sub>1</sub> und f<sub>2</sub> und möchte den Faktor f<sub>1</sub> getrennt in jedem Level des Faktors f<sub>2</sub> auszuwerten. Wie geht das in der Funktion emmeans()? (1 Punkt)*

Sie antworten:

## Teil II.

# Deskriptive Statistik & Explorative Datenanalyse

### 53. Aufgabe

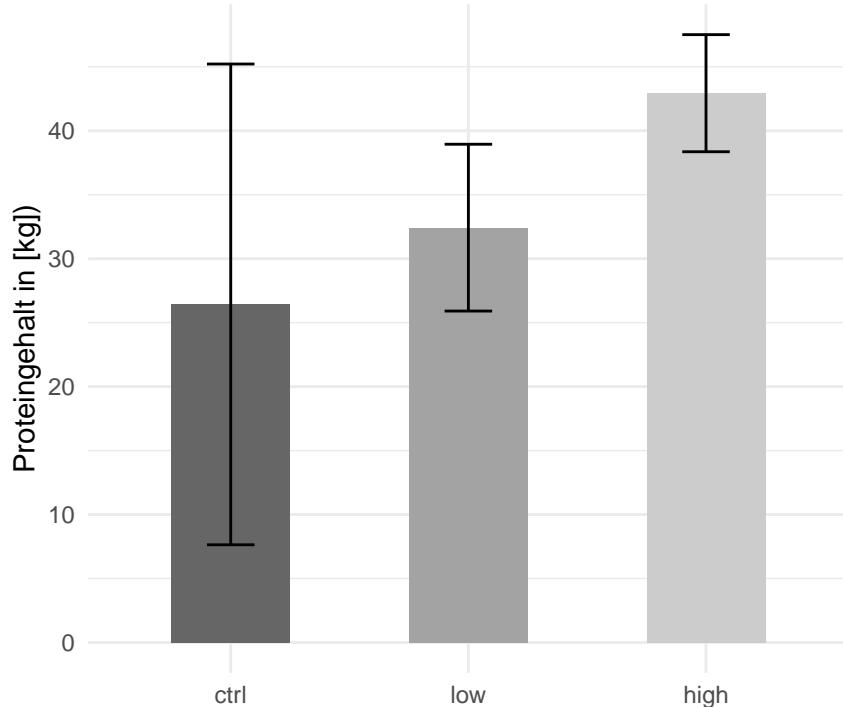
(9 Punkte)

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

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



**Zerforschen des Barplots** Jonas steht vor einem ersten Problem, denn wenn es nach seiner Betreuerin geht, soll er in einem Freilandversuch Erdbeeren auswertet. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Stricken. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Jonas denkt gerne über Stricken nach. Das heißt erstmal überlegen für Jonas. Aus den Boxen wummert Iron Maiden und sein Mund ist verklebt von Snickers. 'Herrlich', denkt Jonas. Die Behandlung werden verschiedene Düngestufen (*ctrl*, *low* und *high*) sein. In seiner Exceldatei wird er den Outcome (*Y*) *Proteingehalt* als *protein* aufnehmen. Vorab soll Jonas aber eimal die folgenden Barplots seiner Betreuerin nachbauen, damit er den R Code schonmal für später vorliegen hat. Damit geht das Problem schon los. Wenn die Erschöpfung nicht wäre, ja dann wäre wohl vieles möglich für Jonas! Aber so..



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

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

## 54. Aufgabe

(8 Punkte)

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

Inhalt folgender Module: Mathematik & Statistik • Statistik • Modellierung landwirtschaftlicher Daten • Angewandte Mathematik & Statistik für Bioverfahrenstechnik



**Visualisierung des Barplots** Yuki steht vor einem ersten Problem, denn wenn es nach seiner Betreuerin geht, soll er in einem Feldexperiment Erdbeeren auswertet. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Am Ende dann doch besser Orchideen. Wunderbar. Eine echte Ablenkung für Yuki. Die Behandlung waren verschiedene Substratarten (*torf*, *40p60n* und *70p30n*). In seiner Exceldatei hat er den Outcome (Y) *Trockengewicht* als *drymatter* aufgenommen. Nun soll Yuki 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. Eine echte Herausforderung für ihn war schon immer die Faulheit gewesen. Ein leidiges Lied. Aber egal. Einfach mal raus um zu Boldern. Ohne Ziel und Uhr. Das ist was für Yuki.

treatment	drymatter
<i>torf</i>	29.4
<i>40p60n</i>	34.4
<i>70p30n</i>	27.6
<i>70p30n</i>	27.3
<i>70p30n</i>	19.5
<i>70p30n</i>	22.7
<i>40p60n</i>	28.5
<i>40p60n</i>	33.3
<i>40p60n</i>	51.5
<i>torf</i>	37.4
<i>40p60n</i>	33.2
<i>torf</i>	43.2

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

## 55. Aufgabe

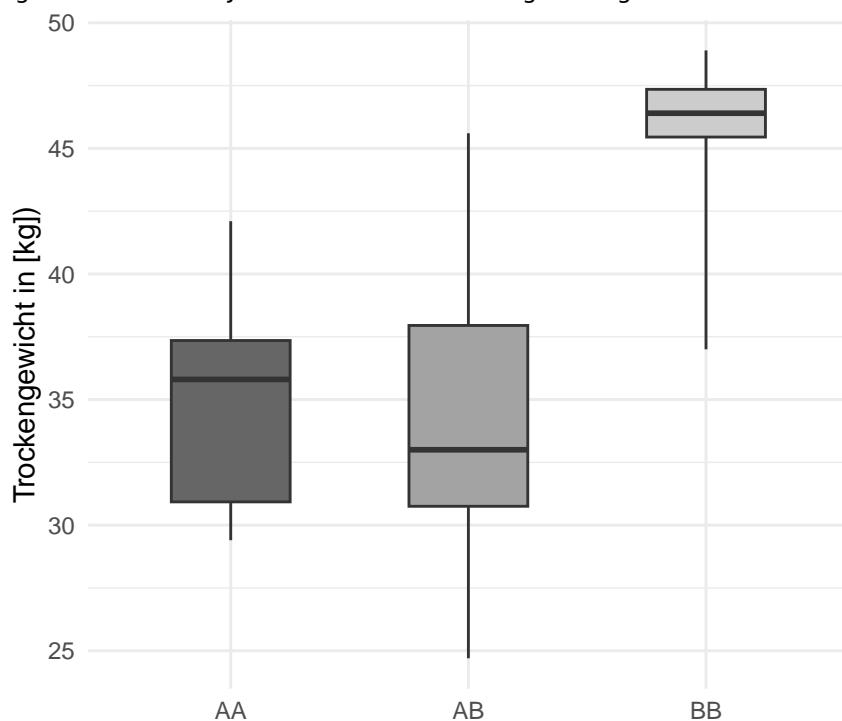
(9 Punkte)

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

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



**Zerforschen des Boxplots** Eine echte Herausforderung für sie war schon immer die Wut gewesen. Ein leidiges Lied. Deshalb gilt anzuschauen, was andere vor einem gemacht haben. Für Tina ist es eine Möglichkeit schneller ans Ziel zu gelangen. Tina soll in ihrer Hausarbeit Brokoli untersuchen. Die Behandlung in ihrer Hausarbeit werden verschiedene Genotypen (AA, AB und BB) sein. Erheben wird Tina als Messwert ( $Y$ ) *Trockengewicht* benannt als *drymatter* in ihrer Exceldatei. Von ihrer Betreuerin erhält sie nun folgende Abbildung von Boxplots, die sie erstmal zur Übung nachbauen soll, bevor sie mit dem eigentlichen Versuch beginnt. Aber nur in passender Atmosphäre! Schon dutzende Male gesehen: Indiana Jones. Aber immer noch großartig zusammen mit Katjes.



Leider kennt sich Tina mit der Erstellung von Boxplots 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 aus der obigen Abbildung der drei Boxplots! *Bachten Sie die korrekte Darstellungsform der statistischen Maßzahlen!* (2 Punkte)
3. Beschriften Sie *einen* der Boxplots mit den gängigen statistischen Maßzahlen! (2 Punkte)
4. Erstellen Sie einen beispielhaften Datensatz, aus dem die drei Boxplots *möglicherweise* erstellt wurden, im R üblichen Format! (2 Punkte)
5. 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 • Modellierung landwirtschaftlicher Daten • Angewandte Mathematik & Statistik für Bioverfahrenstechnik



**Visualisierung des Boxplots** Steffen steht vor einem ersten Problem, denn wenn es nach seinem Betreuer geht, soll er in einem Freilandversuch Kartoffeln auswertet. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Klemmbausteine. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Steffen denkt gerne über Klemmbausteine nach. Die Behandlung waren verschiedene Lichtstufen (*none* und *600lm*). In seiner Exceldatei hat er den Endpunkt (*Y*) *Protein Gehalt* als *protein* aufgenommen. Nun soll Steffen die Daten einmal als Boxplots in einer Präsentation visualisieren, damit seinem Betreuer wieder klar wird, was er eigentlich nochmal gemacht hat und was für ein Ergebnis in einem statistischen Test zu erwarten wäre. Anhand von Boxplots lässt sich eine Aussage über die Normalverteilung von *Y* treffen. Wäre da nicht noch etwas. Wenn die Romantik nicht wäre, ja dann wäre wohl vieles möglich für Steffen! Aber so.. Aber egal. Steffen will später nochmal raus um zu Ringen. Druck ablassen, dass muss er auch.

Behandlung	Messwert
600lm	26.0
none	13.0
none	15.9
none	25.9
600lm	44.8
600lm	34.3
none	31.8
none	19.2
none	20.0
none	18.7
600lm	36.5
600lm	33.6
none	27.9
none	18.8
600lm	39.7
600lm	18.7
600lm	34.7
600lm	36.8
none	22.7
600lm	37.9

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

## 57. Aufgabe

(10 Punkte)

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



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



**Visualisierung des Scatterplots** Wenn es nach Mark ginge, wäre er schon längst fertig mit seiner Abschlussarbeit. Geht es aber nicht. Aus den Boxen wummert Andrea Berg und sein Mund ist verklebt von Marzipankugeln. 'Herrlich', denkt Mark. In seiner Abschlussarbeit hatte er ein Kreuzungsexperiment im Wendland durchgeführt. Nach der Meinung seiner Betreuerin sieht das jedoch etwas anders aus. Jetzt soll er doch noch eine explorative Datenanalyse für den Zusammenhang zwischen mittlerer Anzahl an weißen Blutkörperchen [LEU/ml] und Gewichtszuwachs in der 1LW in Milchvieh durchführen. Wie nervig! Eine echte Herausforderung für ihn war schon immer die Unsicherheit gewesen. Ein leidiges Lied. Da zwei kontinuierliche Variablen vorliegen, geht die explorative Datenanalyse leider nicht mit Boxplots oder Barplots. Dann was anderes. Wenn Columbo läuft, dann ist der Hamster nicht mehr da. Aber jetzt braucht er mal Entspannung!

Gewichtszuwachs in der 1LW	Mittlerer Anzahl an weißen Blutkörperchen [LEU/ml]
1.0	3.9
5.5	6.3
0.5	0.5
4.0	4.7
10.5	8.0
9.0	6.9
3.5	4.3
12.0	11.4

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

1. Erstellen Sie eine Visualisierung für die Datentabelle. Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Schätzen Sie eine Grade durch die Punkte! **(1 Punkt)**
3. Beschriften Sie die Grade 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 Grade verlaufen und welche Werte würden die statistischen Maßzahlen annehmen? **(2 Punkt)**

## 58. Aufgabe

(10 Punkte)

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



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



**Visualisierung des Mosaicplots** Das Verrückte ist, dass die Schlange Harry Potter wirklich liebt. Das ist Steffen sehr recht, denn er braucht Entspannung. Aber Ablenkung hilft nur begrenzt. 'Uff!', denkt sich Steffen. Jetzt hat er doch tatsächlich zwei kategoriale Variablen in seiner Hausarbeit gemessen. Zum einen die Behandlung Herbizideinsatz [ja/nein] und zum anderen die Messung Frischegewicht über Zielwert [ja/nein] im Kontext von Maiss. Hierfür hat er ein Feldexperiment im Oldenburger Land durchgeführt. Jetzt möchte Steffen die Daten einmal in einer explorativen Datenanalyse darstellen. Danach kann er dann über den passenden statistischen Test nachdenken. Dabei unterstützt sein Betreuer diesen Ansatz bevor es in der Datenanalyse weiter geht. So schön wie so gut. Eine echte Herausforderung für ihn war schon immer die Romantik gewesen. Ein leidiges Lied.

Herbizideinsatz	Frischegewicht über Zielwert
ja	nein
ja	ja
ja	ja
ja	ja
nein	ja
nein	ja
ja	ja
ja	nein
ja	ja
nein	ja
ja	ja
nein	ja
nein	ja
ja	nein
ja	ja

Herbizideinsatz	Frischegewicht über Zielwert
nein	ja
ja	ja
nein	ja
nein	ja
ja	nein
ja	ja
ja	ja
ja	ja
nein	ja
ja	ja
nein	ja
ja	ja
nein	ja

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

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

## 59. 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** In einem Gespräch mit ihrer Betreuerin wird Paula gebeten seine Daten aus einem Stallexperiment mit Milchvieh in einem Histogramm darzustellen. Paula schmeißt noch eine Handvoll Smarties in ihren Rachen. Im Hintergrund klirrt leise der Spiegel zum Sound von White Lies. In ihrem Experiment hat er die dunklen Pigmentstörungen erst fotografiert und dann ausgezählt. Laut ihrer Betreuerin soll das Histogramm helfen, die Verteilung der die dunklen Pigmentstörungen zu bestimmen. Es wäre einfacher, wenn da nicht noch was wäre. Paula und der Perfektionismus, eine unendliche Geschichte mit kniffligen Wendungen. Paula nickt im Takt von White Lies und bemerkt dabei gar nicht was die Ratte schon wieder anstellt.

Die dunklen Pigmentstörungen: 5, 2, 5, 11, 10, 5, 2, 4, 3, 3, 4, 4, 3, 4, 5, 3, 4, 5, 5, 6, 4, 2, 5, 3, 7, 2, 2, 8, 8, 6

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

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

## 60. 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** In ihrer Hausarbeit möchte Tina gerne die Daten aus einem Gewächshausexperiment mit Kartoffeln in einem Histogramm darstellen. Das Histogramm erlaubt ihr dabei Rückschlüsse auf die Verteilung über den Endpunkt (Y) zu treffen. Aus den Boxen wummert Tocotronic und ihr Mund ist verklebt von Katjes. 'Herrlich', denkt Tina. In seinem Experiment hat Tina die mittleren Läsionen auf den Blättern gezählt. Es wäre einfacher, wenn da nicht noch was wäre. Tina und die Wut, eine unendliche Geschichte mit kniffligen Wendungen. Wenn Tocotronic ertönt, dann sucht die Spinne schleunigst Schutz unter dem Sofa. Tina schüttelt den Kopf.

Die mittleren Läsionen auf den Blättern: 9.9, 12.8, 9.3, 9.9, 8.6, 9.5, 5.3, 9.4, 10.6, 11, 13.3, 8.7, 9, 5.7, 9.6, 10.1, 5.9, 9.9, 8.1, 12.1, 12.3, 11.3

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

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

## 61. Aufgabe

(10 Punkte)

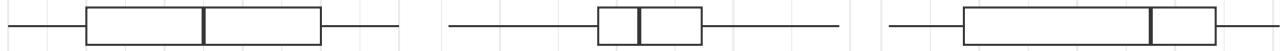
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Visualisierung von Verteilungen** 'Ich glaube, dass es sich hier wieder um so ein kryptisches Lernziel handelt, was nicht so gleich klar ist.', meint Alex und streichelt sanft die Katze. Das Tier versucht dem strammen Griff zu entkommen, gibt aber auf. Tina sieht sich sehr genau die drei liegenden Boxplots an. 'Du weißt doch wie es heißt, *Frei ist, wer missfallen kann.*<sup>1</sup>', merkt Alex nickend an. Das Ziel ist es zu verstehen, wie eine Verteilung anhand eines Boxplots bewertet werden kann. Alex und die Gefälligkeit machen die Sache nicht einfacher.



Jetzt brauchen Alex und Tina 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 Striche! **(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 Mathematik & 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 Tina laut aus dem Skript vor. Jonas war kurz eingenickt und wird mit einem Stoß geweckt. 'Reiz dich zusammen und iss noch ein paar Katjes das hilft mir immer. Alleine komme ich hier nicht weiter.', tadeln Tina Jonas etwas zu forscht. 'War ne lange Nacht', mault Jonas. Beide sollen in ihrer Abschlussarbeit einen statistischen Test interpretieren und versuchen die Grundlagen zu wiederholen. Jonas war auf einem Konzert von Iron Maiden.

Leider kennen sich Tina und Jonas mit der Grundgesamtheit und der Stichprobe überhaupt nicht aus. Daher sind Sie gefragt!

1. 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)**
  - a) Nennen Sie das statistische Verfahren um von einer Grundgesamtheit auf eine Stichprobe zu gelangen! **(1 Punkt)**
  - b) Nennen Sie ein konkretes Beispiel zur Durchführung um von einer Grundgesamtheit auf eine Stichprobe zu gelangen! **(1 Punkt)**
  - c) Benennen Sie die statistische Eigenschaft, die zwischen Grundgesamtheit und Stichprobe vorliegen muss! **(1 Punkt)**
2. 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 Mathematik & Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Das Nullritual - Die statistische Testtheorie** 'Das Känguruh hat mir gerade zugezwinkert. Das macht mir Angst', bemerkt Alex. Es reicht ja schon ein Problem. Alex und die Gefälligkeit, eine unendliche Geschichte mit kniffligen Wendungen. Da braucht es nicht noch Wahnvorstellungen. 'Ich glaube nicht, das Känguruhs zwinkern. Aber was Wichtigeres. Wo hast du eigentlich meine Schnapspralinen hingetan? Wir haben nur noch Gummibärchen von dir. Bäh!', antwortet Nilufar. Beide sind im Zoo und wollen sich von der statistischen Testtheorie ablenken lassen. Eigentlich wollte ja Alex stoppen wie lange Tiere pinkeln<sup>2</sup>, scheiterte aber an einer Oma mit Stock, die die beiden beschimpfte.

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

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

$H_0$  falsch    $H_0$  wahr    $\alpha$ -Fehler    $\beta$ -Fehler

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

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

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

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

## 64. Aufgabe

(9 Punkte)

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



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

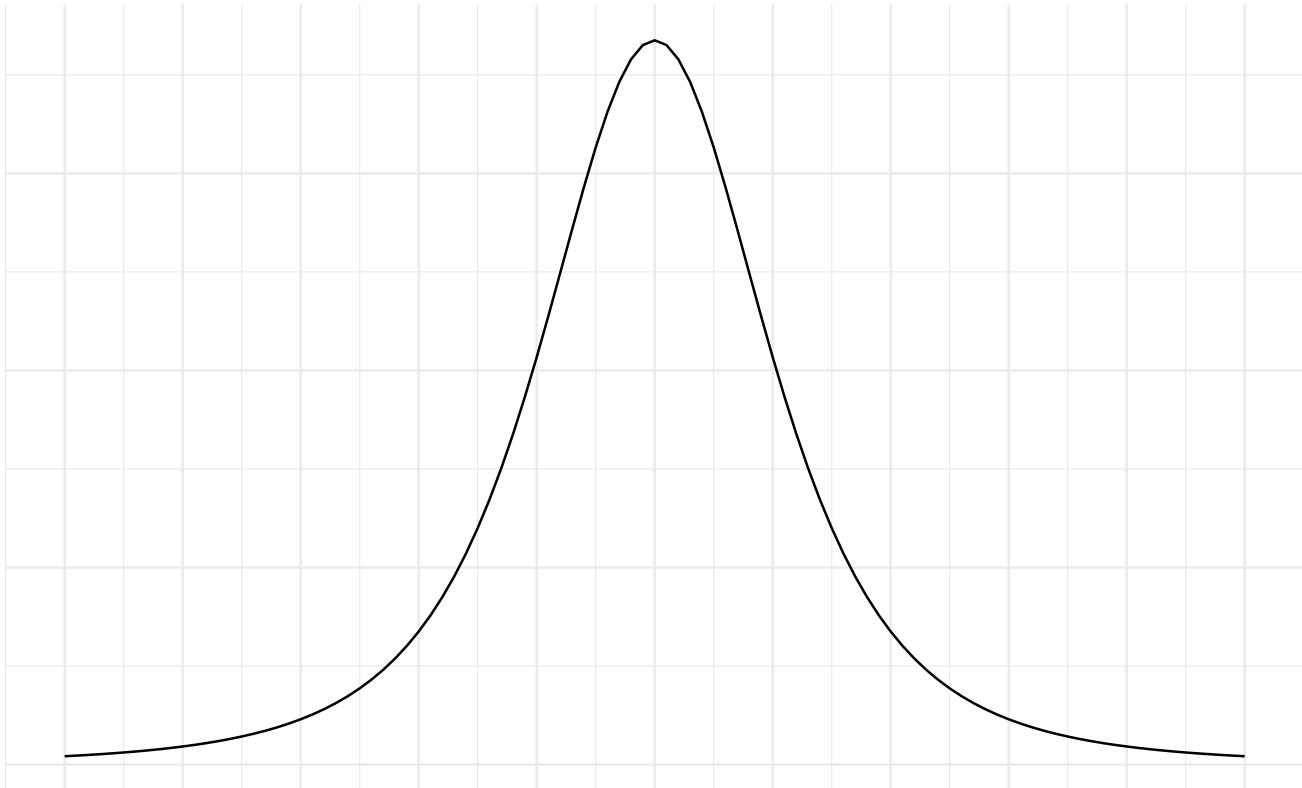


**Visualisierung der Teststatistik  $T_D$  und dem p-Wert** Jessica und Steffen wollten eigentlich einen Flug nach Mallorca buchen, sind jetzt aber dann doch dazu übergegangen nochmal die Aufgaben für die Statistikklausur durchzugehen. 'Kannst du mir nochmal an einer Visualisierung erklären, wie der Zusammenhang zwischen der Teststatistik aus den Daten  $T_D$  und dem p-Wert ist? Ich habe hier zig Fachbegriffe, kriege die abr nicht zusammen...', fragt Jessica. Steffen zuckt mit den Schultern. So genau hatte Steffen jetzt auch nicht aufgepasst. Da hilft dann eventuell das YouTube Video weiter. Jessica mapmt Schokobons und fragt sich, was das alles soll.

Leider kennen sich Jessica und Steffen mit der Visualisierung der Teststatistik  $T_D$  und dem p-Wert überhaupt nicht aus und brauchen dahr 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 „95%“! **(1 Punkt)**
4. Zeichnen Sie  $T_{\alpha=5\%}$  in die Abbildung! **(1 Punkt)**
5. Zeichnen Sie das Signifikanzniveau  $\alpha$  in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Zeichnen Sie  $+T_D$  in die Abbildung! **(1 Punkt)**
7. Zeichnen Sie einen signifikant p-Wert in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**



## 65. Aufgabe

(10 Punkte)



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

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



## 66. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Zusammenhang zwischen dem Effekt, der Streuung sowie der Fallzahl** An einem kalten Dezembermorgen haben sich Mark und Nilufar zum Lernen verabredet. Eine große Kanne Tee und Berge von Marzipankugeln warten darauf gegessen zu werden. Mark 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  $KI_{1-\alpha}$ ?*

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

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

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

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

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

# Teil IV.

## Der 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 Mathematik & Statistik für Bioverfahrenstechnik



**Berechnung des Student t-Test oder Welch t-Test** Der t-Test. Mark erschaudert. Mark und die Unsicherheit, eine unendliche Geschichte mit kniffligen Wendungen. Ein mächtiges Werkzeug ist der t-Test in den Händen desjenigen, der einen normalverteilten Endpunkt ( $Y$ ) hat. Aber erstmal überhaupt den t-Test rechnen können. Wie sah das Experiment von Mark überhaupt aus? Aus den Boxen wummert Andrea Berg und sein Mund ist verklebt von Marzipankugeln. 'Herrlich', denkt Mark. Mark hat ein Kreuzungsexperiment mit Schweinen durchgeführt um eine neue technische Versuchsanlage zu testen. Bei dem Pilotexperiment mit sehr geringer Fallzahl ( $n_1 = n_2 = 3$ ) wurde die Behandlung Lüftungssystem (keins und vorhanden) an den Schweinen getestet und dabei wurde geschaut, ob der Versuch überhaupt technisch klappen könnte. Gemessen hat Mark dann als Messwert Gewichtszuwachs in der 1LW [%/kg]. Warum der Versuch im Emsland für seine Abschlussarbeit stattfinden musste, ist ihm bis heute ein Rätsel. Egal. Gibt es jetzt einen Zusammenhang zwischen der Behandlung und Gewichtszuwachs in der 1LW [%/kg]?

Lüftungssystem	Gewichtszuwachs
keins	18.1
vorhanden	21.9
keins	17.2
keins	16.3
vorhanden	18.8
vorhanden	18.1

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

1. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
2. Bestimmen Sie die Teststatistik  $T_D$  eines Welch t-Tests! **(3 Punkte)**
3. Treffen Sie mit  $T_{\alpha=5\%} = 1.96$  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 Mark 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 Mathematik & Statistik für Bioverfahrenstechnik



**Berechnung des Student t-Test** Der t-Test. Jessica erschaudert. Wenn der Mangel nicht wäre, ja dann wäre wohl vieles möglich für Jessica! Aber so.. Ein mächtiges Werkzeug ist der t-Test in den Händen desjenigen, der einen normalverteilten Endpunkt (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 einen Versuch in einer Klimakammer mit Erbsen durchgeführt. Dabei wurde die Behandlung Bewässerungstypen (*low* und *high*) an den Erbsen getestet. Gemessen hat Jessica dann als Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD]. 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 Chlorophyllgehalt (SPAD-502Plus) [SPAD]?

Bewässerungstypen	Chlorophyllgehalt
high	37.5
low	27.8
high	37.7
high	38.1
low	32.0
low	37.9
high	32.9
low	30.5
low	34.9
low	34.3
high	36.0
high	46.0
high	33.9
low	32.6
low	30.2
high	35.9
low	27.0
low	27.6
high	35.9

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 Student t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 1.84$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effekt des Student t-Tests! **(1 Punkt)**
6. Wenn Sie einen Unterschied zwischen den Behandlungsgruppen erwarten würden, wie groß wäre dann der mindeste Effekt? Begründen Sie Ihre Antwort! **(2 Punkte)**
7. Formulieren Sie eine Antwort an Jessica 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 Mathematik & Statistik für Bioverfahrenstechnik



**Berechnung des Welch t-Test** Jessica ist im Emsland für einen Versuch mit Maiss. Allein diese Tatsache ist für sie eine Erzählung wert. Wenn der Mangel nicht wäre, ja dann wäre wohl vieles möglich für Jessica! Aber so.. Für ihrer Hausarbeit musste sie ein Gewächshausexperiment mit Maiss durchführen und das sollte laut ihrer Betreuerin an diesem Nichtort besonders gut gelingen. Ablenkung gibt es jedenfalls keine. Gar keine. Alleine sein hilft jetzt aber nur bedingt, denn ihre Behandlung Lichtstufen (*none* und *600lm*) und der Messwert Proteingehalt [g/kg] sollen mit einem t-Test ausgewertet werden. Immerhin weiß sie, dass ihr Messwert einer Normalverteilung folgt. Hm..., was entspannendes wäre gut. Um Rad zu fahren geht Jessica dann später nochmal raus. Echte Entspannung.

Lichtstufen	Proteingehalt
none	42.2
600lm	44.4
600lm	37.5
none	48.7
600lm	44.3
none	48.5
none	35.3
600lm	44.9
none	44.1
600lm	43.2
600lm	38.9
none	41.6
none	47.7
600lm	23.5

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  $T_{\alpha=1\%} = 1.96$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie das 99% Konfidenzintervall. Welche Annahmen haben Sie getroffen? **(2 Punkte)**
6. Nennen Sie den statistischen Grund, warum Sie sich zwischen einem Student t-Test und einem Welch t-Test entscheiden müssen! **(1 Punkt)**
7. Formulieren Sie eine Antwort an Jessica über das Ergebnis Ihrer statistischen Analyse! Beziehen Sie dabei das 99% Konfidenzintervall mit in Ihre Antwort ein! **(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 Mathematik & Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik



**Interpretation des t-Tests in R - die Teststatistik und der p-Wert** Almería. Spanien. Sonne und Strand. Steffen und Tina haben ihren gemeinsamen Auslandsaufenthalt sichtlich genossen. Dann hatte sich auch noch angeboten ihre Abschlussarbeit gemeinsam in Almería durchzuführen. Es hätte sogar noch besser funktioniert, wenn Jonas nicht die Erschöpfung ein paar Mal im Weg gestanden hätte und Tina nicht das Problem gehabt hätte die Romantik zu händeln. Nun müssen jetzt alle Daten in R ausgewertet werden, da R international der Standard in der Datenauswertung ist und die Betreuer in Spanien nur R können. Während beide Jonas Oliven mit Snickers füttern, hoffen Steffen und Tina mehr Informationen von Jonas über die seltsame R Ausgabe des t-Tests. Immerhin erinnern beide sich an die Behandlung Bewässerungstypen (*low* und *high*) und das es um Erdbeeren ging. Im Hintergrund wummert Iron Maiden und Fotos zeigen Jonas mit dem Hobby Stricken.

```
##  
## Two Sample t-test  
##  
## data: Chlorophyllgehalt by Bewässerungstypen  
## t = 4.7065, df = 16, p-value = 0.0002376  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## 5.479422 14.460838  
## sample estimates:  
## mean in group low mean in group high  
## 35.14286 25.17273
```

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Berechnen Sie den Effekt des t-Tests! **(1 Punkt)**
5. Skizzieren Sie eine Abbildung in der Sie  $T_D$ ,  $Pr(D|H_0)$ ,  $A = 0.95$ , sowie  $T_{\alpha=5\%} = |2.12|$  einzeichnen! **(4 Punkte)**
6. Beschriften Sie die Abbildung! **(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** Nilufar und Mark sind bei Tina um sich Hilfe in R zu holen. Im Hintergrund wummert Tocotronic. 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 Tina nicht die Wut im Weg stehen würde und Nilufar nicht das Problem hätte die Erwartung zu handeln. In einer Hausarbeit haben beide zusammen Maiss untersucht. Dabei ging es um den Zusammenhang zwischen der Behandlung Lüftungssystemen und Folientunneln (*ctrl* und *tornado*) und dem Messwert Trockengewicht [kg/ha]. Der Versuch wurde in einem Versuch in einer Klimakammer im Emsland durchgeführt. Nach der Betreuerin ist der Messwert Trockengewicht [kg/ha] normalverteilt und ein t-Test passt daher. Das wird jetzt nicht mehr angezweifel...Tina überlegt, ob sie die beiden nicht noch auf den Film *Indiana Jones* einlädt.

```
##  
## Two Sample t-test  
##  
## data: Trockengewicht by Lüftungssystemen  
## t = -4.7294, df = 18, p-value = 0.0001673  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -21.417882 -8.242118  
## sample estimates:  
## mean in group ctrl mean in group tornado  
## 29.70 44.53
```

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

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

## 72. Aufgabe

(9 Punkte)

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



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



**Interpretation des t-Tests in R - die Visualisierung** 'Mit dem R Paket {emmeans} können wir gleich die Gruppenvergleiche rechnen und uns das *compact letter displac*' ausgeben lassen!', verkündet Nilufar sichtlich stolz. Ein paar Mal hat sie schon die Erwartung gehindert weiterzumachen. 'Nach Meinung der Betreuerin soll es aber nur erstmal ein t-Test sein. Und die Ausgabe ist schon wirr genug.', merkt Yuki an. Yuki und Mark sind bei Nilufar um sich in R helfen zu lassen. Im Hintergrund wummert Deichkind. Mark streichelt zur Beruhigung das Huhn von Nilufar. Die beiden waren 2 Monate im Emsland um einen Versuch mit Puten in einem Leistungssteigerungsversuch durchzuführen. Ziel war es das Outcome (Y) Protein/Fettrate [%/kg] zu bestimmen. Nilufar überlegt, ob sie die beiden nicht noch auf den Film *Star Trek* einlädt oder dann doch lieber raus geht um zu Kicken? Vielleicht will ja Mark mit. Besser als der Film.

```
##  
## Two Sample t-test  
##  
## data: Protein/Fettrate by Genotypen  
## t = 4.3255, df = 17, p-value = 0.0004591  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## 6.862313 19.931020  
## sample estimates:  
## mean in group AA mean in group BB  
## 38.16667 24.77000
```

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

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

### 73. Aufgabe

(11 Punkte)

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



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



**Berechnung des gepaarten t-Test** Yuki und Jessica haben sich dazu entschieden zusammenzuarbeiten. Das sollte alles etwas einfacher machen. Jeder hat zwar ein getrenntes Themenfeld aber den Hauptversuch machen beide gemeinsam. Das hat sich schonmal als gut Idee soweit herausgestellt. In einem Projektbericht sollen beide herausfinden, ob es einen Zusammenhang zwischen Bestandsdichte (*hoch* und *niedrig*) und Protein/Fettrate [%/kg] gibt. Die Besonderheit ist hierbei, dass die Messungen an der gleichen Beobachtung stattfinden. Beide messen also zweimal an den gleichen Fleischrindern. Hier muss dann wohl auf ein normalverteiltes Outcome ( $Y$ ) ein gepaarter t-Test gerechnet werden. Yuki schaut etwas flehentlich zu Jessica. Yuki und die Faulheit, eine unendliche Geschichte mit kniffligen Wendungen.. Steffen denkt derweil angestrengt an David Bowie und wippt leicht mit dem Fuß.

ID	treatment	freshmatter
8	hoch	35.7
7	niedrig	18.2
10	hoch	37.7
5	niedrig	30.5
5	hoch	46.8
2	niedrig	17.0
1	niedrig	13.8
2	hoch	39.6
1	hoch	45.1
3	niedrig	33.1
4	hoch	47.8
6	niedrig	14.4
6	hoch	33.3
3	hoch	40.9
4	niedrig	26.5
8	niedrig	31.2
9	hoch	38.2
7	hoch	38.1

Leider kennen sich Yuki und Jessica mit der Berechnung eines gepaarten t-Tests überhaupt nicht aus. Deshalb brauchen sie beide bei der Berechnung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Bestimmen Sie die Teststatistik  $T_D$  eines gepaarten t-Tests! **(3 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 2.04$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Schätzen Sie den  $p$ -Wert des gepaarten t-Tests ab! Begründen Sie Ihre Antwort mit einer Skizze! **(2 Punkte)**
6. Formulieren Sie eine Antwort an Yuki über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**

## 74. Aufgabe

(10 Punkte)

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



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



**Interpretation des gepaarten t-Tests in R** Tina und Yuki haben sich dazu entschieden zusammenzuarbeiten. Das sollte alles etwas einfacher machen. Jeder hat zwar ein getrenntes Themenfeld aber den Hauptversuch machen beide gemeinsam. Das hat sich schonmal als gut Idee soweit herausgestellt. Wenn die Wut nicht wäre, ja dann wäre wohl vieles möglich für Tina! Aber so.. In einer Hausarbeit sollen beide herausfinden, ob es einen Zusammenhang zwischen Jäten (*Aussaat* und *Ernte*) und Trockengewicht [kg/ha] gibt. Die Besonderheit ist hierbei, dass die Messungen an der gleichen Beobachtung stattfinden. Beide messen also zweimal an den gleichen Brokkoli. Hier muss dann wohl auf einen normalverteilten Messwert (Y) ein gepaarter t-Test gerechnet werden. Leider kennen sich beide nicht sehr gut in R aus. Aber wenigstens haben beide eine Menge an Katjes.

```
##  
## Paired t-test  
##  
## data: Trockengewicht by Jäten  
## t = -2.3923, df = 8, p-value = 0.0437  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -11.1944981 -0.2055019  
## sample estimates:  
## mean difference  
## -5.7
```

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

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

## Teil V.

# Die einfaktorielle & zweifaktorielle ANOVA

### 75. Aufgabe

(11 Punkte)

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



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



**Visualisierung der einfaktoriellen ANOVA** 'Uff... die einfaktorielle ANOVA. Und wir können jetzt anhand der Visualisierung sehen, ob da schon was signifikant ist?', Alex hebt die Augenbraue. 'Ja, können wir. Dafür müssen wir aber erstmal in {ggplot} uns die Daten anschauen. Oder wir zeichnen es flott mit der Hand. Geht auch.', meint Paula dazu. Die beiden hatten sich auf einem Konzert von White Lies kennengelernt. Alex hatte sich in ein Kreuzungsexperiment verschiedene Zandern angeschaut. 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 Jagd auf roter Oktober geguckt. Paula befürwortet das!

Elterlinie	Fettgehalt
Standard	45
Yray	35
Yray	35
Standard	45
Xray	25
Yray	34
Yray	36
Xray	25
Standard	45
Xray	24
Standard	44
Yray	36
Xray	25
Standard	47
Xray	25
Yray	35
Yray	35
Xray	25

Leider kennen sich Alex und Paula 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)**

## 76. Aufgabe

(10 Punkte)

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



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



**Ergebnistabelle der einfaktoriellen ANOVA** 'Uff... die einfaktorielle ANOVA. Und wie füllen wir jetzt die Tabelle der ANOVA aus und schauen, ob da was signifikant ist?', Tina hebt die Augenbraue. 'Das ist eine sehr gute Frage. Ich glaube man kann alles in der Tabelle relativ einfach mit wenigen Informationen berechnen.', meint Steffen dazu. Da hilft die Schlange von Steffen auch nur bedingt. Tina hatte sich in ein Kreuzungsexperiment verschiedene Puten angeschaut. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Ernährungszusatz (*ctrl*, *fedX* und *getIt*) und dem Messwert Fettgehalt [%/kg] gibt. Nachher wollen sich beide noch mit dem Hobby Klemmbausteine von Steffen beschäftigen. Kennt Tina noch nicht, klingt aber interessant.

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

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Ernährungszusatz</b>	2	3145.76			
<b>error</b>		17			
<b>Total</b>	19	3784.55			

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

## 77. Aufgabe

(10 Punkte)

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



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



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

```
## Analysis of Variance Table
##
## Response: Schlachtgewicht
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Genotypen  2 779.43 389.71  28.468 1.409e-06
## Residuals 20 273.79 13.69
```

Leider kennen sich Alex 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! **(2 Punkte)**
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)**

## 78. 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** Steffen und Jessica 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 Schokobons mampfen, die Jessica mitgebracht hat. Nun möchte erstmal ihre Betreuung der Arbeit eine ANOVA Tabelle sehen. Was immer da auch drin zu erkennen sein mag. Steffen schaut Jessica seien erstmla gar nichts. Die beiden waren im Emsland um ein Feldexperiment mit Erdbeeren durchzuführen. Dabei haben Steffen und Jessica den Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] unter der Behandlung Düngestufen (*ctrl, low, mid* und *high*) ermittelt. Später wollen die beiden dann noch raus um Rad zu fahren.

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

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Düngestufen</b>	3	7.09			
<b>Error</b>	30	191.52			

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 unterstehenden Tabelle mit  $T_{\alpha=5\%} = 2.03$ . Begründen Sie Ihre Auswahl! **(3 Punkte)**

Düngestufen	Fallzahl (n)	Mittelwert	Standardabweichung
ctrl	9	3.11	1.54
low	7	4.14	0.90
mid	9	4.22	2.39
high	9	3.56	3.91

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

## 79. 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 Versuch in einer Klimakammer wurden Erdbeeren mit dem Behandlung Düngestufen (*ctrl*, *low*, *mid* und *high*) sowie der Behandlung Bewässerungstypen (*ctrl*, und *high*) untersucht. Es wurde als Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] bestimmt. Nilufar ahnte schon, dass es komplexer wird, als sie mit ihrem Projektbericht angefangen hat. Das es jetzt aber so kompliziert wird, hätte sie jetzt aber auch nicht gedacht. Nilufar kratzt sich am Kopf. Nilufar mampft aus Frust noch eine Handvoll Takis Blue Heat. Eventuell muss sie dann doch nochmal Hilfe in der statistischen Beratung holen. Jetzt versucht sie es aber erstmal selber. Und eigentlich wollte Nilufar doch noch ihrem Hobby nachgehen! Nilufar liebt Hip Hop. Darin kann sie sich wirklich verlieren und immer wieder neu begeistern.

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

1. Formulieren Sie die wissenschaftliche Fragestellung für beide Faktoren separat! **(2 Punkte)**
2. Formulieren Sie die statistischen Hypothesenpaare für beide Faktoren separat! **(2 Punkte)**
3. Füllen Sie die unterstehende zweifaktorielle ANOVA Ergebnistabelle aus! **(3 Punkte)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>Düngestufen</b>	3	334.48			
<b>Bewässerungstypen</b>	1	2.77			
<b>Düngestufen:Bewässerungstypen</b>	3	403.18			
<b>Error</b>	18	232			

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

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

5. Was sagt der Term *Düngestufen:Bewässerungstypen* aus? Interpretieren Sie das Ergebnis! **(2 Punkte)**

## 80. Aufgabe

(12 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** 'Mit der zweifaktoriellen ANOVA lässt sich die Interaktion zwischen den beiden Behandlungen nachweisen!', seine Betreuerin scheint die zweifaktoriellen ANOVA zu verstehen. Warum jetzt er jetzt nochmal alles wiederkäuen muss, wird Steffen echt nicht so klar. Wenn es doch so klar ist? Steffen war im Wendland und hatte dort ein Feldexperiment mit Kartoffeln durchgeführt. Die Komune wo er untergekommen war, war cool gewesen. Nur jetzt muss eben das Experiment fertig ausgewertet werden. Steffen hatte zwei Behandlungen auf Kartoffeln angewendet. Einmal Bewässerungstypen (*ctrl*, *low*, *mid* und *high*) sowie als zweite Behandlung Lüftungssystemen und Folientunneln (*ctrl*, und *tornado*). Gemessen wurde der Messwert (*Y*) Trockengewicht [kg/ha]. Jetzt muss das hier zu einem Ende kommen! Eigentlich wollte Steffen nachher noch einen Film schauen. Irgendwie komisch, wenn er Harry Potter anmacht, dann ist die Schlange eigentlich sofort vor dem Bildschirm und starrt hinein.

```
## Analysis of Variance Table
##
## Response: Trockengewicht
##                               Df Sum Sq Mean Sq F value    Pr(>F)
## Bewässerungstypen            3 99.69 49.844  5.1239  0.014027
## Lüftungssystemen             1 96.70 96.699  9.9404  0.004304
## Bewässerungstypen:Lüftungssystemen 3 618.78 309.391 31.8045 1.786e-07
## Residuals                   24 233.47   9.728
```

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

1. Formulieren Sie die wissenschaftliche Fragestellung für beide Faktoren separat! **(2 Punkte)**
2. Formulieren Sie die statistischen Hypothesenpaare für beide Faktoren separat! **(2 Punkte)**
3. Interpretieren Sie das Ergebnis der zweifaktoriellen ANOVA! Begründen Sie Ihre Antwort! **(3 Punkte)**
4. Zeichnen Sie eine Abbildung, der dem obigen Ergebnis der zweifaktoriellen ANOVA näherungsweise entspricht! **(5 Punkte)**

## 81. 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** Es ist schon kurz nach fünf und Jonas wird langsam nervös. Jonas wollte heute Abend noch seine E-Sport Qualifikation schauen und dann zum Sport. Stattdessen versucht sein Betreuer die Ausgabe der zweifaktoriellen ANOVA zu visualieren und zu überprüfen, ob es mit der Visualisierung der Daten als Boxplots zusammenpasst. Es liegt anscheinend eine signifikante Interaktion vor? Jonas hatte im Wendland ein Freilandversuch mit Spargel durchgeführt. Es gab dabei zwei Behandlungen. Einmal Düngestufen (*ctrl, low, mid* und *high*) sowie als zweite Behandlung Lüftungssystemen und Folientunneln (*ctrl*, und *tornado*). Gemesen wurde der Messwert (Y) Proteingehalt [g/kg]. So kompliziert kann das jetzt doch nicht sein! Eigentlich wollte Jonas nachher noch zum Sport. Um zu Schwimmen geht Jonas dann später nochmal raus. Echte Entspannung.

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

1. Visualisieren Sie folgende mögliche Interaktionen zwischen den Behandlungen! Heben Sie die Positionen der auftretenden Interaktion (farblich) hervor! 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. Schreiben Sie das Modell mit Interaktionsterm in für eine zweifaktorielle ANOVA! **(1 Punkt)**
5. Skizzieren Sie die Funktion in für eine Post-hoc Analyse in {emmeans}! **(1 Punkt)**
6. Wenn eine signifikante Interaktion in den Daten vorliegt, wie ist dann das weitere Vorgehen? Berücksichtigen Sie dabei die Verwendung von {emmeans}! **(2 Punkte)**

## 82. Aufgabe

(10 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 Tina wird langsam nervös. Tina wollte heute Abend noch ihre E-Sport Qualifikation schauen. Hoffentlich kommt sie noch rechtzeitig zum Streamen. Angestrengend krault sie die Spinne. 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 0.052 raus sowie eine F-Statistik  $F_D$  mit 1.78 berechnet. Nach den Boxplots müsste sich eigentlich ein Unterschied zwischen *kontakt* und *standard* ergeben. Der Unterschied ist in {emmeans} auch signifikant mit einem p-Wert von 0.021. Wie kann das sein?', grummelt ihr Betreuer. Tina hatte im Wendland ein Stallexperiment mit Zandern durchgeführt. Dabei wurden die Daten  $D$  erhoben. Es gab dabei eine Behandlungen Bestandsdichte (*standard*, *eng*, *weit* und *kontakt*). Gemesen wurde der Messwert (Y) Gewichtszuwachs in der 1LW. So kompliziert kann das jetzt doch nicht sein! Tina hat schon genug Probleme. Wenn die Wut nicht wäre, dann wäre es einfacher.

### Gegebene Formeln

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

Leider kennen sich Tina und 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. Welche Auswirkung hat die Verletzung der Annahme der Varianzhomogenität für die Teststatistik  $F_D$  der ANOVA? Begründen Sie Ihre Antwort! **(3 Punkte)**
6. Erklären Sie abschließend die Diskrepanz zwischen den Ergebnis der ANOVA und dem paarweisen Gruppenvergleich in {emmeans}! **(2 Punkte)**

## Teil VI.

# Multiple Gruppenvergleiche

### 83. 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** Mark und Jessica untersuchen gemeinsam in ihrer Abschlussarbeit den Messwert Proteingehalt [g/kg] in Erbsen. Es ist einiges schiefgelaufen, wie es immer so passiert. Hauptsächlich waren es Würmer, auch wenn man erstmal bei dem Messwert nicht unbedingt an Würmer denken würde. Aber das ist eine andere Geschichte. Jetzt wollen Mark und Jessica ihre Ergebnisse nochmal mit einer Studie von Meyer et al. (2021) vergleichen und schauen, ob was ähnliches rauskommt. Angeschaut wurde sich als Behandlung Genotypen (00, AA, AB und BB). Es ergab sich dann die folgende Tabelle der rohen p-Werte für die Vergleiche zu Meyer et al. (2021).

Rohen p-Werte	Adjustierte p-Werte	Nullhypothese ablehnen?
0.070		
0.012		
0.340		
0.060		

Leider kennen sich Mark und Jessica 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 Bonferroni-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. Welche Adjustierung wird im Allgemeinen vorgezogen? Die Adjustierung der  $p$ -Werte oder die Adjustierung des Signifikanzniveaus  $\alpha$ ? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 84. Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Statistik • Angewandte Mathematik & 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 Puten beschäftigt. Ernährungszusatz (*ctrl*, *fedX*, *proteinX* und *getIt*) und Schlachtgewicht [kg] wurden dort bestimmt. So richtig schlau, wird er daraus nicht.

Behandlung	Compact letter display
<i>ctrl</i>	BC
<i>fedX</i>	A
<i>proteinX</i>	C
<i>getIt</i>	B

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)**

## 85. 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** Jonas hatte in seine Hausarbeit ein Stallexperiment durchgeführt. Soweit so gut. Dabei hat er sich mit Zandern beschäftigt. Angeblich der neueste heiße Kram... aber das ist wiederum was anderes. So richtig mitgenommen hat Jonas das Thema dann doch nicht. Hat er sich doch mit Ernährungszusatz (*ctrl*, *fedX*, *proteinX* und *getIt*) und Fettgehalt [%/kg] schon eine Menge an Daten angeschaut. Nach seine Betreuerin soll er nun ein CLD bestimmen. Weder weiß er was ein CLD ist, noch war sein erster Gedanke mit Köln und die LGBTQ Community richtig... Als erstes solle er die Gruppen nach absteigender Effektstärke sortieren. Was immer das jetzt bringen soll.

Ernährungszusatz	Fallzahl (n)	Mittelwert	Standardabweichung
<i>ctrl</i>	7	11.57	3.02
<i>fedX</i>	9	10.28	2.79
<i>proteinX</i>	7	6.23	1.45
<i>getIt</i>	9	10.13	1.53

Leider kennen sich Jonas 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! Nehmen Sie einen kritischen Wert  $T_{\alpha=5\%}$  von 1.96 an! **(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 Jonas! **(1 Punkt)**

# Teil VII.

## Der Chi-Quadrat-Test

### 86. Aufgabe

(10 Punkte)

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



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



**Den Chi-Quadrat-Test berechnen** Am Ende war es für Mark in seinem Projektbericht dann doch kein normalverteiltes Outcome. Das was jetzt etwas doof, da er sich auf eine ANOVA gefreut hatte. Dann noch schnell Andrea Berg auf das Ohr und los gehts. Prinzipiell ginge das auch irgendwie, aber nun möchte sein Betreuer gerne einen  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Mark hatte sich in ein Feldexperiment  $n = 170$  Beobachtungen von Maiss angeschaut. Dabei hat er als Behandlung *Pestizideinsatz [ja/nein]* bestimmt und zum anderen die Variable *Frischegewicht über Zielwert [ja/nein]* ermittelt. Jetzt muss Mark mal schauen, wie er das jetzt rechnet. Eigentlich wollte Mark nachher noch zum Sport. Einfach mal raus um zu Reiten. Ohne Ziel und Uhr. Das ist was für Mark.

	38	51	
	43	38	

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

## 87. Aufgabe

(11 Punkte)

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



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



**Den Chi-Quadrat-Test mit Effektmaß berechnen** Am Ende hätte Tina dann doch einen normalverteilten Endpunkt in ihrem Projektbericht nehmen sollen. Vor ihr liegen jetzt die Daten von zwei Variablen als Kategorien oder wie es in so schön heißt, als Faktoren. Aber immerhin, hofft sie das was bei den Daten rausgekommen ist. Dann noch schnell Tocotronic auf das Ohr und los gehts. Gezählt hat Tina einiges mit  $n = 122$  Beobachtungen von Puten. Zum einen hat sie als Behandlung *Automatische Fütterung [ja/nein]* bestimmt und zum anderen die Variable *Gewichtszuwachs erreicht [ja/nein]* ermittelt. Nun möchte ihre Betreuerin gerne einen  $\chi^2$ -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Eigentlich wollte Tina nachher noch einen Film schauen. Wenn Indiana Jones läuft, dann ist die Spinne nicht mehr da. Aber jetzt braucht sie mal Entspannung!

	38	19	
	27	38	

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test! **(3 Punkte)**
4. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem  $\chi^2_{\alpha=5\%} = 5.61$ ! 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)**

## 88. Aufgabe

(10 Punkte)

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



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



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

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

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

f3verband	trifft gar nicht zu	trifft nicht zu	weder noch	trifft zu	trifft voll zu	
ja	4	9	11	29	17	
nein	12	29	13	10	1	

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Berechnen Sie die Teststatistik  $\chi_D^2$  eines Chi-Quadrat-Test! Ignorieren Sie Zellbelegungen kleiner gleich fünf in der Berechnung von  $\chi_D^2$  und Runden Sie den erwarteten Wert E ganzzahlig! **(2 Punkte)**
4. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem  $\chi_{\alpha=5\%}^2 = 11.56$ ! 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)**

## Teil VIII.

# Lineare Regression & Korrelation

## 89. Aufgabe

(10 Punkte)

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

Inhalt folgender Module: Statistik • Angewandte Mathematik & 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 `lm()` aber hier wollen wir, glaube ich, eine Gerade durch die Punkte zeichnen.', merkt Alex an. 'Ich sehe keine Punkte...', antwortet Paula sichtlich übernächtigt. 'Wir müssen die Daten ja auch erst visualisieren!', spricht Alex sehr deutlich und langsam. Die beiden hatten ein Stallexperiment im Wendland mit Lamas durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: mittlerer Eisenkonzentration [Fe/ml] und Schlachtgewicht [kg]. Jetzt wollen sie erstmal schauen, ob es einen Zusammenhang gibt.

Mittlerer Eisenkonzentration [Fe/ml]	Schlachtgewicht [kg]
23.3	22.7
23.8	26.2
22.8	23.9
14.9	19.1
26.9	28.5
20.7	24.0
15.2	17.9
23.0	21.6
17.2	23.3
24.8	27.9
11.0	13.3

Leider kennen sich Alex und Paula 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! Begründen Sie Ihre Antwort mit einer Skizze der Methodik an ausgewählten Punkten! **(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)**

## 90. Aufgabe

(12 Punkte)

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



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



**Interpretation der Ergebnisse einer linearen Regression** 'Ich glaube du bringst da was durcheinander. Wir nutzen zwar auch für die ANOVA die Funktion `lm()` aber hier wollen wir, glaube ich, eine Gerade durch die Punkte zeichnen.', merkt Nilufar an. 'Ich sehe keine Punkte... ich sehe nur zwei Zeilen einer Tabelle und ich glaube du hast gerade was gelöscht.', antwortet Mark sichtlich übernächtigt. 'Wir müssen die Koeffizienten der linearen Regression ja auch erst interpretieren!', spricht Nilufar sehr deutlich und langsam. Die beiden hatten einen Leistungssteigerungsversuch im Oldenburger Land mit Puten durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittlicher Bewegungsscore [Movement/h] und Schlachtgewicht [kg]. Das Bestimmtheitsmaß  $R^2$  hatten die beiden mit 0.2 bestimmt. Jetzt wollen sie erstmal schauen, ob es einen Zusammenhang gibt und das soll mit der Ausgabe möglich sein.

term	estimate	std.error	t statistic	p-value
(Intercept)	3.87	1.97		
Durchschnittlicher Bewegungsscore	0.34	0.20		

Leider kennen sich Nilufar und Mark 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. Formulieren Sie die Regressionsgleichung! **(1 Punkt)**
3. Erstellen Sie eine Visualisierung der `lm()`-Ausgabe. Beachten Sie die Informationen zum Bestimmtheitsmaß  $R^2$  aus dem Aufgabentext! Beschriften Sie die Achsen! **(2 Punkte)**
4. Beschriften Sie die Visualisierung mit den statistischen Maßzahlen der der `lm()`-Ausgabe! **(2 Punkte)**
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)**

## 91. Aufgabe

(11 Punkte)

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



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



**Interpretation der Ergebnisse einer linearen Regression in R** 'Hä? Was ist denn das? Das wird ja immer wilder! Hatten wir das als Aufgabe eine lineare Regression zu rechnen? Wir bauen aus kontinuierlichen Daten eine Abbildung und interpretieren diese dann?', fragt Tina. Mark schaut fragend zurück. 'Keine Ahnung... das ist jetzt jedenfalls keine Abbildung von irgendwas sondern eine R Ausgabe mit ganz wilden Bezeichnungen...', antwortet Mark leicht angespannt. Die beiden hatten ein Freilandversuch in der Uckermark mit Spargel durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Regenwurmdichte [Anzahl/l] und Chlorophyllgehalt (SPAD-502Plus) [SPAD]. Jetzt haben die beiden eigentlich alles zusammen. *Eigentlich...*, denn mit der R Ausgabe haben beide jetzt ein Problem.

```
##  
## Call:  
## Chlorophyllgehalt ~ Durchschnittliche_Regenwurmdichte  
##  
## Residuals:  
##      Min      1Q  Median      3Q      Max  
## -3.2149 -0.9110  0.1736  0.7860  3.0845  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)                2.3844    1.6287   1.464   0.151  
## Durchschnittliche_Regenwurmdichte  0.1054    0.1624   0.649   0.520  
##  
## Residual standard error: 1.495 on 41 degrees of freedom  
## Multiple R-squared:  0.01018, Adjusted R-squared:  -0.01396  
## F-statistic: 0.4216 on 1 and 41 DF,  p-value: 0.5198
```

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

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

## 92. Aufgabe

(10 Punkte)

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



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



**Interpretation der Ergebnisse einer Korrelationsanalyse in R** 'Hä? Was ist denn das? Das wird ja immer wilder! Hatten wir als Aufgabe eine Korrelation zu berechnen? Wir bauen aus kontinuierlichen Daten eine Abbildung und interpretieren diese dann?', fragt sich Mark laut. 'Keine Ahnung... das ist jetzt jedenfalls keine Abbildung von irgendwas sondern eine R Ausgabe mit ganz wilden Bezeichnungen...', denkt er. Aber das hilft jetzt auch nicht. Wenn Columbo läuft, dann ist der Hamster nicht mehr da. Aber jetzt braucht er mal Entspannung! Mark hatte ein Feldexperiment im Emsland mit Maiss durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittliche Regenwurmdichte [Anzahl/l] und Trockengewicht [kg/ha]. Jetzt hat er eigentlich alles zusammen. Eigentlich..., denn mit der R Ausgabe hat Mark jetzt ein Problem. Wenn die Unsicherheit nicht wäre, ja dann wäre wohl vieles möglich für Mark! Aber so..

```
##  
## Pearson's correlation  
##  
## data: Durchschnittliche Regenwurmdichte and Trockengewicht  
## t = 3.1246, df = 8, p-value = 0.01413  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2095603 0.9346933  
## sample estimates:  
## cor  
## 0.7413671
```

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

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

### 93. Aufgabe

(12 Punkte)

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

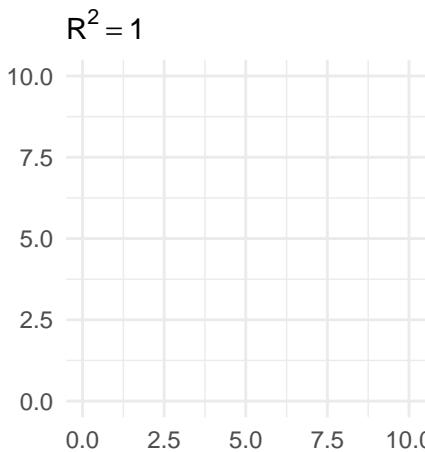


Inhalt folgender Module: Statistik • Angewandte Mathematik & Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

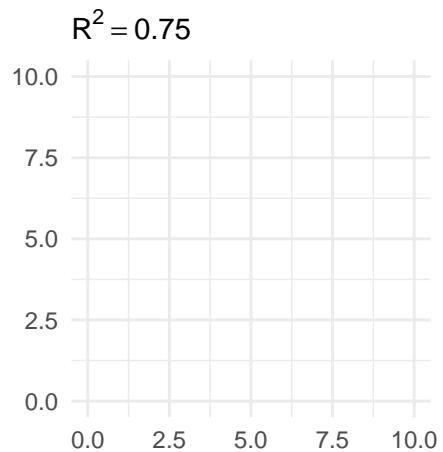


**Visualisierung der Korrelation und des Bestimmtheitsmaßes** Wenn Mission Impossible läuft, dann ist das Meerschweinchen nicht mehr da. Aber jetzt braucht er mal Entspannung! Da hilft dann die Aufgabe auch nur bedingt. 'Hm..., drei leere Abbildungen. Was soll ich da jetzt machen?', fragt sich Jonas und mampft noch ein paar Snickers in sich hinein. Jonas kennt sich nur begrenzt bis gar nicht mit der linearen Regresion und Korrelation aus.

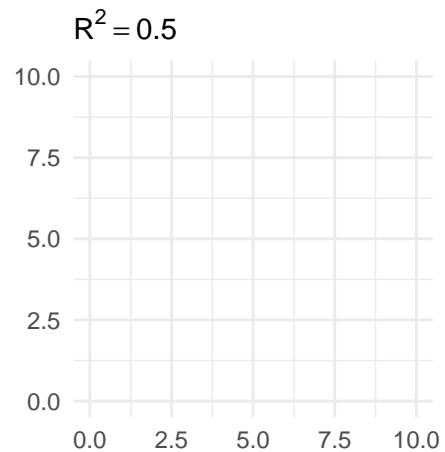
Pearsons  $\rho = 0$



Pearsons  $\rho = 1$



Pearsons  $\rho = 0.75$



Leider kennt sich Jonas mit der Korrelationsanalyse und der linearen Regression überhaupt nicht aus. Deshalb braucht er 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 Punktewolke 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)

## 94. Aufgabe

(12 Punkte)

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

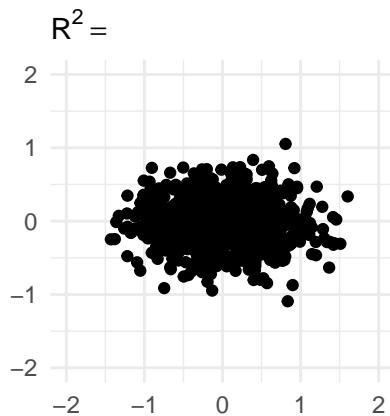


Inhalt folgender Module: Statistik • Angewandte Mathematik & Statistik für Bioverfahrenstechnik • Angewandte Statistik und Versuchswesen • Biostatistik

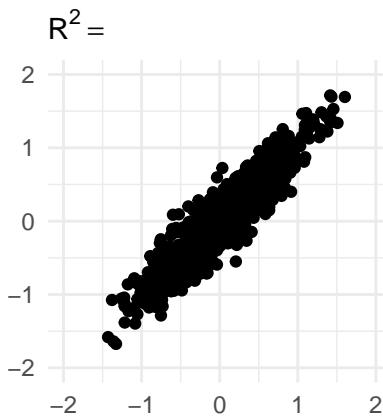


**Schätzen der Korrelation und des Bestimmtheitsmaßes** Heute hat leider Steffen Lerngruppe keine Zeit, so dass dann Steffen alles hier alleine machen muss. Das ist schon recht nervig. Steffen versteht nicht viel von  $\rho$ -Werten oder gar  $R^2$ -Werten. Geschweige denn hat er eine Idee, wie er die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

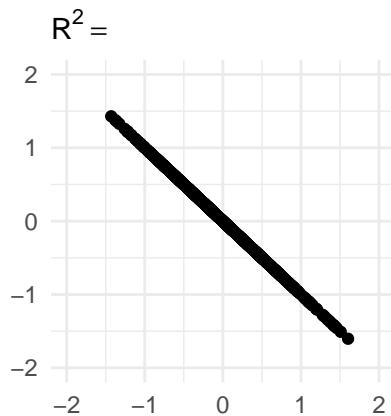
Pearsons  $\rho$  =



Pearsons  $\rho$  =



Pearsons  $\rho$  =



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

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

## 95. 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 Paula so laut, dass die Cafébesucher an den anderen Tischen zu ihr rüberschauen. Paula hatte ein Feldexperiment mit Erdbeeren durchgeführt. Dabei hatte sie dann Anzahl Läsionen auf den Blättern gemessen. Jetzt war es an ihr die Auswertung für ihre Gruppe zu erledigen und dafür sollte Paula sich an dem Regressionskreuz orientieren.

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

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

## 96. Aufgabe

(12 Punkte)

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



Inhalt folgender Module: Modellierung landwirtschaftlicher Daten • Biostatistik



**Interpretation der Ergebnisse eines gemischten Modells in R** Jessica mag die Berge sehr. Also hatte sie sich sofort freiwillig gemeldet als es hieß, dass die Hochschule jemanden braucht um in den Bergen an Drachen was zu untersuchen. Drachen an sich sind ja recht friedfertig, aber die Schreibfähigkeit zu testen, ging den Wesen dann doch gehörig auf den Zeiger. Ein Schrei zerschneidet die Ruhe der Berge. Jessica hatte zusätzlich die Sprunghöhe bestimmt und wollte wissen, ob es hier einen Zusammenhang gibt. Die Messungen wurden an drei Orten *nearby*, *midway* und *far far away* pro Berg durchgeführt. Vielleicht hätte sie nicht so offensiv viel Schokobons vor den Augen der Drachen essen sollen.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: literary_score ~ jump_height + (1 | mountain/direction)
##   Data: dragon_small_tbl
##
## REML criterion at convergence: 2250.7
##
## Scaled residuals:
##   Min     1Q Median     3Q    Max
## -2.6549 -0.6262 -0.0072  0.6467  2.8057
##
## Random effects:
##   Groups            Name        Variance Std.Dev.
##   direction:mountain (Intercept) 14.39     3.794
##   mountain           (Intercept) 352.91    18.786
##   Residual            196.83    14.030
## Number of obs: 273, groups: direction:mountain, 21; mountain, 7
##
## Fixed effects:
##   Estimate Std. Error t value
## (Intercept) 38.35795  25.21992  1.521
## jump_height  0.04956   0.11988  0.413
##
## Correlation of Fixed Effects:
##   (Intr) jump_height
## jump_height -0.958
```

Leider kennt sich Jessica mit der Analyse eines gemischten Modells in R überhaupt nicht aus. Deshalb braucht sie bei der Analyse Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar für den festen Effekt! **(1 Punkt)**
3. Erklären Sie den Begriff fester Effekt im Kontext des gemischten Modells! **(1 Punkt)**
4. Erklären Sie den Begriff zufälliger Effekt im Kontext des gemischten Modells! **(1 Punkt)**
5. Skizzieren Sie das Versuchsdesign! *Nutzen Sie eine reduzierte Auswahl an Faktorkombinationen!* **(1 Punkt)**
6. Visualisieren Sie die Ausgabe des linearen gemischten Modells! *Nutzen Sie eine reduzierte Auswahl an Faktorkombinationen!* **(1 Punkt)**
7. Folgt der Messwert  $y$  einer annähernden Normalverteilung? Begründen Sie Ihre Antwort! **(1 Punkte)**
8. Treffen Sie eine Aussage mit  $T_{\alpha=5\%} = 2.04$  zu Ihrer aufgestellten Nullhypothese! Begründen Sie Ihre Antwort **(2 Punkte)**
9. Interpretieren Sie den Effekt des festen Effekts im Kontext der wissenschaftlichen Fragestellung! **(1 Punkt)**
10. Wieviel Varianz erklärt der zufällige Haupteffekt? **(1 Punkt)**

# Teil IX.

## Experimentelles Design

### 97. Aufgabe

(16 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



**Einfache experimentelle Designs** Alex und Mark sind bei Yuki um sich Hilfe für eine Versuchsplanung in R zu holen. Im Hintergrund läuft viel zu laut London Grammar. Dabei geht es um den Zusammenhang zwischen der Behandlung Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) und dem Messwert Gewichtszuwachs in der 1LW in Zantern. Der Versuch soll in einem Stallversuch im Teuteburgerwald durchgeführt werden. Nach dem Dozenten ist der Messwert Gewichtszuwachs in der 1LW normalverteilt. Die beiden entschieden sich für ein faktorielles Versuchsdesign. Im ersten Schritt überlegt Yuki ein einfaches experimentelles Design zu probieren. Daher entscheiden sich alle drei für ein *Latin square design*. Das sollte für den Anfang erstmal reichen. 'Und jetzt, was machen wir jetzt?', Mark schaut die anderen beiden mit großen Augen an. Die zucken mit der Schulter. Alle mampfen Reese's Peanut Butter Cups.

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Skizzieren Sie das faktorielle Versuchsdesign! **(3 Punkte)**
4. Skizzieren Sie eine Datentabelle für das faktorielle Versuchsdesign in R! **(2 Punkte)**
5. Erstellen Sie das statistische Modell in der in R üblichen Schreibweise für eine ANOVA! Skizzieren Sie die notwendige Funktionen in R! **(3 Punkte)**
6. Skizzieren Sie die weitere Datenanalyse hinsichtlich eines multiplen Gruppenvergleiches in {emmeans}! **(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** Jonas und Tina sind bei Paula um sich Hilfe für eine Versuchsplanaung in R zu holen. Im Hintergrund läuft viel zu laut White Lies. Dabei geht es um den Zusammenhang zwischen der Behandlung Lichtstufen (*none*, 200lm, 400lm und 600lm) sowie Genotypen (AA, und BB) sowie zwei Blöcken und dem Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] in Erbsen. Der Versuch soll in einem Freilandversuch im Oldenburger Land durchgeführt werden. Nach dem Dozenten ist der Messwert Chlorophyllgehalt (SPAD-502Plus) [SPAD] normalverteilt. Die beiden entschieden sich für ein faktorielles Versuchsdesign. Im ersten Schritt überlegt Paula ein komplexeres experimentelles Design zu probieren. Daher entscheiden sich alle drei für ein *Strip plot design oder auch Streifenanlage* mit Berücksichtigung einer Interaktion. Das sollte für den anfang erstmal reichen. 'Und jetzt, was machen wir jetzt?', Tina schaut die anderen beiden mit großen Augen an. Die zucken mit der Schulter. Alle mampfen Smarties.

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistische Hypothesenpaare! **(2 Punkte)**
3. Skizzieren Sie das faktorielle Versuchsdesign! **(3 Punkte)**
4. Skizzieren Sie eine Datentabelle für das faktorielle Versuchsdesign in R! **(2 Punkte)**
5. Erstellen Sie das statistische Modell in der in R üblichen Schreibweise für eine ANOVA! Skizzieren Sie die notwendige Funktionen in R! **(4 Punkte)**
6. Skizzieren Sie die weitere Datenanalyse hinsichtlich eines multiplen Gruppenvergleiches in {emmeans}! **(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

### 99. Aufgabe

(20 Punkte)

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

Inhalt folgender Module: Biostatistik



Die folgende Aufgabe basiert auf einer der zwei folgenden wissenschaftlichen Veröffentlichungen. Für die Prüfung wird die vertiefende Kenntnis der beiden Veröffentlichungen vorausgesetzt.

- Sánchez, M., et al. (2022). Hoverfly pollination enhances yield and fruit quality in mango under protected cultivation. *Scientia Horticulturae*, 304, 111320. [[Link](#)]
- Wu, G., et al. (2004). Arginine nutrition in neonatal pigs. *The Journal of Nutrition*, 134(10), 2783S-2790S. [[Link](#)]

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

**Zerforschen einer wissenschaftlichen Veröffentlichung** Vor dem Start der eigenen Arbeit möchte ihre Betreuerin, dass Yuki einmal die wissenschaftliche Veröffentlichung Wu, G., et al. (2004). *Arginine nutrition in neonatal pigs* sinnvoll zusammenfasst. Dann würde die eigene Arbeit auch leichter von der Hand gehen und Yuki hätte dann schon eine Vorlage. 'Das ist jetzt aber umfangreicher als gedacht!', schnauft sie und runzelt die Stirn. Im Hintergrund spielt viel zu leise London Grammar. Das wird dann vermutlich heute Abend nichts mehr mit ihrem Hobby Orchideen. Das Minischwein schaut mitleidig.

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

1. Erläutern Sie die wissenschaftliche Fragestellung der wissenschaftlichen Veröffentlichung anhand des OCAR Prinzips nach Schimel (2012)<sup>3</sup> (**4 Punkte**)
2. Nennen Sie die untersuchten Endpunkte in der wissenschaftlichen Veröffentlichung! Wie lautet der primäre Endpunkt? (**2 Punkte**)
3. Erstellen Sie das statistische Modell in der in R 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 R 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 ihr Betreuer, dass Tina 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 Tina hätte dann schon eine Vorlage um die eigenen erhobenen Daten in eine Tabelle eintragen zu können. 'Das ist jetzt aber umfangreicher als gedacht!', schnauft sie und runzelt die Stirn als sie in ihren Laptop starrt. Dabei isst sie noch ein paar Katjes. Das wird dann vermutlich heute Abend nichts mehr mit Indiana Jones.

<b><math>f_1</math></b>		<b><math>f_2</math></b>		<b><math>x_1</math></b>		<b><math>y_1</math></b>		<b><math>y_2</math></b>	
<	>	<	>	<	>	<	>	<	>
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	
:		:		:		:		:	
5		2		1.9		9.6		0	

Tina füllt sich mit der Analyse der Daten in der Tabelle 1 überfordert. Deshalb braucht sie bei der Auswertung Ihre Hilfe! Glücklicherweise kennen Sie den wissenschaftlichen Datensatz aus Ihren eigenen Analysen schon im Detail und können sofort helfen.

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 zwei übergeordnete Analysebereiche der Statistik! Nutzen Sie hierfür die Variablennamen der 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 $x_1$ und $y_1$

4. In welchen der übergeordneten Analysebereiche der Statistik gehört die Auswertung Ihres Endpunktes? Begründen Sie Ihre Antwort! (2 Punkte)
5. Skizzieren Sie eine beispielhafte 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 Eigenschaften der Daten müssen Sie für Ihre statistische Analyse im Besonderen achten? (2 Punkte)
9. Welche statistische Maßzahl können Sie aus Ihrer Datenanalyse berichten? (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



Stichworte: The Great Horse-Manure Crisis of 1894<sup>4</sup> • Napoleon

**Herodot – der Schimmel aus Ivenack** Die Lerngruppe *Die Pantoffeltieren* bestehend aus Jonas, Nilufar, Mark und Jessica 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>5</sup>. Jetzt wollen die vier herausfinden: "Konnten die Ivenacker den Apfelschimmel Herodot vor dem Zugriff von Napoleon in der 1000-jährigen Eiche verstecken?"

Helfen Sie der Lerngruppe *Die Pantoffeltieren* bei der Beantwortung der Forschungsfrage! Gehen Sie von einem radialen Wachstum der 1000-jährigen Eiche von 3mm pro Jahr aus. Es ist bekannt, dass die Eiche im Jahr 2022 einen Umfang von 17m in Brusthöhe hatte.

1. Wie groß war der Durchmesser in m der Eiche im Jahr 1820 als Herodot in der Eiche versteckt werden sollte? **(2 Punkte)**
2. Erstellen Sie die lineare Gradengleichung für das radiale Wachstum! **(1 Punkt)**
3. Skizzieren Sie in einer Abbildung einen linearen Zusammenhang und einen exponentiellen Zusammenhang für das Wachstum der 1000-jährigen Eiche! **(1 Punkt)**

Herodot hatte eine Schulterhöhe von 190cm, eine Breite von 85cm sowie eine Länge von 240cm.

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

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

4. Berechnen Sie die Dicke der Eichenwand in cm! Verdeutlichen Sie Ihre Berechnungen an einer aussagekräftigen Skizze für Pferd und Eiche! **(2 Punkte)**
5. Unter einer Dicke der Eichenwand von 15cm bricht die Eiche zusammen. Beantworten Sie die Forschungsfrage! Begründen Sie Ihre Antwort! **(2 Punkte)**

<sup>4</sup>Weitere Informationen: [The Great Horse-Manure Crisis of 1894](#) und [The Great Horse Manure Crisis of 1894 was never mentioned before 2004](#)

<sup>5</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 J bestehend aus Paula, Yuki, Jonas und Steffen hat sich zusammengefunden um den ersten Versuch zu planen. In einem Experiment wollen sie die Wuchshöhe von 252 Sonnenblumen 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 Sonnenblumen einpflanzen und müssen dafür Substrat bestellen. Zum anderen muss die Projektgruppe die Sonnenblumen auch bewegen und in ein Gewächshaus auf rechteckigen Tischen platzieren. Die schmale Tischseite fast ohne Randpflanzen 7 Pflanzen. Die Töpfe für die Keimung haben einen Durchmesser von 10cm und eine Höhe von 8cm. Der Kubikmeterpreis für Torf liegt bei 260 EUR.

Helfen Sie der Projektgruppe J bei der Planung des Versuches!

1. Skizzieren Sie den Versuchsplan auf *drei* Tischen im Gewächshaus! *Skizzieren Sie einen Tisch detaillierter und die restlichen Tische schematisch!* (**2 Punkte**)
2. Berechnen Sie die benötigte Anzahl an Pflanztopfen, wenn Sie Randpflanzen mit berücksichtigen wollen! (**1 Punkt**)
3. Berechnen Sie die benötigte (**a**) Pflanztopffläche in  $m^2$  sowie die (**b**) Tischfläche in  $m^2$  für einen Tisch gegeben der Anzahl an Pflanztopfen 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 Pflanztopfe benötigen! Gehen Sie von *einem Zylinder* für die Pflanztopfe 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** Alex bringt ein neues, tolles Projekt mit in die Lerngruppe *Die Blattläuse* bestehend aus ihm, Steffen, Yuki sowie Tina. 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  $5.5m$ . Die hintere Seite des Rinderstall hat eine Höhe  $h_b$  von  $9m$ . Der Rinderstall hat eine Tiefe  $t$  von  $15m$  und eine Breite  $b$  von  $40m$ . 'Sag mal Alex, ist das eine Matheaufgabe oder rechnen wir hier gerade für dich kostenlos als menschliche Computer Sachen für deinen Betrieb?', fragt Yuki mit erhobenen Augenbrauen. Tina und Steffen 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 Alex eine neue Biogasanlage für seinen Betrieb. Der neue Methantank hat einen Radius  $r$  von  $1m$ . Leider gibt es ein paar bauliche Beschränkungen auf dem Grundstück. Das Fundament des zylindrischen Methantanks kann nur ein Gewicht von maximal  $10t$  aushalten bevor der Tank wegbricht. Alex rechnet eine Sicherheitstoleranz von  $10\%$  ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei  $-80^\circ C$  eine Dichte von  $220kg/m^3$ . Bei  $-100^\circ C$  hat Methan eine Dichte von  $270kg/m^3$ . Alex betreibt seine Anlage bei  $-90^\circ C$ .

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 gefüllten 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

**Aliagatorenbirnen und Blaubeeren** "Sind Sie ein Riesenfautier oder warum kaufen Sie so viele Aliagatorenbirnen?", spricht es hinter Ihnen. Irritiert drehen Sie sich um und blicken in das puderrote Gesicht von Paula. "Wieso?", entfährt es Ihnen und Sie bereuen sogleich die Frage. Sofort werden Sie zu einem Whiteboard voller roter Schnüre geschliffen und müssen folgenden mathematischen untermauerten Argumenten im Lidl über sich ergehen lassen. Da kommen Sie nicht mehr raus, also können Sie auch gleich mitmachen. Das Problem liegt in Chile<sup>6</sup>. Tja, die Deutschen und Südamerika.

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

1. Wenn 4 Blaubeerschalen 7.96 Euro kosten, wie viel kosten 5 Schalen? **(2 Punkte)**
2. Wenn Sie die 5 Blaubeerschalen gekauft haben, wie viele Aliagatorenbirnen zu je 0.99 EUR können Sie sich dann noch für 50 EUR leisten? **(1 Punkt)**

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

- Ein Kilo Strauchtomaten benötigt 190l Wasser. Eine Strauchtomate wiegt 90 - 125g.
- Ein Kilo Salat benötigt 130l Wasser. Ein Salatkopf wiegt 300 - 520g.
- Ein Kilo Avocado benötigt 1100l Wasser. Eine Avocado wiegt 150 - 410g.
- Ein Kilo Blaubeeren benötigt 820l 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 2024 blieben die Erträge von Blaubeeren mit  $7.9 \times 10^4$ t in dem prognostizierten Rahmen. Die Menge reduzierte sich um 6.8%. Die Exporte für Avocados stiegen in dem gleichen Zeitraum um 18.1% auf  $1.9 \times 10^5$ t.

4. Wie viele Hektoliter Wasser hat Chile in dem Exportjahr 2023 exportiert? **(2 Punkte)**

Chile ist eines der wenigen Länder der Welt, die ihr Wasser komplett privatisiert haben. Derzeit sind nur ein Prozent des Wassers des Landes für den häuslichen Verbrauch vorgesehen. In den Dörfern der Anbauregionen versorgen Tankwagen die Bevölkerung jede Woche mit Wasser, es gibt etwa 50 Liter Wasser pro Kopf für den täglichen Bedarf. In Deutschland liegt der Verbrauch bei 8 - 17 Liter pro Spülmaschinenlauf und 35 - 115 Liter pro Waschgang einer Waschmaschine.

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

Das alles hätten Sie nicht von Paula erwartet. Ganz schön viele Informationen wurden da zusammengetragen.

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

<sup>6</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 Alex kostbaren Schlaf prokrastiniert um einem Ernährungswissenschaftler auf YouTube über die Erde als Dampfnudel zu lauschen. Irgendwie passt es dann doch mit der Analogie. Übermüdet müssen Sie darüber nachdenken, warum vor 69 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>7</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  $9.81 \text{ m/s}^2$  an. Im Weiteren hat die Erde einen ungefähren Durchmesser von  $1.2742 \times 10^4 \text{ km}$  und eine mittlere Dichte  $\rho$  von  $5.86 \text{ g/cm}^3$ . Das Gewicht von einem heute lebenden asiatischen Elefanten liegt bei 3t bis 5t und das Gewicht von einem Triceratops bei 6t bis 12t.

1. Welchen Durchmesser müsste die Erde vor 69 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 69 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 69 Millionen Jahren! (2 Punkte)
2. Beantworten Sie die Eingangsfrage mit 1-2 Antwortsätzen! (1 Punkt)

Die Distanz zwischen Sonne und Erde entspricht 1.05 astronomische Einheiten (AE). Die Einheit 1 AE wird mit  $1.55 \times 10^8 \text{ km}$  angegeben. Der *massebehaftete* Sonnenwind besteht aus 84% Wasserstoffkernen mit einer molaren Masse von  $1.01 \text{ g/mol}$ , 11% Heliumkernen mit  $3.92 \text{ g/mol}$  sowie 5% 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  $8 \text{ cm}^{-3}$  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>7</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?'** Man hört schon häufig vieles Geschnatter von höflichen Puten in Mastställen. Enge 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 Yuki, Jonas, Tina und Mark aber als vorsorgliche Puten-Halter:innen nicht<sup>8</sup>. Gemeinsam sind die Vier in einer Projektgruppe gelandet. Betrachten wir also gemeinsam einmal das Platzangebot (eng. space allowance, abk. SA) der Puten 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 Pute plus dem benötigten Radius für das Verhalten  $i$ .
- $R_i$  dem notwendigen Abstand zu den Nachbarn für das Verhalten  $i$ .
- $i$  dem Verhalten: (1) dustbathing, (2) wing/leg stretching, (3) drinking/eating und (4) preening.

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

	Aldridge et al. (2021)	Baxter et al. (2022)	Jabcobs et al. (2019)
dustbathing	34cm; 37cm; 1.2%	43cm; 39cm; 8.1%	14cm; 35cm; 7.6%
wing/leg stretching	45cm; 23cm; 4.1%	22cm; 41cm; 5.2%	21cm; 22cm; 6.3%
drinking/eating	37cm; 17cm; 3.2%	34cm; 40cm; 5.1%	12cm; 34cm; 6.9%
preening	30cm; 35cm; 12.8%	31cm; 31cm; 12.8%	3cm; 18cm; 18.1%

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

1. Skizzieren Sie die Werte  $r_i$ ,  $R_i$  und  $A_i$  für zwei nebeneinander agierende Puten 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 Puten 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>8</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** Alex und Steffen 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 Alex ziemlich plötzlich, so dass Steffen die Oreos aus dem Mund fallen. 'Nein, und ehrlich gesagt bin ich auch ziemlich müde...'. Das ist jetzt aber Alex egal, denn er möchte folgende Sachlage diskutieren. Und Alex 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 45 Tagen die ersten Symptome ein; die ersten Toten sind nach 65 Tagen zu beklagen; nach 110 Tagen rafft die Skorbut eine ganze Schiffsbesatzung dahin«

Ferdinand Magellan stach im Jahre 1519 in See um eine Passage durch den südamerikanischen Kontinent zu finden. Zu seiner Flotte gehörten fünf Schiffe - das Flaggschiff Trinidad, die San Antonio, die Victoria, die Concepción und die Santiago - mit einer Besatzung von insgesamt 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 Skorbut bei der Überquerung des Pazifiks, da durch Erkundungsfahrten weniger Zeit blieb, um wilden Sellerie aufzunehmen. Wilder Sellerie enthält  $4000\mu\text{g}/100\text{g}$  Vitamin C. Der Bedarf liegt bei  $120\text{mg}$  pro Tag für Männer.

3. Berechnen Sie die notwendige Menge in kg an aufzunehmenden wilden Sellerie auf die Concepción für die ununterbrochene Fahrt von drei Monaten und 22 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** Nilufar ist bei Jonas um gemeinsam *Event Horizon – Am Rande des Universums* zu streamen. Das war jetzt nicht die beste Idee. Denn Nilufar kann Horror überhaupt nicht ab. Deshalb flüchtet sie sich in Logik um ihre Emotionen zu bändigen. Jonas mampft ungerührt Snickers. Folgenden Gedankengang nutzt Nilufar um dem Film zu entkommen. Die Sonne hat eine aktuelle, angenommene Masse von  $2 \times 10^{31}$  kg. Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 5000m kollabiert, wird die Sonne 20% der aktuellen Masse verloren haben. Ein Lichtteilchen mit der Masse  $m_f$  und der Fluchtgeschwindigkeit  $v_f$  will dem schwarzen Loch entkommen. An folgende Formeln erinnert sich Nilufar für die kinetische Energie des Lichtteilchens  $E_{kin}$  und der Gravitationsenergie des schwarzen Lochs  $E_{grav}$ <sup>9</sup>.

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

mit

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

Im Folgenden wollen wir Nilufar 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  $2.9 \times 10^8$  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>9</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** Tina und Alex wandern durch den Teutoburgerwald um mal vom Studium runterzukommen. 'Kennst du eigentlich Enrico Fermi?', fragt Tina 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 sie irritiert und interessiert an. Die beiden haben das Problem gepackt. Deshalb wollen Tina 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>10</sup>

Die beiden treffen folgende Annahmen. Eine außerirdische Zivilisation schickt *drei* Voyager 1 ähnliche Sonden mit der Geschwindigkeit von  $6.0523 \times 10^4 \text{ km/h}$  los um sich auf den erreichten Planeten selbst zu replizieren. Nach 750 Jahren ist die Replikation abgeschlossen und wiederum *drei* Sonden werden ausgesendet. Gehen Sie von 3.57 Lichtjahren als mittlerer Abstand der Sterne in der Milchstraße aus. Es gibt  $10^{11}$  Sterne in der Milchstraße. Nehmen Sie eine Lichtgeschwindigkeit von  $2.9 \times 10^8 \text{ m/s}$  an.

1. Skizzieren Sie in einer Abbildung die ersten drei Schritte der Vervielfältigung der Sonden in der Galaxie! Beschriften Sie die Abbildung mit der Dauer und der Anzahl an Sonden für jeden Schritt der Vervielfältigung! **(4 Punkte)**
2. Berechnen Sie die theoretische Anzahl an Vervielfältigungsschritten die benötigt werden um mit *einem einzigen Vervielfältigungsschritt* die gesamten Sterne der Milchstraße mit Sonden zu besuchen! **(2 Punkte)**
3. Berechnen Sie die Dauer, die eine außerirdische Zivilisation annährungsweise benötigt um die gesamten Sterne der Milchstraße mit Sonden zu besuchen! **(2 Punkte)**
4. Bei einem vermuteten Alter der Erde von  $4.3 \times 10^9$  Jahren, wie oft war dann eine Sonde einer außerirdischen Zivilisation schon zu Besuch? Korrigieren Sie Ihre Antwort mit dem Wissen, dass sich die Kontinentalplatten einmal alle  $8 \times 10^7$  Jahre vollständig im Erdinneren umgewandelt haben! **(2 Punkte)**

<sup>10</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 niedersächsischen Pyramidentage an! Sie und Yuki sind auf abenteuerlichen Wegen für den Bau der Pyramiden zuständig. Zu allem Überfluss handelt es sich auch noch eine *Reenactment* Veranstaltung. Thema der diesjährigen Pyramidentage sind die Pyramiden von Meroe, die den Königen und Königinnen des historischen Reiches von Kusch in Nubien, dem heutigen Sudan, als Grabstätten dienten. Die Pyramiden in Meroe fallen durch ihren steilen Winkel von 73 Grad im Vergleich zu den ägyptischen Pyramiden mit 52 Grad auf. Die durchschnittliche Seitenlänge der Grundfläche einer Pyramide beträgt 44 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 44 Königsellen lang sein. Einer der Skla'angestellten hat eine Höhe der Königspyramide von 37.61m berechnet. Dem müssen Sie auf jeden Fall nachgehen. Überprüfen Sie die Berechnung! **(1 Punkt)**
2. Die Außenflächen der Königspyramide soll begrünt werden. Für die Bepflanzung muss eine 5cm dicke Torfschicht auf die Königspyramide 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 2 Sklaven, die Ihnen und Yuki bei dem Befüllen der Königspyramide mit Schutt zu Seite stehen. Leider haben Ihre Sklaven zu allem Überfluss auch noch chronische Rückenschmerzen entwickelt, als die Sklaven von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 90% aus. In eine Schubkarre passen 100 Liter.

3. Wie oft müssen Ihre maladen Sklaven die Rampe mit der Schubkarre zur Spitze der Königspyramide hochfahren um die Pyramide mit Schutt zu füllen? **(1 Punkt)**
4. Berechnen Sie die Länge der Rampe zur Spitze der Königspyramide mit einem Anstellwinkel von  $10^\circ$ ! **(2 Punkte)**
5. Wie weit reicht Ihre Rampe vom Fuß der Königspyramide in die niedersächsische Landschaft? **(2 Punkte)**

Bei der Besichtigung der Königspyramide teilt Ihnen der leicht übergewichtige Pharao (Nebenberuf *Mittelständler*) mit, das die Pyramide zu steil sei und somit nicht in die niedersächsische Landschaft passen würde. Sie müssen nochmal ran.

6. Die Grundfläche der Königspyramide ändert sich nicht. Berechnen Sie die Änderung der Höhe in Königsellen, wenn sich der Anstellwinkel der Königspyramide 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. Alex und Mark schwingen sich auf ihre Cachermobile um mit 17km/h, geleitet von modernster Satellietentechnologie und einem Supercomputer aus dem Jahr 2000 in den Händen, Plastikdosen in der Natur und an sehenswerten Orten zu finden. Alex und Mark wollen diesmal endlich die aufwärts 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 Alex und Mark für die Planung der Route zu Verfügung<sup>11</sup>.

Ort	Cache	Wertung (S T G)
A	GCAFIE9	2.5   4.5   Mikro
B	GCDCNX7	3.5   2.0   Normal
C	GCISRD2	4.5   4.0   Normal
D	GC9GBBB	4.0   1.0   Mikro
E	GC4MPQ5	5.0   1.5   Mikro

Im Weiteren sind den beiden folgende Informationen zu den Entfernungen der Orte zugänglich. Der Entfernungsvektor  $\vec{AB}$  ist 3km. Im Weiteren ist Ihnen der Entfernungsvektor  $\vec{CB}$  mit 5.5km bekannt. Der Entfernungsvektor  $\vec{BE}$  ist das 1.3-fache des Entfernungsvektor  $\vec{CB}$ . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr 25° südlich. Wenn Sie von dem Ort C den Ort B anpeilen, so liegt der Ort D ungefähr 50° 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 B 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 aufwärts 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

$$\text{Suchzeit} = 0.2 + 0.15 \cdot \text{Schwierigkeit}$$

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

<sup>11</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!

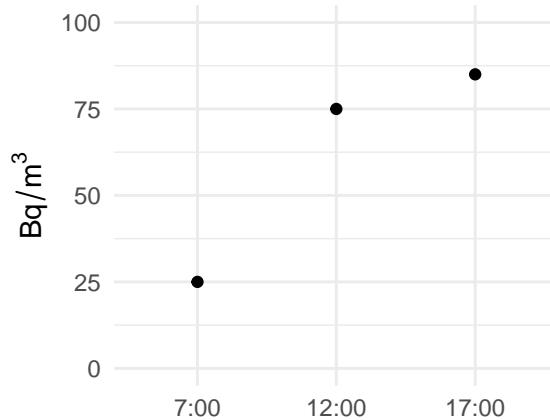


Inhalt folgender Module: Mathematik & Statistik



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>12</sup> vom Dorf will Paula 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 17:00 bestimmt sie dreimal automatisch die Radonbelastung in ihrem Kellerraum in  $Bq/m^3$ . Es ergibt sich folgende Abbildung<sup>13</sup>. Leider helfen die Messwerte Paula überhaupt nicht weiter. Sie müssen also helfen!



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

Radon zerfällt mit einer Halbwertszeit von 2.8d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 160d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

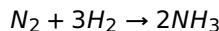
2. Wie lange dauert es in Stunden bis die kritische Radonbelastung von  $320Bq/m^3$  auf unter  $80Bq/m^3$  gefallen ist? (4 Punkte)

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	79.7	28.4	
Sauerstoff	19.5	16.5	
Kohlenstoffdioxid	0.035	11.8	

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

Während Paula ihr etwas pappiges Toastbrot mampfen kommt Paula die Dokumentation über Brot aus Luft in den Sinn. Paula denkt darüber ein wenig nach. Für die Umwandlung von Stickstoff  $N_2$  mit Wasserstoff  $H_2$  zu Ammoniak  $NH_3$  gilt folgende Reaktionsgleichung<sup>14</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>12</sup>Tocotronic - Electric Guitar als passende Untermalung für diese Aufgabe.

<sup>13</sup>Die Quelle der Inspiration für die Aufgabe war folgender Artikel: [Atmende Wand](#)

<sup>14</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 Steffen war nach Ladenschluss bei Penny die Regale einzuräumen. Dabei ist Steffen in der Auslage der Sonderangebote das Necronomicon<sup>15</sup> in die Hände gefallen. Nun ist er eine Magierin der Zeichen geworden! Also eigentlich kann Steffen nur Mathe und das dämliche Necronomicon hat ihn in die Vergangenheit geschleudert... aber gut, was tut man nicht alles im Jahr 705 n. Chr. für den neuen Lehnsherren Fürsten Arthur. Steffen baut natürlich einen Schrottkugelturm um sich den Horden der Finsternis mit genug Schrott erwehren zu können! Steffen stehen zwei mächtige magische Formeln zur Unterstützung zu Verfügung. Leider wird das nicht reichen, deshalb müssen Sie hier auch noch durch Zeit und Raum helfen!

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

mit

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

Als erstes müssen Sie die Höhe des zu bauenden Schrottkugelturmes bestimmen. Hierfür ist wichtig zu wissen, dass sich die Bleitropfen mit einem Gewicht von  $10mg$  zu gleichförmigen Bleikugeln bei einer Geschwindigkeit von  $14m/s$  bilden.

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

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

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

Sie haben jetzt die  $2.3 \times 10^5$  Bleikugeln zusammen. Blei hat eine Dichte von  $11.23\text{g/cm}^3$ .

4. Wie schwer in Kilogramm  $kg$  sind die  $2.3 \times 10^5$  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  $\text{cm}^2$  ist Ihr quadratisches Sieb am oberen Ende des Turms, wenn Sie pro Fall ca. 900 Bleikugeln produzieren wollen und die Bleikugel im Fall  $0.5\text{cm}$  Abstand haben müssen? (1 Punkt)

<sup>15</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 Jessica mit der Koalakuschelschule in Down Under nicht geklappt. War vielleicht auch nicht so die beste Idee... aber dafür hat Jessica eine neue Eingebung! Oder wie es Mike Tyson zugeschrieben wird: »Ich wurde nie niedergeschlagen, ich war immer am Aufstehen!«. Daher macht Jessica jetzt einen Großhandel mit Kaninchenfleisch und damit dem teuersten Fleisch in Australien auf. Moment, hopsen hier nicht, seit Thomas Austin im Jahr 1860 ungefähr 30 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Jessica wird stutzig und fragt Sie, dem mal mathematisch nachzugehen!<sup>16</sup>

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

$$f(t) = 8 \times 10^9 - 1.4 \times 10^9 \cdot 1.7^{-0.2 \cdot t + 3.1}$$

1. Skizzieren Sie die Sättigungsfunktion *annäherungsweise* in einer Abbildung! **(1 Punkt)**
2. Wie viele Kaninchen können nach der Sättigungsfunktion maximal im westlichen Australien leben? Ergänzen Sie den Wert in Ihrer Abbildung! **(2 Punkte)**
3. Wie viele Millionen Kaninchen leben nach der Sättigungsfunktion nach 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 12 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 30% der gesättigten Kaninchenpopulation angewachsen ist? **(2 Punkte)**

Thomas Austin entließ die Kaninchen im äußersten Norden von Australien. Australien hat eine West-Ost-Ausdehnung von 4200km und eine Nord-Süd-Ausdehnung von knapp 3400km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 7.8km pro Jahr aus.

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

Eine jährliche Impfung gegen das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV) kosten 13\$ pro Tier und der durchführende Arzt verlangt ca. 45\$ pro Tier.

6. In Ihrem Stall leben 1000 Mastkaninchen. Mit welchen jährlichen Zusatzkosten für die Impfungen der Kaninchen müssen Sie daher kalkulieren? **(1 Punkt)**

<sup>16</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



**Uckermark. Unendliche Weiten.** Wir schreiben das Jahr 2025. Dies sind die Abenteuer des Esels Frida und Tina. Grünes Gras unter Tinas Füßen und ein strammer Wind im Gesicht, egal wohin sie schaut. Ein schmatzendes Geräusch ertönt unter Tina. Tina sinniert, sollte sie ihre weiten Graslandschaften jetzt schon düngen? Dafür benötigt Tina die *Grünlandtemperatur!* Die Grünlandtemperatur (GLT) ist die Summe aller positiven Tagesmitteltemperaturen seit Jahresbeginn. Ab einer GLT von  $200^\circ$  kann mit der Stickstoffdüngung begonnen werden. Tina sieht nicht ein, Geld für einen Agrarmetrologen zu bezahlen, wenn auch Sie mitrechnen können. Also rechnen Sie beide mit folgenden Informationen zu Monatsmultiplikatoren des GLT-Wertes: Januar mit  $0.4 \times$ , Februar mit  $0.8 \times$  und März mit  $1.1 \times$ . Sie haben noch im letzten Jahr folgende Temperaturen gemessen.

Datum	$C^\circ$
01. Jan 2024	0.3
01. Feb 2024	1.1
01. Mrz 2024	3.5
01. Apr 2024	5.8

1. Erstellen Sie eine Skizze aus den Informationen aus der Temperaturtabelle! **(1 Punkt)**
2. Stellen Sie die linearen Funktionen  $f_1(t)$ ,  $f_2(t)$  und  $f_3(t)$  aus der obigen Temperaturtabelle auf! **(1 Punkt)**
3. Bestimmen Sie die Stammfunktionen  $F_1(t)$ ,  $F_2(t)$  und  $F_3(t)$  für Ihre linearen Funktionen aus der obigen Temperaturtabelle! **(1 Punkt)**
4. Osterglocken beginnen ab einer GLT von  $210^\circ C$  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 Tinas Pink Lady Plantage werden Sie beide auf dem Trecker von einer Gruppe elektrifizierter Rentner abgedrängt. Der Trecker muss wieder aus dem Graben! Frida und die elektrifizierten Rentner ziehen an zwei, separaten Seilen. Dabei zieht Frida mit  $180N$ . Die elektrifizierte Renter bringen eine Kraft von  $190N$  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  $50^\circ$  Winkel mit ihrem Seil an dem Trecker. Welche Kraft wird aufgebracht? **(2 Punkte)**
7. Mit welcher Beschleunigung ziehen Sie den  $1.3t$  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** Jonas, Jessica, Nilufar und Paula sitzen im Bus. Wenn man sich zu spät anmeldet, dann ist die Exkursion nicht so toll. Jonas hatte den Anderen in der Lerngruppe zu spät Bescheid gesagt. 'Was denn, bin ich eure Nanny oder was?!', entfährt es Jonas nachdem die vorwurfsvollen Blicke schon eine Weile auf ihm lasten. Also geht es eben mit Rektor Skinner und Mrs. Krabappel in die Kartonagenfabrik. Wie schon im vorherigen Semester... In der Kartonagenfabrik angekommen erfahren die Vier, dass die Kartons zum Versand von Nägeln nicht hier zusammengebautwerden sondern das sich die Endfertigung in Flint, Michigan befindet. Unter anderem wird dort der berühmte *Doppelt gewellte, 6-mal-gefaltete, 0.5mm, 40-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 40cm und eine Breite von 23cm. Die Kartonagenmaschine in Flint soll dann einen quadratischen Eckenausschnitt der Länge  $x$  falzen.

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

Rektor Skinner möchte sich gerne wieder in seinem Vorgarten aufhalten und nicht die ganze Zeit von Bart mit Erdnüssen beworfen werden. Deshalb möchte er einen geräumigen Teil seines Vorgartens einzäunen. Ein Teil der Umzäunung bildet seine Vorderhauswand. Wegen Lieferschwierigkeiten stehen Rektor Skinner nur 130m 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 130m 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!

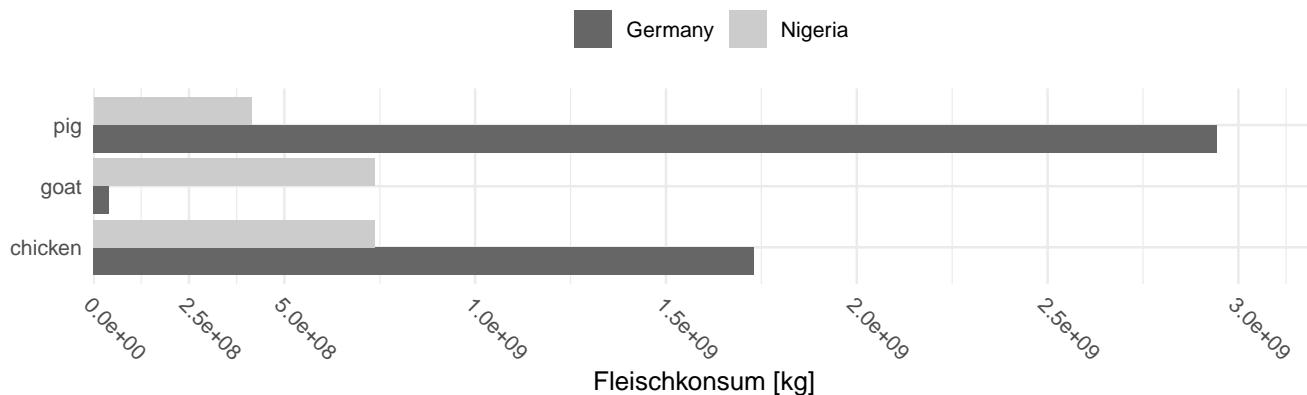


Inhalt folgender Module: Mathematik & Statistik



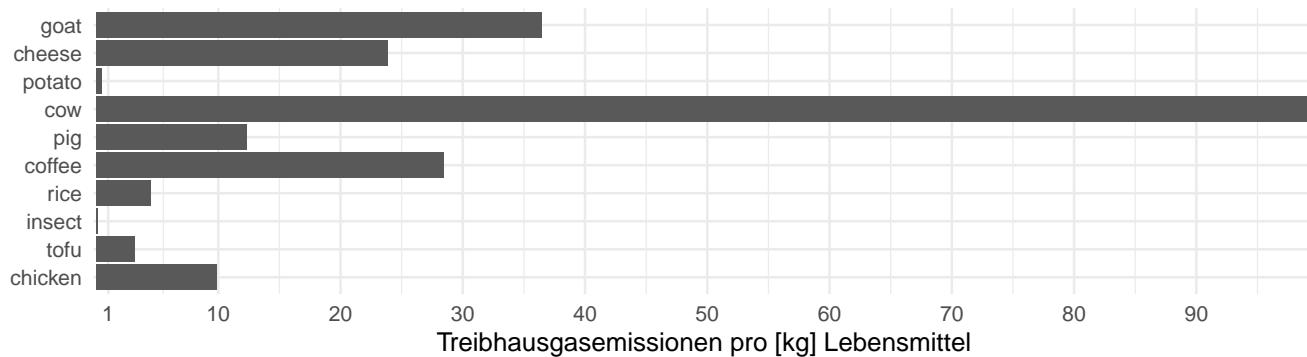
**Ein Pfund Insekten, bitte!** 'Das wird wohl häufiger gehört werden, wenn wir die Menschheit mit Proteinen ausreichend ernähren wollen<sup>17</sup>', merkt Jonas an. Die Lerngruppe um Mark, Alex und Nilufar sind bei Jonas um mal was außergewöhnliches zu essen. Um den Sinn der Nahrungsumstellung zu verdeutlichen, vergleicht Jonas 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 2020 leben ca.  $8.2 \times 10^7$  Menschen in Deutschland und ca.  $1.8 \times 10^8$  Menschen in Nigeria. Mit den Informationen wollen Sie und Jonas mit der Überzeugungsarbeit anfangen und dann eine Prognose für den Fleischkonsum im Jahr 2050 zu treffen.

Im Folgenden ist Abbildung des Fleischkonsums im Jahr 2020 in Deutschland und Nigeria in [kg] einmal dargestellt.



1. Stellen Sie den Fleischkonsum in Deutschland und Nigeria im Jahr 2020 *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)

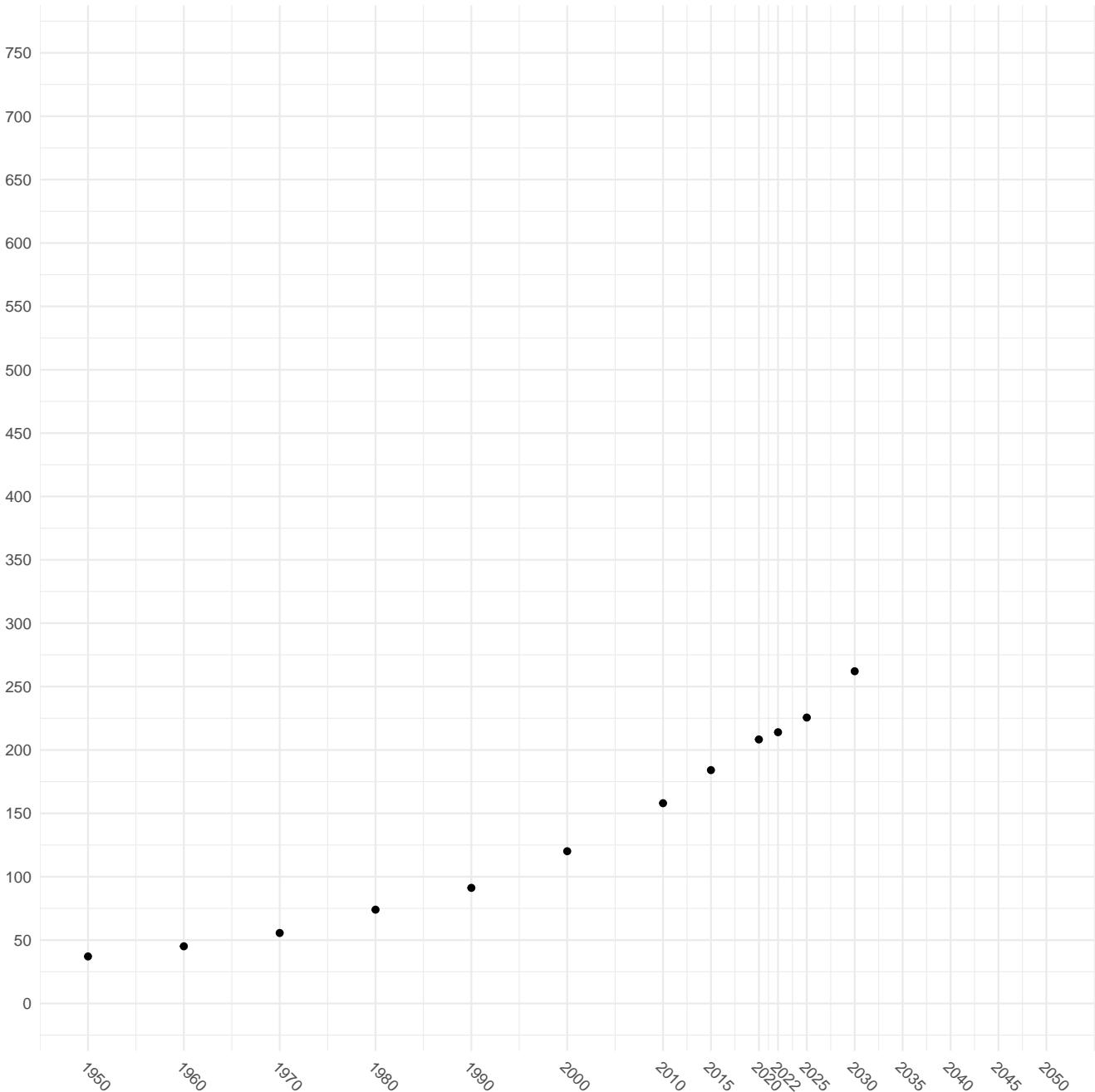
In der nächsten Abbildung finden Sie die CO<sub>2</sub>-Emission in [kg] nach Lebensmittel, die durch die Produktion entsteht, abgebildet.



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 2020 entstehen, dar! Ergänzen Sie auch hier das Verhältnis Nigeria zu Deutschland! (2 Punkte)

<sup>17</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 70%-iger Angleichung der Lebensbedingungen zu Deutschland im Jahr 2020! **(1 Punkt)**
6. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an  $CO_2$  in Nigeria im Jahr 2050 im Vergleich zum Jahr 2020, der sich durch den angeglichenen Fleischkonsum ergibt! **(1 Punkt)**
7. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an  $CO_2$  in Nigeria, wenn die gesamte Proteinaufnahme durch Insekten ersetzt würde! **(1 Punkt)**

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

Die Prävalenz von AIDS bei einem Menschen in Europa wird mit 0.1% angenommen. In 96% der Fälle ist ein HIV-Test positiv, wenn der Patient erkrankt ist. In 0.5% der Fälle ist ein HIV-Test positiv, wenn der Patient *nicht* erkrankt ist und somit gesund ist. 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  $n = 3 \times 10^4$  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>18</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 Alex und Nilufar, dass beim diagnostischen Testen *True Positives (TP)*, *True Negatives (TN)*, *False Positives (FP)* und *False Negatives (FN)* auftreten. Das verstehen beiden so noch nicht und deshalb stellen Sie für Alex und Nilufar den Zusammenhang in einer 2x2 Kreuztabelle dar.

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

<sup>18</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!** Mark, Steffen und Jessica sitzen bei Alex und hören sich etwas über Network-Marketing an. Alex ist jetzt im Network-Marketing tätig. 'Jetzt reicht es. Wir sind eine Lerngruppe und du versuchst uns hier abzuziehen!', poltert Mark und fährt fort, 'Ich erklär dir mal, wie falsch du liegst!'. Eine Möglichkeit, leicht Geld zu verdienen, ist es anderen Menschen für Geld zu versprechen, wie man leicht reich werden kann. Am besten natürlich ohne viel Aufwand und ortsunabhängig. Schnell ein YouTube-Werbevideo gedreht und auf geht es mit unserem Network-Marketing. Aber Moment, wie funktioniert Network-Marketing eigentlich und was hat das alles mit einer Schneeballschlacht zu tun? Mark und Sie wollen hier einmal in die Untiefen des »passiven Einkommens« abtauchen und die Lerngruppe vor Schlimmeren bewahren<sup>19</sup>!

Das Jahr 2023 war das erfolgreichste Jahr in der Geschichte von Up/Down Systems and Networking (UDSysNet). Das Unternehmen steigerte den Umsatz um rund 18 Prozent von 290 Millionen Euro im Jahr 2022. Doch wie viel kommt bei den Partnern an? Laut UDSysNet habe das Unternehmen  $3.8 \times 10^5$  aktive Partner.

1. Berechnen Sie zuerst den Umsatz der Firma UDSysNet 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 30%? **(1 Punkt)**

Das von Alex zu vermarkende Produkt, hinter dem Alex voll steht, kostet 50EUR pro Einheit im Direktverkauf. Die direkte Provision für die erste Stufe beträgt 30%. Für die zweite, dritte und vierte Stufe betragen die indirekten Provisionen jeweils 3.25%, 2.25% und 1.75%. Jeder von Alex angeworbener »Partner« wirbt wiederum vier Partner für sich selbst an. Pro Monat werden im Schnitt zwei Einheiten vom Produkt verkauft. Alex will nun 4100EUR im Monat *passiv* – also durch indirekte Provisionen – erwirtschaften. Kann das klappen? Sie sind zusammen mit Mark skeptisch.

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

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

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

Alex musste zum Einstieg bei UDSysNet Einheiten des Produkts für 1500EUR kaufen. Diese Einheiten kann Alex nur direkt verkaufen. Das ganze Wohnzimmer ist voll davon. Leider musste Alex den Kauf über einen Kredit über 5.1% p.a. über 48 Monate finanzieren. Sie schütteln den Kopf und klären Alex 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>19</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** Steffen, Nilufar und Paula sitzen bei Jonas nachdem sich alle begeistert in der Serie *Stranger Thinks* verloren haben. Alle drei wollen jetzt einmal bei Jonas *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 Jonas nicht so alleine ist.

In dem Spiel hat Steffen nun auf einmal 7 achtseitige Würfel (7d8) zum würfeln in der Hand. Wenn Steffen eine 8 würfelt, hat Steffen einen Erfolg.

1. Berechnen Sie die Wahrscheinlichkeit genau 5 Erfolge zu erzielen! **(2 Punkte)**
2. Berechnen Sie die Wahrscheinlichkeit keinen Erfolg zu erzielen! **(1 Punkt)**

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

3. Würden Sie die Axt mit zwei sechseitigen Würfeln (2d6) als Schaden oder das Schwert mit einem sechseitigen Würfel plus 5 (1d6+5) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

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

4. Erstellen Sie eine 2x2 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 2x2 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 Alex gestern. Die Lerngruppe mit Steffen und Paula ging viel zu lang. Während er wegämmert, kommen in ihm 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 Alex fragt sich, ob er nicht doch lieber bezahlter Guest bei Barbara Salesch hätten sein sollten. Aber Alex braucht das Geld und jetzt heißt es Spielschows farmen! Erstmal eine Kaffemaschine von Mitropa gewinnen. Ein Kandidat gewinnt die Kaffemaschine von Mitropa, wenn nicht alle Kandidaten überbieten (eng. *outbid*). Mit Ihnen bilden Steffen und Paula das Team der drei Kandidaten. Alex braucht dringend Ihre Hilfe in seinen Wahnträumen. Sie gehen wie in *Interception* rein!<sup>20</sup>

Name	$P(\text{win})$	$P(\text{outbid})$
Steffen	0.3	0.11
Paula	0.3	0.076

1. Mit welcher Wahrscheinlichkeit gewinnen Sie die Kaffemaschine von Mitropa, wenn keiner der Kandidaten überbietet? (**1 Punkt**)
2. Wenn Ihre Überbietungswahrscheinlichkeit  $P(\text{outbid})$  bei 0.043 liegt, mit welcher Wahrscheinlichkeit gewinnt *keiner* die Kaffemaschine von Mitropa? (**1 Punkt**)

Glücksrad für Arme auf der Kirmes! Leider hat es für Maren Gilzer nicht gereicht. Deshalb sind Sie jetzt mit Alex auf der Kirmes und spielen mit Catwoman um das große Geld. Das Glücksrad hat 24 Felder. Sie beide drehen das Glücksrad zweimal. Auf 8 Feldern gewinnen Alex und Sie 4000EUR sonst 1500EUR. Ganz schön viel Geld und ganz schön zwißelichtig 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 5500EUR? (**1 Punkt**)

Im Fiebertraum von Alex reisen sie beide im Zug nach Köln um bei »Geh aufs Ganze!« mitzuspielen. Alex 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 einer 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>20</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 Kovariablen Laktationsnummer
- $e_{ijkl}$ : zufälliger Restfehler

Erläutern Sie anhand dieses Beispiels die Begriffe fixer Effekt, Interaktion, zufälliger Effekt und Kovariablen 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.

# Teil XIII.

## Platzhalter

### 125. Hausarbeit

(10 Punkte)

Inhalt folgender Module: Statistik • Angewandte Statistik und Versuchswesen • Modellierung landwirtschaftlicher Daten



Teil der Prüfungsleistung in dem Modul ist eine Hausarbeit zur Datenanalyse in R.

Diese Aufgabe dient in der Klausur als Platzhalter für die erreichten Punkte in der Hausarbeit.

Zum Bestehen des Moduls ist die Abgabe des Berichts nicht notwendig.

Alle Informationen zu der Hausarbeit sowie dem Stichtag der Abgabe finden Sie auf ILIAS.

# Teil XIV.

## Lösungen

Die folgenden Lösungen sind absolut ein Entwurf und können Fehler enthalten. Daher immer mit Vorsicht genießen. Im Weiteren ist das Runden so eine Sache. Deshalb wird es vermutlich auch Abweichungen zu deiner händischen Berechnung geben. Damit leben wir dann aber alle.

### Statistik

```
library(tidyverse)
data_tbl <- read.table(file = pipe("pbpaste"), sep = " ", header = TRUE)
```

### Mathematik

#### Herodot – der Schimmel aus Ivenack

1)	2)	3)	4)	5)
$d_{1820} = 419.93\text{cm}$	$r = 0 + 3\text{mm} * \text{Baumalter}$	$V_{eff} = 1.94\text{m}^3$	$s_{Eiche} = 89.96\text{cm}$	$s_{Platz} = 74.96\text{cm}$

#### Von Töpfen auf Tischen

2)	3a)	3b)	4)	5)
$n = 378$	$A_{Tisch} = 1.52\text{m}^2$	$A_{Topf} = 0.99\text{m}^2$	$V_{Topf} = 237.5\text{cm}^3$	$61.75\text{EUR}$

#### Solar- & Biogasanlagen

2)	3)	4)	5)
$A_{Dach} = 616.12\text{m}^2$	$\rho = 245\text{kg/m}^3$	$V = 36.73\text{m}^3$	$h = 11.69\text{m}$

#### Aliatorenbirnen und Blaubeere

1)	2)
$9.95\text{EUR}$	$n = 90$

3)	Wasserverbrauch	Mittleres Gewicht	Liter pro Stück
Strauchtomate	0.19l/g	107.5g	20.425l
Salat	0.13l/g	410g	53.3l
Avocado	1.1l/g	280g	308l
Blaubeere	0.82l/g	3.35g	2.747l

4a)	4b)	5a)	5b)	6)
$V = 603749600\text{hl}$	$V = 2468290000\text{hl}$	$4 \times$	$0.67 \times$	FAO Stat, Our World in Data

<b>1)</b>	<b>2)</b>	<b>3)</b>	<b>4)</b>	<b>5)</b>	<b>6)</b>
$h = 3.27 \cdot 11.5 = 37.61m$	$V_T = 27.14m^3$	$n = 1927$	$r = 216.59m$	$b' = 201.8m$	$66.54Ellen$

## Die Dampfnudelerde

### Pyramiden bauen

<b>1)</b>	<b>2)</b>	<b>3)</b>	<b>4)</b>	<b>5)</b>	<b>6)</b>
$h = 3.27 \cdot 11.5 = 37.61m$	$V_T = 27.14m^3$	$n = 1927$	$r = 216.59m$	$b' = 201.8m$	$66.54Ellen$