

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

*Nicht bestanden:* ☐

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

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

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

# Klausurfragen der Bio Data Science

**Hochschule Osnabrück**

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

Version vom 26. Februar 2023

## Inhalte der Klausur nach Modul

*Die folgende Einordnung der Themenkomplexe zu den Modulen dient zur Orientierung. Einzelne Klausuren können und werden davon abweichen. Die Angaben hier sind daher ohne Gewähr.*

### *Mathematik & Statistik*

- 1 Frage zur deskriptiven Statistik
- 1 Frage zum t-Test
- 1 Frage zur einfaktoriellen ANOVA
- 1 Frage zur R Programmierung
- 3 Fragen zur Mathematik

### *Statistik*

- 1 Frage zur deskriptiven Statistik
- 1 Frage zum statistischen Testen
- 1 Frage zum t-Test
- 1 Frage zur einfaktoriellen ANOVA
- 1 Frage zur linearen Regression beinhaltend Korrelation und  $R^2$
- 1 Frage zum  $\chi^2$ -Test oder Nicht-Parametrik oder diagnostischen Testen
- 1 Frage zur R Programmierung

### *Angewandte Statistik für Bioverfahrenstechnik*

- 1 Frage zur deskriptiven Statistik
- 1 Frage zum statistischen Testen
- 1 Frage zum t-Test
- 1 Frage zur einfaktoriellen ANOVA
- 1 Frage zur linearen Regression beinhaltend Korrelation und  $R^2$
- 1 Frage zum  $\chi^2$ -Test oder Nicht-Parametrik oder diagnostischen Testen
- 1 Frage zur R Programmierung

### *Angewandte Statistik und Versuchswesen*

- 1 Frage zur deskriptiven Statistik
- 1 Frage zu statistischen Verteilungen
- 1 Frage zum statistischen Testen
- 1 Frage zur zweifaktoriellen ANOVA
- 1 Frage zur linearen Regression beinhaltend Korrelation und  $R^2$
- 1 Frage zu multiplen Gruppenvergleichen
- 1 Frage zur R Programmierung

### *Biostatistik*

- 1 Frage zu der deskriptiven Statistik
- 1 Frage zum statistischen Testen
- 1 Frage zu der zweifaktoriellen ANOVA
- 1 Frage zu der multiplen Regression
- 1 Frage zu der R Programmierung
- 2 Fragen abhängig von der Vertiefung (Pflanze- oder Nutztierwissenschaften)

## 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.

## Ergebnis der Klausur

\_\_\_\_\_ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.  
\_\_\_\_\_ von 55 Punkten sind aus dem Rechen- und Textteil erreicht.  
\_\_\_\_\_ von 75 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

<b>Punkte</b>	<b>Note</b>
72.0 - 75.0	1,0
68.0 - 71.5	1,3
64.5 - 67.5	1,7
60.5 - 64.0	2,0
57.0 - 60.0	2,3
53.5 - 56.5	2,7
49.5 - 53.0	3,0
46.0 - 49.0	3,3
42.0 - 45.5	3,7
37.5 - 41.5	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	4	8	6	8	7	12	10

- Es sind \_\_\_\_ von 55 Punkten erreicht worden.

## 1 Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.89, 0.34, 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 1, 1, 0.004 und 0.12. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.2225, 0.085, 3e-04 und 0.0075. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 3.56, 1.36, 0.004 und 0.12. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 1, 0.004 und 0.12. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.2225, 0.085, 3e-04 und 0.0075. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.

## 2 Aufgabe

(2 Punkte)

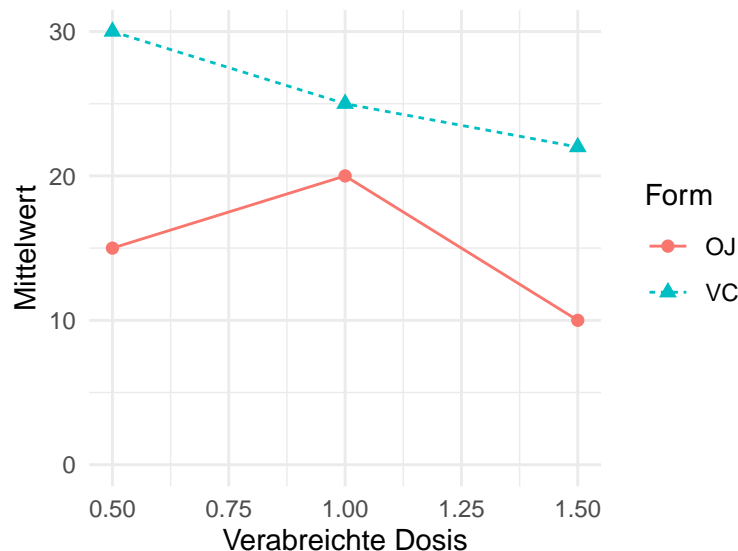
Der Datensatz PlantGrowth enthält das Gewicht von Pflanzen, die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden. Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein  $\eta^2 = 0.27$ . Welche Aussage ist richtig?

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

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

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

## 5 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 liegt Varianzhomogenität vor.
- C ☐ Es handelt sich um ein unbalanciertes Design
- D ☐ Es handelt sich um ein balanciertes Design.
- E ☐ Es liegt Varianzheterogenität vor.

## 6 Aufgabe

(2 Punkte)

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

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

## 7 Aufgabe

(2 Punkte)

Welche Aussage über den Korrelationskoeffizienten nach Pearson ist richtig?

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

## 8 Aufgabe

(2 Punkte)

Berechnen Sie den Mittelwert und Standardabweichung von  $y$  mit 10, 19, 15, 5 und 7.

- A ☐ Es ergibt sich 11.2 +/- 33.2
- B ☐ Es ergibt sich 11.2 +/- 5.76
- C ☐ Es ergibt sich 11.2 +/- 2.88
- D ☐ Es ergibt sich 12.2 +/- 2.88
- E ☐ Es ergibt sich 10.2 +/- 16.6

## 9 Aufgabe

(2 Punkte)

Berechnen Sie den Median, das 1<sup>st</sup> Quartile sowie das 3<sup>rd</sup> Quartile von  $y$  mit 9, 12, 7, 19, 17, 0, 9, 13, 15, 20 und 63.

- A ☐ Es ergibt sich 13 [9, 19]
- B ☐ Es ergibt sich 13 +/- 9
- C ☐ Es ergibt sich 17 [10, 20]
- D ☐ Es ergibt sich 13 +/- 19
- E ☐ Es ergibt sich 17 +/- 9

## 10 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. Cohen's  $d$  findet auch Outlier, ist aber ein veraltetes Konzept in der Statistik.
- 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 und Cohen's  $d$  um standardisierte Outlier für Gruppenvergeliche zu erhalten.

## 11 Aufgabe

(2 Punkte)

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

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

## 12 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 nicht. 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 unterscheiden, liegen vermutlich Outlier in den Daten vor. Wir untersuchen den Datensatz nach auffälligen Beobachtungen.



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

### 13 Aufgabe

(2 Punkte)

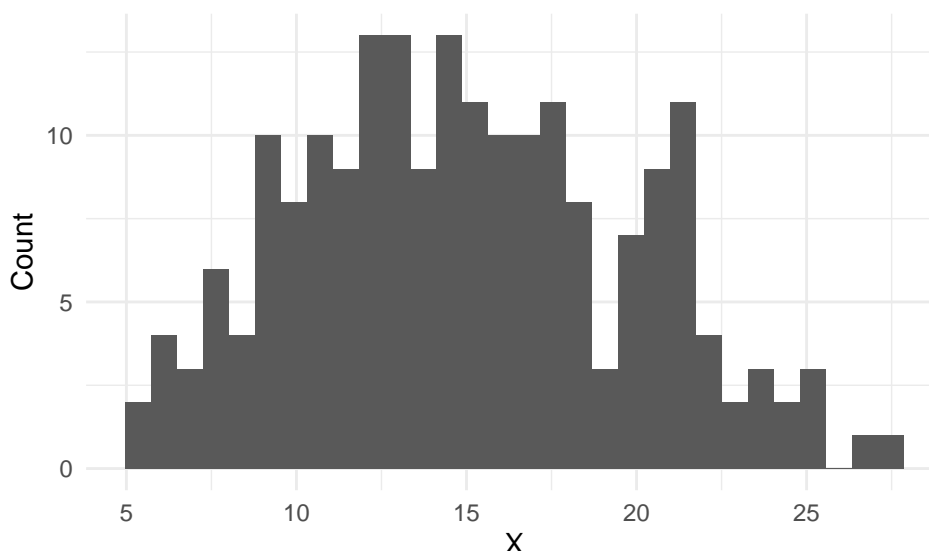
Nach einer simplen linearen Regression zur Untersuchung vom Einfluss der  $CO_2$ -Konzentration [ $\mu g$ ] im Wasser auf das Wachstum von Wasserlinsen [ $kg$ ] erhalten Sie einen  $\beta_1$  Koeffizienten von 0.00001 und einen hoch signifikanten  $p$ -Wert mit  $2.3 \cdot 10^{-9}$ . Warum sehen Sie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

- ☐ A Das Gewicht und die  $CO_2$ -Konzentration korrelieren sehr stark, deshalb wird der  $\beta_1$  Koeffizient sehr klein.
- ☐ B Die Fallzahl ist zu klein angesetzt. Je kleiner die Fallzahl ist, desto höher ist die Teststatistik und damit auch der  $p$ -Wert kleiner.
- ☐ C 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.
- ☐ D 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.
- ☐ E 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.

### 14 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 Binomial-Verteilung mit  $\text{Binom}(10)$ .
- D** ☐ Es handelt sich um eine Normalverteilung mit  $N(15, 5)$ .
- E** ☐ Es handelt sich um eine Poisson-Verteilung mit  $\text{Pois}(15)$ .

## 15 Aufgabe

(2 Punkte)

Sie schätzen zwei lineare Regressionsmodelle zur Analyse von Zähldaten. Modell 1 mit einer Poisson Verteilung und Modell 2 mit einer Quasi-Poisson Verteilung. Welche Aussage zu einer geschätzten Overdispersion von 2.62 ist richtig?

- A** ☐ Bei einer geschätzten Overdispersion höher als 1.5 ist von Overdispersion in den Daten auszugehen. Daher wird die Varianz systematisch unterschätzt, was zu kleineren p-Werten führt. Daher gibt es mehr signifikante Ergebnisse als es in Wirklichkeit gibt. Daher ist das Modell 2 die bessere Wahl.
- B** ☐ Bei einer geschätzten Overdispersion höher als 1.5 ist von keiner Overdispersion in den Daten auszugehen. Dennoch sind die p-Werte zu klein, dass diese p-Werte natürlich entstanden sein könnten. Die p-Werte müssen adjustiert werden.
- C** ☐ Das vergleichen von verschiedenen Modellen muss erst über ein AIC Kriterium erfolgen. Die Abschätzung über die Overdispersion ist nicht notwendig. Die Varianzen werden später in einer ANOVA adjustiert. Die Confounder Adjustierung.
- D** ☐ Bei einer geschätzten Overdispersion höher als 1.5 ist von Overdispersion in den Daten auszugehen. Daher wird die Varianz systematisch unterschätzt, was zu höheren p-Werten führt. Daher gibt es weniger signifikante Ergebnisse als es in Wirklichkeit gibt. Daher ist das Modell 1 die bessere Wahl.
- E** ☐ Bei einer geschätzten Overdispersion höher als 1.5 ist von Overdispersion in den Daten auszugehen. Daher wird die Varianz systematisch überschätzt, was zu höheren p-Werten führt. Daher gibt es mehr signifikante Ergebnisse als es in Wirklichkeit gibt. Daher ist das Modell 1 die bessere Wahl.

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

## 17 Aufgabe

(2 Punkte)

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

- A ☐ Ein prädiktives 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 prädiktives Modell wird auf einem Trainingsdatensatz trainiert und anschliessend über eine explorative Datenanalyse validiert. Signifikanzen über  $\beta_i$  können hier nicht festgestellt werden.
- C ☐ Ein prädiktives Modell basiert auf einem Trainingsdatensatz und einem Testdatensatz. Auf dem Trainingsdatensatz wird das Modell trainiert und auf dem Testdatensatz validiert.
- D ☐ Ein prädiktives Modell schliesst grundsätzlich lineare Modell aus. Es muss ein Graph gefunden werden, der alle Punkte beinhaltet. Erst dann kann das  $R^2$  berechnet werden.
- E ☐ Ein prädiktives Modell benötigt mindestens eine Fallzahl von über 100 Beobachtungen und darf keine fehlenden Werte beinhalten. Die Varianzkomponenten müssen homogen sein.

## 18 Aufgabe

(2 Punkte)

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Kühen durch. Bei 3 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 0.2, da es sich um eine Chancenverhältnis handelt.
- B ☐ Es ergibt sich ein Risk ratio von 0.2, da es sich um ein Anteil handelt.
- C ☐ Es ergibt sich ein Risk ratio von 0.25, da es sich um eine Chancenverhältnis handelt.
- D ☐ Es ergibt sich ein Risk ratio von 4, da es sich um ein Anteil handelt.
- E ☐ Es ergibt sich ein Risk ratio von 0.25, da es sich um ein Anteil handelt.

## 19 Aufgabe

(2 Punkte)

Welche Aussage über die nicht-parametrische Statistik ist richtig?

- A ☐ Die nicht-parametrische Statistik basiert auf dem Schätzen von Parametern aus einer festgelegten Verteilung. Daher gibt es auch direkt zu interpretierenden Effektschätzer.
- B ☐ Die nicht-parametrische Statistik basiert auf Rängen. Daher gibt es auch direkt zu interpretierenden Effektschätzer.
- C ☐ 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.
- 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 ist ein Vorgänger der parametrischen Statistik und wurde wegen dem Mangel an Effektschätzern nicht mehr ab 1960 genutzt.

## 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 erlaubt erst die Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich.
- C** ☐ Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte.
- D** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- 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 **R** auswerten wollen. Welcher folgende Grund ist richtig?

- A** ☐ Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- B** ☐ Die Spracherkennung von **R** ist nicht in der Lage Deutsch zu verstehen.
- C** ☐ Alle Funktionen und auch Anwendungen sind in **R** in englischer Sprache. Die Nutzung von deutschen Wörtern ist nicht schick und das ist zu vermeiden.
- D** ☐ Programmiersprachen können nur englische Begriffe verarbeiten. Zusätzliche Pakete können zwar geladen werden, aber meist funktionieren diese Pakete nicht richtig. Deutsch ist International nicht bedeutend genug.
- E** ☐ Im Allgemeinen haben Programmiersprachen Probleme mit Umlauten und Sonderzeichen, die in der deutschen Sprache vorkommen. Eine Nutzung der englischen Sprache umgeht dieses Problem auf einfache Art.

## 22 Aufgabe

(2 Punkte)

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

- A** ☐ Wir lesen die Daten ein und mutieren die Daten. Dabei ist wichtig, dass wir nicht das Paket `tidyverse` nutzen, da dieses Paket veraltet ist. Über die Funktion `library(tidyverse)` entfernen wir das Paket von der Analyse.
- B** ☐ Wir 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.
- C** ☐ 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.

- D** ☐ 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.
- E** ☐ Wir lesen die Daten über eine generische Funktion `read()` ein und müssen dann die Funktion `ggplot()` nur noch installieren. Dann haben wir die Abbildungen als `*.png` vorliegen.

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

## 24 Aufgabe

(2 Punkte)

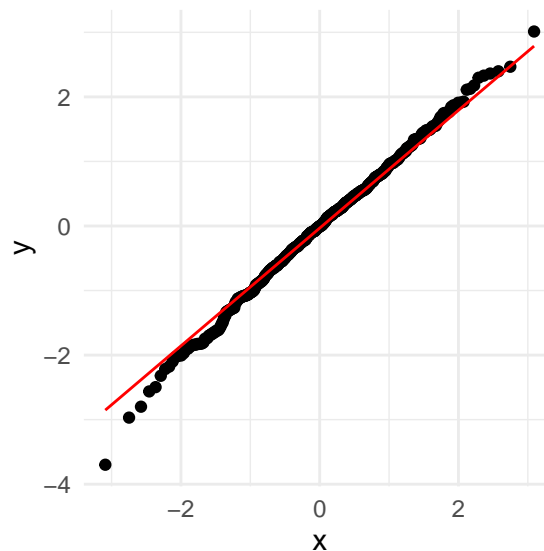
Welche Aussage über das *generalisierte lineare Modell (GLM)* ist richtig?

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

## 25 Aufgabe

(2 Punkte)

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

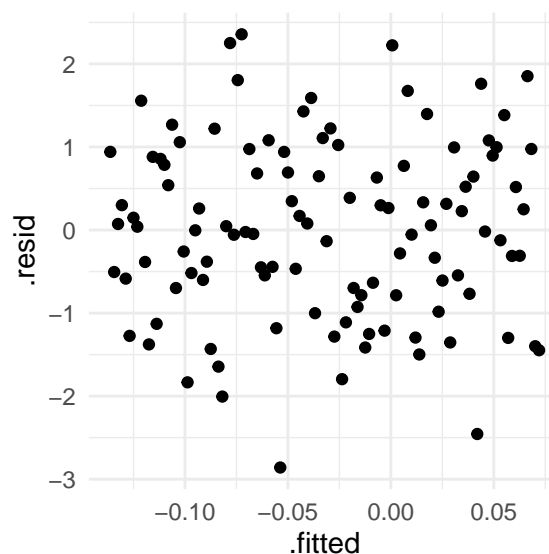


- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden und Korrelation ist negativ.
- B** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- C** ☐ Die Annahme der normalverteilten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

## 26 Aufgabe

(2 Punkte)

Sie rechnen eine lineare 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 erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Diagonalen.
- 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 nicht erfüllt. Es ist kein Muster zu erkennen.
- E** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Es ist ein Muster zu erkennen.

## 27 Aufgabe

(2 Punkte)

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

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

## 28 Aufgabe

(2 Punkte)

Das Falsifikationsprinzip besagt...

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

## 29 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 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.
- B** ☐ 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.

- C** ☐ Im Rahmen eines langen Disputs zwischen Neyman und Fischer wurde  $\alpha = 5\%$  festgelegt. Leider werden die Randbedingungen und Voraussetzungen an statistische Modelle heute immer wieder ignoriert.
- D** ☐ Auf einer Statistikkonferenz in Genf im Jahre 1942 wurde dieser Cut-Off nach langen Diskussionen festgelegt. Bis heute ist der Cut Off aber umstritten, da wegen dem 2. Weltkrieg viele Wissenschaftler nicht teilnehmen konnten.
- E** ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.

### 30 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  $1 - \beta$  wird auf 80% gesetzt. Damit liegt die Wahrscheinlichkeit für die  $H_0$  bei 20%.
- C** ☐ Es gilt  $\alpha + \beta = 1$  und somit liegt  $\beta$  meist bei 95%.
- D** ☐ Die Power ist nicht in der aktuellen Testtheorie mehr vertreten. Wir rechnen nur noch mit dem Fehler 1. Art.
- E** ☐ Die Power  $1 - \beta$  wird auf 80% gesetzt. Alle statistischen Tests sind so konstruiert, dass die  $H_A$  mit 80% "bewiesen wird".

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

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



### 33 Aufgabe

(2 Punkte)

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

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

### 34 Aufgabe

(2 Punkte)

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

- A ☐ Aktuell werden zu grosse Datensätze für die gängige Statistik gemessen. Daher wendet man maschinelle Lernverfahren für kausale Modelle an. Hier ist die Relevanz gleich Signifikanz.
- B ☐ Relevanz und Signifikanz haben nichts miteinander zu tun. Daher gibt es auch keinen Zusammenhang zwischen hoher Fallzahl ( $n > 10000$ ) und einem signifikanten Test. Ein Effekt ist immer relevant und somit signifikant.
- C ☐ Aktuell werden immer grössere Datensätze erhoben. Eine erhöhte Fallzahl führt automatisch auch zu mehr signifikanten Ergebnissen, selbst wenn die eigentlichen Effekte nicht relevant sind.
- D ☐ 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 immer grössere Datensätze erhoben. Dadurch wird auch die Varianz immer höher was automatisch zu mehr signifikanten Ergebnissen führt.

### 35 Aufgabe

(2 Punkte)

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

- A ☐ Das  $\Delta$ . Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da  $\Delta$  antiproportional zum p-Wert ist, bedeutet auch ein hohes  $\Delta$  ein sehr kleinen p-Wert.
- B ☐ Der p-Wert. Durch den Vergleich mit  $\alpha$  lässt sich über die Signifikanz entscheiden und der  $\beta$ -Fehler erlaubt über die Power eine Einschätzung der Relevanz.
- C ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- D ☐ Das Konfidenzintervall. Durch die Visualisierung des Konfidenzintervalls kann eine Relevanzschwelle vom Anwender definiert werden. Zusätzlich erlaubt das Konfidenzintervall auch eine Entscheidung über die Signifikanz.
- E ☐ 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.

### 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 die Effekte des  $p$ -Wertes mit den Effekten der Signifikanzschwelle unter der Annahme der Nullhypothese.
- D ☐ 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.
- 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)

Welche Aussage über den t-Test ist richtig?

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

### 38 Aufgabe

(2 Punkte)

Welche Aussage über den Welch t-Test ist richtig?

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

### 39 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 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.
- B ☐ Die ANOVA testet auf der gesamten Fallzahl. Es wäre besser die ANOVA auf der gleichen Fallzahl wie die einzelnen t-Tests zu rechnen.
- C ☐ Es gibt einen Fehler in der Varianzstruktur. Daher kann die ANOVA nicht richtig sein und paarweise t-Tests liefern das richtige Ergebnis.
- D ☐ Der Fehler liegt in den t-Tests. Wenn eine ANOVA signifikant ist, dann muss zwangsweise auch ein t-Test signifikant sein.
- E ☐ 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.

### 40 Aufgabe

(2 Punkte)

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

- A ☐ Der gepaarte t-Test wird genutzt, wenn die Differenzen der Beobachtungen verbunden sind und wir dadurch die Unabhängigkeit nicht mehr vorliegen haben.
- B ☐ Der gepaarte t-Test nutzt die Varianz der Beobachtungen jeweils paarweise und bildet dafür eine verbundene Stichprobe. Dieser Datensatz  $d$  dient dann zur Differenzbildung.
- C ☐ 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.
- D ☐ 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.
- 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.

## 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](#)

## 41 Aufgabe

**(10 Punkte)**

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



Sie haben folgende Zahlenreihe  $y$  vorliegen  $y = \{16, 18, 19, 17, 12, 14, 11, 15, 14, 16, 13\}$ . Berechnen Sie folgende deskriptive Maßzahlen. Geben Sie Formeln und Rechenwege mit an!

1. Die Range oder Spannweite **(2 Punkte)**
2. Die Standardabweichung **(2 Punkte)**
3. Den Interquartileabstand **(2 Punkte)**
4. Das 3rd Quartile **(2 Punkte)**
5. Den Median **(2 Punkte)**

## 42 Aufgabe

(8 Punkte)

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



Sie haben folgende Zahlenreihe  $y$  vorliegen  $y = \{17, 14, 16, 12, 19, 12, 22\}$ .

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!



Nach einem Gewächshausexperiment mit drei Bewässerungstypen (*low*, *mid* und *high*) ergibt sich die folgende Datentabelle mit dem gemessenen Frischgewicht (*freshmatter*).

water_type	freshmatter
mid	31
low	13
high	6
low	15
high	11
mid	32
low	27
low	29
high	7
mid	31
low	18

1. Zeichnen Sie in *einer* Abbildung die Barplots für die Bewässerungstypen! 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 der Bewässerungstypen erwarten würden, wie sehen dann die Barplots aus? **(1 Punkt)**

## 44 Aufgabe

(6 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! **(4 Punkte)**
2. Achten Sie auf die entsprechende Skalierung der beiden Verteilungen in der ersten Abbildung! **(2 Punkte)**

$N(1, 3)$  und  $N(3, 3)$



$\text{Pois}(10)$





## 45 Aufgabe

(6 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  $s_1 \neq s_2 \neq s_3 \neq s_4$ ! **(2 Punkte)**
2. Beschriften Sie die Normalverteilungen mit den entsprechenden Parametern! **(2 Punkte)**
3. Liegt Varianzhomogenität oder Varianzheterogenität vor? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 46 Aufgabe

(6 Punkte)

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



Nach einem Experiment zählen Sie folgende Anzahl an Läsionen auf den Blättern von Sonnenblumen nach einer durchgestandenen Infektion.

6, 3, 4, 1, 6, 1, 6, 2, 5, 3, 4, 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)**

## 47 Aufgabe

(8 Punkte)

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



Nach einem Experiment zählen Sie folgende Trockengewichte von Sonnenblumen nach einer durchgestandenen Infektion.

10.8, 10.2, 11.5, 11.8, 7.8, 11.2, 11.3, 11, 10.6, 12.7, 11.6, 9.8, 10.7, 9.6, 8.5

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)**

## 48 Aufgabe

(9 Punkte)

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



Nach einem Feldexperiment mit zwei Düngestufen (A und B) ergibt sich die folgende Datentabelle mit dem gemessenen Trockengewicht (*drymatter*).

trt	drymatter
A	15.4
B	20.3
A	16.4
A	16.4
B	12.2
A	15.7
B	18.3
A	18.8
A	19.5
B	18.8
B	15.1
A	14.8
B	14.3
A	16.4
A	17.6
B	14.0
B	17.0
A	15.0
B	17.4
A	21.1
B	17.6
B	14.9

1. Zeichnen Sie in *einer* Abbildung die beiden Boxplots für die zwei Düngestufen A und B! 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 Düngestufen erwarten würden, wie sehen dann die beiden Boxplots aus? **(1 Punkt)**

## 49 Aufgabe

(9 Punkte)

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



Nach einem Feldexperiment mit mehreren Düngestufen stellt sich die Frage, ob die Düngestufe *low* im Bezug auf das Trockengewicht normalverteilt sei. Sie erhalten folgende Datentabelle.

fertilizer	drymatter
low	25
low	27
low	29
low	26
low	24
low	27
low	23
low	25
low	25
low	24
low	24

1. Zeichnen Sie eine passende Abbildung in der Sie visuell überprüfen können, ob eine Normalverteilung des Trockengewichts vorliegt! **(4 Punkte)**
2. Beschriften Sie die Achsen und ergänzen Sie die statistischen Maßzahlen. **(3 Punkte)**
3. Entscheiden Sie, ob eine Normalverteilung vorliegt. Begründen Sie Ihre Antwort. **(2 Punkte)**

## 50 Aufgabe

(4 Punkte)

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



1. Zeichnen Sie über den untenstehenden Boxplot die entsprechende zugehörige Verteilung! **(2 Punkte)**
2. Zeichnen Sie unter den untenstehenden Boxplot die entsprechende zugehörige Beobachtungen! **(2 Punkte)**



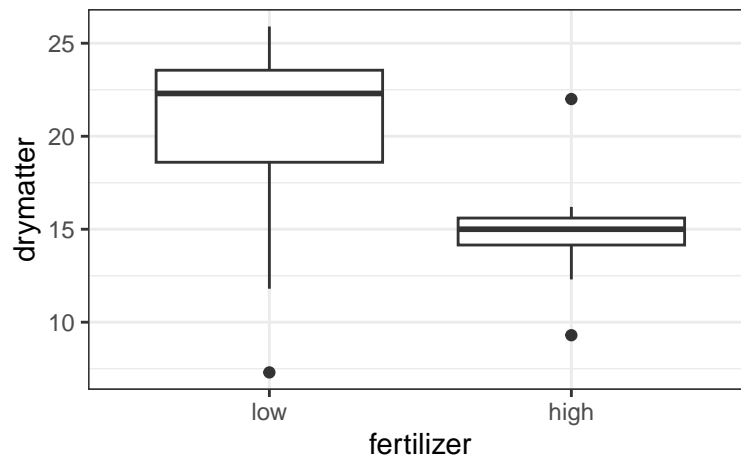
## 51 Aufgabe

(9 Punkte)

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



In einem Experiment mit zwei Düngestufen für den Ertrag von Kichererbsen ergibt sich folgende Abbildung.



- Tragen Sie in die untenstehende Tabelle die gängigen Maßzahlen des Boxplots und die abgeschätzten Werte aus den obigen Boxplots ein! **(4 Punkte)**

Statistische Maßzahl	Abgeschätzter Wert	
	low	high

- Ergänzen Sie den Mittelwert für beide Level des Düngers in die Abbildung der Boxplots! Begründen Sie Ihre Antwort! **(2 Punkte)**
- Ergänzen Sie in der untenstehenden Tabelle die  $p$ -Werte für den Shapiro-Wilk-Test auf Normalverteilung und den Levene-Test auf Varianzhomogenität. Beachten Sie die unterschiedliche, angenommene Fallzahl  $n_g$  der beiden Level des Düngers! **(3 Punkte)**

Fallzahl	Shapiro-Wilk-Test	Levene-Test
$n_g = 5$		
$n_g = 20$		
$n_g > 50$		

## 52 Aufgabe

(6 Punkte)

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



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

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

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)**



## 53 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.

pesticide	infected
RoundUp	yes
OutEx	no
RoundUp	yes
RoundUp	no
OutEx	no
OutEx	no
RoundUp	yes
OutEx	yes
OutEx	yes
RoundUp	yes
OutEx	yes
RoundUp	yes
OutEx	yes
OutEx	no
OutEx	no
OutEx	no
RoundUp	yes
RoundUp	yes
RoundUp	no
OutEx	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)**

## 54 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
24	24
25	22
23	27
25	21
23	20
25	23
25	13

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)**

## Statistisches Testen

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

- Kapitel 3 - Falsifikationsprinzip
- Kapitel 19 - Die Testentscheidung
- Kapitel 20 - Die Testtheorie

## 55 Aufgabe

(6 Punkte)



1. Erklären Sie den Zusammenhang zwischen Stichprobe und Grundgesamtheit an einem Schaubild! **(3 Punkte)**
2. Was ist der Unterschied zwischen  $\mu$  und  $\sigma$  und  $\bar{y}$  und  $s$  im Kontext der Stichprobe und Grundgesamtheit? **(2 Punkte)**
3. Warum müssen wir überhaupt zwischen einer Stichprobe und einer Grundgesamtheit unterscheiden? **(1 Punkt)**

## 56 Aufgabe

(8 Punkte)



Geben ist folgende 2x2 Kreuztabelle.


1. Tragen Sie folgende Fachbegriffe korrekt in die 2x2 Kreuztabelle ein! **(4 Punkte)**

- (Unbekannte) Wahrheit
- $H_0$  wahr
- $H_0$  falsch
- $H_0$  abgelehnt
- $H_0$  beibehalten
- Testentscheidung
- $\alpha$ -Fehler
- $\beta$ -Fehler
- Richtige Entscheidung
- 5%
- 20%

2. In der Analogie des Feuermelders, wie lautet der  $\alpha$ -Fehler? **(1 Punkt)**

3. In der Analogie des Feuermelders, wie lautet der  $\beta$ -Fehler? **(1 Punkt)**

4. Wenn der Feuermelder einmal pro Tag messen würde, wie oft würde der Feuermelder mit einem  $\alpha$  von 5% in einem Jahr Alarm schlagen? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 57 Aufgabe

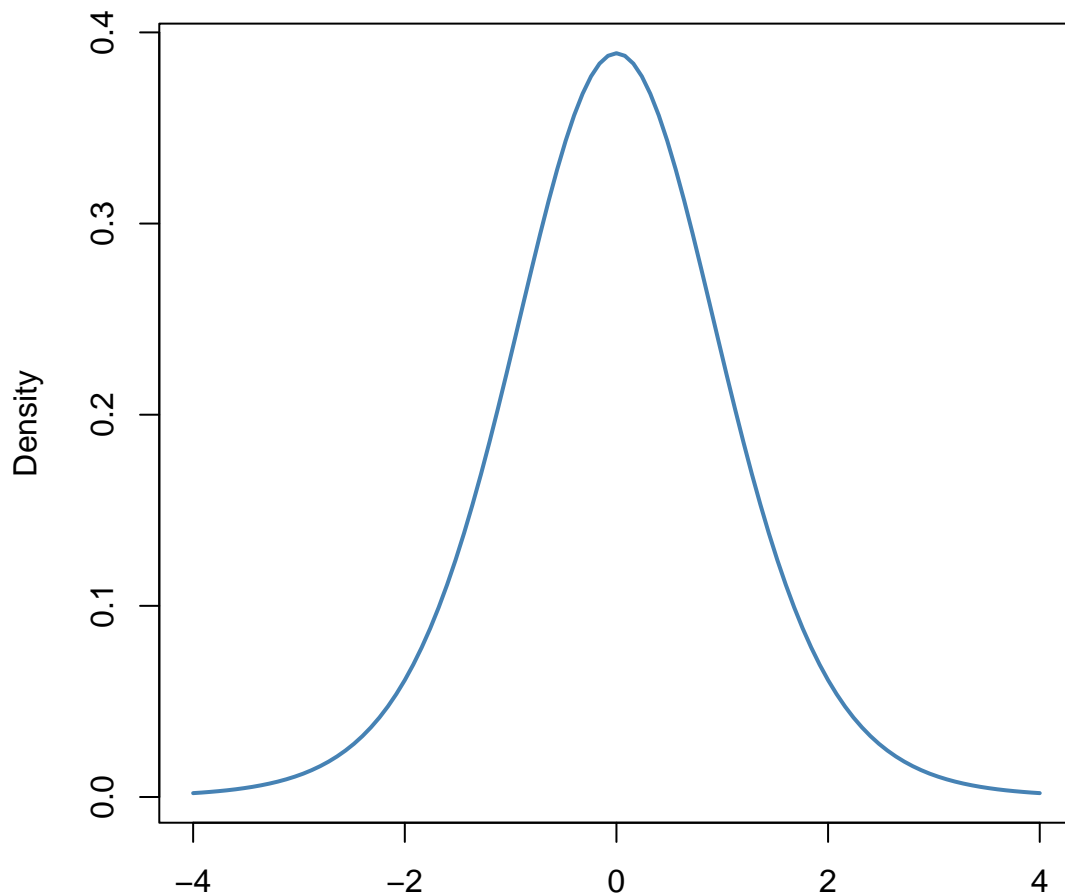
(8 Punkte)



Im folgenden ist eine t-Verteilung abgebildet. Ergänzen Sie die Abbildung wie folgt.

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

**t Distribution**



## 58 Aufgabe

(10 Punkte)



Sie rechnen einen t-Test für Gruppenvergleiche. Sie schätzen den Unterschied zwischen dem mittleren Befall mit Parasiten.

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



## 59 Aufgabe

(6 Punkte)



Gegeben ist die vereinfachte Formel für den Zweistichproben t-Test mit der gepoolten Standardabweichung  $s_p$  und gleicher Gruppengrösse  $n_g$  der beiden Sample.

$$T = \frac{\bar{y}_1 - \bar{y}_2}{s_p \cdot \sqrt{\frac{2}{n_g}}}$$

Welche Auswirkung hat die Änderungen der jeweiligen statistischen Masszahl auf den T-Wert und damit auf die *vermutliche* Signifikanz? Füllen Sie hierzu die untenstehende Tabelle aus! **(6 Punkte)**

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





Sie haben folgende Aussage gegeben.

### **Bin ich im Sommer?**

1. Erklären Sie den Gedankengang der Testtheorie sowie des Falsifikationsprinzips an der Aussage! **(4 Punkte)**
2. Erklären Sie Ihre Entscheidung zu der Aussage! **(3 Punkte)**
3. Schätzen Sie den p-Wert zu der Aussage ab! **(1 Punkt)**

## Der t-Test

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

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

## 61 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	23
RoundUp	13
RoundUp	16
RoundUp	18
RoundUp	18
GoneEx	18
RoundUp	18
RoundUp	17
GoneEx	18
GoneEx	17
GoneEx	18
GoneEx	19

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 große wäre dann die Teststatistik  $T_{calc}$ ? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 62 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	17
ProGain	16
FatDown	19
FatDown	16
FatDown	18
ProGain	13
FatDown	16
ProGain	13
ProGain	17
FatDown	17
ProGain	16
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% Konfidentintervall unter der Verwendung von  $s_p$  und der gemittelten Fallzahl über die beiden Gruppen! **(3 Punkte)**
6. Nennen Sie statistischen Grund, warum Sie sich zwischen einem Student t-Test und einem Welch t-Test entscheiden müssen! **(1 Punkt)**

## 63 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	1	16
2	16	8
3	-5	17
4	14	10
5	14	10
6	9	13
7	12	12

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  $p$ -Werte aus Ihrem berechneten  $T_{calc}$  ab! Begründen Sie Ihre Antwort mit einer Skizze! **(2 Punkte)**

## 64 Aufgabe

(10 Punkte)



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

```
##
## Two Sample t-test
##
## data:  freshmatter by N
## t = 0.24167, df = 10, p-value = 0.8139
## alternative hypothesis: true difference in means between group high and group low is not equal to 0
## 95 percent confidence interval:
##  -5.401576  6.715861
## sample estimates:
## mean in group high  mean in group low
##           16.85714           16.20000
```

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 eine Abbildung in der Sie  $T_{calc}$ ,  $Pr(D|H_0)$ ,  $A = 0.95$ , sowie  $T_{\alpha=5\%} = |2.23|$  einzeichnen! **(4 Punkte)**
4. Beschriften Sie die Abbildung entsprechend! **(2 Punkte)**

## 65 Aufgabe

(8 Punkte)



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

```
##  
## Two Sample t-test  
##  
## data:  drymatter by Fe  
## t = 2.531, df = 12, p-value = 0.02637  
## alternative hypothesis: true difference in means between group high and group low is not equal to 0  
## 95 percent confidence interval:  
##  0.6559652  8.7726062  
## sample estimates:  
## mean in group mid mean in group trt1  
##          17.85714          13.14286
```

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 das sich ergebende 95% Konfidenzintervall! **(2 Punkte)**
4. Beschriften Sie die Abbildung und das 95% Konfidenzintervall entsprechend! **(2 Punkte)**

## 66 Aufgabe

(8 Punkte)



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

```
##
##  Welch t-Test
##
## data:  drymatter by Fe
## t = 1.7721, df = 12, p-value = 0.1018
## alternative hypothesis: true difference in means between group high and group low is not equal to 0
## 95 percent confidence interval:
##  -0.5712961  5.5490739
## sample estimates:
## mean in group high  mean in group low
##           18.88889           16.40000
```

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 Boxplots! Welche Annahmen an die Daten haben Sie getroffen? Begründen Sie Ihre Antwort! **(4 Punkte)**



## 67 Aufgabe

(8 Punkte)



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

```
##  
## Paired t-test  
##  
## data: waterintake by infusion  
## t = -2.0252, df = 4, p-value = 0.1128  
## alternative hypothesis: true mean difference is not equal to 0  
## 95 percent confidence interval:  
## -8.535517 1.335517  
## sample estimates:  
## mean difference  
## -3.6
```

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.

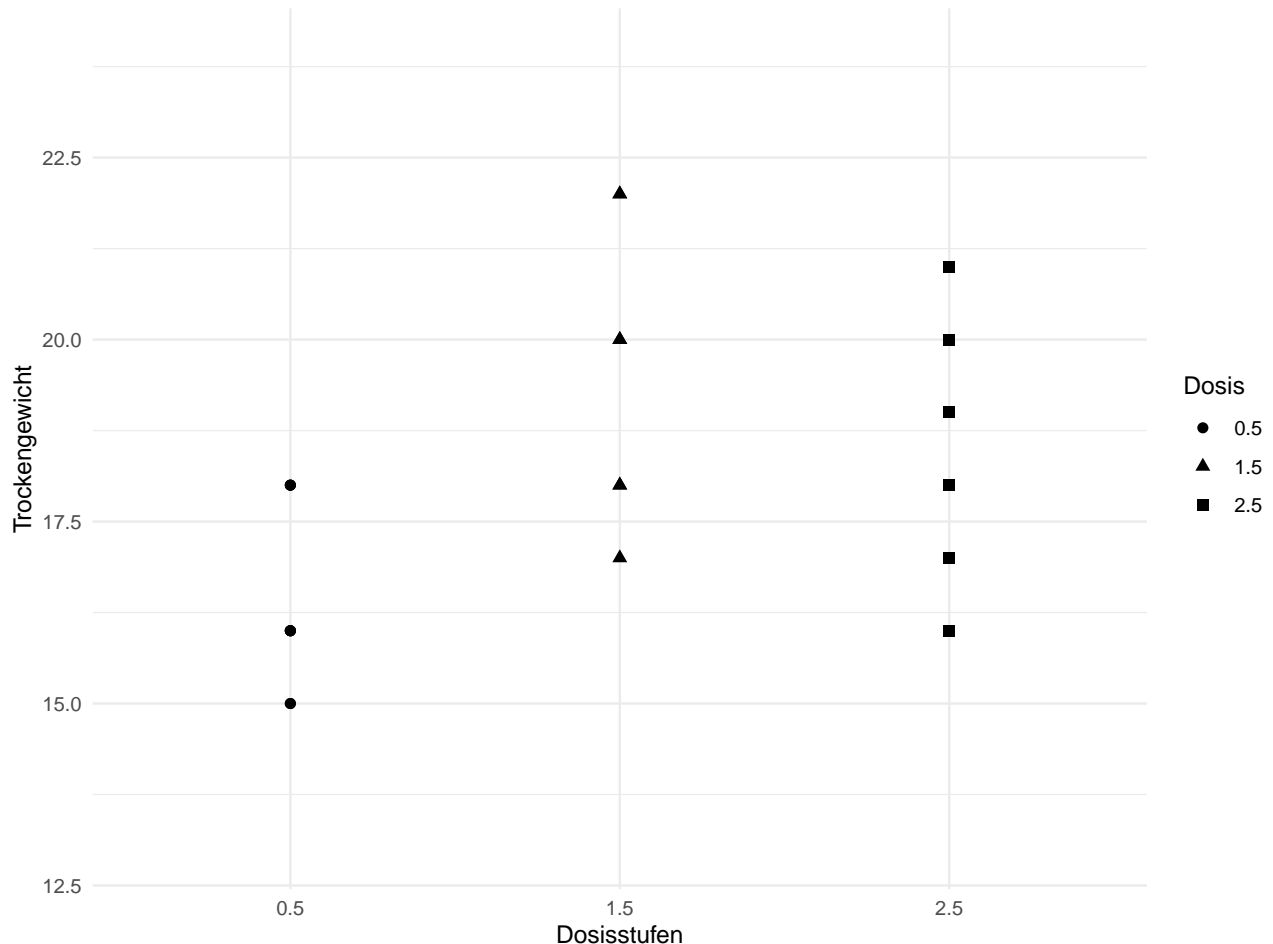
- [Kapitel 23 - Die ANOVA](#)

## 68 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)**

## 69 Aufgabe

(13 Punkte)



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  $F_{\alpha=5\%} = 3.44$  ab. Begründen Sie Ihre Antwort! **(2 Punkte)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>group</b>	2	158.42			
<b>Residuals</b>	22	42.54			

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	14.78	1.09
trt1	7	20.71	1.38
trt2	9	19.22	1.64

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

## 70 Aufgabe

(9 Punkte)



Der Datensatz *crop\_tbl* enthält das Trockengewicht der Maispflanzen (*drymatter*), die unter drei verschiedenen Düngerbedingungen erzielt wurden. Die Düngerbedingungen sind in dem Faktor *trt* mit den Faktorstufen *low*, *trt1* und *trt2* codiert. Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: drymatter
##           Df Sum Sq Mean Sq F value Pr(>F)
## trt         2  7.714   3.8571   0.9033  0.4228
## Residuals  18 76.857   4.2698
```

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)**

## 71 Aufgabe

(6 Punkte)



Sie haben ein Experiment mit drei Behandlungen (A, B und C) und vier Blöcken (I, II, III und IV) durchgeführt. Insgesamt haben Sie die Wuchshöhe von zwölf Sonnenblumen bestimmt. Im Folgenden sehen Sie die Wuchshöhen in [cm] aus dem Experiment.

##	[1]	134	123	86	95	97	97	143	156	144	166	145	157
----	-----	-----	-----	----	----	----	----	-----	-----	-----	-----	-----	-----

Erstellen Sie vier Zeichnungen des experimentellen Designs und beachten Sie folgende Angaben zu der Quelle der erklärten Varianz.

1. Ordnen Sie die Pflanzen so in den vier Blöcken und drei Behandlungen an,
  - (1) dass die Blöcke *kaum* Varianz erklären. **(1 Punkt)**
  - (2) dass die Blöcke *viel* Varianz erklären. **(1 Punkt)**
  - (3) dass die Behandlungen *kaum* Varianz erklären. **(1 Punkt)**
  - (4) dass die Behandlungen *viel* Varianz erklären. **(1 Punkt)**
2. Wenn Sie ein geplantes Experiment durchführen, wie viel Varianz soll dann von den Blöcken und den Behandlungen jeweils erklärt werden? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 72 Aufgabe

(12 Punkte)



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 5 Vitamin-C-Dosen *dose* über eine von 2 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 kritischen F-Werten mit  $F_{supp} = 4.08$  und  $F_{dose} = 2.61$  sowie  $F_{supp:dose} = 2.61$  ab. Begründen Sie Ihre Antwort! **(4 Punkte)**


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>supp</b>	1	32.17			
<b>dose</b>	4	297.16			
<b>supp:dose</b>	4	126.98			
<b>Residuals</b>	40	426.69			

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)**

## 73 Aufgabe

(8 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 50 Mastschweinen durchgeführt, wobei jedes Tier eine von drei Selen-Dosen *dose* (0.5 ng/Tag, 1 ng/Tag und 5 ng/Tag) über eine von zwei Verabreichungsmethoden *form* erhielt (Wasser oder Festnahrung). Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: gain
##      Df Sum Sq Mean Sq F value    Pr(>F)
## dose      2   55.78    27.889    1.1908  0.32682
## form      1  123.92   123.916    5.2909  0.03361
## dose:form  2  269.17   134.584    5.7464  0.01175
## Residuals 18 421.57    23.421
```

1. Stellen Sie die statistische  $H_0$  und  $H_A$  Hypothese für die obige zweifaktorielle ANOVA für den Faktor *form* 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)**



## 74 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)**

## 75 Aufgabe

(6 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 keine und eine mittelere Interaktion zwischen den Faktoren  $f_1$  und  $f_2$ ! **(2 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? Benennen Sie zwei mögliche Vorgehen! **(2 Punkte)**

## 76 Aufgabe

(7 Punkte)



Sie rechnen eine einfaktorielle ANOVA mit einem Faktor  $f_1$  mit fünf 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 zweiten und dritten 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! **(2 Punkte)**
2. Visualisieren Sie die Berechnung von  $F_{calc}$  am obigen Beispiel! **(2 Punkte)**
3. Erklären Sie das Ergebnis der obigen einfaktoriellen ANOVA unter der Berücksichtigung der Annahmen an eine ANOVA! **(3 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](#)

## 77 Aufgabe

(10 Punkte)

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



Nach einem Experiment ergibt sich die folgende 2x2 Datentabelle mit einem Pestizid (ja/nein), dargestellt in den Zeilen. Im Weiteren mit dem infizierten Pflanzenstatus (ja/nein) in den Spalten. Insgesamt wurden  $n = 106$  Pflanzen untersucht.

	Erkrankt (ja)	Erkrankt (nein)	
Pestizid (ja)	24	21	
Pestizid (nein)	23	38	

1. Ergänzen Sie die Tabelle um die Randsummen! **(1 Punkt)**
2. Formulieren Sie die Fragestellung! **(1 Punkt)**
3. Formulieren Sie das Hypothesenpaar! **(2 Punkte)**
4. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test auf der 2x2 Tafel. Geben Sie Formeln und Rechenweg mit an! **(4 Punkte)**
5. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem  $\chi^2_{\alpha=5\%} = 3.841$ ! **(1 Punkt)**
6. Skizzieren Sie eine 2x2 Tabelle mit  $n = 30$  Pflanzen in dem *vermutlich* die Nullhypothese nicht abgelehnt werden kann! **(1 Punkt)**

## 78 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  $n = 140$  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)			68
Medikament (nein)			72
	72	68	140

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)**

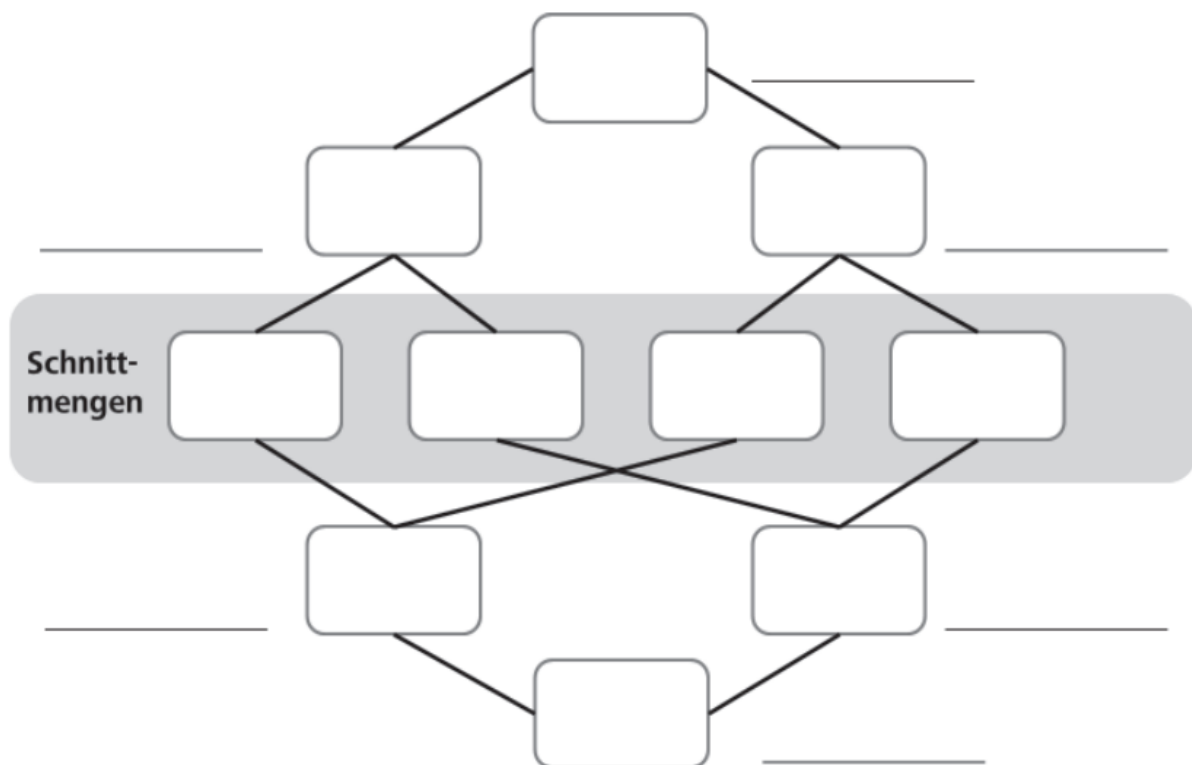
## 79 Aufgabe

(11 Punkte)



Die Prävalenz von Klauenseuche bei Wollschweinen wird mit 3% angenommen. In 80% der Fälle ist ein Test positiv, wenn das Wollschwein erkrankt ist. In 7.5% der Fälle ist ein Test positiv, wenn das Wollschwein *nicht* erkrankt ist und somit gesund ist. Sie werten 2000 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)**



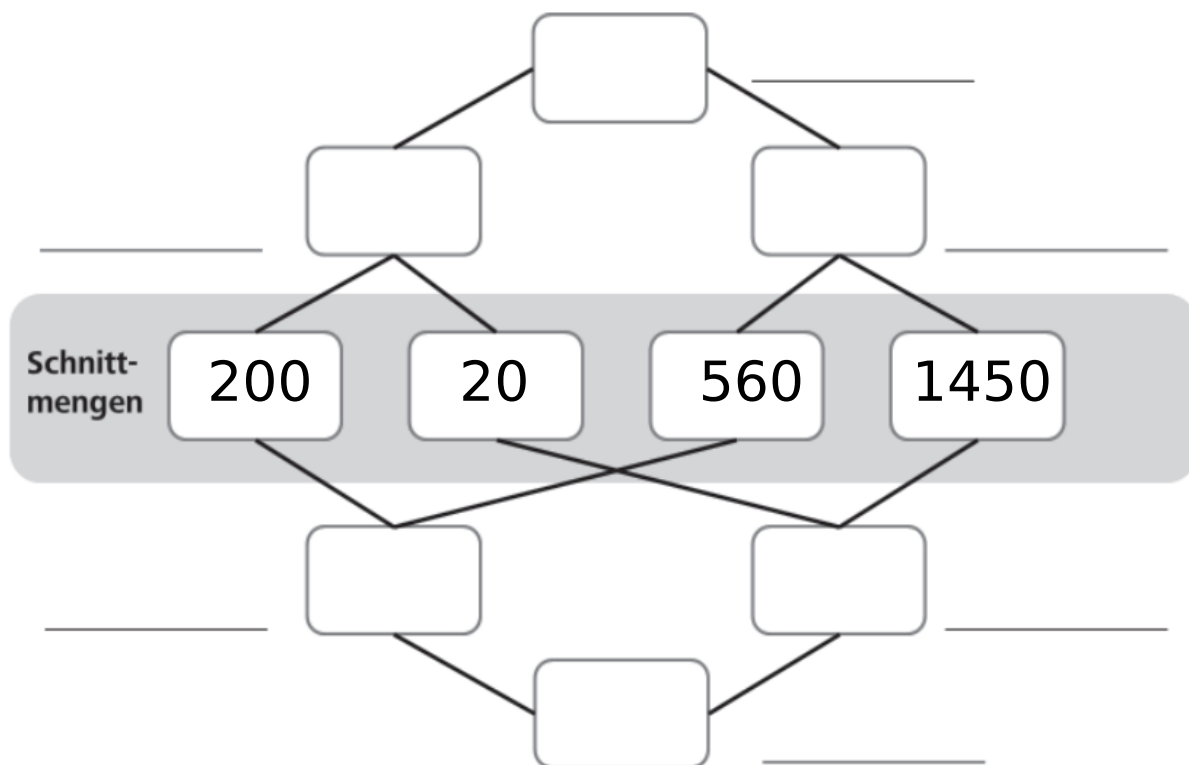
## 80 Aufgabe

(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)**





## 81 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)**

## Simple linear Regression

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](#)

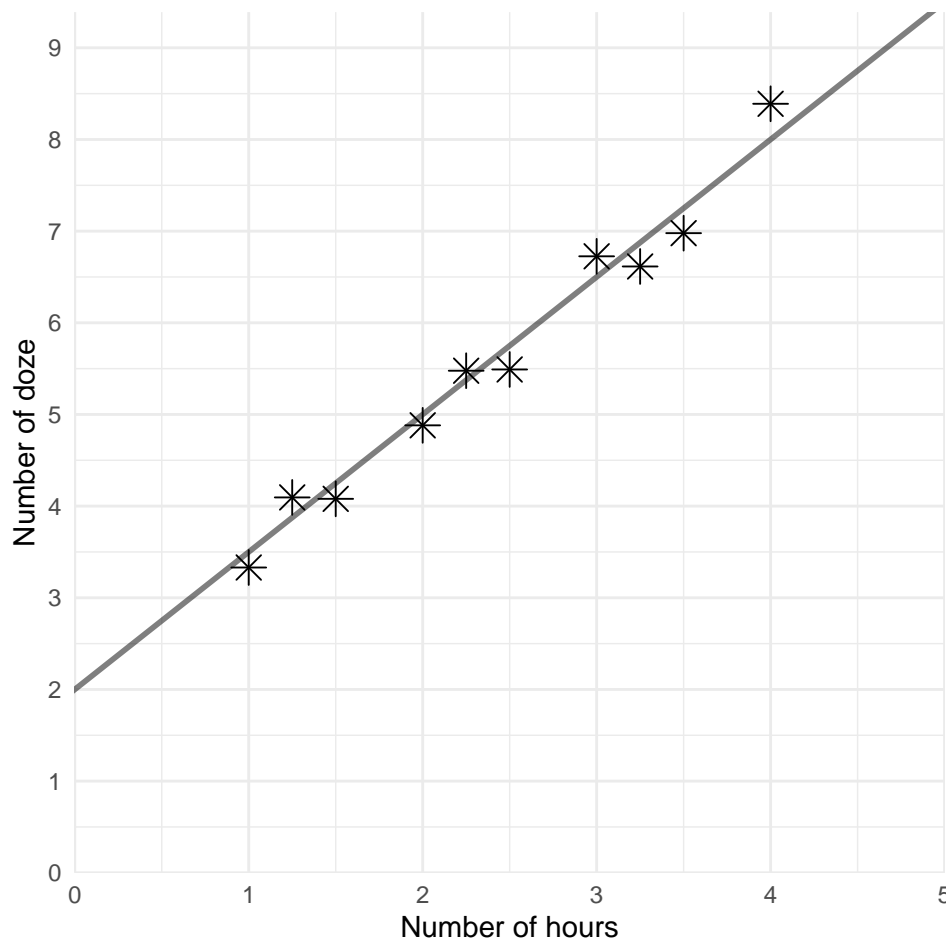
## 82 Aufgabe

(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 Gerade 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 Gerade aussehen? **(1 Punkt)**

## 83 Aufgabe

(10 Punkte)

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



In einem Stallexperiment mit  $n = 40$  Ferkeln wurde der Gewichtszuwachs unter bestimmten Lichtverhältnissen gemessen. Sie erhalten den  Output der Funktion `tidy()` einer simplen Gaussian linearen Regression sieben Wochen nach der ersten Messung.

term	estimate	std.error	t statistic	p-value
(Intercept)	26.38	2.31		
light	1.41	0.23		

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 Gerade 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)**


## 84 Aufgabe

(6 Punkte)



Sie erhalten folgende R Ausgabe der Funktion `lm()` nach einem Experiment mit zwei Behandlungen (A und B) sowie dem Ertragsgewicht von Weizen.

```
##  
## Call:  
## lm(formula = weight ~ trt, data = data_tbl)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -5.4000 -0.8571  0.1429  1.2571  2.6000   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   16.4000     1.0204   16.073 1.8e-08      
## trtB           -0.5429     1.3360   -0.406  0.693      
##  
## Residual standard error: 2.282 on 10 degrees of freedom  
## Multiple R-squared:  0.01624, Adjusted R-squared:  -0.08213   
## F-statistic: 0.1651 on 1 and 10 DF,  p-value: 0.6931
```

1. Ist die Annahme der Normalverteilung an das Outcome *rsp* erfüllt? Begründen Sie die Antwort! **(2 Punkte)**
2. Wie groß ist der Effekt des *Trt*? Liegt ein signifikanter Effekt vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Schreiben Sie das Ergebnis der  Ausgabe in einen Satz nieder, der die Information zum Effekt und der Signifikanz enthält! **(2 Punkte)**

## 85 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



## 86 Aufgabe

(9 Punkte)

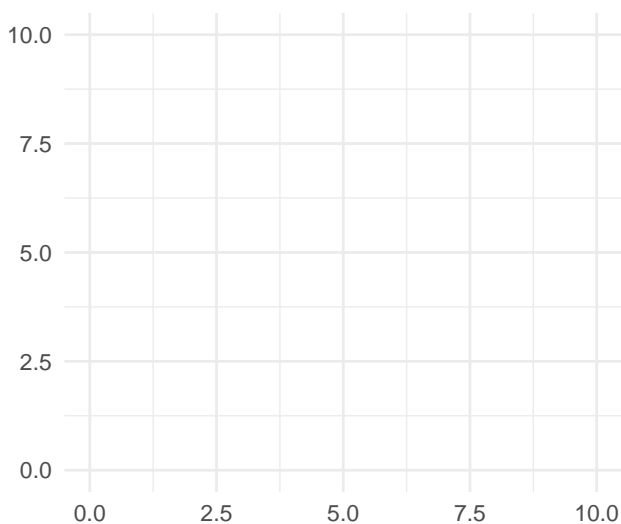


Im folgenden sehen Sie drei leere Scatterplots. Füllen Sie diese Scatterplots nach folgenden Anweisungen.

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)**

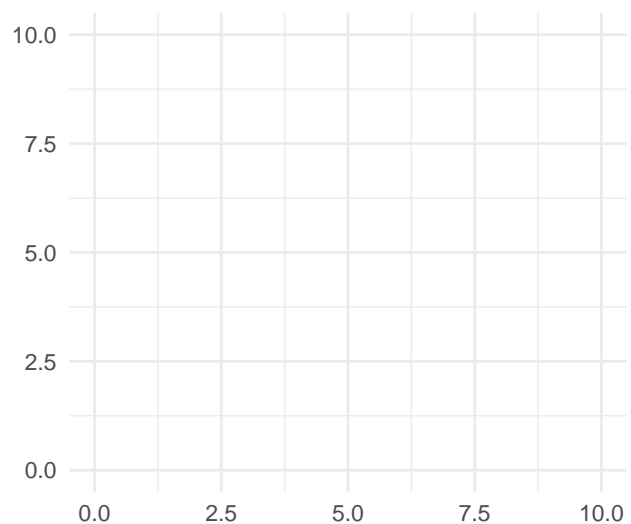
Pearsons  $\rho = 0$

$R^2 = 0.25$



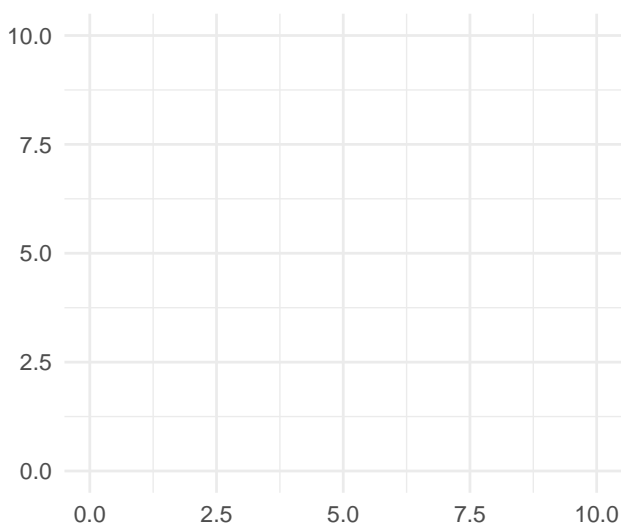
Pearsons  $\rho = 0.75$

$R^2 = 0.75$



Pearsons  $\rho = -0.5$

$R^2 = 0.5$



## 87 Aufgabe

(9 Punkte)

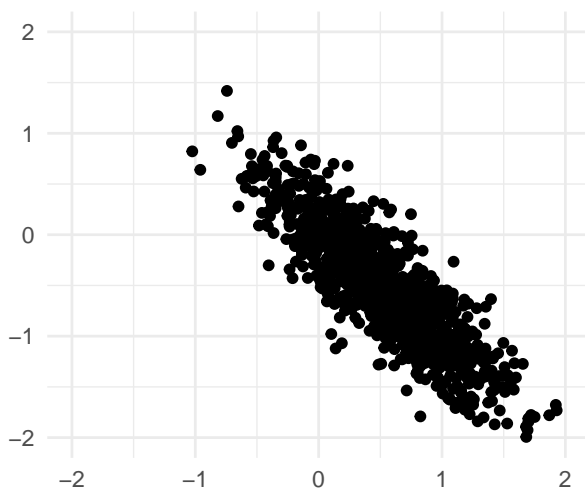


Im folgenden sehen Sie vier Scatterplots. Ergänzen Sie die Überschriften der jeweiligen Scatterplots.

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)**

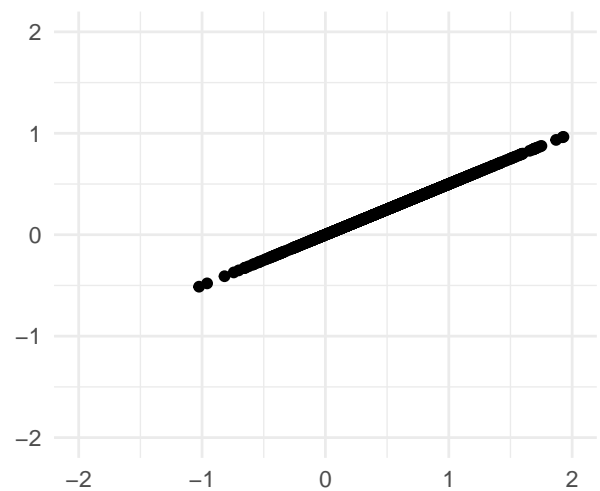
Pearsons  $\rho =$

$R^2 =$



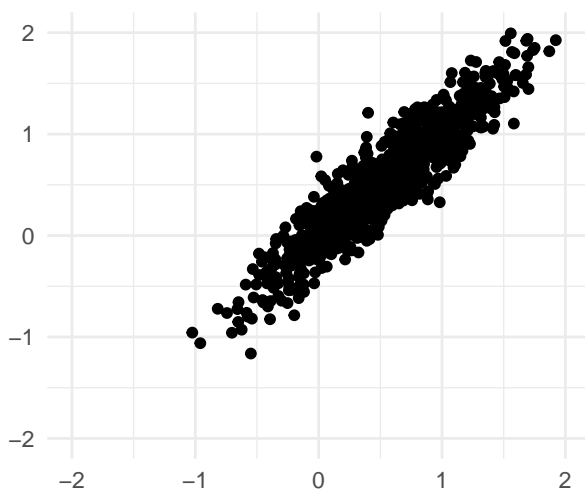
Pearsons  $\rho =$

$R^2 =$



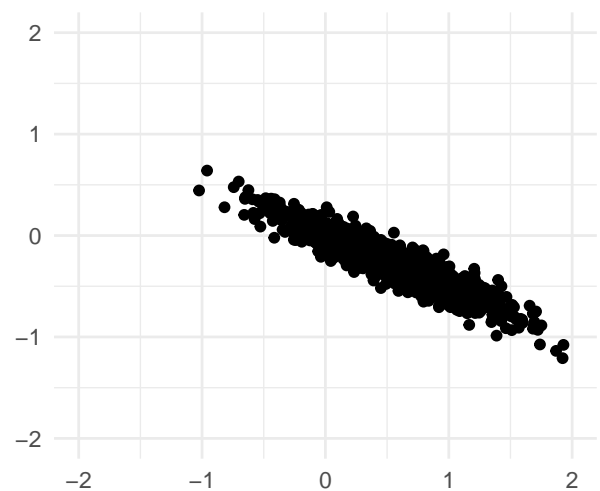
Pearsons  $\rho =$

$R^2 =$



Pearsons  $\rho =$

$R^2 =$





## 88 Aufgabe

(12 Punkte)



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

```
##  
## Kendall's correlation  
##  
## data:  drymatter and water  
## t = -0.45905, df = 8, p-value = 0.6584  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.7174596  0.5220863  
## sample estimates:  
##          cor  
## -0.1602012
```

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` and `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 die Teststatistik und den p-Wert! Beschriften Sie die Abbildung! **(3 Punkte)**

## 89 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 3 outlier fulfilling the normality assumption.



Residual plot violating the normality assumption.



## 90 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.



## 91 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	26.4	10.7	26.7	
2	29.4	12.4	29.4	
3	35.0	16.6	35.8	
4	19.2	6.4	20.1	
5	16.1	4.6	17.3	
6	17.9	5.3	18.3	
7	37.4	17.7	37.5	
8	19.0	4.9	17.6	
9	36.3	17.4	37.1	
10	31.2	12.5	29.5	
11	31.0	12.7	29.7	

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)**

## Multiple lineare Regression

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

- Kapitel 35 - Multiple lineare Regression
- Kapitel 40 - Gaussian Regression
- Kapitel 41 - Poisson Regression
- Kapitel 43 - Logistische Regression
- Kapitel 44 - Lineare gemischte Modelle

## 92 Aufgabe

(12 Punkte)



1. Zeichnen Sie in die drei untenstehenden, leeren Abbildungen die Zeile des Regressionskreuzes der Binomialverteilung. 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)**



The image shows three identical empty grid areas, each consisting of 10 columns and 10 rows of light gray lines, intended for the student to draw and write their answers to the questions above.

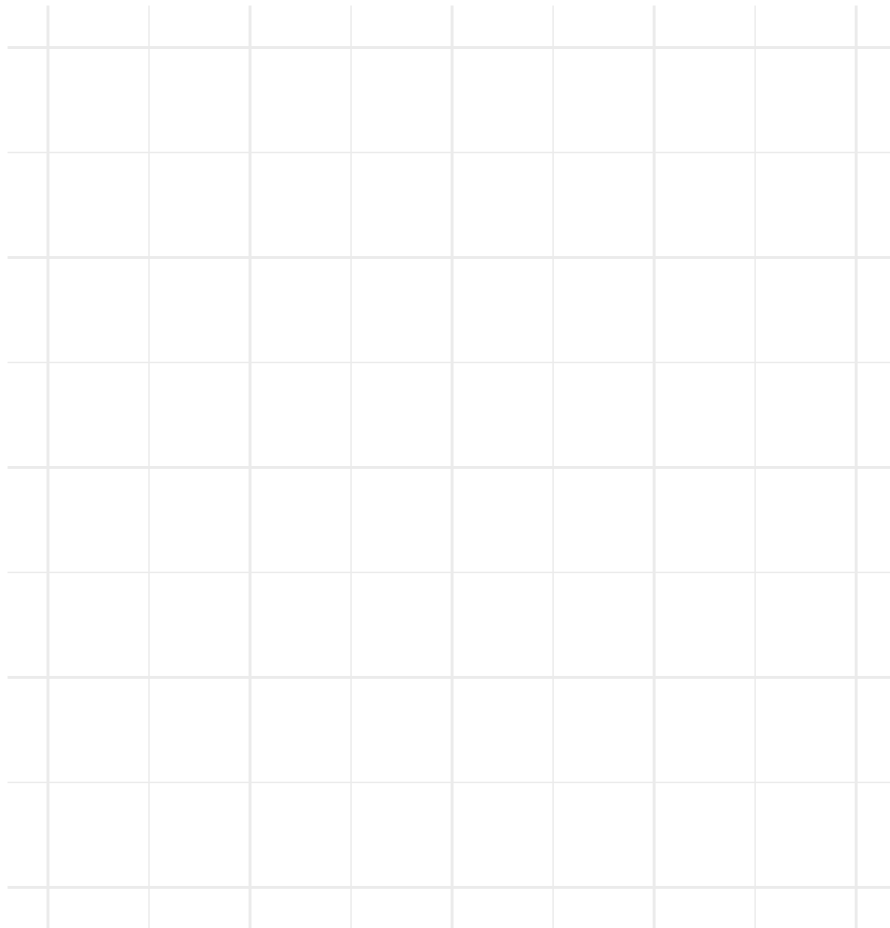
## 93 Aufgabe

(9 Punkte)



Ein Feldexperiment wurde mit  $n = 200$  Pflanzen durchgeführt. Folgende Einflussvariablen ( $x$ ) wurden erhoben: block, S und rainfall. 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 *Gaussian* 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 *starke* 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)**




## 94 Aufgabe

(9 Punkte)



Maispflanzen sollen auf die ertragssteigernde Wirkung von verschiedenen Einflussfaktoren untersucht werden. Gemessen wurde als Outcome die Trockenmasse in  $\text{kg/m}^2$ . Dafür wurde für jede Maispflanze gemessen wieviel Wasser ( $\text{l/m}^2$ ) die Pflanze erhalten hat oder ob die Pflanze ein neuartiges Lichtregime (0 = alt, 1 = neu) erhalten hatte. Zusätzlich wurde die Anzahl an Nematoden im Boden bestimmt sowie der Eisen- und Phosphorgehalt ( $\mu\text{g/kg}$ ) des Bodens. Es ergibt sich folgender Auszug aus den Daten.

water	light	P	Fe	drymatter	nematodes
10.14	0	11.46	106.09	74.66	16
9.36	0	10.77	96.48	66.41	10
9.26	1	12.85	96.77	68.20	11
10.72	1	11.97	103.93	70.41	9

Sie rechnen nun eine Gaussian lineare Regression auf den Daten und erhalten folgenden  Output.

```
##
## Call:
## lm(formula = reformulate(response = "drymatter", termlabels = wanted_vec),
##     data = data_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7533 -2.1480  0.2115  1.6842  7.4021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.00026    8.71947  -0.344   0.731
## water       -0.33538    0.20855  -1.608   0.111
## Fe           0.75556    0.08428   8.965 1.47e-14
## nematodes    0.04640    0.09058   0.512   0.610
##
## Residual standard error: 2.744 on 103 degrees of freedom
## Multiple R-squared:  0.4539, Adjusted R-squared:  0.438
## F-statistic: 28.54 on 3 and 103 DF, p-value: 1.633e-13
```

1. Sind die Residuals approximativ Normalverteilt? Begründen Sie Ihre Antwort! **(3 Punkte)**
2. Welche der Einflussfaktoren sind signifikant? Begründen Sie Ihre Antwort! **(3 Punkte)**
3. Interpretieren Sie die Spalte *estimate* im Bezug auf den Ertrag in Trockenmasse der Maispflanzen! **(3 Punkte)**




## 95 Aufgabe

(10 Punkte)




In verschiedenen Flüssen (*stream*) wurde die Anzahl an Knochenhechten (*longnose*) gezählt. Daneben wurden noch andere Eigenschaften der entsprechenden Flüsse gemessen. Es ergibt sich folgender Auszug aus den Daten.

stream	longnose	so4	maxdepth	acerage	no3
BUSH_CR	32	11.31	120	19046	2.93
MILL_CR	6	15.77	53	1097	1.71
NORTH_BR	23	7.36	121	3836	1.32

Sie rechnen nun eine Poisson lineare Regression auf den Daten und erhalten folgenden  Output.

```
##
## Call:
## glm(formula = reformulate(response = "longnose", termlabels = wanted_vec),
##      family = quasipoisson, data = data_tbl)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -9.377  -4.494  -2.234   1.413  15.689
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.149e+00  5.401e-01   3.979 0.000228
## so4          -2.423e-02  2.427e-02  -0.998 0.322945
## maxdepth     1.122e-02  4.155e-03   2.701 0.009465
## acerage      1.380e-05  1.737e-05   0.795 0.430659
## no3          2.450e-01  7.071e-02   3.465 0.001112
##
## (Dispersion parameter for quasipoisson family taken to be 35.09122)
##
##      Null deviance: 2248.1  on 53  degrees of freedom
## Residual deviance: 1521.5  on 49  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

1. Erklären Sie warum eine Quasipoisson-Regression gerechnet wurde! **(2 Punkte)**
2. Erklären Sie den Effekt der alternativen Verwendung einer Poisson-Regression auf den obigen  Output! **(2 Punkte)**
3. Können Sie die *Estimate* der einzelnen Einflussvariablen direkt interpretieren? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Interpretieren Sie den Effekt von *so4* auf die Anzahl an Knochenhechten! Liegt ein signifikanter Effekt vor? Begründen Sie Ihre Antwort! **(4 Punkte)**


## 96 Aufgabe

(10 Punkte)




Auf einer Erdbeerplantage treten unerwartet häufig infizierte Erdbeerpflanzen auf. In einem Versuch sollen verschiedene Einflussfaktoren auf den Infektionsstatus betrachtet werden. Dafür wurde für jede Erdbeerpflanze gemessen, wieviel Wasser die Pflanze erhalten hat oder ob die Pflanze ein neuartiges Lichtregime erhalten hatte. Zusätzlich wurde die Anzahl an Nematoden im Boden bestimmt. Es ergibt sich folgender Auszug aus den Daten.

	infected	water	light	nematodes
1		9.72	0	3
0		9.97	0	1
0		10.43	0	3
0		10.83	1	3

Sie rechnen nun eine logistische lineare Regression auf den Daten und erhalten folgenden  Output.

```
## # A tibble: 3 x 4
##   term          std.error statistic p.value
##   <chr>          <dbl>      <dbl>   <dbl>
## 1 (Intercept)    0.394      0.250   0.803
## 2 light          0.467      0.889   0.374
## 3 nematodes      0.147     -0.429   0.668
```

1. Die Spalte *estimate* wurde gelöscht. Berechnen Sie die Werte der Spalte *estimate* aus den  Output! **(2 Punkte)**
2. Welche Einflussfaktoren sind protektiv, welche ein Risiko? Berechnen Sie hierfür zunächst das OR aus der Spalte *estimate*! **(4 Punkte)**
3. Interpretieren Sie die Spalte *estimate* im Bezug auf den Infektionsstatus der Erdbeerpflanzen! **(2 Punkte)**
4. Was ist der Unterschied zwischen einem OR und einem RR? Geben Sie ein numerisches Beispiel! **(2 Punkte)**


## 97 Aufgabe

(11 Punkte)



In einem Experiment zur Steigerung der Milchleistung (*gain*) in *dl/h* von Kühen wurden zwei Arten von Musik in den Ställen gespielt. Zum einen ruhige Musik (*calm*) und eher flotte Musik (*pop*). Die Messungen wurden an jeder Kuh (*subject*) wiederholt durchgeführt. Darüber hinaus wurden verschiedene Ställe (*barn*) mit der Musik bespielt.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: gain ~ attitude + (1 | subject) + (1 | barn)
## Data: data_tbl
##
## REML criterion at convergence: 794.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.2084 -0.6583 -0.1083  0.6299  3.3271
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## barn     (Intercept)         195.0    13.96
## subject  (Intercept)      4120.8    64.19
## Residual                    656.5    25.62
## Number of obs: 83, groups: barn, 7; subject, 6
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    202.66      27.02    7.500
## attitudepop    -21.30       5.63   -3.783
##
## Correlation of Fixed Effects:
##              (Intr)
## attitudepop -0.103
```

1. Ist die Annahme der Normalverteilung an das Outcome *gain* erfüllt? Begründen Sie Ihre Antwort! **(2 Punkte)**
2. Wie groß ist der Effekt der Musikart *attitude*? Liegt ein signifikanter Effekt vor? Schätzen Sie den p-Wert mit einem kritischen t-Wert von  $T_k = 1.96$  ab. Begründen und visualisieren Sie Ihre Antwort und Entscheidung! **(3 Punkte)**
3. Was ist der Unterschied zwischen einem „random“ und „fixed“ Effekt. Gehen Sie in der Begründung Ihrer Antwort auf dieses konkrete Beispiel ein! **(3 Punkte)**
4. Wie groß ist die Varianz, die durch die zufälligen Effekte erklärt wird? **(1 Punkt)**
5. Schreiben Sie das Ergebnis der  Ausgabe in einen Satz nieder, der die Information zum Effekt und der Signifikanz enthält! **(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](#)

## 98 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
12	10					
10	13					
12	10					
9	10					
9	13					
9	15					
8	11					
11	11					
9	10					
7	13					
12	11					
10	9					
11	12					
8	14					
14	13					

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

$$\text{Wert } \mu_W = \frac{n!_0 \cdot (n!_0 + 1)}{4}! \quad \textbf{(2 Punkte)}$$

3. Berechnen Sie anschließend den z-Wert mit  $z = \frac{W - \mu_W}{15.93}$ ! **(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)**

## 99 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	6.4
Kontrolle	8.1
Kontrolle	9.3
RootsGoneX	15.2
Kontrolle	8.5
RootsGoneX	15.2
Kontrolle	8.3
RootsGoneX	14.9
RootsGoneX	14.1
Kontrolle	9.7
Kontrolle	9.2

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

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

- 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)

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

## 100 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
Dosis 40	9.5			
Dosis 40	8.6			
Kontrolle	9.1			
Dosis 40	10.2			
Kontrolle	8.9			
Dosis 40	8.6			
Dosis 20	9.9			
Dosis 40	8.7			
Kontrolle	7.3			
Kontrolle	4.5			
Kontrolle	8.5			
Kontrolle	10.4			
Kontrolle	7.7			
Dosis 20	11.9			
Dosis 20	8.6			
Dosis 20	9.2			

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)**

## 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](#)



## 101 Aufgabe

(8 Punkte)



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


Vergleich	Raw p-val	Adjusted p-val	Reject $H_0$
dose 10 - ctrl	0.760		
dose 15 - ctrl	0.012		
dose 20 - ctrl	0.001		
dose 40 - ctrl	0.060		

1. Füllen Sie die Spalte „adjustierte p-Werte“ 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)**

## 102 Aufgabe

(9 Punkte)



In einem Experiment 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 0.0309697 0.0000000 0.0000073
## high 0.0309697 1.0000000 0.0000003 0.0041507
## low  0.0000000 0.0000003 1.0000000 0.0023112
## mid  0.0000073 0.0041507 0.0023112 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	5.16	2.78
high	9	7.56	2.09
low	9	14.36	1.87
mid	9	10.84	2.18

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! **(2 Punkte)**
4. Interpretieren Sie das *Compact letter display (CLD)*! **(2 Punkte)**

## 103 Aufgabe

(8 Punkte)



In einem Experiment 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
## "ab"    "a"    "c"    "bc"    "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)**

## R Programmierung

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

- **Kapitel 9ff. Programmieren in R**

In der Klausur zu dem Modul **Mathematik & Statistik** wird *eine Aufgabe* aus den folgenden Aufgaben zur R Programmierung ausgewählt.

In der Klausur zu dem Modul **Statistik** wird *eine Aufgabe* aus den folgenden Aufgaben zur R Programmierung ausgewählt.

In der Klausur zu dem Modul **Angewandte Statistik für Bioverfahrenstechnik** wird *eine Aufgabe* aus den folgenden Aufgaben zur R Programmierung ausgewählt.

In der Klausur zu dem Modul **Angewandte Statistik und Versuchswesen** wird *eine Aufgabe* aus den folgenden Aufgaben zur R Programmierung ausgewählt.

In der Klausur zu dem Modul **Biostatistik** wird *eine Aufgabe* aus den folgenden Aufgaben zur R Programmierung ausgewählt.

## 104 Aufgabe

(8 Punkte)



Bearbeiten Sie folgenden Aufgaben mit Bezug zu  !

1. Erklären Sie den Pipe-Operator am Beispiel der Berechnung des Mittelwertes mit der Funktion `mean` und den Zahlen 14, 15, 8, 12 und 10! **(2 Punkte)**
2. Erklären Sie den Unterschied zwischen einer Funktion und einem Objekt in R an einem Beispiel! **(2 Punkte)**
3. Erklären Sie den Vorteil der Verwendung der Funktion `p_load()` an einem Beispiel. Was ist das alternative Vorgehen zu der Verwendung der Funktion? **(2 Punkte)**
4. Erklären Sie die Verwendung des Operators `::` am Beispiel der Funktion `select()` und `p_load()`! **(2 Punkte)**

## 105 Aufgabe

(8 Punkte)



Bearbeiten Sie folgenden Aufgaben mit Bezug zu  !

1. Erklären Sie den Pfeil-Operator am Beispiel eines Zahlenvektors mit der Funktion `c()` und den Zahlen 8, 5, 13, 8 und 14! **(2 Punkte)**
2. Erklären Sie den Nutzen des R Paketes `conflicted` am Beispiel der Funktion `select()`! **(2 Punkte)**
3. Erklären Sie den Unterschied zwischen einer `library` und einem `package` in R an einem Beispiel! **(2 Punkte)**
4. Erklären Sie den Unterschied zwischen `"mean"`, `mean` und `mean()`! **(2 Punkte)**

## 106 Aufgabe

(10 Punkte)



Sie wollen eine explorative Datenanalyse auf dem folgenden, in R schon geladenen, Datensatz `leaf_tbl` durchführen.

```
leaf_tbl

## # A tibble: 10 x 3
##   treatment block  leaf
##   <dbl> <int> <dbl>
## 1       1     1     9
## 2       1     2     9
## 3       1     3     9
## 4       1     4    13
## 5       1     5    12
## 6       2     1    13
## 7       2     2    10
## 8       2     3     4
## 9       2     4    13
## 10      2     5    10
```

1. Welche R Pakete benötigen Sie für die explorative Datenanalyse? **(2 Punkte)**
2. Skizzieren Sie den R Code für die Erstellung eines Barplots unter der Verwendung des Pipe-Operators! **(4 Punkte)**
3. Nehmen Sie an, dass Sie die Funktion `as_factor()` verwenden. Wozu benötigen Sie die Funktion? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Erläutern Sie Ihr weiteres Vorgehen nachdem Sie eine explorative Datenanalyse durchgeführt haben! **(2 Punkte)**

## 107 Aufgabe

(8 Punkte)



Sie wollen einen multiplen, paarweisen Gruppenvergleich auf dem folgenden, in R schon geladenen, Datensatz `leaf_tbl` durchführen.

```
leaf_tbl

## # A tibble: 25 x 3
##   treatment block  leaf
##   <fct>      <fct> <dbl>
## 1 1      1      1      12
## 2 1      2      2      11
## 3 1      3      3       9
## 4 1      4      4       7
## 5 1      5      5      10
## 6 2      1      1       8
## 7 2      2      2      12
## 8 2      3      3      12
## 9 2      4      4      13
## 10 2     5      5      11
## # ... with 15 more rows
```

1. Welche R Pakete benötigen Sie für den multiplen Vergleich? **(2 Punkte)**
2. Skizzieren Sie den R Code für die Erstellung einer Berechnung eines multiplen Vergleiches unter der Verwendung des Pipe-Operators! Nutzen Sie hierfür folgende Funktionen in der passenden Reihenfolge: `emmeans()`, `cld()`, `lm()`, `anova()`, `ggplot()`! **(4 Punkte)**
3. Erklären Sie den Unterschied zwischen der Funktion `contrast()` und `cld()`! **(2 Punkte)**



## 108 Aufgabe

(9 Punkte)



Sie wollen einen Datensatz aus Excel in R laden. In Ihrem Experiment haben Sie die Sorten eins bis vier sowie die Blöcke I bis IV vorliegen. Sie messen die Outcomes freshmatter, Fe und drymatter an drei verschiedenen Messterminen.

1. Welches R Paket benötigen Sie für das Einlesen einer Excel Datei? **(1 Punkt)**
2. Skizzieren Sie den sich ergebenden Datensatz als Datentabelle im Long-Format, so dass Sie die Daten erfolgreich R laden können! **(3 Punkte)**
3. Skizzieren Sie den R Code, den Sie benötigen um die Daten aus Excel in R zu laden! Nutzen Sie hierfür die Funktion `pivot_longer()` aus dem R Paket `tidyr`! **(3 Punkte)**
4. Skizzieren Sie die Anwendung der Funktion `mutate()`! Begründen Sie die Anwendung! **(2 Punkte)**

## Mathematik

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

In der Klausur zu dem Modul **Mathematik & Statistik** werden drei Aufgaben aus den folgenden Aufgaben zur Mathematik ausgewählt.

## 109 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  $0.8\text{mm}$  pro Jahr aus. Es ist bekannt, dass die Eiche im Jahr 2022 einen Umfang von  $11\text{m}$  in Bruthöhe hatte.

1. Wie groß war der Durchmesser der Eiche im Jahr 1805 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  $180\text{cm}$ , eine Breite von  $85\text{cm}$  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 *mühsam* 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  $20\text{cm}$  bricht die Eiche zusammen. Beantworten Sie die Forschungsfrage! Begründen Sie Ihre Antwort! **(2 Punkte)**



## 110 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 120 Maispflanzen 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 Maispflanzen einpflanzen und müssen dafür Substrat bestellen. Zum anderen müssen Sie die Maispflanzen auch bewegen und in ein Gewächshaus platzieren. Die Töpfe für die Keimung haben einen Durchmesser von 10cm und eine Höhe von 7cm. Der Kubikmeterpreis für Torf liegt bei 310 EUR.

1. Skizzieren Sie den Versuchsplan auf *vier* Tischen im Gewächshaus! **(2 Punkte)**
2. Berechnen Sie die benötigten Pflanztöpfe, wenn Sie Randpflanzen mit berücksichtigen wollen! **(1 Punkt)**
3. Welche Fläche in  $m^2$  gegeben der Anzahl an Pflanztöpfen benötigen Sie im Gewächshaus am Anfang der Keimungsphase? **(2 Punkte)**
4. Berechnen Sie die benötigte Menge an Torf, die Sie für das Befüllen der Pflanztöpfe benötigen! Gehen Sie von *einem Zylinder* für die Pflanztöpfe aus! **(2 Punkte)**
5. Berechnen Sie die Kosten in EUR für Ihre Torfbestellung! **(1 Punkt)**
6. Nach dem Befüllen der Pflanztöpfe haben Sie noch Torf übrig. Welche Menge an Torf hätten Sie benötigt, wenn Sie mit einem *halbierten Kegel* gerechnet hätten? **(2 Punkte)**



## 111 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 60 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 Skorbuterkrankung auf einem Schiff der Flotte dar! Beschriften Sie die Achsen! **(4 Punkte)**
2. Schätzen Sie die Überlebenswahrscheinlichkeit nach 90 Tagen aus Ihrer Abbildung ab! **(1 Punkt)**

Der Chronist an Bord der Trinidad, Antonio Pigafetta, schrieb in seinem Bericht „[...] Um nicht Hungers zu sterben, aßen wir das Leder, mit dem die große Rahe zum Schutz der Taue umwunden war.“ Insbesondere die Mannschaft der Concepción erlitt große Verluste durch die Skorbut bei der Überquerung des Pazifiks, da durch Erkundungsfahrten weniger Zeit blieb, um wilden Sellerie aufzunehmen. Wilder Sellerie enthält  $8000\mu\text{g}/100\text{g}$  Vitamin C. Der Bedarf liegt bei  $105\text{mg}$  pro Tag für Männer.

3. Berechnen Sie die notwendige Menge an aufzunehmenden wilden Sellerie auf die Concepción für die ununterbrochene Fahrt von drei Monate und 20 Tage über den Pazifik! **(3 Punkte)**
4. Skizzieren Sie die Überlebenszeitkurve für die Concepción im Vergleich zu der Überlebenszeitkurve der Trinidad! Beschriften Sie die Achsen! **(2 Punkte)**





## 112 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  $2 \times 10^{28}$  kg. Wenn die Sonne nun am Ende ihrer Lebenszeit zu einem schwarzen Loch mit dem Radius von 5000 m kollabiert, wird die Sonne 30% der aktuellen Masse verloren haben. Ein Lichtteilchen mit der Masse  $m_f$  und der Fluchtgeschwindigkeit  $v_f$  will dem schwarzen Loch entkommen. 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  $6.674 \cdot 10^{-11} \frac{\text{m}^3}{\text{kg} \cdot \text{s}^2}$

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  $2.9 \times 10^8$  m/s aus. Kann das Lichtteilchen der Gravitation des schwarzen Lochs entkommen? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Stellen Sie den Zusammenhang zwischen dem sich verringernden Radius  $r$  des schwarzen Lochs bei gleichbleibender Masse  $m_s$  und der notwendigen Fluchtgeschwindigkeit  $v_f$  in einer Abbildung dar! **(2 Punkte)**
6. Erklären Sie in diesem Zusammenhang den Begriff *Singularität*! **(2 Punkte)**



## 113 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 *fünf* Voyager 1 ähnliche Sonden mit der Geschwindigkeit von  $6.0523 \times 10^4 \text{ km/h}$  los um sich auf den erreichten Planeten selbst zu replizieren. Nach 750 Jahren ist die Replikation abgeschlossen und wiederum *fünf* 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 drei Schritte der Vervielfältigung der Sonden in der Galaxie! Beschriften Sie die Abbildung mit der Dauer und der Anzahl an Sonden für jeden Schritt der Vervielfältigung! **(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  $10^8$  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)**



## 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 Things* verloren haben, wollen Sie bei einer Ihrer Freundinnen 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* 3 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 zwölfseitigen Würfeln ( $2d12$ ) als Schaden oder das Schwert mit einem vierseitigen Würfel plus 7 ( $1d4+7$ ) als Schaden bevorzugen? Begründen Sie Ihre Antwort mathematisch! **(1 Punkt)**

Jetzt wird es immer wilder, da 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  $Pr(B) = 0.75$ . 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

(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 Schweinestall montieren lassen. Sie messen Ihren Stall und finden folgende Maße wieder. Die vordere Seite des Schweinestall hat eine Höhe  $h_v$  von  $6m$ . Die hintere Seite des Schweinestall hat eine Höhe  $h_b$  von  $8m$ . Der Schweinestall hat eine Tiefe  $t$  von  $12m$  und eine Breite  $b$  von  $30m$ .

1. Skizzieren Sie den Schweinestall auf dem die Solaranlage montiert werden soll! Ergänzen Sie die Angaben für die Höhen  $h_v$ ,  $h_b$ , die Tiefe  $t$  und die Breite  $b$  des Stalls! **(2 Punkte)**
2. Berechnen Sie die Fläche der schrägen, neuen Solaranlage auf dem Schweinestall! **(3 Punkte)**

Ebenfalls 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  $5t$  aushalten bevor der Tank wegbricht. Sie rechnen eine Sicherheitstoleranz von 20% ein beinhaltend das Gewicht des Methantanks. In flüssiger Form hat Methan bei  $-80^\circ C$  eine Dichte von  $220kg/m^3$ . Bei  $-100^\circ C$  hat Methan eine Dichte von  $270kg/m^3$ . Sie betreiben Ihre Anlage bei  $-90^\circ C$ .

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)**





## 116 Aufgabe

(10 Punkte)

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



**Pyramiden bauen** Es stehen die niedersächsischen 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 70 Grad im Vergleich zu den ägyptischen Pyramiden mit 53 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. Die Königspyramide von Meroe soll eine Seitenlänge der Grundfläche, von 38 Königsellen haben. Wie hoch wird die Königspyramide sein? **(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 2 Sklaven, die Ihnen bei dem Befüllen der Pyramide mit Schutt zu Seite stehen. Leider haben Ihre Sklaven zu allem Überfluss auch noch chronische Schulterschmerzen entwickelt, als sie von der anstehenden Aufgabe erfahren haben. Gehen Sie daher von einer Effizienz der Sklaven von 90% aus. In eine Schubkarre passen 110 Liter.

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

Bei der Besichtigung der Pyramide teilt Ihnen der leicht übergewichtige Pharaos (Nebenberuf *Versicherungsverteter*) mit, das die Pyramide zu flach sei und somit nicht in die niedersächsische 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  $7^\circ$  ändert! **(2 Punkte)**





## 117 Aufgabe

(12 Punkte)

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



**Geocaching – Von Satelliten und Plastikdosen** Es ist Wochenende und das Wetter ist sweet. Sie schwingen sich auf Ihr Cachermobil um mit 17km/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. Sie starten an dem Ort E.

Ort	Cache	Wertung (S T G)
A	GCH7Q5D	5.0   3.5   Klein
B	GCE9OZN	1.0   1.0   Klein
C	GCLZ2WT	4.0   4.0   Mikro
D	GC8BAEE	3.0   2.5   Normal
E	GC9JRAI	2.5   4.5   Klein

Im Weiteren sind Ihnen folgende Informationen zu den Entfernungen der Orte zugänglich. Der Entfernungsvektor  $\vec{AC}$  ist 3km. Im Weiteren ist Ihnen der Entfernungsvektor  $\vec{CB}$  mit 6km bekannt. Der Entfernungsvektor  $\vec{BE}$  ist das 1.2-fache des Entfernungsvektor  $\vec{CB}$ . Wenn Sie von dem Ort A den Ort C anpeilen, so liegt der Ort B ungefähr  $40^\circ$  nördlich. Wenn Sie von dem Ort C den Ort B anpeilen, so liegt der Ort D ungefähr  $50^\circ$  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 nördlich von B. Die Strecke zwischen A und E ist nicht passierbar.

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

1. An der höchsten Schwierigkeit müssen Sie angeln. Ihre Angel ist ausgefahren 7m lang. Erreichen Sie einen Cache in der Höhe von 8.7m? Berechnen Sie dazu Ihre maximale mögliche Angelhöhe! **(1 Punkt)**
2. Erstellen Sie eine Matrix  $M$  der Terrain- und Schwierigkeitswertungen der Caches! **(1 Punkt)**
3. Wie viele Wochenenden benötigen Sie, um die Matrix  $M$  mit mindestens einem gefundenen Cache zu füllen, wenn Sie pro Tag drei Caches schaffen? **(1 Punkt)**
4. Welche Strecke in km legen Sie bei der Bewältigung der abwärts Schwierigkeitschallenge zurück? **(6 Punkte)**
5. Gehen Sie von einer zusätzlichen Suchzeit in Stunden für die Caches an den jeweiligen Orten zur reinen Reisezeit mit Ihrem Cachermobil aus. Die Suchzeit in Stunden für *jeden einzelnen* Cache wird durch die Funktion

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

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





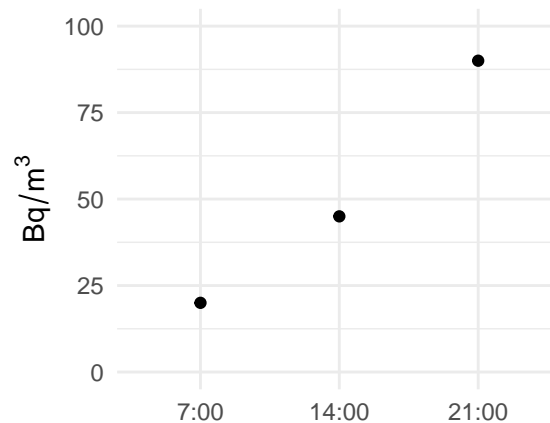
## 118 Aufgabe

(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 21:00 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  $300 Bq/m^3$  in Ihrem ungelüfteten Kellerraum erreicht haben? **(2 Punkte)**

Radon zerfällt mit einer Halbwertszeit von 2.8d zu Polonium. Polonium wiederum zerfällt mit einer Halbwertszeit von 143d zu Blei. Nur Radon und Polonium tragen zur radioaktiven Strahlenbelastung bei.

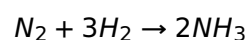
2. Wie lange dauert es in Stunden bis Ihre kritische Radonbelastung von  $300 Bq/m^3$  auf unter  $90 Bq/m^3$  gefallen ist? **(4 Punkte)**

Folgende Tabelle enthält die Informationen zur Zusammensetzung der normalen Umgebungsluft.

	Vol-%	M [g/mol]	ppm
Stickstoff	78.1	28.1	
Sauerstoff	21.3	15.8	
Kohlenstoffdioxid	0.029	12.1	

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

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? **(3 Punkte)**







## 119 Aufgabe

(10 Punkte)

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



**Armee der Finsternis** Ihr Studentenjob war nach Ladenschluss bei IKEA die Regale einzuräumen. Dabei ist Ihnen in der Auslage der Sonderangebote das Necronomicon<sup>1</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 653 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  $10mg$  zu gleichförmigen *Bleikugeln* bei einer Geschwindigkeit von  $12m/s$  bilden.

1. Wie hoch müssen Sie den Schrottkugelturm bauen lassen, damit sich runde Bleikugeln durch die Fallgeschwindigkeit von  $12m/s$  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  $6.1 \times 10^4$  Bleikugeln zusammen. Blei hat eine Dichte von  $10.32g/cm^3$ .

4. Wie schwer in Kilogramm  $kg$  sind die  $6.1 \times 10^4$  produzierten Bleikugeln, die Sie jetzt auf die Burgmauer transportieren müssen? **(1 Punkt)**

Am Ende müssen Sie noch die Produktion von dem Bleischrott im Turm optimieren.

5. Wie groß in  $cm^2$  ist Ihr quadratisches Sieb am oberen Ende des Turms, wenn Sie pro Fall ca. 1100 Bleikugeln produzieren wollen und die Bleikugel im Fall  $0.8cm$  Abstand haben müssen? **(1 Punkt)**

<sup>1</sup>Ein wirklich gefährliches Buch ist: *Du bist genug: Vom Mut, glücklich zu sein* von Fumitake Koga und Ichiro Kishimi





## 120 Aufgabe

(10 Punkte)

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



**Armee der Kaninchen** Leider hat es mit Ihrer Surfschule 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 22 Kaninchen entlassen hat, Millionen von Kaninchen rum? Wieso ist das Kaninchenfleisch dann so exklusiv? Dem wollen wir mal mathematisch nachgehen...

Unter australischen Bedingungen liegt die Zahl der Jungen eines Kaninchenweibchens pro Jahr bei ca. 11 Jungtieren. Trotz fehlender Fressfeinde liegt die Mortalität der Säuglinge bei ca. 40%. Ein Weibchen erreicht nach einem halben Jahr die Geschlechtsreife.

1. Wie viele Millionen Kaninchen würden bei einem ungebremsten Wachstum *im elften Jahr* geboren? **(2 Punkte)**
2. Die durchschnittliche, australische Lebenserwartung eines Kaninchens liegt bei 9 Jahren. Wie viele Kaninchen bevölkern bei einem ungebremsten Wachstum den australischen Kontinent bevor die ersten, australischen Kaninchen an Altersschwäche sterben würden? **(2 Punkte)**

Forscherinnen fand folgende Sättigungsfunktion für das Wachstum der 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 40 Jahren auf dem australischen Kontinent? **(1 Punkt)**

Das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV), transmittiert von Stechmücken, töten 99.7% der Kaninchenpopulation innerhalb weniger Wochen.

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

Thomas Austin entließ die Kaninchen im äußersten Süden von Australien. Australien hat eine West-Ost-Ausdehnung von 4200km und eine Nord-Süd-Ausdehnung von knapp 3600km. Die Kaninchen breiten sich radial mit einer Geschwindigkeit von 11.5km pro Jahr aus.

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

Eine jährliche Impfung gegen das Myxoma Virus und das Rabbit Haemorrhagic Disease Virus (RHDV) kosten 11\$ und der durchführende Arzt verlangt ca. 40\$.

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





## 121 Aufgabe

(10 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 Hängebauchschweins Fridolin und Ihnen. Grünes Gras unter Ihren Füßen und ein strammer Wind im Gesicht, egal wohin Sie schauen. Ein schmatzendes Geräusch unter Ihnen.

1. Skizzieren Sie den Versuchsplan auf Tischen im Gewächshaus! **(2 Punkte)**

Auf dem Weg zu Ihrer Pink Ladyplantage wurden Sie mit Ihrem Trecker von einer Gruppe elektrifizierter Renter abgedrängt. Der Trecker muss wieder aus dem Graben.

<https://www.ingenieurkurse.de/technische-mechanik-statik/einzelkraefte-mit-gemeinsamen-angriffspunkt/re-analytisch-bestimmen/kraefte-mit-unterschiedlicher-wirkungslinie/kraefte-mit-unterschiedlicher-wirkungslinie>

<https://www.sofatutor.com/physik/videos/kraefteparallelogramm-rechnerische-ermittlung-von-betrag-und-richtung-einer-resultierenden-kraft>

resultierende kraft berechnen 3 kräfte

4. Wie lange in Jahren dauert es bis eine Kaninchenpopulation nach einer Viruspandemie wieder auf **(2 Punkte)**





## 122 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
GLEICH	
UND	
KLEINER GLEICH	
ODER	

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
  - $A$  tritt ein, aber  $B$  tritt nicht ein
  - $A$  tritt ein oder  $B$  tritt ein
  - $A$  tritt nicht ein
3. Ergänzen Sie zu den Venndiagrammen die mathematische Notation! (2 Punkte)
4. Erstellen Sie die numeric  Ausgabe für die Suche nach der Zeichenfolge AT in folgenden DNA Sequenzen! (2 Punkte)
- AAATTTAT
  - ATATACAC
  - ACTGGTAA
  - ACACACAC
5. Geben ist das  Objekt  $A$  beinhaltend die Zahl 12. Erklären Sie den Unterschied zwischen dem Ausdruck  $A == 12$  und dem Ausdruck  $A != 12$  in ! Wie lautet die Ausgabe von  in beiden Fällen? (2 Punkte)



## 123 Aufgabe

(10 Punkte)

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



Sie haben den Vektor  $A = \{6, 2, 10\}$  und den Vektor  $B = \{-8, 3, 1\}$  gegeben.

1. Addieren Sie die Vektoren  $A$  und  $B$ ! **(1 Punkt)**
2. Transponieren Sie den Vektor  $B$ ! **(1 Punkt)**
3. Multiplizieren Sie den Vektor  $A$  mit 9! **(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 lineares Gleichungssystem bestehend aus drei Gleichungen in die Matrixschreibweise! **(2 Punkte)**

```
## [1] "5x1 + 8x2 + 2x3 = 2"
## [1] "4x1 + 8x2 + 10x3 = 2"
## [1] "2x1 + 7x2 + 10x3 = 3"
```

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

```
##      [,1] [,2] [,3]
## [1,]   11    3    9
## [2,]    3    6    7
## [3,]    6    5    4
## [4,]    6    5    6
```

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

