

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

*Nicht bestanden:* ☐

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

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

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

**Studierende der Fakultät Agrarwissenschaften und Landschaftsarchitektur (AuL)**

# **Klausurfragen der Bio Data Science**

**für Pflichtmodule**

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

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

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

24. Dezember 2023

### Erlaubte Hilfsmittel für die Klausur

- Normaler Taschenrechner ohne Möglichkeit der Kommunikation mit anderen Geräten - also ausdrücklich kein Handy!
- Eine DIN A4-Seite als beidseitig, selbstgeschriebene, handschriftliche Formelsammlung - keine digitalen Ausdrucke.
- **You can answer the questions in English without any consequences.**

### Ergebnis der Klausur

\_\_\_\_\_ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.

\_\_\_\_\_ von 67 Punkten sind aus dem Rechen- und Textteil erreicht.

\_\_\_\_\_ von 87 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
83.0 - 87.0	1,0
79.0 - 82.5	1,3
74.5 - 78.5	1,7
70.5 - 74.0	2,0
66.0 - 70.0	2,3
61.5 - 65.5	2,7
57.5 - 61.0	3,0
53.0 - 57.0	3,3
49.0 - 52.5	3,7
43.5 - 48.5	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.**
- Es werden nur Antworten berücksichtigt, die in dieser Tabelle angekreuzt sind!

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

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

## Rechen- und Textaufgaben

- Die Tabelle wird vom Dozenten ausgefüllt.

Aufgabe	11	12	13	14	15	16	17
Punkte	7	7	9	12	12	10	10

- Es sind \_\_\_\_ von 67 Punkten erreicht worden.

## ANOVA

### 1 Aufgabe

(2 Punkte)

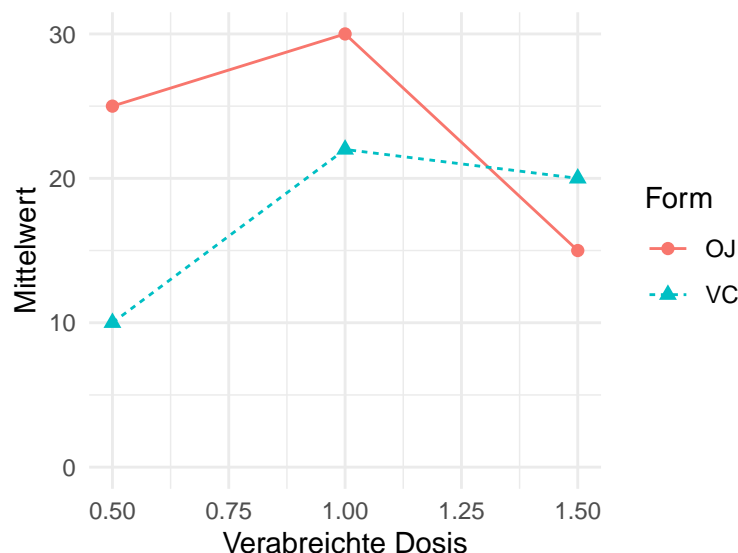
Der Datensatz PlantGrowth enthält das Gewicht von Pflanzen, die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden. Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein  $\eta^2 = 0.29$ . Welche Aussage ist richtig?

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

### 2 Aufgabe

(2 Punkte)

Die folgende Abbildung enthält die Daten aus einer Studie zur Bewertung der Wirkung von Vitamin C auf das Zahnwachstum bei Meerschweinchen. Der Versuch wurde an 60 Schweinen durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0.5, 1 und 1.5 mg/Tag) über eine von zwei Verabreichungsmethoden mit Orangensaft (OJ) oder Ascorbinsäure (VC) erhielt.



Welche Aussage ist richtig im Bezug auf eine zweifaktorielle ANOVA?

- A** ☐ Eine starke Interaktion liegt vor. Die Geraden laufen parallel und schneiden sich nicht.
- B** ☐ Keine Interaktion ist zu erwarten. Die Geraden der Verabreichungsmethode laufen parallel und mit ähnlichen Abständen.
- C** ☐ Eine leichte Interaktion ist zu erwarten. Die Geraden schneiden sich noch nicht, aber die Abstände unterscheiden sich stark.
- D** ☐ Eine starke Interaktion ist zu erwarten. Die Geraden schneiden sich und die Abstände sind nicht gleichbleibend.
- E** ☐ Keine Interaktion liegt vor. Die Geraden scheiden sich und laufen nicht parallel.

### 3 Aufgabe

(2 Punkte)

Eine einfaktorielle ANOVA berechnet eine Teststatistik um zu die Nullhypothese abzulehnen. Welche Aussage über die Teststatistik der ANOVA ist richtig?

- A** ☐ Die ANOVA berechnet die F-Statistik indem die MS der Behandlung durch die MS des Fehlers geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese nicht abgelehnt werden.
- B** ☐ Die ANOVA berechnet die F-Statistik aus den SS Behandlung geteilt durch die SS Fehler.
- C** ☐ 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 abgelehnt werden.
- D** ☐ Die ANOVA berechnet die F-Statistik indem die MS des Fehlers durch die MS der Behandlung geteilt werden. Wenn die F-Statistik sich der 1 annähert kann die Nullhypothese nicht abgelehnt werden.
- E** ☐ 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.

### 4 Aufgabe

(2 Punkte)

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

## Deskriptive Statistik & Explorative Datenanalyse

### 5 Aufgabe

(2 Punkte)

Berechnen Sie den Mittelwert und Standardabweichung von y mit 14, 11, 7, 8 und 16.

- A** ☐ Es ergibt sich 11.2 +/- 14.7
- B** ☐ Es ergibt sich 12.2 +/- 1.915
- C** ☐ Es ergibt sich 11.2 +/- 3.83
- D** ☐ Es ergibt sich 11.2 +/- 1.915
- E** ☐ Es ergibt sich 10.2 +/- 7.35

## 6 Aufgabe

(2 Punkte)

Berechnen Sie den Median, das 1<sup>st</sup> Quartile sowie das 3<sup>rd</sup> Quartile von  $y$  mit 31, 14, 23, 20, 16, 6 und 63.

- A ☐ Es ergibt sich 20 +/- 14
- B ☐ Es ergibt sich 25 [15, 32]
- C ☐ Es ergibt sich 20 +/- 31
- D ☐ Es ergibt sich 20 [14, 31]
- E ☐ Es ergibt sich 25 +/- 14

## 7 Aufgabe

(2 Punkte)

Die empfohlene Mindestanzahl an Beobachtungen für einen Boxplot sind...

- A ☐ mindestens 20 Beobachtungen.
- B ☐ 5 und mehr Beobachtungen.
- C ☐ 1 Beobachtung.
- D ☐ 10 Beobachtungen.
- E ☐ 2-5 Beobachtungen.

## 8 Aufgabe

(2 Punkte)

Nachdem Sie in einem Experiment die Daten  $D$  erhoben haben, berechnen Sie den Mittelwert und den Median. Der Mittelwert  $\bar{y}$  und der Median  $\tilde{y}$  unterscheiden sich. Welche Aussage ist richtig?

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

## Lineare Regression & Korrelation

### 9 Aufgabe

(2 Punkte)

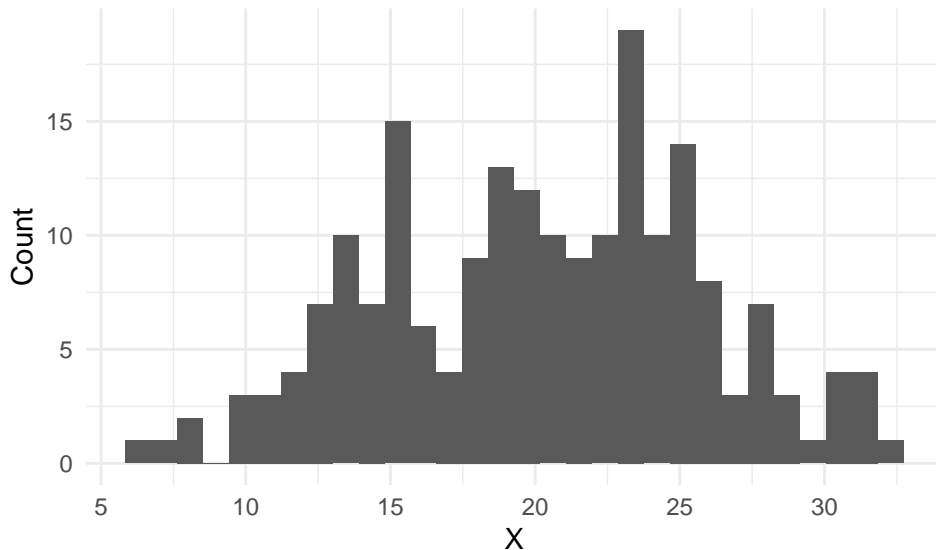
Der Barplot stellt folgende statistische Maßzahlen in einer Abbildung dar. Damit gehört der Barplot zu einem der am meisten genutzten statistischen Verfahren zur Visualisierung von Daten.

- A ☐ Den Median und die Standardabweichung.
- B ☐ Den Mittelwert und die Varianz.
- C ☐ Den Mittelwert und die Standardabweichung.
- D ☐ Den Median und die Quartile.
- E ☐ Den Mittelwert sowie den Median und die Streuung.

## 10 Aufgabe

(2 Punkte)

In dem folgenden Histogramm von  $n = 200$  Pflanzen ist welche Verteilung mit welchen korrekten Verteilungsparametern dargestellt?



- A** ☐ Eine rechtsschiefe, multivariate Normalverteilung.
- B** ☐ Eine Standardnormalverteilung mit  $N(0,1)$ .
- C** ☐ Es handelt sich um eine Normalverteilung mit  $N(20, 5)$ .
- D** ☐ Es handelt sich um eine Binomial-Verteilung mit  $\text{Binom}(10)$ .
- E** ☐ Es handelt sich um eine Poisson-Verteilung mit  $\text{Pois}(20)$ .

## 11 Aufgabe

(2 Punkte)

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

- A** ☐ Ein kausales Modell möchte die Zusammenhänge von  $X$  auf  $Y$  modellieren. Hierbei geht es um die Effekte von  $X$  auf  $Y$ . Man sagt, wenn  $X$  um 1 ansteigt ändert sich  $Y$  um einen Betrag  $\beta$ .
- B** ☐ Ein kausales Modell benötigt mindestens eine Fallzahl von über 100 Beobachtungen und darf keine fehlenden Werte beinhalten. Die Varianzkomponenten müssen homogen sein.
- C** ☐ Ein kausales Modell schliesst grundsätzlich lineare Modell aus. Es muss ein Graph gefunden werden, der alle Punkte beinhaltet. Erst dann kann das  $R^2$  berechnet werden.
- D** ☐ Ein kausales Modell basiert auf einem Trainingsdatensatz und einem Testdatensatz. Auf dem Trainingsdatensatz wird das Modell trainiert und auf dem Testdatensatz validiert.
- E** ☐ Ein kausales Modell wird auf einem Trainingsdatensatz trainiert und anschliessend über eine explorative Datenanalyse validiert. Signifikanzen über  $\beta_i$  können hier nicht festgestellt werden.

## 12 Aufgabe

(2 Punkte)

In einer linearen Regression werden die  $\epsilon$  oder Residuen geschätzt. Welcher Verteilung folgen die Residuen bei einer optimalen Modellierung?

- A** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(0, s^2)$ .
- B** ☐ Die Residuen sind binomialverteilt.
- C** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(\bar{y}, s^2)$ .
- D** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(0, 1)$ .
- E** ☐ Die Residuen folgen einer Poissonverteilung mit  $\text{Pois}(0)$ .

### 13 Aufgabe

(2 Punkte)

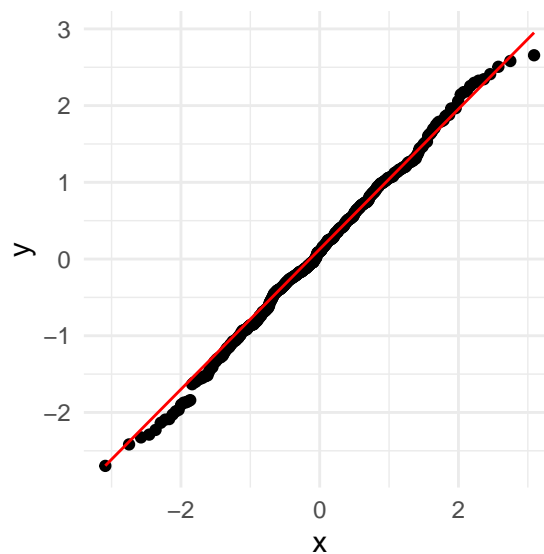
Welche Aussage über das *generalisierte lineare Modell (GLM)* ist richtig?

- A** ☐ Das GLM erlaubt auch weitere Verteilungsfamilien für das Y bzw. das Outcome in einer linearen Regression zu wählen.
- B** ☐ Das GLM erlaubt auch nicht normalverteilte Residuen in der Schätzung der Regressionsgrade.
- C** ☐ Das GLM ist eine allgemeine Erweiterung der linearen Regression auf die Normalverteilung.
- D** ☐ Das GLM ist eine Vereinfachung des LM in R. Mit dem GLM lassen polygonale Regressionen rechnen.
- E** ☐ Das GLM ist ein faktisch maschineller Lernalgorithmus, der selbstständig die Verteilungsfamilie für Y wählt.

### 14 Aufgabe

(2 Punkte)

Sie rechnen in eine linearen Regression und erhalten folgenden QQ Plot. Welche Aussage ist richtig?



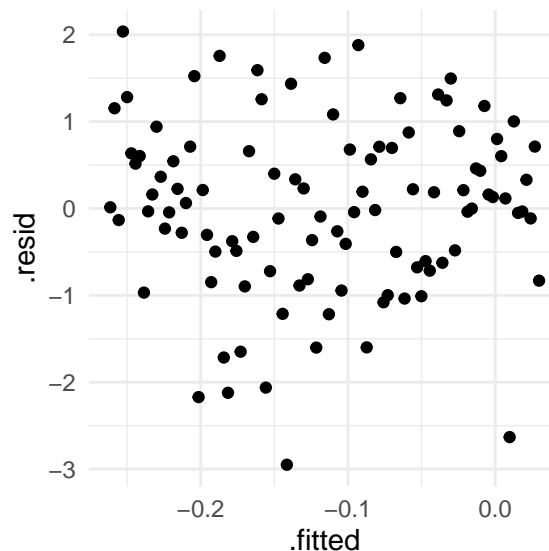
- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- B** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden.
- C** ☐ Die Annahme der normalverteilten Residuen ist 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** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden und Korrelation ist negativ.

### 15 Aufgabe

(2 Punkte)

Sie rechnen eine linearen Regression und erhalten folgenden Residual Plot. Welche Aussage ist richtig?





- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Kein Muster ist zu erkennen und keine Outlier zu beobachten.
- B** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Es ist kein Muster zu erkennen.
- C** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Vereinzelt Punkte liegen oberhalb bzw. unterhalb der Geraden um die 0 Linie weiter entfernt. Ein klares Muster ist zu erkennen.
- D** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Diagonalen.
- E** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Es ist ein Muster zu erkennen.

## 16 Aufgabe

(2 Punkte)

Welche Aussage über den Korrelationskoeffizienten nach Kendall ist richtig?

- A** ☐ Der Korrelationskoeffizienten nach Kendall wird genutzt, wenn der Korrelationskoeffizienten zwischen -1 und 1 liegt. Dann sind die Residuen normalverteilt.
- B** ☐ Der Korrelationskoeffizienten nach Kendall wird genutzt, wenn das Outcome Y nicht normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen 0 und 1.
- C** ☐ Der Korrelationskoeffizienten nach Kendall wird genutzt, wenn das Outcome Y nicht normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen -1 und 1.
- D** ☐ Der Korrelationskoeffizienten nach Kendall wird genutzt, wenn das Outcome Y normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen 0 und 1.
- E** ☐ Der Korrelationskoeffizienten nach Kendall wird genutzt, wenn das Outcome Y normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen -1 und 1.

## Vermischte Themen

## 17 Aufgabe

(2 Punkte)

Nach einer simplen linearen Regression zur Untersuchung vom Einfluss der  $CO_2$ -Konzentration [ $\mu g$ ] im Wasser auf das Wachstum von Wasserlinsen [ $kg$ ] erhalten Sie einen  $\beta_1$  Koeffizienten von 0.00001 und einen hoch signifikanten  $p$ -Wert mit  $2.3 \cdot 10^{-9}$ . Warum sehen Sie so einen kleinen Effekt bei einer so deutlichen Signifikanz?

- A** ☐ Die Einheit der  $CO_2$ -Konzentration ist zu klein gewählt. Dadurch sehen wir den sehr kleinen  $p$ -Wert. Der  $p$ -Wert und die Einheit von der  $CO_2$ -Konzentration hängen zusammen.

- B** ☐ Das Gewicht und die  $CO_2$ -Konzentration korrelieren sehr stark, deshalb wird der  $\beta_1$  Koeffizient sehr klein.
- C** ☐ Die Einheit der  $CO_2$ -Konzentration ist zu klein gewählt. Die Erhöhung der  $CO_2$ -Konzentration um 1 führt nur zu einem sehr winzigen Anstieg im Gewicht der Wasserlinsen. Die Einheit muss besser gewählt werden.
- D** ☐ Die Fallzahl ist zu klein angesetzt. Je kleiner die Fallzahl ist, desto höher ist die Teststatistik und damit auch der  $p$ -Wert kleiner.
- E** ☐ Die Fallzahl ist zu hoch angesetzt. Je höher die Fallzahl ist, desto kleiner ist die Teststatistik und damit ist dann auch der  $p$ -Wert sehr klein.

## 18 Aufgabe

(2 Punkte)

Sie rechnen eine simple logistische Regression. Welche Aussage betreffend der Konfidenzintervalle ist für die logistische Regression richtig?

- A** ☐ Wenn die 0 im Konfidenzintervall enthalten ist, kann die Nullhypothese abgelehnt werden.
- B** ☐ Wenn die Relevanzschwelle mit enthalten ist, kann die Nullhypothese abgelehnt werden.
- C** ☐ Wenn die 1 im Konfidenzintervall enthalten ist, kann die Nullhypothese nicht abgelehnt werden.
- D** ☐ Wenn die 0 im Konfidenzintervall enthalten ist, kann die Nullhypothese nicht abgelehnt werden.
- E** ☐ Wenn die Konfidenzintervalle den  $p$ -Wert der Regression enthalten, kann die Nullhypothese abgelehnt werden.

## 19 Aufgabe

(2 Punkte)

Welche Aussage über die nicht-parametrische Statistik ist richtig?

- A** ☐ Die nicht-parametrische Statistik ist ein Vorgänger der parametrischen Statistik und wurde wegen dem Mangel an Effektschätzern nicht mehr ab 1960 genutzt.
- B** ☐ Die nicht-parametrische Statistik basiert auf dem Schätzen von Parametern aus einer festgelegten Verteilung. Daher gibt es auch direkt zu interpretierenden Effektschätzer.
- C** ☐ Die nicht-parametrische Statistik basiert auf Rängen. Daher gibt es auch direkt zu interpretierenden Effektschätzer.
- D** ☐ Die nicht-parametrische Statistik basiert auf dem Schätzen von Parametern aus einer a priori festgelegten Verteilung. Daher gibt es auch direkt zu interpretierenden Effektschätzer.
- E** ☐ Die nicht-parametrische Statistik basiert auf Rängen. Daher wird jeder Zahl ein Rang zugeteilt. Nur auf den Rängen wird die Auswertung gerechnet. Daher gibt es auch keinen direkt zu interpretierenden Effektschätzer.

## 20 Aufgabe


(2 Punkte)



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

- A** ☐ Randomisierung war bis 1952 bedeutend, wurde dann aber in Folge besserer Rechnerleistung nicht mehr verwendet. Aktuelle Statistik nutzt keine Randomisierung mehr.
- B** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- C** ☐ Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte.
- D** ☐ Randomisierung erlaubt erst die Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich.
- E** ☐ Randomisierung sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.

## 21 Aufgabe


(2 Punkte)

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

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

## 22 Aufgabe

(2 Punkte)

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

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

## 23 Aufgabe

(2 Punkte)

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

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

## 24 Aufgabe

(2 Punkte)

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

- A** ☐ Untereinander unabhängig. Sollten die Mütter verwandt sein, so ist die Varianzstruktur ähnlich und muss modelliert werden.
- B** ☐ Untereinander abhängig. Die Ferkel stammen von einem Muttertier und haben vermutlich eine ähnliche Varianzstruktur.
- C** ☐ Untereinander unabhängig. Die Ferkel sind eigenständig und benötigen keine zusätzliche Behandlung.
- D** ☐ Untereinander abhängig, wenn die Mütter ebenfalls miteinander verwandt sind. Erst die Abhängigkeit 2. Grades wird in der Statistik modelliert.
- E** ☐ Untereinander stark korreliert. Die Ferkel sind von einer Mutter und somit miteinander korreliert. Dies wird in der Statistik jedoch meist nicht modelliert.

## 25 Aufgabe

(2 Punkte)

Welche Aussage über Cook's d und Cohen's d ist richtig?

- A** ☐ Wir nutzen Cook's d um Outlier in den Daten zu finden und Cohen's d um standardisierte Outlier für Gruppenvergeliche zu erhalten.
- B** ☐ Wir nutzen Cook's d um Outlier in den Daten zu finden und Cohen's d um einen nicht standardisierten Effektschätzer für Gruppenvergeliche zu erhalten.
- C** ☐ Wir nutzen Cohen's d um Outlier in den Daten zu finden und Cook's d um einen standardisierten Effektschätzer für Gruppenvergeliche zu erhalten.
- D** ☐ Wir nutzen Cook's d um Outlier in den Daten zu finden und Cohen's d um einen standardisierten Effektschätzer für Gruppenvergeliche zu erhalten.
- E** ☐ Wir nutzen Cook's d um Outlier in den Daten zu finden. Cohen's d findet auch Outlier, ist aber ein veraltetes Konzept in der Statistik.

## 26 Aufgabe

(2 Punkte)

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Kühen durch. Bei 4 Tieren finden Sie eine Erkrankung der Klauen vor und 12 Tiere sind gesund. Welche Aussage über den Risk ratio Effektschätzer ist richtig?

- A** ☐ Es ergibt sich ein Risk ratio von 0.25, da es sich um eine Chancenverhältnis handelt.
- B** ☐ Es ergibt sich ein Risk ratio von 0.25, da es sich um ein Anteil handelt.
- C** ☐ Es ergibt sich ein Risk ratio von 0.33, da es sich um ein Anteil handelt.
- D** ☐ Es ergibt sich ein Risk ratio von 3, da es sich um ein Anteil handelt.
- E** ☐ Es ergibt sich ein Risk ratio von 0.33, da es sich um eine Chancenverhältnis handelt.

## Multiple Gruppenvergleiche

### 27 Aufgabe

(2 Punkte)

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

- A** ☐ Aktuell werden immer grössere Datensätze erhoben. Dadurch wird auch die Varianz immer höher was automatisch zu mehr signifikanten Ergebnissen führt.

- B** ☐ Aktuell werden immer grössere Datensätze erhoben. Eine erhöhte Fallzahl führt automatisch auch zu mehr signifikanten Ergebnissen, selbst wenn die eigentlichen Effekte nicht relevant sind.
- C** ☐ 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.
- D** ☐ Big Data ist ein Problem der parametrischen Statistik. Parameter lassen sich nur auf kleinen Datensätzen berechnen, da es sich sonst nicht mehr um eine Stichprobe im engen Sinne der Statistik handelt.
- E** ☐ Aktuell werden zu grosse Datensätze für die gängige Statistik gemessen. Daher wendet man maschinelle Lernverfahren für kausale Modelle an. Hier ist die Relevanz gleich Signifikanz.

## Statistische Testtheorie

### 28 Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.01, 0.02, 0.89, 0.21, 0.001 und 0.03. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.06, 0.12, 5.34, 1.26, 0.006 und 0.18. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.06, 0.12, 1, 1, 0.006 und 0.18. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.06, 0.12, 1, 1, 0.006 und 0.18. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 0.83% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0017, 0.0033, 0.1483, 0.035, 2e-04 und 0.005. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0017, 0.0033, 0.1483, 0.035, 2e-04 und 0.005. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 0.83% verglichen.

### 29 Aufgabe

(2 Punkte)

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

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

### 30 Aufgabe

(2 Punkte)

Das Falsifikationsprinzip besagt...

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

### 31 Aufgabe

(2 Punkte)

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

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

### 32 Aufgabe

(2 Punkte)

Welche Aussage über die Power ist richtig?

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

### 33 Aufgabe

(2 Punkte)

Beim statistischen Testen wird signal mit noise zur Teststatistik T verrechnet. Welche der Formel berechnet korrekt die Teststatistik T?

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

### 34 Aufgabe

(2 Punkte)

In der Theorie zur statistischen Testentscheidung kann „ $H_0$  beibehalten obwohl die  $H_0$  falsch ist“ in welche richtige Analogie gesetzt werden?

- A ☐ In die Analogie eines brennenden Hauses ohne Rauchmelder: *House without noise*.
- B ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.
- C ☐ In die Analogie eines Rauchmelders: *Alarm with fire*.
- D ☐ In die Analogie eines Rauchmelders: *Fire without alarm*, dem  $\beta$ -Fehler.
- E ☐ In die Analogie eines Rauchmelders: *Alarm without fire*, dem  $\alpha$ -Fehler.

### 35 Aufgabe

(2 Punkte)

Welche statistische Masszahl erlaubt es *Relevanz* mit *Signifikanz* zuverbinden? Welche Aussage ist richtig?

- A ☐ 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.
- B ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- C ☐ Das Konfidenzintervall. Durch die Visualisierung des Konfidenzintervalls kann eine Relevanzschwelle vom Anwender definiert werden. Zusätzlich erlaubt das Konfidenzintervall auch eine Entscheidung über die Signifikanz.
- D ☐ Das  $\Delta$ . Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da  $\Delta$  antiproportional zum p-Wert ist, bedeutet auch ein hohes  $\Delta$  ein sehr kleinen p-Wert.
- E ☐ Die Teststatistik. Durch den Vergleich von  $T_c$  zu  $T_k$  ist es möglich die  $H_0$  abzulehnen. Die Relevanz ergibt sich aus der Fläche rechts vom dem  $T_c$ -Wert.

## Statistische Tests für Gruppenvergleiche

### 36 Aufgabe

(2 Punkte)

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

- A ☐ Wir machen eine Aussage über die individuelle Wahrscheinlichkeit des Eintretens der Nullhypothese  $H_0$ .
- B ☐ Wir vergleichen mit dem p-Wert und dem Signifikanzniveau  $\alpha$  Wahrscheinlichkeiten und damit die absoluten Werte auf einem Zahlenstrahl, wenn die  $H_0$  gilt.
- C ☐ 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.
- D ☐ Wir vergleichen die Effekte des p-Wertes mit den Effekten der Signifikanzschwelle unter der Annahme der Nullhypothese.
- E ☐ Wir vergleichen mit dem p-Wert und dem Signifikanzniveau  $\alpha$  Wahrscheinlichkeiten und damit die Flächen unter der Kurve der Teststatistik, wenn die  $H_0$  gilt.

### 37 Aufgabe

(2 Punkte)

Die Mindestanzahl an Beobachtungen für eine Zelle der Vierfeldertafel bei der Nutzung eines Chi-Quadrat-Testes ist...

- A ☐ 5 Beobachtungen
- B ☐ 1 Beobachtung
- C ☐ 0 Beobachtungen
- D ☐ 2 Beobachtungen
- E ☐ 10 Beobachtungen

### 38 Aufgabe

(2 Punkte)

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

- A** ☐ Beim gepaarten t-Test kombinieren wir die Vorteile des Student t-Test für Varianzhomogenität mit den Vorteilen des Welch t-Test für Varianzheterogenität. Wir bilden dafür die Differenz der Einzelbeobachtungen.
- B** ☐ Der gepaarte t-Test wird gerechnet, wenn die Beobachtungen nicht unabhängig voneinander sind. Wir messen wiederholt an dem gleichen Probanden oder Tier oder Pflanze. Wir bilden die Differenzen um den gepaarten t-Test rechnen zu können.
- C** ☐ Der gepaarte t-Test wird genutzt, wenn die Differenzen der Beobachtungen verbunden sind und wir dadurch die Unabhängigkeit nicht mehr vorliegen haben.
- D** ☐ Der gepaarte t-Test nutzt die Varianz der Beobachtungen jeweils paarweise und bildet dafür eine verbundene Stichprobe. Dieser Datensatz  $d$  dient dann zur Differenzbildung.
- E** ☐ Der gepaarte t-Test wird gerechnet, wenn die Beobachtungen abhängig voneinander sind. Wir messen jede Beobachtung nur einmal und berechnen dann die Differenz zu dem Mittel der anderen Beobachtungen.

### 39 Aufgabe

(2 Punkte)

Welche Aussage über den Welch t-Test ist richtig?

- A** ☐ Der Welch t-Test ist die veraltete Form des Student t-Test und wird somit nicht mehr verwendet.
- B** ☐ Der Welch t-Test vergleicht die Mittelwerte von zwei Gruppen unter der strikten Annahme von Varianzhomogenität.
- C** ☐ Der Welch t-Test wird angewendet, wenn Varianzheterogenität zwischen den beiden zu vergleichenden Gruppen vorliegt.
- D** ☐ Der Welch t-Test ist ein Post-hoc Test der ANOVA und basiert daher auf dem Vergleich der Varianz.
- E** ☐ Der Welch t-Test vergleicht die Varianz von zwei Gruppen.

### 40 Aufgabe

(2 Punkte)

Nach einem Experiment mit fünf Weizensorten ergibt eine ANOVA ( $p = 0.041$ ) einen signifikanten Unterschied für den Ertrag. Sie führen anschließend die paarweisen t-Tests für alle Vergleiche der verschiedenen Weizensorten durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der  $\alpha$ -Schwelle. Sie schauen sich auch die rohen, unadjustierten p-Werte an und finden hier als niedrigsten p-Wert  $p_{3-2} = 0.053$ . Welche Aussage ist richtig?

- A** ☐ Die adjustierten p-Werte deuten in die richtige Richtung. Zusammen mit den nicht signifikanten rohen p-Werten ist von einem Fehler in der ANOVA auszugehen.
- B** ☐ Es gibt einen Fehler in der Varianzstruktur. Daher kann die ANOVA nicht richtig sein und paarweise t-Tests liefern das richtige Ergebnis.
- C** ☐ Die ANOVA testet auf der gesamten Fallzahl. Es wäre besser die ANOVA auf der gleichen Fallzahl wie die einzelnen t-Tests zu rechnen.
- D** ☐ Die ANOVA testet auf der gesamten Fallzahl. Die einzelnen t-Tests immer nur auf einer kleineren Subgruppe. Da mit weniger Fallzahl weniger signifikante Ergebnisse zu erwarten sind, kann eine Diskrepanz zwischen der ANOVA und den paarweisen t-Tests auftreten.
- E** ☐ Der Fehler liegt in den t-Tests. Wenn eine ANOVA signifikant ist, dann muss zwangsweise auch ein t-Test signifikant sein.



## 41 Aufgabe

(2 Punkte)

Welche Aussage über den t-Test ist richtig?

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

## Deskriptive Statistik & Explorative Datenanalyse

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- Kapitel 15 - Deskriptive Statistik
- Kapitel 16 - Visualisierung von Daten
- Kapitel 18 - Verteilung von Daten

## 42 Aufgabe

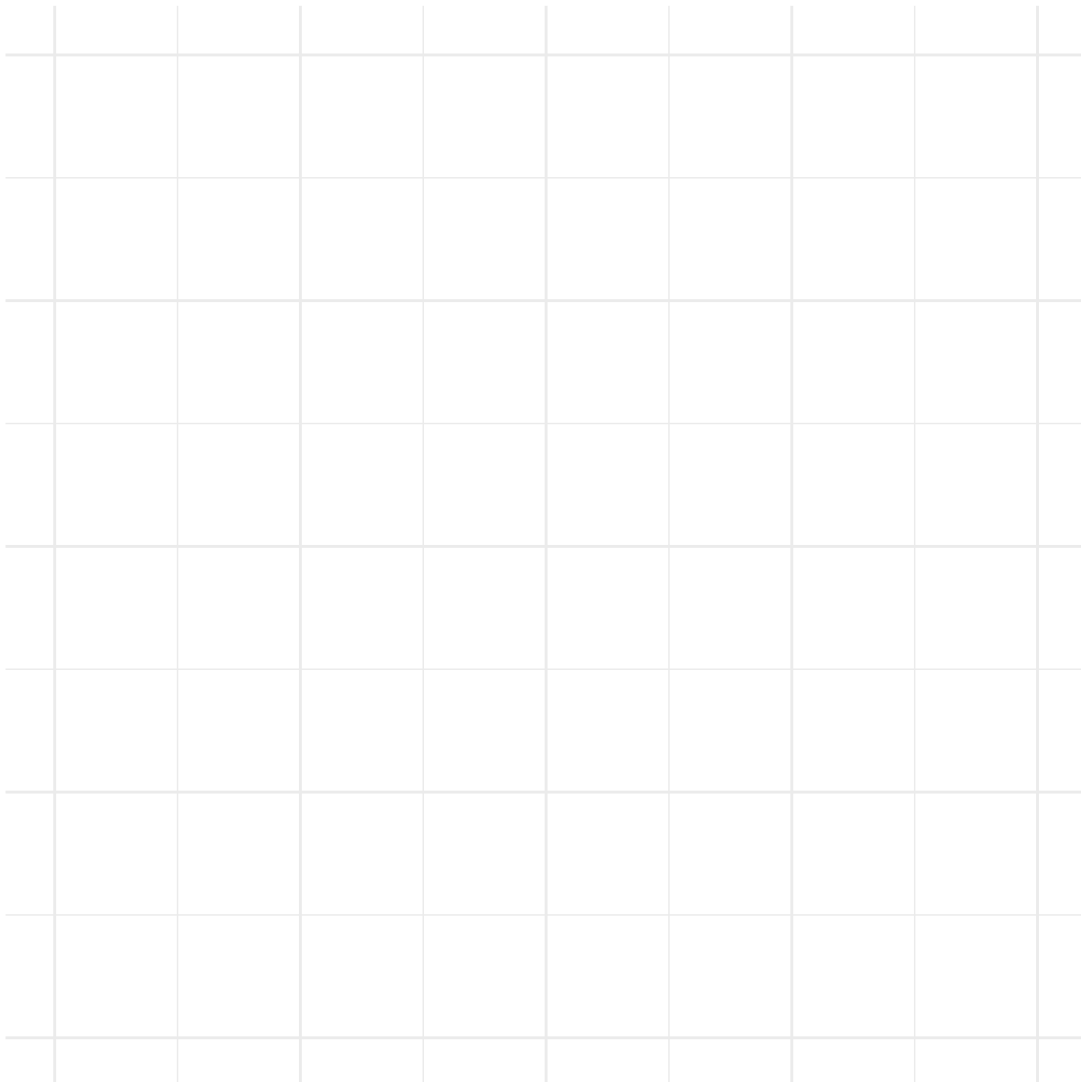
(8 Punkte)

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



Sie haben folgende Zahlenreihe  $y$  vorliegen  $y = \{16, 12, 17, 18, 14, 17\}$ .

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



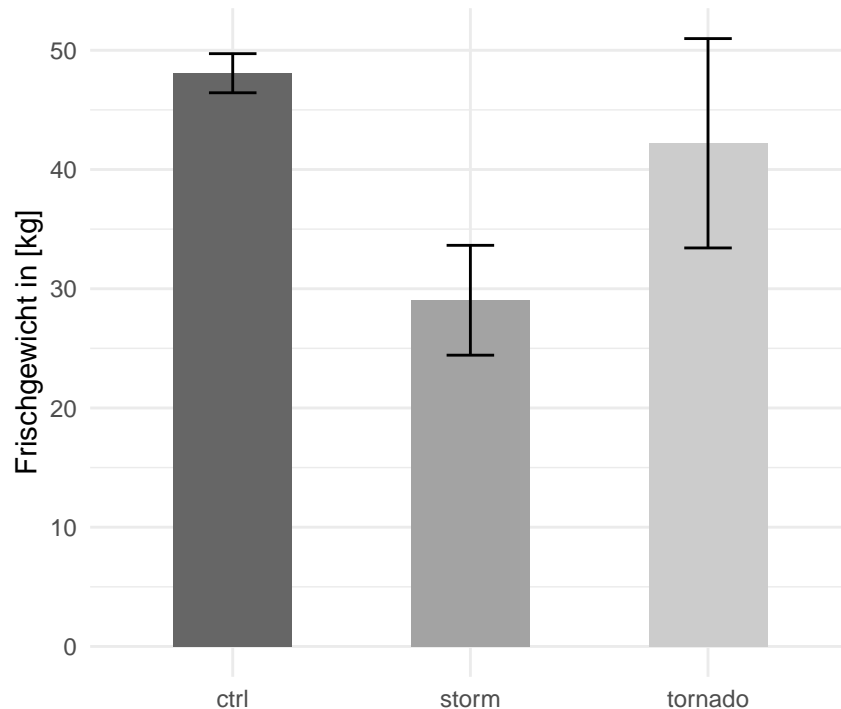
### 43 Aufgabe


(7 Punkte)

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



In einem Gewächshausexperiment mit drei Lüftungssystemen und Folientunneln (*ctrl*, *storm* und *tornado*) als Behandlung (*treatment*) ergibt sich die folgende Barplots mit dem gemessenen Frischgewicht (*fresh-matter*) von Kartoffel.



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

#### 44 Aufgabe

(7 Punkte)

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



In einem Gewächshausexperiment mit drei Düngestufen (*ctrl*, *low* und *high*) als Behandlung (*treatment*) ergibt sich die folgende Datentabelle mit dem gemessenen Frischgewicht (*freshmatter*) von Erdbeeren.

treatment	freshmatter
high	50.7
ctrl	11.9
low	26.2
high	55.5
high	31.0
high	44.1
ctrl	30.8
ctrl	25.2
low	32.0
low	28.8
ctrl	31.2
ctrl	30.6

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

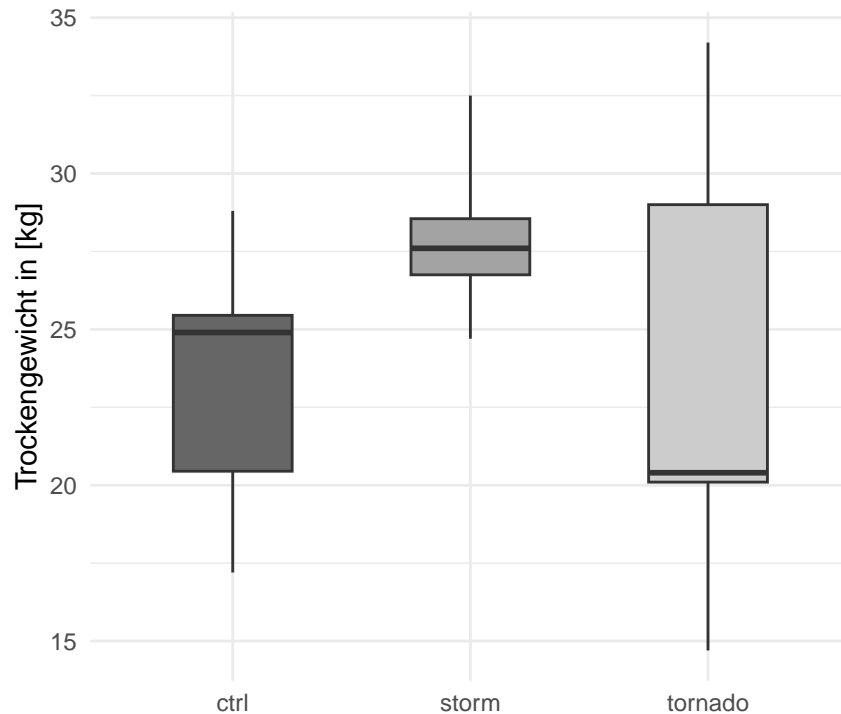
## 45 Aufgabe


(9 Punkte)

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



Nach einem Freilandexperiment mit zwei Lüftungssystemen und Folientunneln (*ctrl*, *storm* und *tornado*) als Behandlung (*treatment*) ergibt sich die folgende Boxplots mit dem gemessenen Trockengewicht (*drymatter*) von Erbsen.



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

## 46 Aufgabe

(9 Punkte)

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



Nach einem Freilandexperiment mit zwei Bewässerungstypen (*low* und *high*) als Behandlung (*treatment*) ergibt sich die folgende Datentabelle mit dem gemessenen Trockengewicht (*drymatter*) von Lauch.

treatment	freshmatter
high	40.5
low	40.6
high	44.4
high	32.1
low	39.5
low	47.3
low	31.8
low	34.7
high	45.9
low	42.6
low	36.6
high	40.2
high	60.7
high	35.7
low	27.6
high	34.0
high	49.7
high	63.0
high	33.6

1. Zeichnen Sie in *einer* Abbildung die beiden Boxplots für die zwei Behandlungen von Lauch! Beschriften Sie die Achsen entsprechend! **(6 Punkte)**
2. Beschriften Sie *einen* der beiden Boxplots mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen den Behandlungen von Lauch erwarten würden, wie sehen dann die beiden Boxplots aus? **(1 Punkt)**

## 47 Aufgabe

(6 Punkte)



Nach einer Bonitur von Lauch mit einer Kontrolle und drei Pestiziden (*ctrl*, *pestKill*, *roundUp* und *zeroX*) ergibt sich die folgende Datentabelle mit den Boniturnoten (*grade*).

pesticide	grade
ctrl	11
roundUp	3
pestKill	3
ctrl	7
ctrl	5
roundUp	4
zeroX	1
zeroX	1
zeroX	1
pestKill	2
roundUp	4
pestKill	3
zeroX	3
roundUp	4

1. Zeichnen Sie in *einer* Abbildung die Dotplots für die vier Pestizidlevel! Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Ergänzen Sie die Dotplots mit der gängigen statistischen Maßzahl! **(1 Punkt)**
3. Wenn Sie *keinen Effekt* zwischen den Pestizidlevel erwarten würden, wie sehen dann die Dotplots aus? **(1 Punkt)**



## 48 Aufgabe

(8 Punkte)

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



Nach einem Feldexperiment mit zwei Pestiziden (*RoundUp* und *OutEx*) ergibt sich die folgende Datentabelle mit dem jeweiligen beobachteten Infektionsstatus bei Spargel.

pesticide	infected
RoundUp	no
OutEx	yes
RoundUp	yes
OutEx	yes
RoundUp	yes
RoundUp	no
OutEx	yes
OutEx	yes
OutEx	yes
OutEx	no
RoundUp	yes
RoundUp	yes
OutEx	yes
RoundUp	yes
OutEx	no
RoundUp	yes
OutEx	no
RoundUp	yes

1. Stellen Sie in einer 2x2 Tafel den Zusammenhang zwischen dem Pestizid und dem Infektionsstatus dar!  
(4 Punkte)
2. Zeichnen Sie den zugehörigen Mosaic-Plot. Berechnen Sie das Verhältnis pro Spalte! (2 Punkte)
3. Wenn das Pestizid keine Auswirkung auf den Infektionsstatus hätte, wie sehe dann der Mosaic-Plot aus?  
(2 Punkte)

## 49 Aufgabe

(10 Punkte)

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



In einem Feldexperiment für die Bodendurchlässigkeit wurde der Niederschlag pro Parzelle sowie der durchschnittliche Ertrag gemessen. Es ergibt sich folgende Datentabelle.

water	drymatter
20	20
11	25
18	29
23	37
7	27
7	29
16	14
13	21
19	29

1. Erstellen Sie den Scatter-Plot für die Datentabelle. Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Zeichnen Sie eine Gerade durch die Punkte! **(1 Punkt)**
3. Beschriften Sie die Gerade mit den gängigen statistischen Maßzahlen! Geben Sie die numerischen Zahlenwerte mit an! **(3 Punkte)**
4. Wenn kein Effekt von dem Niederschlag auf das Trockengewicht vorhanden wäre, wie würde die Gerade verlaufen und welche Werte würden die statistischen Maßzahlen annehmen? **(2 Punkt)**

## 50 Aufgabe

(6 Punkte)

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



Nach einem Gewächshausexperiment zählen Sie folgende Anzahl an Läsionen auf den Blättern von Maiss nach einer durchgestandenen Infektion der Pflanze.

3, 4, 4, 2, 2, 0, 8, 3, 3, 3, 4, 5, 6, 4, 4, 6, 5, 2, 2, 1, 4, 6, 3, 1, 3, 7, 3, 4, 0, 6, 2, 2, 3

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

## 51 Aufgabe

(8 Punkte)

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



Nach einem Freilandexperiment bestimmen Sie folgende Trockengewichte von Erdbeeren nach einer durchgestandenen Infektion der Pflanzen.

10, 11.5, 9.5, 12.2, 11.3, 8.3, 10.7, 8.5, 9, 12.2, 9.4, 9.6, 12, 7.1, 10.2, 10.5, 10.7, 8.6, 7.5, 10.2, 11.4, 5.7, 7.5

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)**

## 52 Aufgabe

(9 Punkte)

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

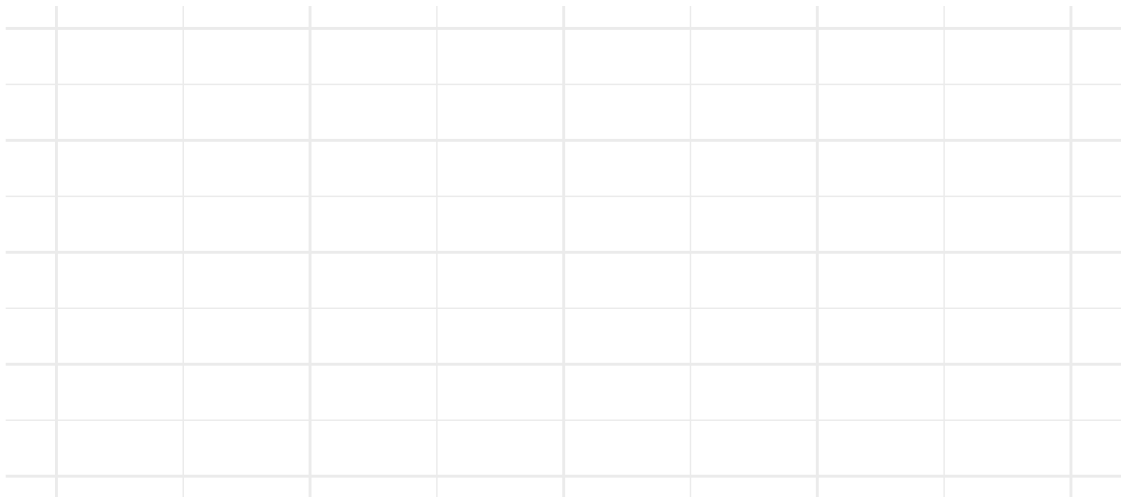


1. Skizzieren Sie in die unten stehenden, freien Abbildungen die Verteilungen, die sich nach der Abbildungsüberschrift ergeben! **(6 Punkte)**
2. Beschriften Sie die Achsen der Abbildungen entsprechend! **(1 Punkt)**
3. Achten Sie auf die entsprechende Skalierung der beiden Verteilungen in den Abbildungen!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

$N(0, 4)$  und  $N(2, 4)$



$\text{Pois}(1)$  und  $\text{Pois}(15)$



### 53 Aufgabe

(9 Punkte)

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



1. Skizzieren Sie 4 Normalverteilungen *in einer Abbildung* mit  $\bar{y}_1 \neq \bar{y}_2 \neq \bar{y}_3 \neq \bar{y}_4$  und  $s_1 \neq s_2 \neq s_3 \neq s_4$ ! **(3 Punkte)**
2. Beschriften Sie die Normalverteilungen mit den entsprechenden Parametern! **(2 Punkte)**
3. Ergänzen Sie die Bereiche in der 68% und 95% der Beobachtungen fallen! Beschriften Sie die Grenzen der Bereiche mit der statistischen Maßzahl! **(2 Punkte)**
4. Liegt Varianzhomogenität oder Varianzheterogenität vor? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 54 Aufgabe

(8 Punkte)

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



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



## Statistisches Testen

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- [Kapitel 19 - Die Testentscheidung](#)
- [Kapitel 20 - Die Testtheorie](#)



## 55 Aufgabe

(9 Punkte)



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

## 56 Aufgabe

(9 Punkte)



1. Tragen Sie folgende Fachbegriffe zur statistischen Testtheorie korrekt in eine passende Kreuztabelle ein! **(3 Punkte)**
  - Testentscheidung
  - 20%
  - $\beta$ -Fehler
  - $H_0$  beibehalten
2. Ergänzen Sie die Kreuztabelle um vier weitere, passende Fachbegriffe zur statistischen Testtheorie! **(2 Punkte)**
3. In der Analogie des Feuermelders, wie lautet der  $\alpha$ -Fehler? **(1 Punkt)**
4. In der Analogie des Feuermelders, wie lautet der  $\beta$ -Fehler? **(1 Punkt)**
5. Wenn der Feuermelder einmal pro Tag messen würde, wie oft würde der Feuermelder mit einem  $\alpha$  von 5% in einem Monat Alarm schlagen? Begründen Sie Ihre Antwort! **(2 Punkte)**

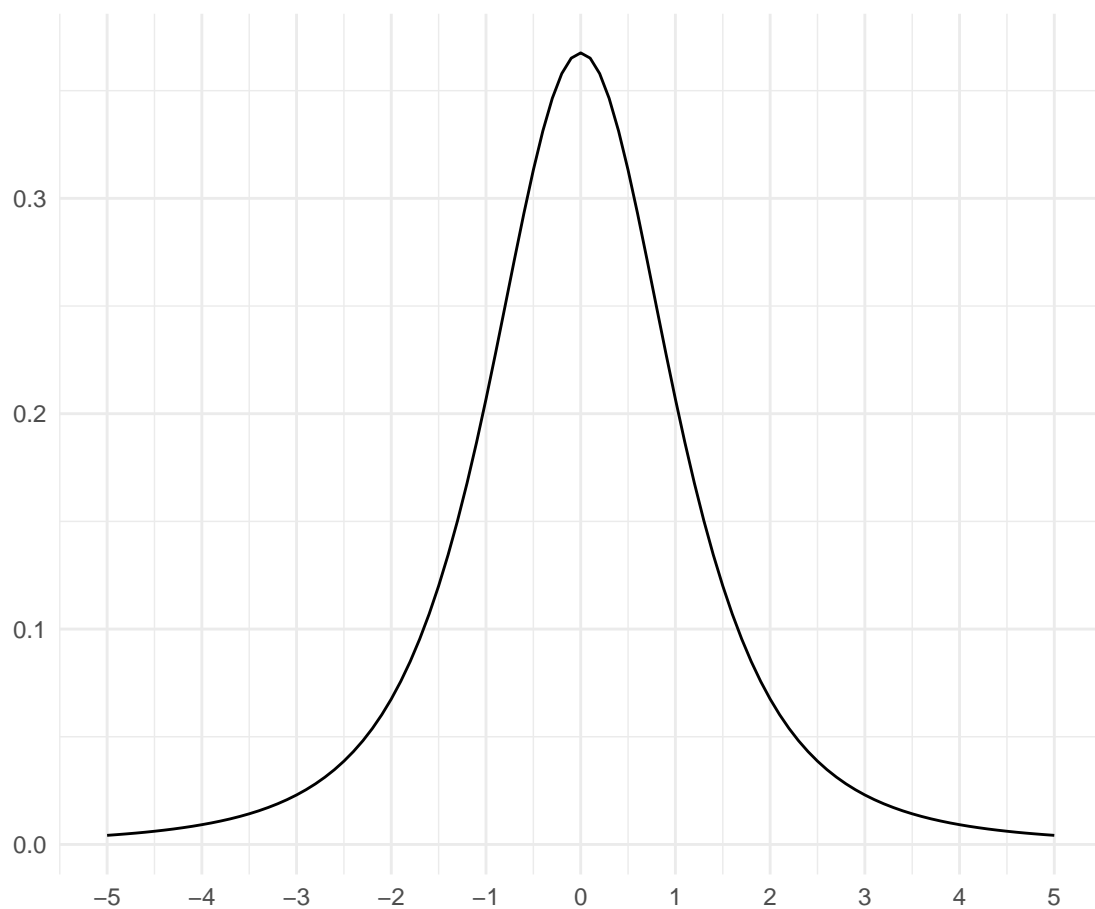
## 57 Aufgabe

(9 Punkte)



Im Folgenden ist die t-Verteilung unter der Annahme der Gültigkeit der Nullhypothese abgebildet.

1. Ergänzen Sie die Beschriftung der x-Achse! **(1 Punkt)**
2. Ergänzen Sie „ $\bar{y}_1 = \bar{y}_2$ “! **(1 Punkt)**
3. Ergänzen Sie „ $A = 0.95$ “! **(1 Punkt)**
4. Zeichnen Sie  $T_{\alpha=5\%}$  in die Abbildung! **(1 Punkt)**
5. Zeichnen Sie das Signifikanzniveau  $\alpha$  in die Abbildung! **(2 Punkte)**
6. Zeichnen Sie  $-T_{calc}$  in die Abbildung! **(1 Punkt)**
7. Zeichnen Sie einen signifikant p-Wert in die Abbildung! **(2 Punkte)**





Sie rechnen einen t-Test für Gruppenvergleiche der Mittelwerte. Sie schätzen den Unterschied zwischen dem mittleren Befall mit Parasiten zu einer unbehandelten Kontrolle.

1. Beschriften Sie die untenstehende Abbildung mit der Signifikanzschwelle! Begründen Sie Ihre Antwort! **(2 Punkte)**
2. Ergänzen Sie eine *in den Kontext passende* Relevanzschwelle! Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie in die untenstehende Abbildung sechs einzelne Konfidenzintervalle (a-f) mit den jeweiligen Eigenschaften! **(6 Punkte)**
  - (a) Ein 95% Konfidenzintervall mit höherer Fallzahl  $n$  in der Stichprobe als der Rest der 95% Konfidenzintervalle
  - (b) Ein signifikantes, relevantes 90% Konfidenzintervall.
  - (c) Ein 95% Konfidenzintervall mit niedriger Fallzahl  $n$  in der Stichprobe als der Rest 95% der Konfidenzintervalle
  - (d) Ein signifikantes, nicht relevantes 95% Konfidenzintervall
  - (e) Ein signifikantes, relevantes 95% Konfidenzintervall
  - (f) Ein nicht signifikantes, nicht relevantes 95% Konfidenzintervall



## 59 Aufgabe

(10 Punkte)



Gegeben ist die Formel für den Student t-Test. Welche Auswirkung hat die Änderungen der jeweiligen statistischen Masszahl  $\Delta$ ,  $s$  und  $n$  auf die Teststatistik  $T_{calc}$ , den p-Wert  $Pr(D|H_0)$  sowie dem Konfidenzintervall  $KI_{1-\alpha}$ ?

1. Visualisieren Sie den Zusammenhang zwischen der Teststatistik  $T_{calc}$  und dem p-Wert  $Pr(D|H_0)$  für sich ändernde  $T_{calc}$ -Werte! **(2 Punkte)**
2. Füllen Sie die untenstehende Tabelle aus! **(6 Punkte)**

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

3. Visualisieren Sie ein 95%-iges Konfidenzintervall im Vergleich zu einem 99%-igen Konfidenzintervall! Begründen Sie Ihre Antwort mathematisch! **(2 Punkte)**

## Der t-Test

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- [Kapitel 22 - Der t-Test](#)

## 60 Aufgabe

(12 Punkte)

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



Nach einem Experiment mit zwei Pestiziden (*RoundUp* und *GoneEx*) ergibt sich die folgende Datentabelle mit dem gemessenen Trockengewicht (*drymatter*) von Weizen.

pesticide	drymatter
RoundUp	17
GoneEx	16
RoundUp	17
GoneEx	13
RoundUp	8
RoundUp	10
GoneEx	11
GoneEx	12
GoneEx	16
GoneEx	12
GoneEx	15
RoundUp	16

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik  $T_{calc}$  eines Student t-Tests für den Vergleich der beiden Pestizide! **(5 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 2.04$  und dem berechneten  $T_{calc}$  eine Aussage zur Nullhypothese! **(2 Punkte)**
5. Wenn Sie *einen* Unterschied zwischen den beiden Pestiziden erwarten würden, wie groß wäre dann die Teststatistik  $T_{calc}$ ? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 61 Aufgabe

(12 Punkte)

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



Nach einem Experiment mit zwei Futtermitteln (*FatDown* und *ProGain*) an Puten ergibt sich die folgende Datentabelle mit dem gemessenen Gewichtszunahmen nach fünf Wochen Mast.

feed	weight
FatDown	16
ProGain	16
ProGain	13
FatDown	17
FatDown	13
FatDown	16
ProGain	14
ProGain	14
ProGain	13
FatDown	14
ProGain	14
ProGain	15

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik  $T_{calc}$  eines Welch t-Tests für den Vergleich der beiden Futtermittel! **(4 Punkte)**
4. Treffen Sie mit  $T_{\alpha=10\%} = 1.36$  und dem berechneten  $T_{calc}$  eine Aussage zur Nullhypothese! **(1 Punkt)**
5. Berechnen Sie das 90% Konfidenzintervall unter der Verwendung von  $s_p$  und der gemittelten Fallzahl über die beiden Gruppen! **(3 Punkte)**
6. Nennen Sie den statistischen Grund, warum Sie sich zwischen einem Student t-Test und einem Welch t-Test entscheiden müssen! **(1 Punkt)**



## 62 Aufgabe

(11 Punkte)

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



Das Gewicht von Küken wurde *vor* der Behandlung mit STARTex und 1 Woche *nach* der Behandlung gemessen. Es ergibt sich die folgende Datentabelle.


animal_id	before	after
1	13	17
2	20	14
3	12	16
4	9	23
5	8	12
6	14	13
7	15	19
8	6	16
9	17	15

1. Formulieren Sie die Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik  $T_{calc}$  eines gepaarten t-Tests für den Vergleich der beiden Zeitpunkte! **(4 Punkte)**
4. Treffen Sie mit  $T_{\alpha=5\%} = 2.04$  und dem berechneten  $T_{calc}$  eine Aussage zur Nullhypothese! **(2 Punkte)**
5. Schätzen Sie den  $p$ -Wert aus Ihrem berechneten  $T_{calc}$  ab! Begründen Sie Ihre Antwort mit einer Skizze! **(2 Punkte)**

## 63 Aufgabe

(10 Punkte)



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

```
##  
## Two Sample t-test  
##  
## data:  freshmatter by N  
## t = -1.5541, df = 8, p-value = 0.1588  
## alternative hypothesis: true  is not equal to [condensed]  
## 95 percent confidence interval:  
##  -7.948353  1.548353  
## sample estimates:  
## mean in group ctrl  mean in group low  
##                16.8                20.0
```

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

## 64 Aufgabe

(8 Punkte)



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

```
##
## Two Sample t-test
##
## data: waterintake by infusion
## t = -0.45698, df = 14, p-value = 0.6547
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -3.614854 2.345013
## sample estimates:
## mean in group ctrl mean in group trt1
##          17.22222          17.85714
```

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

## 65 Aufgabe

(9 Punkte)



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

```
##  
## Two Sample t-test  
##  
## data: drymatter by Fe  
## t = 0.125, df = 16, p-value = 0.9021  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -1.773249 1.995471  
## sample estimates:  
## mean in group mid mean in group trt2  
## 17.11111 17.00000
```

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

## 66 Aufgabe

(8 Punkte)



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

```
##
## Paired t-test
##
## data:  drymatter by Fe
## t = 1.676, df = 4, p-value = 0.169
## alternative hypothesis: true  is not equal to [condensed]
## 95 percent confidence interval:
##  -3.282932 13.282932
## sample estimates:
## mean difference
##                5
```

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

## Die ANOVA

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

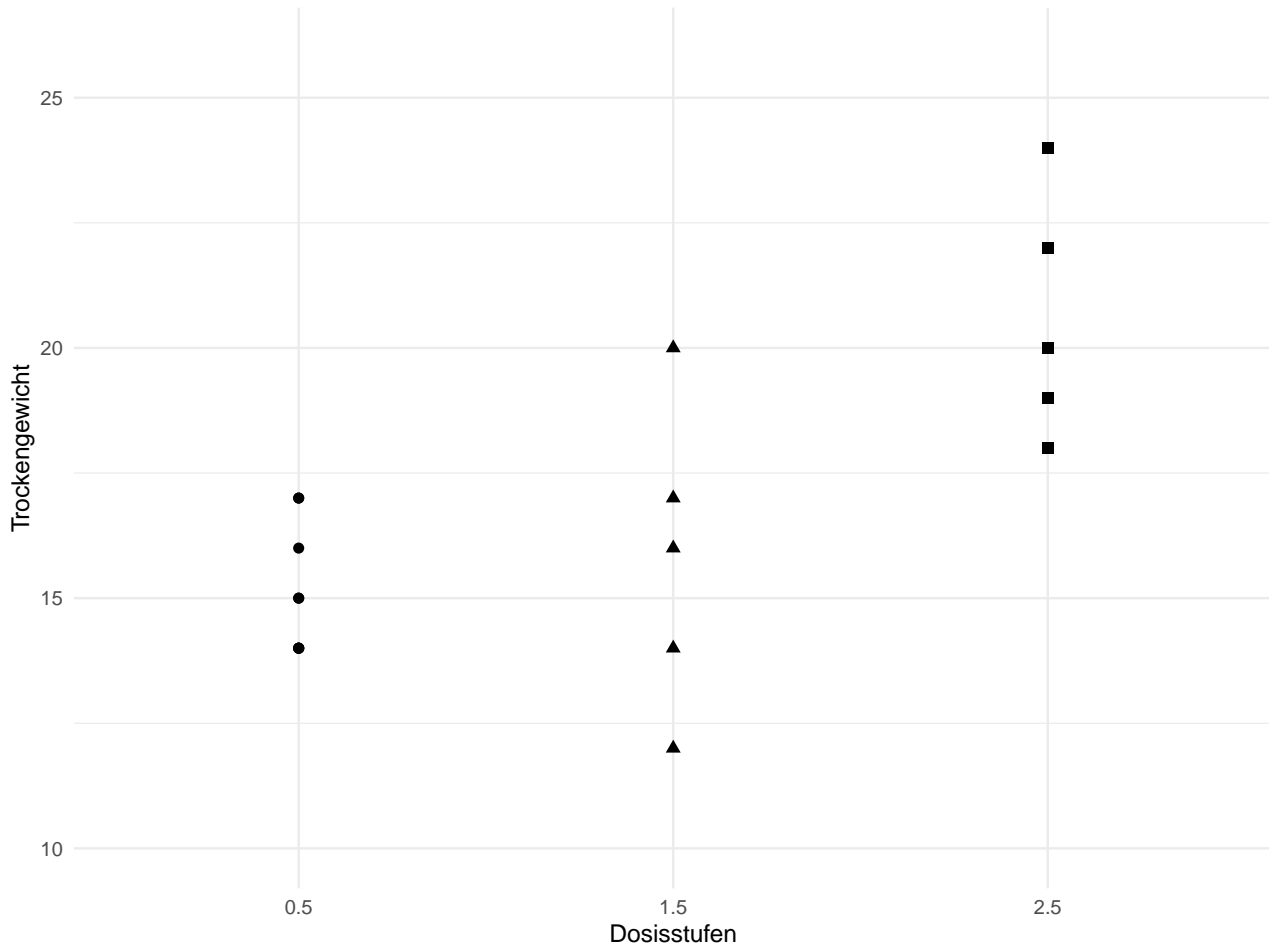
- [Kapitel 23 - Die ANOVA](#)

## 67 Aufgabe

(8 Punkte)



In einem Experiment wurde der Ertrag von Erbsen unter drei verschiedenen Pestizid-Dosen 0.5 g/l, 1.5 g/l und 2.5 g/l gemessen. Unten stehenden sehen Sie die Visualisierung des Datensatzes.



1. Zeichnen Sie folgende statistischen Maßzahlen in die Abbildung ein! Beschriften Sie die statistischen Maßzahlen! **(6 Punkte)**
  - Total (grand) mean:  $\beta_0$
  - Mittelwerte der Dosen:  $\bar{y}_{0.5}$ ,  $\bar{y}_{1.5}$  und  $\bar{y}_{2.5}$
  - Effekt der einzelnen Level der Dosen:  $\beta_{0.5}$ ,  $\beta_{1.5}$ , und  $\beta_{2.5}$
  - Residuen oder Fehler:  $\epsilon$
2. Liegt ein *vermutlicher* signifikanter Unterschied zwischen den Dosisstufen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 68 Aufgabe

(13 Punkte)

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



Der Datensatz `plant_growth_tbl` enthält das Gewicht der Kohlköpfe (*weight*), die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden – dem Faktor *group* mit den Faktorstufen *ctrl*, *trt1*, *trt2*.

1. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus mit den gegebenen Informationen von Df und Sum Sq! **(4 Punkte)**
2. Schätzen Sie den p-Wert der Tabelle mit der Information von  $F_{\alpha=5\%} = 3.49$  ab. Begründen Sie Ihre Antwort! **(2 Punkte)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	182.28			
Residuals	20	33.46			

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

group	n	mean	sd
ctrl	9	15.11	1.17
trt1	7	19.86	1.21
trt2	7	21.57	1.51


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



## 69 Aufgabe

(9 Punkte)



Der Datensatz *crop\_tbl* enthält das Outcome *height* für ein Experiment mit Maispflanzen, welches unter drei verschiedenen Düngerbedingungen erzielt wurden. Die Düngerbedingungen sind in dem Faktor *trt* mit den Faktorstufen *A*, *trt1* und *trt2* codiert. Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: height
##           Df Sum Sq Mean Sq F value Pr(>F)
## trt        2 14.399   7.1994   2.2125 0.1355
## Residuals 20 65.079   3.2540
```

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

## 70 Aufgabe

(12 Punkte)

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



Der Datensatz *tooth\_tbl* enthält Daten aus einer Studie zur Bewertung der Wirkung von Vitamin C auf das Zahnwachstum bei Meerschweinchen. Der Versuch wurde an verschiedenen Schweinen durchgeführt, wobei jedes Tier eine von 3 Vitamin-C-Dosen *dose* über eine von 1 Verabreichungsmethoden *supp* erhielt. Die Zahnlänge wurde als normalverteiltes Outcome gemessen.

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


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>supp</b>	0	255.37			
<b>dose</b>	2	85.94			
<b>supp:dose</b>	0	NA			
<b>Residuals</b>	NA	NA			

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


## 71 Aufgabe

(10 Punkte)



Der Datensatz *pig\_gain\_weight\_tbl* enthält Daten aus einer Studie zur Bewertung der Wirkung vom Vitamin Selen auf das Wachstum bei Mastschweinen. Der Versuch wurde an 15 Mastschweinen durchgeführt, wobei jedes Tier eine von drei Selen-Dosen *dose* (0.5 ng/Tag, 1 ng/Tag und 10 ng/Tag) über eine von zwei Verabreichungsmethoden *form* erhielt (Wasser oder Festnahrung). Sie erhalten folgende Ausgabe in .

```
## Analysis of Variance Table
##
## Response: fat_perc
##      Df Sum Sq Mean Sq F value    Pr(>F)
## dose      2 239.48  119.739   6.0681 0.0096748
## form      1   37.52   37.519   1.9014 0.1848207
## dose:form  2 424.17  212.086  10.7481 0.0008481
## Residuals 18 355.18   19.732
```

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

## 72 Aufgabe

(8 Punkte)



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

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

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

## 73 Aufgabe

(8 Punkte)



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

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

## 74 Aufgabe

(9 Punkte)



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

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

## Multiple Gruppenvergleiche

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- [Kapitel 20.3 Adjustierung für multiple Vergleiche](#)
- [Kapitel 31.7 Compact letter display](#)

## 75 Aufgabe

(8 Punkte)

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



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

Vergleich	Raw p-value	Adjusted p-value	Reject Null
dose 30 - ctrl	0.340		
dose 15 - ctrl	0.020		
dose 80 - ctrl	0.002		
dose 90 - ctrl	0.070		

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




## 76 Aufgabe

(9 Punkte)

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



In einem Experiment für den Proteingehalt von Wasserlinsen in g/l mit vier Dosisstufen (ctrl, low, mid und high) erhalten Sie folgende Matrix als  Ausgabe mit den rohen, unadjustierten  $p$ -Werten.

```
##          ctrl          high          low          mid
## ctrl 1.0000000 0.0363436 0.6985898 0.0004566
## high 0.0363436 1.0000000 0.0148344 0.0949929
## low  0.6985898 0.0148344 1.0000000 0.0001514
## mid  0.0004566 0.0949929 0.0001514 1.0000000
```

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

trt	n	mean	sd
ctrl	9	11.73	0.77
high	9	8.82	2.09
low	9	12.25	3.18
mid	9	6.52	4.11


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

## 77 Aufgabe

(8 Punkte)

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



In einem Experiment für den Proteingehalt von Wasserlinsen in g/l mit fünf Dosisstufen (A, B, C, D und E) erhalten Sie folgendes *Compact letter display (CLD)* als  Ausgabe aus den rohen, unadjustierten  $p$ -Werten.

```
##   A   B   C   D   E
## "a" "b" "b" "b" "b"
```

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

## Der $\chi^2$ -Test & Der diagnostische Test

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- [Kapitel 28 - Der  \$\chi^2\$ -Test](#)
- [Kapitel 29 - Der diagnostische Test](#)

## 78 Aufgabe

(11 Punkte)

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



Nach einem Feldexperiment ergibt sich die folgende 2x2 Datentabelle mit einem Pestizid (ja/nein) der Marke KillAll, dargestellt in den Zeilen, und dem infizierten Pflanzenstatus (ja/nein) von Kohl, dargestellt in den Spalten. Insgesamt wurden  $n = 126$  Pflanzen untersucht.

	Erkrankt (ja)	Erkrankt (nein)	
Pestizid (ja)	38	21	
Pestizid (nein)	23	44	

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

## 79 Aufgabe

(7 Punkte)

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



Gegeben sind folgende Randsummen in einer 2x2 Kreuztabelle aus einem Experiment mit  $n = 136$  Sauen. In dem Experiment wurde gemessen, ob eine Sau nach einer Behandlung mit einem Medikament (ja/nein) mehr als 30 Ferkel pro Jahr bekommen konnte (ja/nein).

	>30 Ferkel (ja)	≤30 Ferkel (nein)	
Medikament (ja)			71
Medikament (nein)			65
	77	59	136

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

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

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

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


## 80 Aufgabe

(10 Punkte)



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

```
##           Infiziert
## Gruppe      yes no
## Papageien-Tulpe   4 12
## Zwerg-Tulpe      10  3
```

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

```
##
## Fisher's Exact Test for Count Data
##
## data:  mat
## p-value = 0.009221
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.01244701 0.70309883
## sample estimates:
## odds ratio
##  0.1102808
```

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

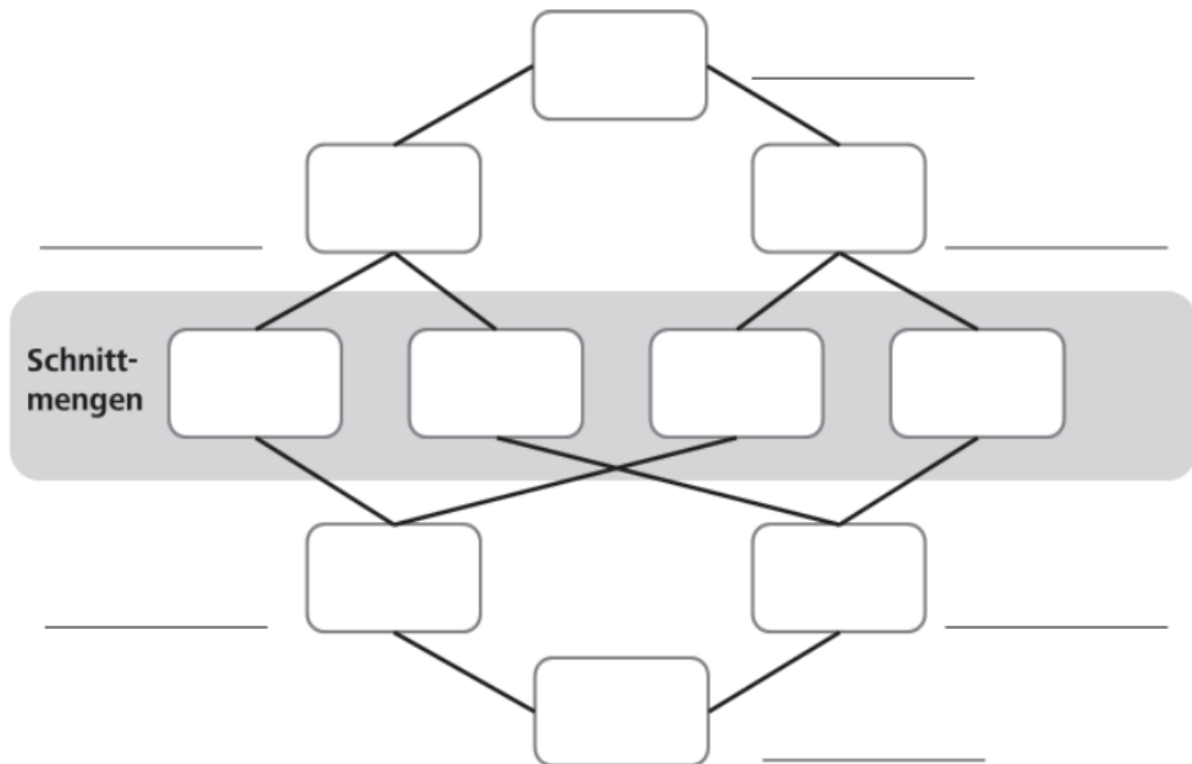
## 81 Aufgabe

**(11 Punkte)**



Die Prävalenz von Klauenseuche bei Wollschweinen wird mit 4% angenommen. In 75% der Fälle ist ein Test positiv, wenn das Wollschwein erkrankt ist. In 8% der Fälle ist ein Test positiv, wenn das Wollschwein *nicht* erkrankt ist und somit gesund ist. Sie werten 4000 Wollschweine mit einem diagnostischen Test auf Klauenseuche aus.

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



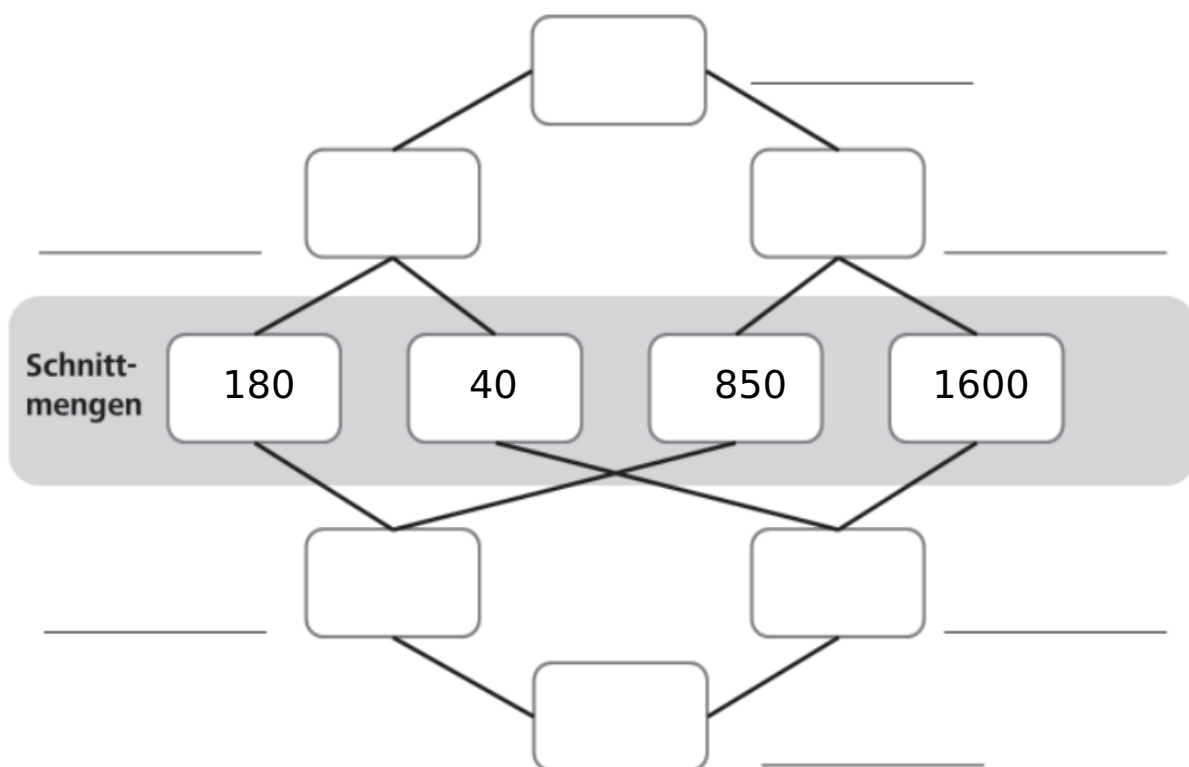
## 82 Aufgabe

(12 Punkte)



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

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





### 83 Aufgabe

(10 Punkte)



Beim diagnostischen Testen erhalten Sie *True Positives (TP)*, *True Negatives (TN)*, *False Positives (FP)* und *False Negatives (FN)*. Erklären Sie den Zusammenhang wie folgt.

1. Tragen Sie *TP*, *TN*, *FP* und *FN* in eine 2x2 Kreuztabelle ein. Beschriften Sie die Tabelle entsprechend! **(2 Punkte)**
2. Visualisieren Sie *TP*, *TN*, *FP* und *FN* in einer Abbildung. Beschriften Sie die Abbildung und die Achsen entsprechend! **(4 Punkte)**
3. Erklären Sie an einem numerischen Beispiel und der Abbildung die Berechnung der Prävalenz! **(2 Punkte)**
4. Erklären Sie an einem Schaubild den Unterschied zwischen Inzidenz und Prävalenz! **(2 Punkte)**

## Nicht parametrische Tests

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- [Kapitel 25 - Der Wilcoxon-Mann-Whitney-Test](#)
- [Kapitel 26 - Der Kruskal-Wallis-Test](#)

## 84 Aufgabe

(12 Punkte)

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



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

Vorher	Nachher	Differenz	Vorzeichen	Rang	Positiv Rang	Negativ Rang
11	11					
10	11					
6	16					
10	11					
11	11					
12	13					
10	11					
9	12					
8	12					
8	11					
10	13					
15	11					
8	13					
7	13					
10	10					

1. Ergänzen Sie die obige Tabelle mit den notwendigen Informationen, die Sie benötigen um einen Wilcoxon-Vorzeichen-Rang-Test zu rechnen! **(4 Punkte)**
2. Bestimmen Sie die Teststatistik  $W$  mit  $W = \min(T_-; T_+)$  und berechnen Sie den erwarteten Wert  $\mu_W = \frac{n_{10} \cdot (n_{10} + 1)}{4}$ ! **(2 Punkte)**
3. Berechnen Sie anschließend den z-Wert mit  $z = \frac{W - \mu_W}{12.748}$ ! **(2 Punkte)**
4. Liegt mit einer Signifikanzschwelle von  $z_{\alpha=5\%} = 1.96$  ein Unterschied zwischen den beiden Zeitpunkten vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie die Effektstärke mit  $r = |\frac{z}{\sqrt{n}}|$  und interpretieren Sie die Effektstärke! **(2 Punkte)**

## 85 Aufgabe

(8 Punkte)

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



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

Treatment	Count
Kontrolle	4.5
Kontrolle	9.2
RootsGoneX	10.4
RootsGoneX	13.5
Kontrolle	5.4
RootsGoneX	11.5
Kontrolle	2.6
RootsGoneX	10.2
Kontrolle	6.6
RootsGoneX	13.9
Kontrolle	5.8
RootsGoneX	12.6

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

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

2. Geben Sie eine Aussage über die Signifikanz von  $U_c$  durch  $z = \frac{U_c - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$  und dem kritischen Wert von  $z_{\alpha=5\%} = 1.96$ . Begründen Sie Ihre Antwort! (2 Punkte)

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

## 86 Aufgabe

(10 Punkte)

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



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

Treatment	Count	Rang Kontrolle	Rang Dosis 20	Rang Dosis 40
Kontrolle	8.5			
Dosis 40	10.6			
Dosis 20	13.6			
Dosis 20	11.3			
Dosis 20	11.3			
Kontrolle	10.5			
Dosis 20	10.5			
Kontrolle	12.4			
Dosis 20	13.0			
Kontrolle	8.4			
Dosis 40	11.5			
Kontrolle	13.3			
Dosis 40	9.4			
Dosis 40	9.6			
Dosis 40	7.6			
Dosis 40	6.9			
Kontrolle	7.5			

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

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

## Simple lineare Regression & Korrelation

Mehr Informationen zu den Aufgaben in den folgenden Kapiteln aus dem Skript Bio Data Science.

- [Kapitel 32 - Simple lineare Regression](#)
- [Kapitel 33 - Maßzahlen der Modelgüte](#)
- [Kapitel 34 - Korrelation](#)

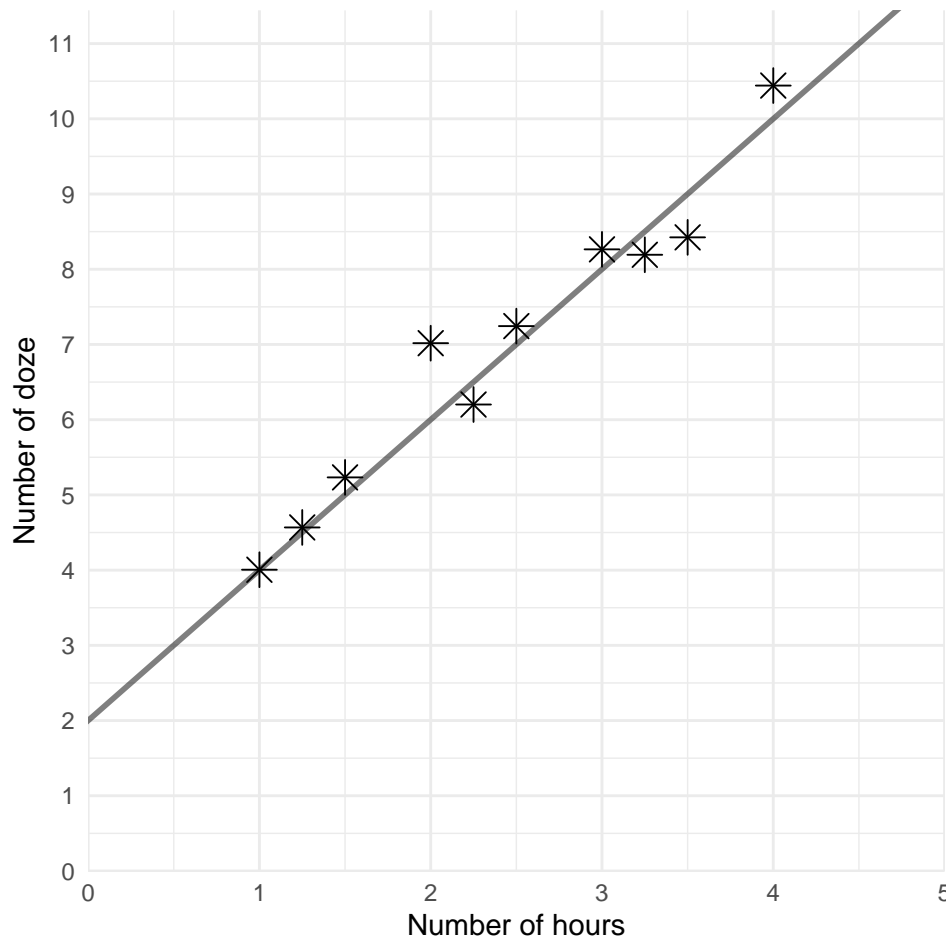
## 87 Aufgabe

(7 Punkte)

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



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




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

## 88 Aufgabe

(10 Punkte)

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



In einem Stallexperiment mit  $n = 120$  Ferkeln wurde der Gewichtszuwachs in kg unter ansteigender Lichteinstrahlung in nm gemessen. Sie erhalten den  Output einer simplen Gaussian linearen Regression sieben Wochen nach der ersten Messung.

term	estimate	std.error	t statistic	p-value
(Intercept)	4.05	1.07		
light	-0.11	0.10		


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




## 89 Aufgabe

(10 Punkte)



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

```
##  
## Call:  
## lm(formula = weight ~ water, data = data_tbl)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -3.889 -1.111  0.000   1.111  3.111   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  15.8889     0.6443   24.661 3.7e-14      
## waterB       -2.7778     0.9112   -3.049 0.00766      
##  
## Residual standard error: 1.933 on 16 degrees of freedom  
## Multiple R-squared:  0.3674, Adjusted R-squared:  0.3279   
## F-statistic: 9.294 on 1 and 16 DF,  p-value: 0.007662
```

1. Ist die Annahme der Normalverteilung an das Outcome *water* erfüllt? Begründen Sie die Antwort! **(2 Punkte)**
2. Wie groß ist der Effekt der Wassergabe? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Liegt ein signifikanter Effekt vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Erklären Sie *kurz* den Begriff R-squared! Was sagt Ihnen der Wert 0.3674 aus? **(2 Punkte)**
5. Schreiben Sie das Ergebnis der  Ausgabe in zwei Sätzen auf, der die Information zum Effekt und der Signifikanz enthält! **(2 Punkte)**

## 90 Aufgabe

(8 Punkte)



1. Skizzieren Sie in die unten stehenden, freien Abbildungen ein kausales und ein prädiktives Modell mit  $n = 9$  Beobachtungen! **(4 Punkte)**
2. Beachten Sie bei der Erstellung der Skizze, ob ein Effekt von X vorliegt oder nicht! **(2 Punkte)**
3. Beschriften Sie die Abbildung mit „Trainingsdaten“ und „Testdaten“! **(2 Punkte)**

Causal model with no effect of X



Predictive model with no effect of X



## 91 Aufgabe

(9 Punkte)

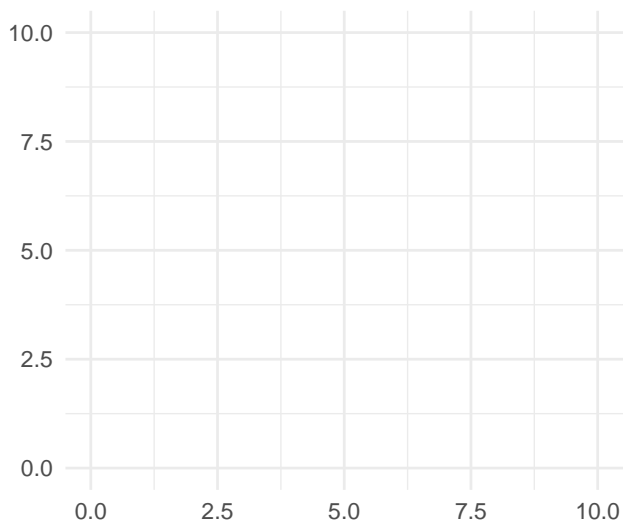


Im folgenden sehen Sie drei leere Scatterplots. Füllen Sie diese Scatterplots nach folgenden Anweisungen.

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

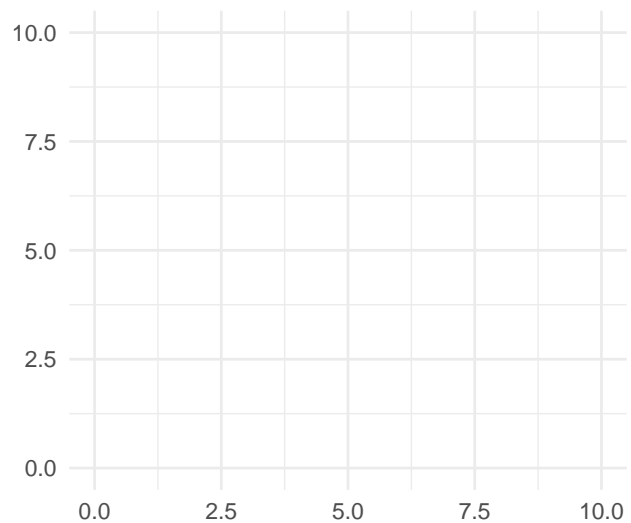
Pearsons  $\rho = 0.75$

$R^2 = 0.5$



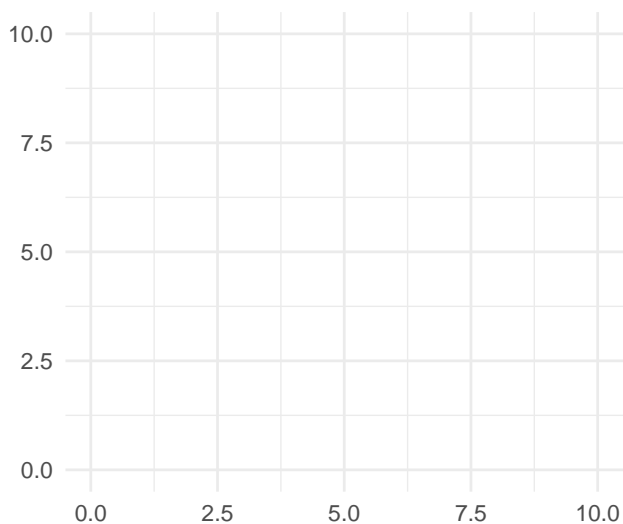
Pearsons  $\rho = 0.25$

$R^2 = 0.75$



Pearsons  $\rho = -0.75$

$R^2 = 1$



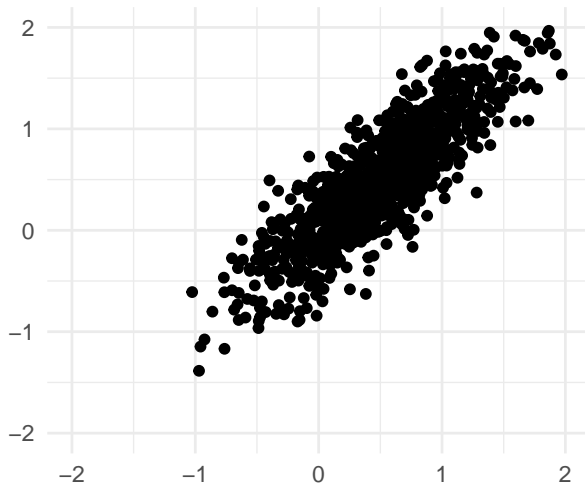


Im folgenden sehen Sie vier Scatterplots. Ergänzen Sie die Überschriften der jeweiligen Scatterplots.

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

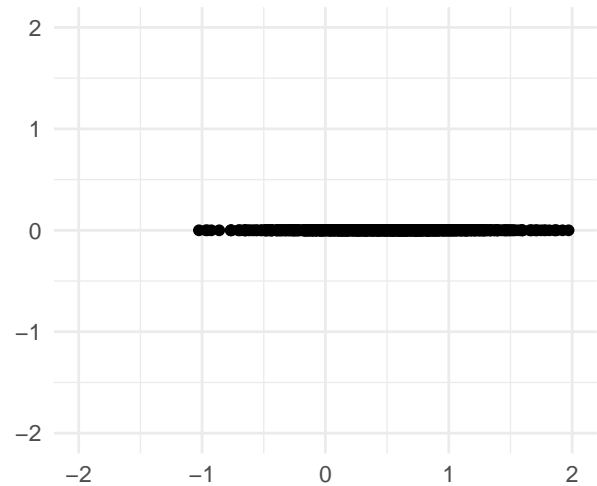
Pearsons  $\rho =$

$R^2 =$



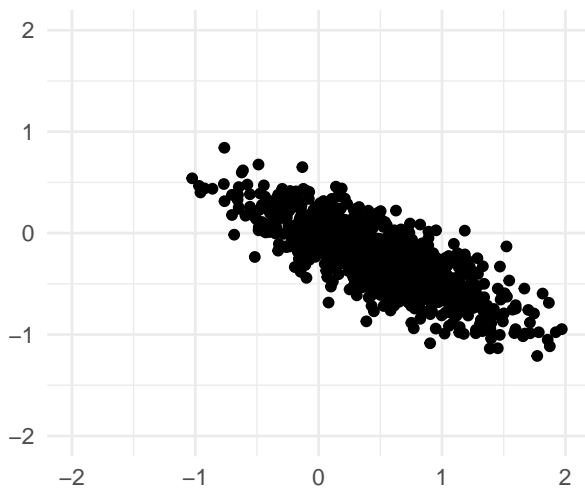
Pearsons  $\rho =$

$R^2 =$



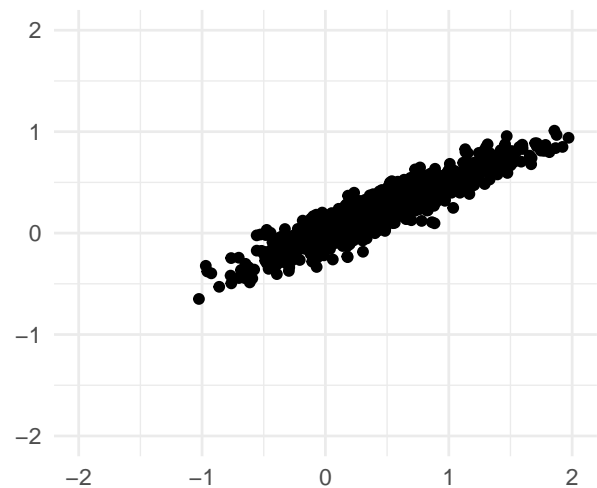
Pearsons  $\rho =$

$R^2 =$



Pearsons  $\rho =$

$R^2 =$



## 93 Aufgabe

(12 Punkte)



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

```
##  
## Spearman's correlation  
##  
## data: drymatter and water  
## t = 5.1014, df = 8, p-value = 0.0009283  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
##  0.5451103 0.9700411  
## sample estimates:  
##      cor  
## 0.8745703
```

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

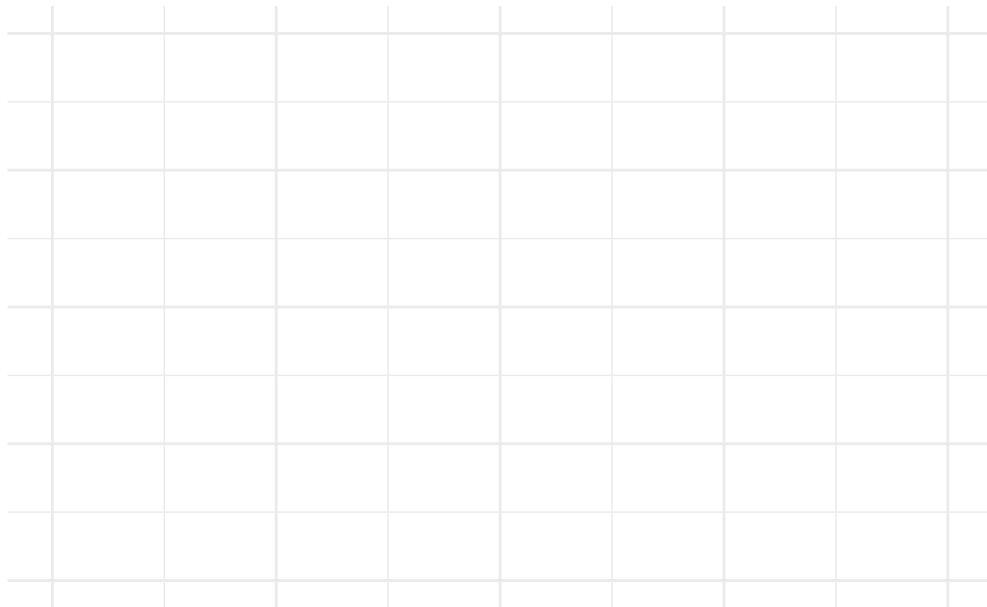
## 94 Aufgabe

(6 Punkte)



1. Skizzieren Sie in die unten stehenden, freien Abbildungen die Abbildung, die sich nach der Überschrift ergibt! **(4 Punkte)**
2. Beschriften Sie die Achsen entsprechend! **(2 Punkte)**

Residual plot with 1 outlier fulfilling the normality assumption.



Residual plot violating the normality assumption.



## 95 Aufgabe

(7 Punkte)



1. Skizzieren Sie in die unten stehenden, freien Abbildungen die Abbildung, die sich nach der Überschrift ergibt! **(4 Punkte)**
2. Beschriften Sie die Achsen entsprechend! **(3 Punkte)**

QQ plot fullfiling the normality assumption.



QQ plot violating the normality assumption.



**96 Aufgabe****(10 Punkte)**

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

.id	drymatter	water	.fitted	.resid
1	19.9	6.1	20.8	
2	18.1	4.3	18.7	
3	32.5	14.1	30.4	
4	24.2	11.6	27.4	
5	28.2	12.5	28.4	
6	29.8	12.0	27.8	
7	26.4	8.0	23.1	
8	28.7	14.2	30.5	
9	24.9	10.0	25.4	

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



## 97 Aufgabe

(12 Punkte)





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

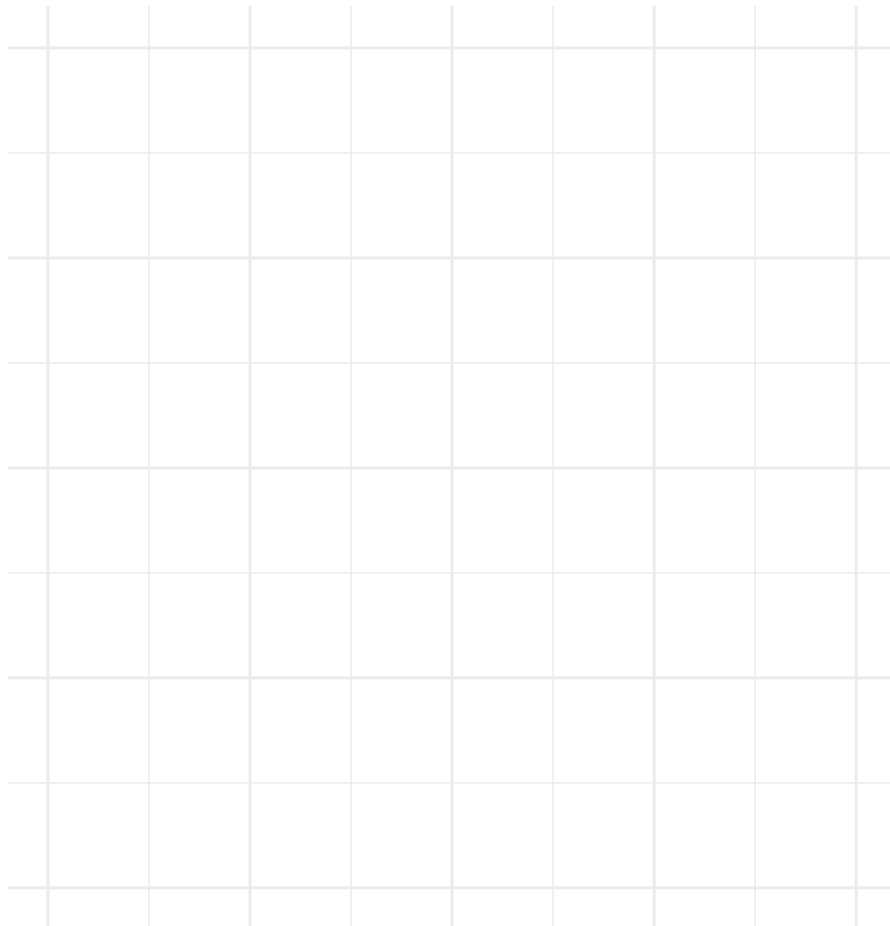

## 98 Aufgabe

(9 Punkte)



Ein Feldexperiment wurde mit  $n = 200$  Pflanzen durchgeführt. Folgende Einflussvariablen ( $x$ ) wurden erhoben: N, center und height. Als mögliche Outcomevariablen stehen Ihnen nun folgende gemessene Endpunkte zu Verfügung: drymatter, yield, count, quality\_score und dead.

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



## Mathematik

Mehr Informationen zu den Aufgaben in den [Skript Mathematik](#) und den entsprechenden Kapiteln.

## 99 Aufgabe

(12 Punkte)

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



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

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

Gehen Sie von einem radialen Wachstum der 1000-jährigen Eiche von  $1\text{mm}$  pro Jahr aus. Es ist bekannt, dass die Eiche im Jahr 2022 einen Umfang von  $14\text{m}$  in Brusthöhe hatte.

1. Wie groß war der Durchmesser der Eiche im Jahr 1810 als Herodot in der Eiche versteckt werden sollte? **(3 Punkte)**
2. Skizzieren Sie in einer Abbildung einen linearen Zusammenhang und einen exponentiellen Zusammenhang für das Wachstum der 1000-jährigen Eiche. Erklären Sie die Auswirkungen der Entscheidung für linear oder exponentiell auf Ihre Berechnungen! **(2 Punkte)**

Herodot hatte eine Schulterhöhe von  $195\text{cm}$ , eine Breite von  $95\text{cm}$  sowie eine Länge von  $250\text{cm}$ .

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

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

4. Berechnen Sie die Dicke der Eichenwand in  $\text{cm}$ ! Verdeutlichen Sie Ihre Berechnungen an einer aussagekräftigen Skizze für Pferd und Eiche! **(3 Punkte)**
5. Unter einer Dicke der Eichenwand von  $15\text{cm}$  bricht die Eiche zusammen. Beantworten Sie die Forschungsfrage! Begründen Sie Ihre Antwort! **(2 Punkte)**

## 100 Aufgabe

(10 Punkte)

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



**Von Töpfen auf Tischen** In einem Experiment wollen Sie die Wuchshöhe von 200 Sonnenblumen bestimmen. Bevor Sie überhaupt mit dem Experiment beginnen können, gibt es aber ein paar Abschätzungen über die Kosten und den Aufwand zu treffen. Zum einen müssen Sie die Sonnenblumen einpflanzen und müssen dafür Substrat bestellen. Zum anderen müssen Sie die Sonnenblumen auch bewegen und in ein Gewächshaus platzieren. Die Töpfe für die Keimung haben einen Durchmesser von 8.5cm und eine Höhe von 10cm. Der Kubikmeterpreis für Torf liegt bei 270 EUR.

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

## 101 Aufgabe

(10 Punkte)

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



**Entschuldigung, ist das Ihre Feder in meinem Auge?** So hört man häufiger höfliche Gänse in Mastställen sagen. Das ist natürlich etwas ungünstig, den dann kommt es zu Picken und Kannibalismus. Denn wenn der Nachbar nervt, dann muss zu Maßnahmen gegriffen werden. Kennt jeder aus einer mittelmäßigen Wohngemeinschaft. Das wollen wir aber als vorsorgliche Gänse-Halter:innen aber nicht<sup>1</sup>. Betrachten wir also einmal das Platzangebot (eng. *space allowance*, abk. *SA*) der Gänse für vier Tätigkeiten und versuchen die notwendige Fläche zu optimieren. Wie immer gibt es dafür 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<sub>i</sub>* dem benötigten Platz für ein Verhalten *i*.
- *PB<sub>i</sub>* dem Anteil des Auftretens eines Verhaltens *i*.
- *r<sub>i</sub>* dem benötigten Radius für das Verhalten *i*.
- *R<sub>i</sub>* dem notwendigen Abstand zu den Nachbarn für das Verhalten *i*.
- *i* dem Verhalten: (1) walking, (2) wingflapping, (3) sitting und (4) foraging incl. scratching.

In der folgenden Tabelle 1 sind die Werte für *r<sub>i</sub>*, *R<sub>i</sub>* und *PB<sub>i</sub>* für ein spezifisches Verhalten *i* aus drei wissenschaftlichen Veröffentlichungen dargestellt.

	Aldridge et al. (2021)	Baxter et al. (2022)	Jacobs et al. (2019)
walking	34cm; 21cm; 52.2%	35cm; 21cm; 25.4%	34cm; 22cm; 25.4%
wingflapping	30cm; 31cm; 1.8%	35cm; 24cm; 5.2%	34cm; 25cm; 1.8%
sitting	42cm; 26cm; 0.6%	36cm; 25cm; 0.8%	28cm; 25cm; 0.2%
foraging incl. scratching	40cm; 23cm; 1%	33cm; 16cm; 1%	44cm; 25cm; 0.1%

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

<sup>1</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](#)

## 102 Aufgabe

(10 Punkte)

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



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

»Ohne ausreichende Zufuhr von Vitamin C stellen sich nach 50 Tagen die ersten Symptome ein; die ersten Toten sind nach 72 Tagen zu beklagen; nach 100 Tagen rafft die Skorbut eine ganze Schiffsbesatzung dahin«

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

1. Stellen Sie den Verlauf der Anzahl an Matrosen auf einem Schiff der Flotte in der Form einer Überlebenszeitkurve dar! Beschriften Sie die Achsen entsprechend! **(3 Punkte)**
2. Schätzen Sie die Überlebenswahrscheinlichkeit nach 100 Tagen aus Ihrer Abbildung ab! **(2 Punkte)**

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  $5000\mu\text{g}/100\text{mg}$  Vitamin C. Der Bedarf liegt bei  $105\text{mg}$  pro Tag für Männer.

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

## 103 Aufgabe

(12 Punkte)

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



**Event Horizon – Am Rande des Universums** Die Sonne hat eine aktuelle, angenommene Masse von  $2 \times 10^{27} \text{ kg}$ . Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 5000m kollabiert, wird die Sonne 45% der aktuellen Masse verloren haben. Ein Lichtteilchen mit der Masse  $m_f$  und der Fluchtgeschwindigkeit  $v_f$  will dem schwarzen Loch entkommen. Sie haben folgende Formeln für die kinetische Energie des Lichtteilchens  $E_{kin}$  und der Gravitationsenergie des schwarzen Lochs  $E_{grav}$  gegeben.

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

mit

- $m_f$ , gleich der Masse [kg] des fliehenden Objektes
- $m_s$ , gleich der Masse [kg] des stationären Objekts
- $r_s$ , gleich dem Radius [m] des stationären Objekts
- $G$ , gleich der Gravitationskonstante mit  $6.274 \cdot 10^{-11} \text{ m}^3 (\text{kg} \cdot \text{s}^2)^{-1}$

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

1. Geben Sie die Formel für die Fluchtgeschwindigkeit  $v_f$  an! **(2 Punkte)**
2. Überprüfen Sie Ihre umgestellte Formel nach  $v_f$  anhand der Einheiten! **(2 Punkte)**
3. Berechnen Sie die notwendige Fluchtgeschwindigkeit  $v_f$  des Lichtteilchens mit den angegebenen Informationen! **(2 Punkte)**
4. Gehen Sie von einer Lichtgeschwindigkeit von  $2.7 \times 10^8 \text{ m/s}$  aus. Kann das Lichtteilchen der Gravitation des schwarzen Lochs entkommen? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Stellen Sie den Zusammenhang zwischen dem sich verringernden Radius  $r$  des schwarzen Lochs bei gleichbleibender Masse  $m_s$  und der notwendigen Fluchtgeschwindigkeit  $v_f$  in einer Abbildung dar! **(2 Punkte)**
6. Erklären Sie in diesem Zusammenhang den Begriff *Ereignishorizont*! **(2 Punkte)**



## 104 Aufgabe

(10 Punkte)

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



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

Wir treffen folgende Annahmen. Eine außerirdische Zivilisation schickt vier Voyager 1 ähnliche Sonden mit der Geschwindigkeit von  $5.9256 \times 10^4 \text{ km/h}$  los um sich auf den erreichten Planeten selbst zu replizieren. Nach 750 Jahren ist die Replikation abgeschlossen und wiederum vier Sonden werden ausgesendet. Gehen Sie von 5.16 Lichtjahren als mittlerer Abstand der Sterne in der Milchstraße aus. Es gibt  $2 \times 10^{11}$  Sterne in der Milchstraße. Nehmen Sie eine Lichtgeschwindigkeit von  $2.7 \times 10^8 \text{ m/s}$  an.

1. Skizzieren Sie in einer Abbildung die ersten vier 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! **(2 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! **(1 Punkt)**
3. Berechnen Sie die Dauer, die eine außerirdische Zivilisation annäherungsweise benötigt um die gesamten Sterne der Milchstraße mit Sonden zu besuchen! **(3 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)**
5. Skizzieren Sie in einer Abbildung den Zusammenhang zwischen Zeit  $t$  und Raum  $r$ . Ergänzen Sie den Geschwindigkeitsvektor  $\vec{v}_t$  und  $\vec{v}_r$  einer ruhenden Sonde, einer mit 50% Lichtgeschwindigkeit und mit 99% Lichtgeschwindigkeit fliegender Sonde! **(1 Punkt)**
6. Warum ist die Lichtgeschwindigkeit die maximale mögliche Geschwindigkeit? Begründen Sie Ihre Antwort anhand der Abbildung! **(1 Punkt)**

## 105 Aufgabe

(10 Punkte)

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



**Solar- & Biogasanlagen** Um die Energiekosten Ihres Betriebes zu senken, wollen Sie eine Solaranlage auf den Rinderstall montieren lassen. Sie messen Ihren Stall und finden folgende Maße wieder. Die vordere Seite des Rinderstall hat eine Höhe  $h_v$  von  $5m$ . Die hintere Seite des Rinderstall hat eine Höhe  $h_b$  von  $11m$ . Der Rinderstall hat eine Tiefe  $t$  von  $12m$  und eine Breite  $b$  von  $60m$ .

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 planen Sie eine neue Biogasanlage für Ihren Betrieb. Der neue Methantank hat einen Radius  $r$  von  $1m$ . Leider gibt es ein paar bauliche Beschränkungen auf dem Grundstück. Ihr Fundament des zylindrischen Methantanks kann nur ein Gewicht von maximal  $12t$  aushalten bevor der Tank wegbricht. Sie rechnen eine Sicherheitstoleranz von 25% ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei  $-80^\circ C$  eine Dichte von  $235kg/m^3$ . Bei  $-100^\circ C$  hat Methan eine Dichte von  $300kg/m^3$ . Sie betreiben Ihre Anlage bei  $-85^\circ C$ .

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

## 106 Aufgabe

(10 Punkte)

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



**Pyramiden bauen** Es stehen die oldenburgischen Pyramidentage an und Sie sind auf abenteuerlichen Wegen für den Bau der Pyramiden zuständig. Zu allem Überfluss handelt es sich auch noch eine *Reenactment* Veranstaltung. Thema der diesjährigen Pyramidentage sind die Pyramiden von Meroe, die den Königen und Königinnen des historischen Reiches von Kusch in Nubien, dem heutigen Sudan, als Grabstätten dienten. Die Pyramiden in Meroe fallen durch ihren steilen Winkel von 73 Grad im Vergleich zu den ägyptischen Pyramiden mit 60 Grad auf. Die durchschnittliche Seitenlänge der Grundfläche einer Pyramide beträgt 38 Königsellen. Eine Königselle misst 52.4cm.

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

1. Bei der Königspyramide von Meroe soll eine Seitenlänge der Grundfläche 38 Königsellen lang sein. Welche Höhe der Königspyramide in m ergibt sich?? **(1 Punkt)**
2. Die Außenflächen der Pyramide soll begrünt werden. Für die Bepflanzung muss eine 4cm dicke Torfschicht auf die Pyramide aufgebracht werden. Berechnen Sie die ungefähre Menge an benötigten Torf in  $m^3$ ! **(2 Punkte)**

Wie in jedem guten *Reenactment* gibt es viel Oberschicht, aber nur 4 Sklaven, die Ihnen bei dem Befüllen der Pyramide mit Schutt zu Seite stehen. Leider haben Ihre Sklaven zu allem Überfluss auch noch chronische Rückenschmerzen entwickelt, als sie von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 85% aus. In eine Schubkarre passen 100 Liter.

3. Wie oft müssen Ihre maladen Sklaven die Rampe mit der Schubkarre zur Spitze der Pyramide hochfahren um die Pyramide mit Schutt zu füllen? **(1 Punkt)**
4. Berechnen Sie die Länge der Rampe zur Spitze der Pyramide mit einem Anstellwinkel von  $14^\circ$ ! **(2 Punkte)**
5. Wie weit reicht Ihre Rampe vom Fuß der Pyramide in die oldenburgische Landschaft? **(2 Punkte)**

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

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

## 107 Aufgabe

(12 Punkte)

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



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

Ort	Cache	Wertung (S T G)
A	GCNPCOA	3.0   2.0   Mikro
B	GCHPDQ8	4.5   3.5   Klein
C	GCRDD3R	1.0   4.5   Mikro
D	GCFOCSG	5.0   4.0   Klein
E	GCALDCK	1.5   5.0   Klein

Im Weiteren sind Ihnen folgende Informationen zu den Entfernungen der Orte zugänglich. Der Entfernungsvektor  $\overrightarrow{AB}$  ist 4km. Im Weiteren ist Ihnen der Entfernungsvektor  $\overrightarrow{CB}$  mit 6.5km bekannt. Der Entfernungsvektor  $\overrightarrow{BE}$  ist das 2.1-fache des Entfernungsvektor  $\overrightarrow{CB}$ . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr  $25^\circ$  südlich. Wenn Sie von dem Ort C den Ort B anpeilen, so liegt der Ort D ungefähr  $45^\circ$  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 E Ihre Cachertour.

1. Lösen Sie diese Aufgabe mit Hilfe einer aussagekräftigen Skizze der Orte und Caches. Bezeichnen Sie die Strecken und die Winkel Ihrer Skizze entsprechend! **(2 Punkte)**
2. Welche Strecke in km legen Sie bei der Bewältigung der abwärts Schwierigkeitschallenge zurück? **(4 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

$$\text{Suchzeit} = 0.15 + 0.18 \cdot \text{Schwierigkeit}$$

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

4. An der höchsten Schwierigkeit müssen Sie angeln. Ihre Angel ist ausgefahren 8m lang. Erreichen Sie einen Cache in der Höhe von 10.1m? Berechnen Sie dazu Ihre maximale mögliche Angelhöhe! Welche Annahmen mussten Sie treffen um die Aufgabe zu lösen? **(3 Punkte)**

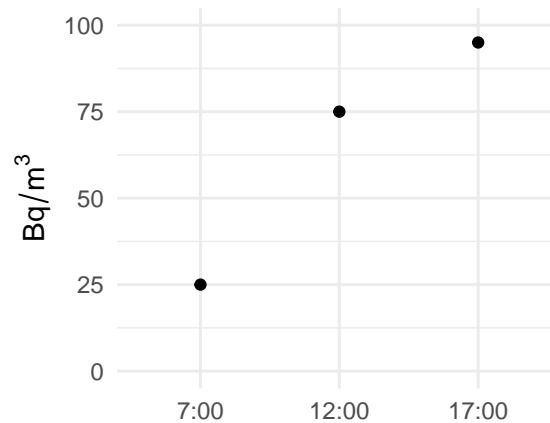
## 108 Aufgabe

(10 Punkte)

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



**Die atmende Wand und Brot aus Luft** Sie wollen das Ausmaß der Radonbelastung in ihrem Keller bestimmen und lüften daher nicht. Während einer Messperiode von 7:00 Uhr bis 17:00 bestimmen Sie dreimal automatisch die Radonbelastung in Ihrem Kellerraum in  $Bq/m^3$ . Es ergibt sich folgende Abbildung.



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

Radon zerfällt mit einer Halbwertszeit von 3.7d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 135d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

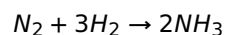
2. Wie lange dauert es in Stunden bis Ihre kritische Radonbelastung von  $320 Bq/m^3$  auf unter  $80 Bq/m^3$  gefallen ist? **(4 Punkte)**

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	77.1	28.1	
Sauerstoff	21.3	16.5	
Kohlenstoffdioxid	0.029	12.5	

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

Für die Umwandlung von Stickstoff  $N_2$  mit Wasserstoff  $H_2$  zu Ammoniak  $NH_3$  gilt folgende Reaktionsgleichung:



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

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

## 109 Aufgabe

(10 Punkte)

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



**Armee der Finsternis** Ihr Studentenjob war nach Ladenschluss bei Aldi die Regale einzuräumen. Dabei ist Ihnen in der Auslage der Sonderangebote das Necronomicon<sup>2</sup> in die Hände gefallen. Nun sind Sie eine Magierin der Zeichen geworden! Also eigentlich können Sie nur Mathe und das dämliche Necronomicon hat Sie in die Vergangenheit geschleudert... aber gut, was tut man nicht alles im Jahr 992 n. Chr. für den neuen Lehnsherren Fürsten Arthur. Sie bauen natürlich einen Schrottkugelturm um sich den Horden der Finsternis mit genug Schrott erwehren zu können! Ihnen stehen zwei mächtige magische Formeln zur Unterstützung zu Verfügung.

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

mit

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

Als erstes müssen Sie die Höhe des zu bauenden Schrottkugelturmes bestimmen. Hierfür ist wichtig zu wissen, dass sich die Bleitropfen mit einem Gewicht von  $20mg$  zu gleichförmigen Bleikugeln bei einer Geschwindigkeit von  $10m/s$  bilden.

1. Wie hoch müssen Sie den Schrottkugelturm bauen lassen, damit sich runde Bleikugeln durch die Fallgeschwindigkeit von  $10m/s$  bilden? **(3 Punkte)**

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

2. Nennen Sie die beiden geometrischen Formen aus denen sich näherungsweise ein Tropfen zusammensetzt! Erstellen Sie eine beschriftete Skizze des Tropfens! **(2 Punkte)**
3. Sie messen eine Länge des Tropfens von  $3.5mm$ . Die Löcher im Sieb erlauben ein Tropfendurchmesser von  $1.7mm$ . 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  $10.32g/cm^3$ .

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  $cm^2$  ist Ihr quadratisches Sieb am oberen Ende des Turms, wenn Sie pro Fall ca. 900 Bleikugeln produzieren wollen und die Bleikugel im Fall  $0.8cm$  Abstand haben müssen? **(1 Punkt)**

<sup>2</sup>Ein wirklich gefährliches Buch ist: *Du bist genug: Vom Mut, glücklich zu sein* von Fumitake Koga und Ichiro Kishimi

## 110 Aufgabe

(10 Punkte)

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



**Armee der Kaninchen** Leider hat es mit Ihrer Faultierpension in Down Under nicht geklappt. War vielleicht auch nicht so die beste Idee... aber dafür haben Sie eine Neue! Oder wie es Mike Tyson zugeschrieben wird: „Ich wurde nie niedergeschlagen, ich war immer am Aufstehen!“. Daher machen Sie jetzt einen Großhandel mit Kaninchenfleisch und damit dem teuersten Fleisch in Australien auf. Moment, hopsen hier nicht, seit Thomas Austin im Jahr 1875 ungefähr 32 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Dem wollen wir mal mathematisch nachgehen!<sup>3</sup>

Unter australischen Bedingungen liegt die Zahl der Jungtiere eines Kaninchenweibchens pro Jahr bei ca. 9 Tieren pro Wurf. Ein Weibchen hat im Durchschnitt 7 Würfe pro Jahr. Trotz fehlender Fressfeinde liegt die Mortalität der Säuglinge bei 35%.

1. Wie viele Millionen Kaninchen würden bei einem ungebremsten Wachstum *im vierten Jahr* geboren? Berechnen Sie hierfür zuerst die Anzahl an geborenen Weibchen im vierten Jahr! **(2 Punkte)**
2. Die durchschnittliche, australische Lebenserwartung eines weiblichen Kaninchens liegt bei 8 Jahren. Wie viele weibliche Kaninchen bevölkern bei einem ungebremsten Wachstum den australischen Kontinent bevor die ersten Weibchen an Altersschwäche sterben würden? **(2 Punkte)**

Forscherinnen fand folgende Sättigungsfunktion für das Wachstum der gesamten Kaninchenpopulation.

$$f(t) = 1.1 \times 10^{10} - 9 \times 10^8 \cdot 2^{-0.1 \cdot t + 3.4}$$

3. Wie viele Millionen Kaninchen leben nach der Sättigungsfunktion nach 14 Jahren auf dem australischen Kontinent? **(1 Punkt)**

Das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV), transmittiert von Stechmücken, töten 99.9% der Kaninchenpopulation innerhalb weniger Wochen.

4. Wie lange in Jahren dauert es bis eine Kaninchenpopulation nach einer Viruspandemie wieder auf 50% der gesättigten Kaninchenpopulation angewachsen ist? **(2 Punkte)**

Thomas Austin entließ die Kaninchen im äußersten Westen von Australien. Australien hat eine West-Ost-Ausdehnung von 4100km und eine Nord-Süd-Ausdehnung von knapp 3600km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 9.8km pro Jahr aus.

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

Eine jährliche Impfung gegen das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV) kosten 11\$ pro Tier und der durchführende Arzt verlangt ca. 45\$ pro Tier.

6. In Ihrem Stall leben 1200 Mastkaninchen. Mit welchen jährlichen Zusatzkosten für die Impfungen der Kaninchen müssen Sie daher kalkulieren? **(1 Punkt)**

<sup>3</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...](#)

## 111 Aufgabe

(12 Punkte)

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



**Ostfriesland. Unendliche Weiten.** Wir schreiben das Jahr 2023. Dies sind die Abenteuer des Hängebauschweins Frida und Ihnen. Grünes Gras unter Ihren Füßen und ein strammer Wind im Gesicht, egal wohin Sie schauen. Ein schmatzendes Geräusch ertönt unter Ihnen. Sie sinnieren, sollten Sie Ihre weiten Graslandschaften jetzt schon düngen? Dafür benötigen Sie die *Grünlandtemperatur*! Die Grünlandtemperatur (GLT) ist die Summe aller positiven Tagesmitteltemperaturen seit Jahresbeginn. Ab einer GLT von  $200^\circ$  kann mit der Stickstoffdüngung begonnen werden. Sie sehen nicht ein, Geld für einen Agrarmetrologen zu bezahlen. Also rechnen Sie mit folgenden Informationen zu Monatsmultiplikatoren des GLT-Wertes: Januar mit  $0.5\times$ , Februar mit  $0.8\times$  und März mit  $1.2\times$ . Sie haben noch im letzten Jahr folgende Temperaturen gemessen.

Datum	C°
01. Jan 2022	0.1
01. Feb 2022	1.2
01. Mrz 2022	2.7
01. Apr 2022	4.3

1. Erstellen Sie eine Skizze aus den Informationen aus der Temperaturtabelle! **(1 Punkt)**
2. Stellen Sie die linearen Funktionen  $f_1(t)$ ,  $f_2(t)$  und  $f_3(t)$  aus der obigen Temperaturtabelle auf! **(1 Punkt)**
3. Bestimmen Sie die Stammfunktionen  $F_1(t)$ ,  $F_2(t)$  und  $F_3(t)$  für Ihre linearen Funktionen aus der obigen Temperaturtabelle! **(1 Punkt)**
4. Osterglocken beginnen ab einer GLT von  $190^\circ\text{C}$  zu blühen. An welchem Tag im 1. Quartal des Jahres 2022 war dies der Fall? *Ignorieren Sie ein eventuelles Schaltjahr in Ihrer Berechnung.* **(4 Punkte)**

Auf dem Weg zu Ihrer Jonagoldplantage wurden Sie mit Ihrem Trecker von einer Gruppe elektrifizierter Rentner abgedrängt. Der Trecker muss wieder aus dem Graben! Frida und die elektrifizierten Rentner ziehen an zwei, separaten Seilen. Dabei zieht Frida mit  $180\text{N}$ . Die elektrifizierten Rentner bringen eine Kraft von  $190\text{N}$  auf.

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

5. Im ersten Versuch legen Sie das Seil für Frida lotrecht über einen Ast oberhalb des Treckers. Die Rentner ziehen in einer geraden Linie über die Böschung hinweg am anderen Seil. Welche Kraft wird aufgebracht? **(2 Punkte)**
6. Im zweiten Versuch ziehen Frida und die Rentner mit einem  $40^\circ$  Winkel mit ihrem Seil an dem Trecker. Welche Kraft wird aufgebracht? **(2 Punkte)**
7. Mit welcher Beschleunigung ziehen Sie den  $1.2\text{t}$  schweren Trecker *jeweils* aus dem Graben, wenn  $F = m \cdot a$  gilt? **(1 Punkt)**



## 112 Aufgabe

(10 Punkte)

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



**In der Kartonagenfabrik** Wenn man sich zu spät anmeldet, dann ist die Exkursion nicht so toll. Also geht es mit Rektor Skinner und Mrs. Krabappel in die Kartonagenfabrik. Wie schon im vorherigen Semester... In der Kartonagenfabrik angekommen erfahren Sie, dass die Kartons zum Versand von Nägeln nicht hier zusammengebaut werden sondern das sich die Endfertigung in Flint, Michigan befindet. Unter anderem wird dort der berühmte *Doppelt gewellte, 4-mal-gefaltete, 0.7mm, 60-cm-Karton* durch Falzung hergestellt. Beim letzten Mal war Rektor Skinner die Stimmung zu schlecht und deshalb geht es erst nach Hause, wenn ein paar Aufgaben gelöst sind. Martin gefällt das. An dem Vorrat an Zigaretten von Mrs. Krabappel meinen Sie wenig Zuversicht zu erkennen.

Jetzt heißt es Kartons optimieren. Der nun zu optimierende, flache Karton hat eine Länge von 60cm und eine Breite von 21cm. Die Kartonagenmaschine in Flint soll dann einen quadratischen Eckenausschnitt der Länge  $x$  falzen.

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

Rektor Skinner möchte sich gerne wieder in seinem Vorgarten aufhalten und nicht die ganze Zeit von Bart mit Erdnüssen beworfen werden. Deshalb möchte er einen geräumigen Teil seines Vorgartens einzäunen. Ein Teil der Umzäunung bildet seine Vorderhauswand. Wegen Lieferschwierigkeiten stehen Rektor Skinner nur 110m Zaun zu Verfügung. Sie wollen nun die maximale Fläche des abgeschirmten Vorgartens in Abhängigkeit der Seitenlängen bei der Verwendung von 110m 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)**

## 113 Aufgabe

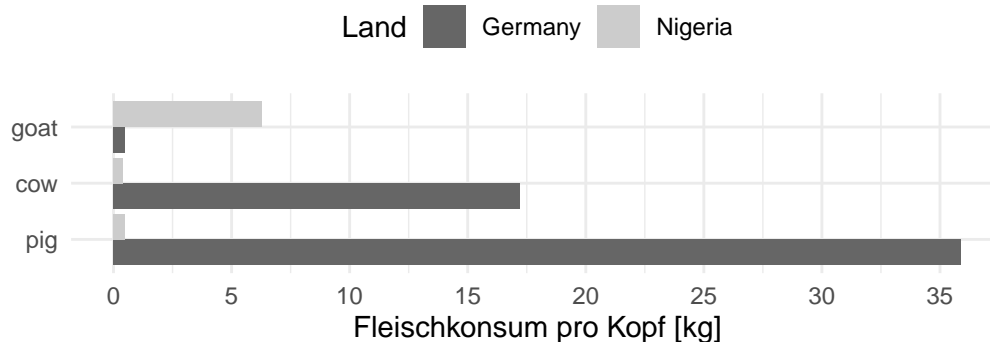
(10 Punkte)

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



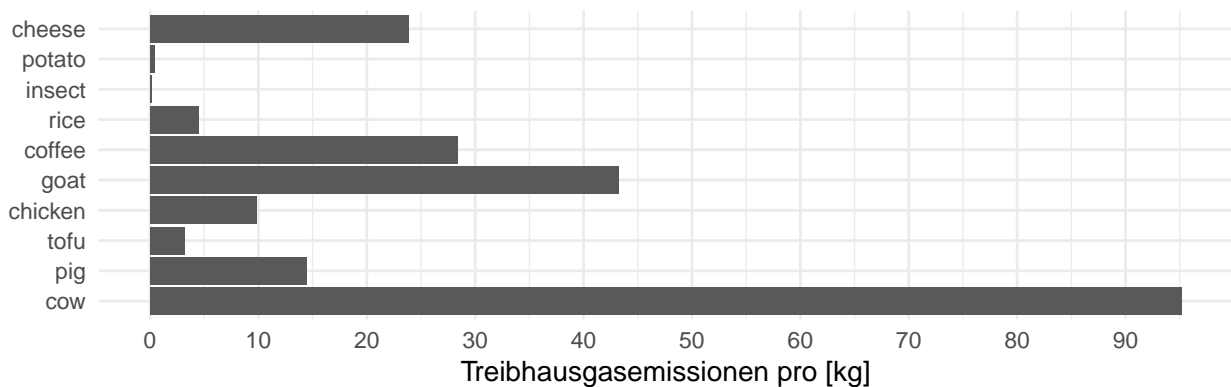
**Ein Pfund Insekten, bitte!** Das wird wohl häufiger gehört werden, wenn wir die Menschheit mit Proteinen ausreichend ernähren wollen<sup>4</sup>. Schauen wir uns dazu einmal den Vergleich Deutschland zu Nigeria an. Nigeria hat eine der am schnellsten wachsenden Bevölkerungen der Welt und wird vermutlich im Jahr 2100 zu den Top 5 der bevölkerungsreichsten Länder zählen. Im Jahr 2023 leben ca.  $8 \times 10^7$  Menschen in Deutschland und ca.  $1.79 \times 10^8$  Menschen in Nigeria. Mit den Informationen wollen wir anfangen und dann eine Prognose für 2050 zu treffen.

Im folgenden ist Abbildung des Fleischkonsums im Jahr 2023 in Deutschland und Nigeria einmal dargestellt.



1. Stellen Sie den Fleischkonsum in Deutschland und Nigeria im Jahr 2023 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 d.h. ins Verhältnis setzen! **(1 Punkt)**

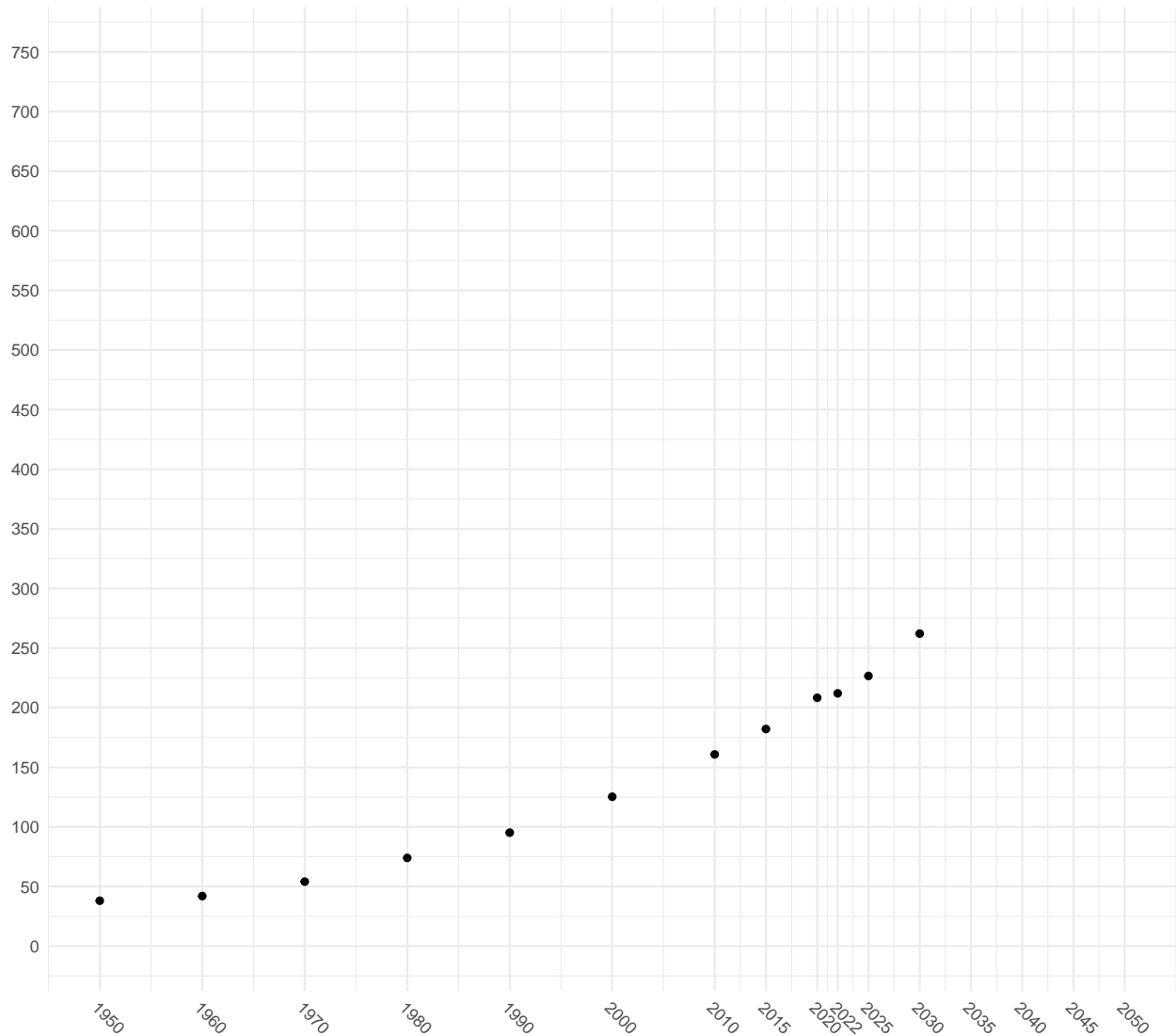
In der nächsten Abbildung finden Sie die CO<sub>2</sub>-Emission nach Lebensmittel abgebildet, die durch die Produktion entsteht.



3. Ergänzen Sie Ihre erstellte Tabelle die Treibhausgasemissionen an CO<sub>2</sub>, die durch den Fleischkonsum in Deutschland und Nigeria im Jahr 2023 entstehen! **(1 Punkte)**

<sup>4</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 in Nigeria von 1950 bis ins Jahr 2030 fortgeführt.



4. Schätzen Sie graphisch die zu erwartende Bevölkerung in Nigeria im Jahr 2050, die sich anhand der Informationen aus der Abbildung ergibt!
  - a) Ohne Berücksichtigung der Covid-19-Pandemie! **(2 Punkte)**
  - b) Unter Berücksichtigung der Covid-19-Pandemie! **(1 Punkt)**
5. Berechnen Sie den geschätzten Fleischkonsum von Nigeria im Jahr 2050 unter der Annahme 80%-iger Angleichung der Lebensbedingungen zu Deutschland im Jahr 2023! **(1 Punkt)**
6. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an  $\text{CO}_2$  in Nigeria im Jahr 2050 im Vergleich zum Jahr 2023, der sich durch den angeglichenen Fleischkonsum ergibt! **(1 Punkt)**
7. Berechnen Sie die prozentuale Steigerung der Treibhausgasemissionen an  $\text{CO}_2$  in Nigeria, wenn die gesamte Proteinaufnahme durch Insekten ersetzt würde! **(1 Punkt)**

## 114 Aufgabe

(10 Punkte)

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



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

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

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

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

3. Würden Sie die Axt mit zwei achtseitigen Würfeln ( $2d8$ ) als Schaden oder das Schwert mit einem sechseitigen Würfel plus 2 ( $1d6+2$ ) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

Jetzt wird es immer wilder, da Sie sich jetzt überlegen müssen, wie wahrscheinlich es ist, dass Ihr Rettungswurf gegen den zaubernden Hexer funktioniert. Sie haben folgende Wahrscheinlichkeiten gegeben. Die Wahrscheinlichkeit für das Ereignis  $A$ , der Rettungswurf ist erfolgreich, ist  $Pr(A) = 0.6$ , die Wahrscheinlichkeit für das Ereignis  $B$ , der Zauberwurf des Hexers ist erfolgreich, ist  $Pr(B) = 0.7$ . Sie haben mitgezählt und festgestellt, dass in 45 von 100 Fällen Ihr Rettungswurf bei einem erfolgreichen Zauber funktioniert hat.

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

## 115 Aufgabe

(12 Punkte)

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



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

Name	$P(\text{win})$	$P(\text{outbid})$
Günther	0.2	0.02
Thorsten	0.2	0.043

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

Glücksrad für Arme auf der Kirmes! Leider hat es für Maren Gilzer nicht gereicht. Deshalb sind Sie jetzt auf der Kirmes und spielen mit Catwoman um das große Geld. Das Glücksrad hat 24 Felder. Sie drehen das Glücksrad zweimal. Auf 10 Feldern gewinnen Sie 3000EUR sonst 1500EUR. Ganz schön viel Geld und ganz schön zwielichtig hier...

3. Skizzieren Sie das Glücksrad und ergänzen Sie die Wahrscheinlichkeiten! **(1 Punkt)**
4. Zeichnen Sie das zugehörige Baumdiagramm für das zweimalige Drehen! Ergänzen Sie die Wahrscheinlichkeiten und die entsprechenden Ereignisse **(2 Punkte)**
5. Mit welcher Wahrscheinlichkeit gewinnen Sie 4500EUR? **(1 Punkt)**

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

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

## 116 Aufgabe

(12 Punkte)

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



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

Die Prävalenz von AIDS bei einem Menschen in Europa wird mit 0.01% angenommen. In 95% der Fälle ist ein HIV-Test positiv, wenn der Patient erkrankt ist. In 1% der Fälle ist ein HIV-Test positiv, wenn der Patient *nicht* erkrankt ist und somit gesund ist. Sie stutzen. Wie wahrscheinlich ist es denn eigentlich an AIDS erkrankt zu sein ( $K^+$ ), wenn Sie einen positiven AIDS-Test vorliegen haben ( $T^+$ )? Gehen Sie für die folgenden Berechnungen von  $n = 4 \times 10^4$  Patienten mit einem diagnostischen Test für AIDS aus. Sie nehmen sich also einen Kuli und fangen an auf der historischen Ausgabe des Spiegels zu rechnen<sup>5</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 zuder AIDS Erkrankung und dem AIDS-Test! **(2 Punkte)**
4. Füllen Sie den Häufigkeitsdoppelbaum mit den sich ergebenden, absoluten Patientenzahlen  $n$  aus! **(2 Punkte)**

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

5. Tragen Sie  $TP$ ,  $TN$ ,  $FP$  und  $FN$  in eine 2x2 Kreuztabelle ein. Beschriften Sie die Tabelle entsprechend! **(1 Punkt)**
6. 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)**
7. 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>5</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

## 117 Aufgabe

(12 Punkte)

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



**Network-Marketing oder Schneeballschlacht!** Eine Möglichkeit, leicht Geld zu verdienen, ist es anderen Menschen für Geld zu versprechen, wie man leicht reich werden kann. Am besten natürlich ohne viel Aufwand und ortsunabhängig. Dann wollen wir mal loslegen. Schnell ein YouTube-Werbevideo gedreht und auf geht es mit unserem Network-Marketing. Aber Moment, wie funktioniert Network-Marketing eigentlich und was hat das alles mit einer Schneeballschlacht zu tun? Wir wollen hier einmal in die Untiefen des „passiven Einkommens“ abtauchen<sup>6</sup>.

Das Jahr 2021 war das erfolgreichste Jahr in der Geschichte von KH Gesund und Schön Components (KH-GSC). Das Unternehmen steigerte den Umsatz um rund 10 Prozent von 310 Millionen Euro im Jahr 2020. Doch wie viel kommt bei den Partnern an? Laut KH-GSC habe das Unternehmen  $3.3 \times 10^5$  aktive Partner.

1. Berechnen Sie zuerst den Umsatz der Firma KH-GSC im Jahr 2021! **(1 Punkt)**
2. Wie viel von dem Umsatz im Jahr 2021 wird im Durchschnitt von jedem aktiven Partner erwirtschaftet? **(1 Punkt)**
3. Welche *monatlicher* Umsatz ergibt sich dadurch im Durchschnitt für jeden aktiven Partner bei einer direkten Provision von 20%? **(1 Punkt)**

Ihr zu vermarktendes Produkt, hinter dem Sie voll stehen, kostet 150EUR pro Einheit im Direktverkauf. Die direkte Provision für die erste Stufe beträgt 25%. Für die zweite, dritte und vierte Stufe betragen die indirekten Provisionen jeweils 1.75%, 0.75% und 0.25%. Jeder Ihrer angeworbenen „Partner“ wirbt wiederum drei Partner für sich selbst an. Pro Monat werden im Schnitt fünf Einheiten vom Produkt verkauft. Sie wollen nun 4100EUR im Monat *passiv* erwirtschaften.

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

Stufe	Anzahl Partner	Umsatz/Stufe	Provision
1	0 (Sie selber)		
2	1		
3			
4			

5. Wie viele Partner müssen auf der 2 Stufe anwerben um Ihr passives Einkommen zu erreichen? Stellen Sie den Zusammenhang graphisch dar! **(3 Punkte)**

Sie mussten zum Einstieg bei KH-GSC Einheiten des Produkts für 6000EUR kaufen. Diese Einheiten können Sie nur direkt verkaufen. Leider mussten Sie den Kauf über einen Kredit über 4.5% p.a. über 48 Monate finanzieren.

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

<sup>6</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](#)

## 118 Aufgabe





(10 Punkte)

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



1. Ergänzen Sie die logischen Operatoren in  in die untenstehende Tabelle! (2 Punkte)

Ausdruck	Operator
UND	
ODER	
GROESSER GLEICH	
GLEICH	

2. Visualisieren Sie folgende logische Aussagen zu der Menge  $A$  und der Menge  $B$  als Mengendiagramme bzw. Venndiagramme! (2 Punkte)
- Entweder  $A$  tritt ein oder  $B$  tritt ein oder keins von beiden
  - $B$  ist Teilmenge von  $A$
  - $A$  tritt ein, aber  $B$  tritt nicht ein
  - $A$  tritt ein und  $B$  tritt ein
3. Ergänzen Sie zu den Venndiagrammen die mathematische Notation! (2 Punkte)
4. Erstellen Sie die numeric  Ausgabe für die Suche nach der Zeichenfolge GT in folgenden DNA Sequenzen! (2 Punkte)
- ACTGGTAA
  - AAATTTAT
  - ATATACAC
  - AACCAACC
5. Geben ist das  Objekt  $A$  beinhaltend die Zahl 8. Erklären Sie den Unterschied zwischen dem Ausdruck  $A == 8$  und dem Ausdruck  $A != 8$  in ! Wie lautet die Ausgabe von  in beiden Fällen? (2 Punkte)



## 119 Aufgabe

(10 Punkte)

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



Sie haben den Vektor  $A = \{7, 10, -3\}$  und den Vektor  $B = \{1, 5, 1\}$  gegeben.

1. Addieren Sie die Vektoren  $A$  und  $B$ ! **(1 Punkt)**
2. Transponieren Sie den Vektor  $A$ ! **(1 Punkt)**
3. Multiplizieren Sie den Vektor  $A$  mit 10! **(1 Punkt)**
4. Erstellen Sie eine stochastische Matrix mit  $m \times n$  Dimensionen! Legen Sie  $m$  und  $n$  vorher sinnvoll fest! **(1 Punkt)**
5. Übersetzen Sie folgendes linearen Gleichungssystem bestehend aus drei Gleichungen in die Matrixschreibweise! **(2 Punkte)**

```
## [1] "4x1 + 7x2 + 2x3 = 2"
## [1] "4x1 + 7x2 + 9x3 = 1"
## [1] "4x1 + 7x2 + 2x3 = 3"
```

6. Multiplizieren Sie folgende gegebene Matrix mit dem Vektor  $B$ ! **(2 Punkte)**

```
##      [,1] [,2] [,3]
## [1,]    8    6    8
## [2,]    6    4    5
## [3,]    8    5    5
## [4,]    7    5    8
```

7. Skizzieren Sie den entsprechenden  Code für die Matrixmultiplikation! **(2 Punkte)**

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

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

## 120 Aufgabe

(6 Punkte)

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

## 121 Aufgabe

(8 Punkte)

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

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

mit

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

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

## 122 Aufgabe

(6 Punkte)

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