

Name: _____

Nicht bestanden: ☐

Vorname: _____

Matrikelnummer: _____

Endnote: _____

Studierende der Fakultät Agrarwissenschaften und Landschaftsarchitektur (AuL)

Probeklausur Bio Data Science

für Pflichtmodule

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

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

26. Februar 2025

Erlaubte Hilfsmittel für die Klausur

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

Ergebnis der Klausur

_____ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.

_____ von 79 Punkten sind aus dem Rechen- und Textteil erreicht.

_____ von 99 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
94.5 - 99.0	1,0
89.5 - 94.0	1,3
85.0 - 89.0	1,7
80.0 - 84.5	2,0
75.0 - 79.5	2,3
70.0 - 74.5	2,7
65.0 - 69.5	3,0
60.5 - 64.5	3,3
55.5 - 60.0	3,7
49.5 - 55.0	4,0

Es ergibt sich eine Endnote von _____.

Multiple Choice Aufgaben

- Pro Multiple Choice Frage ist *genau* eine Antwort richtig.
- **Übertragen Sie Ihre Kreuze in die Tabelle auf dieser Seite.**
- 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	10	20	8	10	11	10	10

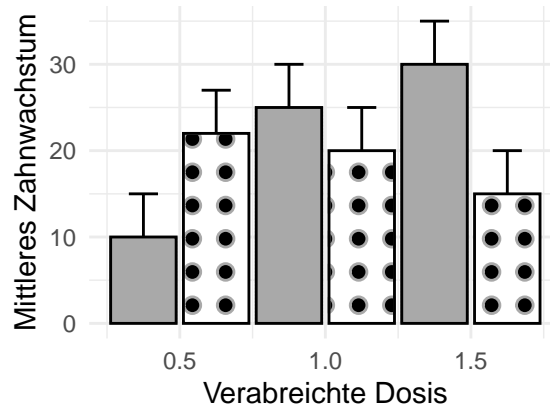
- Es sind ____ von 79 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Ein Versuch wurde an 65 Tieren durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0.5, 1 und 1.5 mg/Tag) über eine von zwei Verabreichungsmethoden erhielt. Die folgende Abbildung enthält die Daten aus diesem Versuch zur Bewertung der Wirkung von Vitamin E auf das Zahnwachstum bei Hasen. Welche Aussage ist richtig, wenn Sie eine zweifaktorielle ANOVA rechnen?



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

2 Aufgabe

(2 Punkte)

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

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

- A ☐ Wir lesen 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 ☐ Für eine explorativen Datenanalyse (EDA) in R müssen wir als erstes die Daten über `read_excel()` einlesen. Danach müssen wir schauen, dass wir die Zeilen richtig über `mutate()` transformiert haben. Insbesondere müssen Variablen mit kontinuierlichen Werten in einen Faktor umgewandelt werden. Am Ende nutzen wir die Funktion `ggplot()` für die eigentliche EDA.
- 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.

3 Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ Das R Paket `{lm}`. Das Paket `{lm}` erstellt selbstständig Konfidenzintervalle und entsprechende p -Werte. Da wir in dem Paket nicht adjustieren müssen, ist es bei Anwendern sehr beliebt.
- B** ☐ Das R Paket `{ggplot}`. Wir erhalten hier sofort eine Visualisierung der Daten. Anhand der Visualisierung lässt sich eine explorative Datenanalyse durchführen, die gleichwertig zu einem Posthoc-Test ist.
- C** ☐ Das R Paket `{hmisc}` erlaubt die Durchführung eines multiplen Gruppenvergleichs aus verschiedenen Modellen heraus. Aus einem `hmisc` Objekt lässt sich recht einfach das CLD erstellen und so über Barplots eine schnelle Interpretation der statistischen Auswertung durchführen.
- D** ☐ Das R Paket `{emmeans}` erlaubt die Durchführung eines multiplen Gruppenvergleichs. Aus einem `{emmeans}` Objekt lässt sich recht einfach das CLD erstellen und so über Barplots eine schnelle Interpretation der statistischen Auswertung durchführen.
- E** ☐ Das R Paket `{emmeans}` erlaubt die Durchführung eines multiplen Gruppenvergleichs. Aus einem `emmeans` Objekt lässt sich leider kein CLD erstellen. Dennoch ist das Paket einfach zu bedienen und wird deshalb genutzt. Die Interpretation der statistischen Auswertung wird über einen Barplot abgebildet.

4 Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Sie haben ein Feldexperiment mit Erbsen durchgeführt und wollen nun in einer simplen linearen Regression den Einfluss der CO_2 -Konzentration in $[\mu g]$ im Wasser auf den absoluten Proteingehalt in $[kg]$ untersuchen. Sie erhalten einen β_{CO_2} Koeffizienten von 6.9×10^{-7} und einen p -Wert mit 0.00032. Welche Aussage zu der Signifikanz und dem Effekt ist richtig?

- A** ☐ Die Fallzahl ist zu hoch angesetzt. Je höher die Fallzahl ist, desto kleiner ist die Teststatistik und damit ist dann auch der p -Wert sehr klein. Es sollte über eine Reduzierung der Fallzahl nachgedacht werden. Dann sollte der Effekt zum p -Wert passen.
- B** ☐ Wenn der Effekt β_{CO_2} sehr klein ist, dann kann es an einer falsch gewählten Einheit liegen. Der Anstieg von einer Einheit in X führt ja zu einer Änderung von β_{CO_2} in y . Daher ist hier mit einer anderen Einheit in den Daten zu rechnen, so dass wir hier einen besser formatierten Effekt sehen. Der p -Wert stammt aus einer einheitslosen Teststatistik.
- C** ☐ Manchmal ist die Einheit der Einflussvariable X zu groß gewählt, so dass der Anstieg von 1 Einheit in X zu einer zu großen Änderung in y führt. Daher kann der Effekt β_{CO_2} sehr klein wirken, da der p -Wert wird auf einer einheitslosen Teststatistik bestimmt wird.
- D** ☐ Das Gewicht und die CO_2 -Konzentration korrelieren sehr stark, deshalb wird der β_{CO_2} Koeffizient sehr klein. Mit einer ANOVA kann für die Korrelation korrigiert werden und der Effektschätzer passt dann zum p -Wert.
- E** ☐ Die Einheit der 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 antiproportional zusammen.

5 Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ Anhand des 95%-Konfidenzintervalls lässt sich wie folgt eine Entscheidung treffen. Ist der Wert kleiner dem Signifikanzniveau α dann kann die Nullhypothese abgelehnt werden.
- B** ☐ Anhand des 95%-Konfidenzintervalls lässt sich wie folgt eine Entscheidung treffen. Liegt der Wert in dem Signifikanzniveauintervall α dann kann die Nullhypothese abgelehnt werden.
- C** ☐ Das Signifikanzniveauintervall α ist gleich 5% und damit muss das berechnete Intervall unter dem Signifikanzniveauintervall α liegen, dann kann die Nullhypothese nicht abgelehnt werden.
- D** ☐ Wir betrachten das Ganze Intervall des 95%-Konfidenzintervalls. Liegt die Null mit in dem Intervall, dann kann die Nullhypothese abgelehnt werden.
- E** ☐ Der kritische Wert $T_{\alpha=5\%}$ ist tabelarisch festgelegt und gegeben. Ist T_D größer als der kritische Wert, kann die Nullhypothese abgelehnt werden

6 Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

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

- A** ☐ Nach der Erstellung eines Boxplots schauen wir, ob der Median in der Mitte der Box liegt. Dabei ist der Median als dicke Linie dargestellt und die Box ist das IQR.
- B** ☐ Einen Boxplot. Das IQR muss über alle Behandlungen zusammen mit den Whiskers ungefähr gleich aussehen.
- C** ☐ Einen Barplot. Die Mittelwerte müssen alle auf einer Höhe liegen. Die Fehlerbalken haben hier keine Informationen.
- D** ☐ Einen Dotplot. Die Punkte müssen sich wie an einer Perlenschnur aufreihen. Eine Abweichung führt zur Ablehnung der Annahme einer Varianzhomogenität.
- E** ☐ In einer explorativen Datenanalyse nutzen wir den Violinplot. Dabei sollte der Bauch am Rand liegen. Dann können wir von einer Varianzhomogenität ausgehen.

7 Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Neben der Mittelwertsdifferenz als Effektschätzer bei normalverteilten Endpunkten wird auch häufig der Effektschätzer Risk ratio bei binären Endpunkten verwendet. Welche Aussage über den Effektschätzer Risk ratio ist im folgenden Beispiel zur Behandlung von Klaueninfektionen bei Schweinen richtig? Dabei sind 3 Tiere krank und 12 Tiere sind gesund.

- A** ☐ Es ergibt sich ein Risk ratio von 4, da es sich um ein Anteil handelt.
- B** ☐ Das Verhältnis von Chancen Risk ratio ergibt ein Chancenverhältnis von 0.25.
- C** ☐ Da es sich um ein Chancenverhältnis handelt ergibt sich ein Risk ratio von 5.
- D** ☐ Es ergibt sich ein Risk ratio von 0.25, da es sich um ein Anteil handelt.
- E** ☐ Der Anteil der Kranken wird berechnet. Da es sich um ein Anteil handelt ergibt sich ein Risk ratio von 0.2.

8 Aufgabe

(2 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik

Sie führen paarweise t-Tests für alle Vergleiche der verschiedenen Rapssorten in Ihrem Experiment durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der α -Schwelle. Ihr Experiment beinhaltet fünf Rapssorten und eine ANOVA ergibt $p = 0.045$ für den Ertrag. Sie schauen sich auch die rohen, unadjustierten p-Werte an und finden hier als niedrigsten p-Wert $p_{3-2} = 0.052$. Welche Aussage ist richtig?

- A** ☐ Der Fehler liegt in den t-Tests. Wenn eine ANOVA signifikant ist, dann muss zwangsweise auch ein t-Test signifikant sein.
- 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** ☐ Die ANOVA testet auf der gesamten Fallzahl. Die einzelnen t-Tests immer nur auf einer kleineren Subgruppe. Da mit weniger Fallzahl weniger signifikante Ergebnisse zu erwarten sind, kann eine Diskrepanz zwischen der ANOVA und den paarweisen t-Tests auftreten.
- E** ☐ Das Beispiel kann so nicht auftreten, da die ANOVA und die t-Tests algorithmisch miteinander verknüpft sind.

9 Aufgabe

(2 Punkte)

Inhalt folgender Module: Biostatistik

Welche Aussage über die *Power* ist richtig?

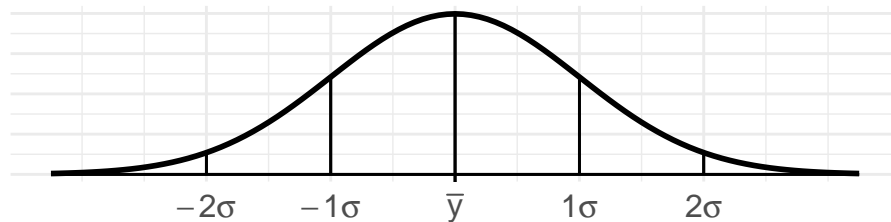
- A** ☐ Die Power wird berechnet und ist keine Eigenschaft des Tests. Die Power wird auf 80% gesetzt und beschreibt mit welcher Wahrscheinlichkeit H_0 bewiesen wird
- B** ☐ Die Power ist nicht in der aktuellen Testtheorie mehr vertreten. Wir rechnen nur noch mit dem Fehler 1. Art.
- C** ☐ Die Power wird nicht berechnet sondern ist eine Eigenschaft des Tests. Die Power wird auf 80% gesetzt und beschreibt mit welcher Wahrscheinlichkeit H_A bewiesen wird
- D** ☐ Alle statistischen Tests sind so konstruiert, dass die H_A mit 20% bewiesen wird. Die Power ist $1 - \beta$ mit β gleich 80% gesetzt.
- E** ☐ Die Power beschreibt die Wahrscheinlichkeit die H_A abzulehnen. Wir testen die Power jedoch nicht.

10 Aufgabe

(2 Punkte)

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

Die Normalverteilung wird durch den Mittelwert und die Standardabweichung definiert. Welche Aussage im Bezug zur Standardabweichung in einer Normalverteilung ist richtig?



- A** ☐ Die Fläche unter der Kurve ist 1, wenn die Nullhypothese falsch ist. Wenn die Nullhypothese gilt, dann ist die Fläche $1 - \alpha$. Somit ergibt sich auch eine Standardabweichung σ mit α gleich 0.05 in beiden Fällen.
- B** ☐ Die Fläche links von -2σ ist der p-Wert mit $Pr(D|H_0)$ in der obigen Abbildung.
- C** ☐ Die Fläche unter der Kurve entspricht dem Signifikanzniveau α von 5%. Damit ist die Standardabweichung σ gleich 1 in der obigen Abbildung.
- D** ☐ Die Fläche zwischen -1σ und 1σ ist 0.95 und 95% der Beobachtungen liegen somit zwischen $\bar{y} \pm \sigma$ in der obigen Verteilung.
- E** ☐ Es liegen 95% der Beobachtungen zwischen $\bar{y} \pm 2\sigma$. Angezeigt durch die Fläche zwischen -2σ und 2σ in der obigen Verteilung.

11 Aufgabe

(10 Punkte)

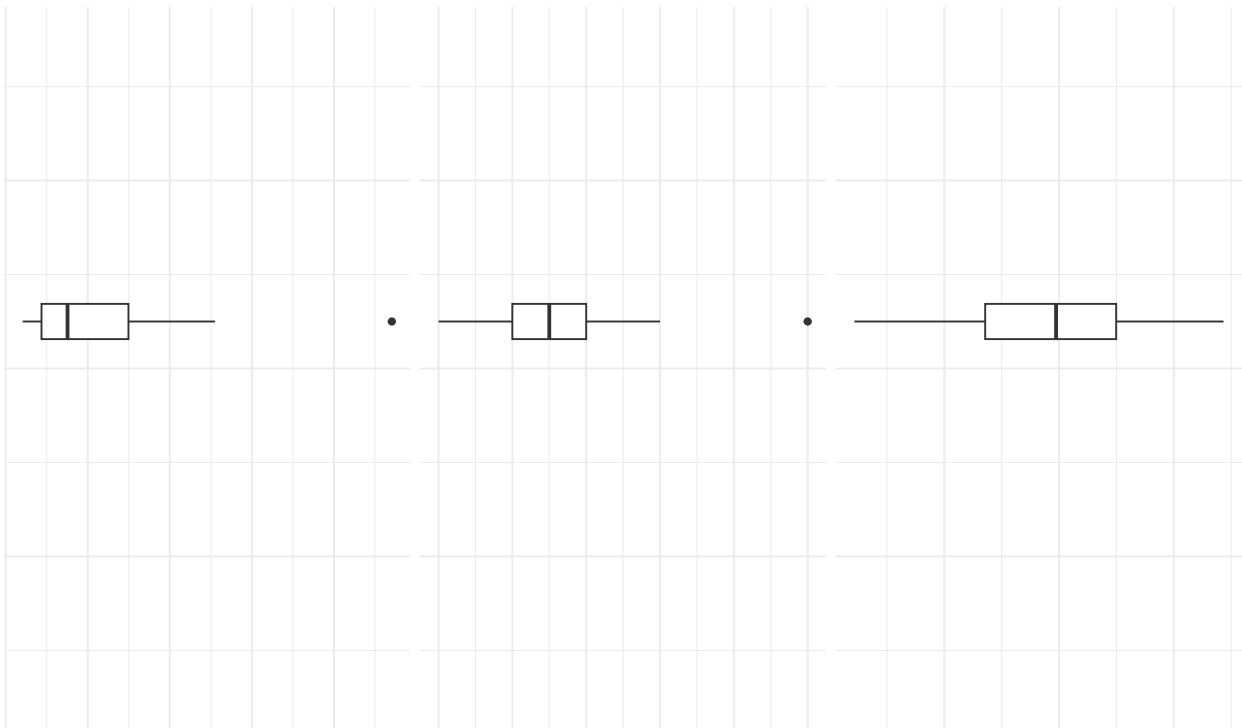
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



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



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

1. Zeichnen Sie über die Boxplots die entsprechende zugehörige Verteilung! **(3 Punkte)**
2. Zeichnen Sie unter die Boxplots die entsprechende zugehörige Beobachtungen als Stiche! **(3 Punkte)**
3. Wie viel Prozent der Beobachtungen fallen in das IQR? Ergänzen Sie die Abbildung entsprechend um den Bereich! **(2 Punkte)**
4. Wie viel Prozent der Beobachtungen fallen in $\bar{y} \pm 1s$ und $\bar{y} \pm 2s$ unter der Annahme einer Normalverteilung? **(2 Punkte)**

¹Oschmann, A. (2024) Mädchen stärken: Stärken fördern, Selbstwert erhöhen und liebevoll durch Krisen begleiten. Goldegg Verlag

12 Aufgabe

(20 Punkte)

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



Inhalt folgender Module: Biostatistik



Zerforschen eines wissenschaftlichen Datensatzes Unter einem langen Schnaufen starrt Mark auf den wissenschaftlichen Datensatz in der Tabelle 1 in seinem Laptop. Insgesamt wurden n Beobachtungen erhoben. 'Worum geht es denn eigentlich in diesem Datensatz?', fragt er sich kopfschüttelnd und mampft noch ein paar Marzipankugeln. Mark soll die Datentabelle nutzen um das eigene Experiment zu planen und eine Blaupause zu haben. Als eine Vorlage sozusagen, die er nur noch ausfüllen muss. Daher möchte seine Betreuerin, dass er einmal die Daten sinnvoll zusammenfasst. Das sollte dann doch etwas aufwendiger werden. Das wird dann vermutlich heute Abend nichts mehr mit Columbo.

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


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

Beantworten Sie die folgenden Fragen anhand eines selbst gewählten Beispiels!

Allgemeiner Aufgabenteil

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

Spezieller Aufgabenteil für die Variablen x_1 und y_2

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

13 Aufgabe


(8 Punkte)

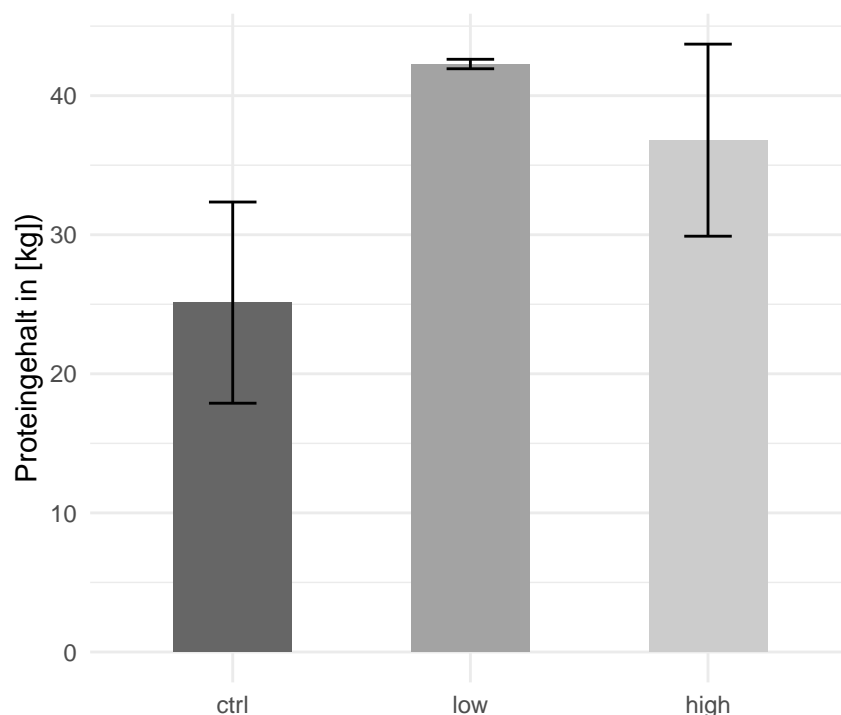
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!





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



Zerforschen des Barplots Barplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Mark nicht gemerkt, welche statistischen Maßzahlen für einen Barplot erhoben werden müssen. Besser wäre was anderes gewesen. Mark liebt Geocaching. Darin kann er sich wirklich verlieren und immer wieder neu begeistern. Das ist in soweit doof, da nach seiner Betreuer erstmal ein Barplot nachgebaut werden soll, bevor es mit seiner Hausarbeit losgeht. Dann hat er schonmal den  Code vorliegen und nachher geht dann alles schneller. Na dann mal los. Mark schafft sich die nötige Stimmung. Mark streichelt liebevoll der Hamster. Der Kopf ist in seinem Schloß vergraben um den Klang von Andrea Berg zu dämpfen. In der Behandlung für Brokoli werden verschiedene Düngeufen (*ctrl*, *low* und *high*) sein. Erfasst wird als Outcome (Y) *Proteingehalt*. Mark soll dann *protein* in seiner Exceldatei eintragen.



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

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

14 Aufgabe

(10 Punkte)

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

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



Visualisierung des Compact Letter Displays (CLD) Yuki hatte in seinem Projektbericht ein Stallexperiment durchgeführt. Soweit so gut. Dabei hat er sich mit Hühnern beschäftigt. Angeblich der neueste heiße Kram... aber das ist wiederum was anderes. So richtig mitgenommen hat Yuki das Thema dann doch nicht. Hat er sich doch mit Bestandsdichte (*effizient*, *standard*, *weit* und *kontakt*) und Fettgehalt [%/kg] schon eine Menge an Daten angeschaut. Nach seiner Betreuerin soll er nun ein CLD bestimmen. Weder weiß er was ein CLD ist, noch war sein erster Gedanke mit Köln und die LGBTQ Community richtig...

Behandlung	Compact letter display
effizient	B
standard	C
weit	A
kontakt	B

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie die statistischen Hypothesen! **(1 Punkt)**
3. Zeichnen Sie die sich anhand des *Compact letter display (CLD)* ergebenden Barplots! **(2 Punkte)**
4. Ergänzen Sie das *Compact letter display (CLD)* zu den Barplots! **(1 Punkt)**
5. Erklären Sie *einen* Vorteil und *einen* Nachteil des *Compact letter display (CLD)*! **(2 Punkte)**
6. Erstellen Sie eine Matrix mit den paarweisen *p*-Werten eines Student t-Tests, die sich näherungsweise aus dem *Compact letter display (CLD)* ergeben würde! Begründen Sie Ihre Antwort! **(3 Punkte)**

15 Aufgabe

(12 Punkte)

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




Inhalt folgender Module: Biostatistik



Visualisierung des Regressionskreuzes Gelangweilt schaut das Känguruh auf und schmeißt sich eine Schnapspraline in den Rachen. 'Ich dachte nur in Bayern hat man es mit Kreuzen...', stöhnt es gelangweilt. Nilufar hätte einfach nicht die Tür aufmachen sollen ohne zu schauen, ob es wirklich ihr Lerngruppe ist. Jetzt hat sie den Salat oder die Schnapspraline. 'Du bist angesapnnt', bemerkt das Känguruh. 'Ich glaube nicht, dass sowas hilft.' Nilufar klappt den Laptop zu und flüchtet auf die Toilette. Was hatte ihr Gruppe nochmal gemacht? Genau ein Gewächshausexperiment mit Brokkoli und es wurde Chlorophyllgehalt unter Zielwert [ja/nein] gemessen. Das Känguru begann damit seine Musiksammlung mit Deichkind extrem laut aufzudrehen.

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

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

16 Aufgabe

(10 Punkte)

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



Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



Die zweifaktorielle ANOVA in R In ein Stallexperiment wurden Zandern mit der Behandlung Genotypen (AA, AB und BB) sowie der Behandlung Elterlinie (*ctrl*, und *Xray*) untersucht. Es wurde als Messwert Protein/Fettrate [%/kg] bestimmt. Jetzt starrt Paula mit auf die R Ausgabe einer zweifaktoriellen ANOVA. Leider starrt ihre Betreuerin in der gleichen Art Paula zurück an. Das wird ein langer Nachmittag, denkt sie sich und kreuselt ihren Mund. 'Und was machen wir jetzt?' entfährt es ihr überrascht entnervt. Immerhin war geht es ja um die Projektbericht. Paula hätte doch nichts mit Zandern machen sollen. Zandern – was soll das auch bedeutendes sein? Eigentlich wollte Paula nachher noch einen Film schauen. Das Verrückte ist, dass die Ratte Jagd auf roter Oktober wirklich liebt. Das ist Paula sehr recht, denn sie braucht Entspannung.

```
## Analysis of Variance Table
##
## Response: Protein/Fettrate
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Genotypen      2 563.53  281.764  24.4496 7.39e-06
## Elterlinie      1   0.23    0.235   0.0204 0.888095
## Genotypen:Elterlinie 2 188.15   94.075   8.1632 0.002998
## Residuals     18 207.44   11.524
```

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Interpretieren Sie das Ergebnis der einfaktoriellen ANOVA! **(3 Punkte)**
4. Zeichnen Sie eine Abbildung, der dem obigen Ergebnis der zweifaktoriellen ANOVA näherungsweise entspricht! **(5 Punkte)**

17 Aufgabe

(10 Punkte)

Inhalt folgender Module: Angewandte Statistik und Versuchswesen • Biostatistik



Visualisierung des 95% Konfidenzintervalls 'Okay, für was war jetzt nochmal das 95% Konfidenzintervall gut?', fragt Jonas und schaut in das leere Gesicht von Yuki. 'Keine Ahnung. Irgendwas mit Relevanz und Effekt oder Signifikanz. Da kannst du irgendwie was verbinden. Keine Ahnung warum', entgegnet Yuki. 'Wir haben doch als Messwert *Energieverbrauch der Klimakammer* erhoben.', stellt Jonas fest. Jetzt haben beide das Problem, die möglichen 95% Konfidenzintervalle zu interpretieren.

Leider kennen sich Jonas und Yuki mit der Visualisierung des 95% Konfidenzintervall überhaupt nicht aus.

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

