

Nicht bestanden: ☐

Endnote: _____

Erlaubte Hilfsmittel

- Normaler Taschenrechner ohne Möglichkeit der Kommunikation mit anderen Geräten! Ausdrücklich kein Handy!
- Eine DIN A4-Seite als beidseitig, selbstgeschriebene, handschriftliche Formelsammlung. Keine digitalen Ausdrücke!
- **Die Verwendung eines roten Farbstiftes ist nicht gestattet! Korrekturfarbe!**
- *You can answer the questions in English without any consequences.*

Endnote

_____ von 20 Punkten sind aus den Multiple Choice Aufgaben erreicht.
_____ von 79 Punkten sind aus den Rechen- und Textaufgaben 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.

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

- 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	8	10	9	10	12	10	20

- Es sind ____ von 79 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

Nachdem Sie in einem Feldexperiment zu Leistungssteigerung von Brokoli durchgeführt haben, berechnen Sie den Mittelwert und den Median. Der Mittelwert ist \bar{y} gleich 21.2 t/ha und als Median ergibt sich \check{y} gleich 9.2 t/ha. Welche Aussage ist richtig?

- A ☐ Wenn sich der Mittelwert und der Median unterscheiden, liegen vermutlich keine Outlier in den Daten vor.
- B ☐ Der Mittelwert und der Median unterscheiden sich. Damit haben Sie ein schwerwiegendes Effektmaßproblem. Sie können die Annahme an die Daten nicht ablehnen.
- C ☐ Nach der Betrachtung der Werte unterscheiden sich Mittelwert und Median nicht. Sie haben Varianzhomogenität vorliegen. Sie können künstlich Outlier zufügen um die Daten auszuwerten.
- D ☐ Der Mittelwert und der Median unterscheiden sich. Daher müssen Sie schauen, ob Sie einen Outlier in den Daten vorliegen haben.
- E ☐ 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.

2 Aufgabe

(2 Punkte)

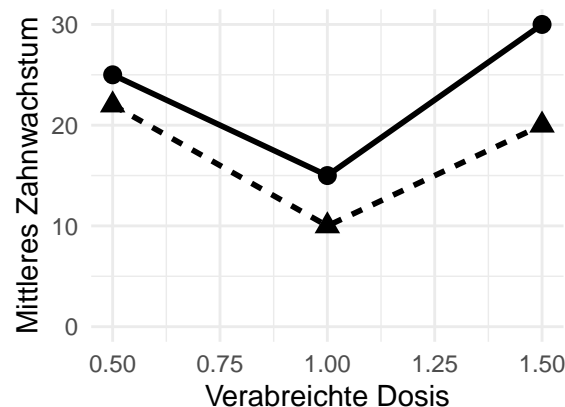
Sie führen paarweise t-Tests für alle Vergleiche der verschiedenen Rapssorten in Ihrem Experiment durch. Nach der Adjustierung für multiples Testen ist kein p-Wert unter der α -Schwelle. Ihr Experiment beinhaltet fünf Rapssorten und eine ANOVA ergibt $p = 0.048$ für den Ertrag. Sie schauen sich auch die rohen, unadjustierten p-Werte an und finden hier als niedrigsten p-Wert $p_{3-2} = 0.052$. Welche Aussage ist richtig?

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

3 Aufgabe

(2 Punkte)

In einer Studie zur Bewertung der Wirkung von Vitamin D auf das Zahnwachstum bei Hasen entstand folgende Abbildung. Der Versuch wurde an 61 Tieren durchgeführt, wobei jedes Tier eine von drei Vitamin-C-Dosen (0.5, 1 und 1.5 mg/Tag) über eine von zwei Verabreichungsmethoden erhielt. Welche Aussage ist im Bezug auf eine zweifaktorielle ANOVA richtig?



- A ☐ Keine Korrelation liegt vor ($p \geq 0.05$).

- B** ☐ Eine negative Interaktion liegt vor ($\rho \geq 0.5$).
- C** ☐ Keine Interaktion liegt vor ($\rho > 0.05$).
- D** ☐ Mit ($\rho \leq 0.05$) liegt eine mittlere bis starke Interaktion vor.
- E** ☐ Die Koeffizienten sind positiv ($\beta_0 > 0; \beta_1 > 0$).

4 Aufgabe

(2 Punkte)

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

- A** ☐ Als Kulturkonstante hat $\alpha = 5\%$ den Rang einer Naturkonstante und wurde nach langer Diskussion in der UN im Jahre 1983 festgesetzt. Damals auch schon mit der Zustimmung der UdSSR.
- B** ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.
- C** ☐ Da Wissenschaftler eine Schwelle für die statistische Testentscheidung benötigen wurde α in einer großen Konferenz 1945 gewählt. Damit ist $\alpha = 5\%$ eine Kulturkonstante mit einem Rang einer Naturkonstante.
- 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** ☐ In der Wissenschaft gibt es neben der Naturkonstante, die sich aus der Beobachtung der Welt ergibt, noch die Kulturkonstante, die von einer Gruppe Menschen selbstgewählt wird. Dabei ist $\alpha = 5\%$ eine Kulturkonstante und wurde somit eher zufällig gewählt.

5 Aufgabe

(2 Punkte)

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

- A** ☐ Die ANOVA berechnet die T-Statistik aus der Multiplikation der MS Behandlung mit der MS der Fehler. Wenn die F-Statistik genau 0 ist, kann die Nullhypothese nicht abgelehnt werden.
- B** ☐ Die ANOVA berechnet die F-Statistik indem die MS des Fehlers durch die MS der Behandlung geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese abgelehnt werden.
- C** ☐ Wenn die F-Statistik höher ist als der kritische Wert kann die Nullhypothese nicht abgelehnt werden. Die F-Statistik ist die Differenz der MS der Behandlung durch die MS des Fehlers.
- D** ☐ Die ANOVA berechnet die T-Statistik indem den Mittelwertsunterschied der Gruppen simultan durch die Standardabweichung der Gruppen teilt. Wenn die T-Statistik höher als 1.96 ist, kann die Nullhypothese abgelehnt werden.
- E** ☐ Die ANOVA berechnet die F-Statistik indem die MS der Behandlung durch die MS des Fehlers geteilt werden. Wenn die F-Statistik sich der 0 annähert kann die Nullhypothese nicht abgelehnt werden.

6 Aufgabe

(2 Punkte)

In einer linearen Regression kann es vorkommen, dass der Effekt repräsentiert durch den β Koeffizienten nicht so richtig von der Größenordnung zu dem p-Wert passen will. So liefert eine Untersuchung des Einflusses von der PO_2 -Konzentration in $[\mu g]$ im Wasser auf den absoluten Proteingehalt in $[kg]$ an Brokkoli folgende Effekte und p-Werte: $1e-04$ als p-Wert und einen β_{PO_2} Koeffizienten von 7.4×10^{-6} . Welche Aussage ist richtig?

- A** ☐ Die Fallzahl ist zu klein angesetzt. Je kleiner die Fallzahl ist, desto höher ist die Teststatistik und damit auch der p-Wert kleiner. Wir brauchen also mehr Fallzahl um den geringen Effekt noch signifikant zu kriegen.
- B** ☐ Manchmal ist die Einheit der Einflussvariable X zu klein gewählt, so dass der Anstieg von 1 Einheit in X zu einer zu kleinen Änderung in y führt. Daher kann der Effekt β_{PO_2} sehr klein wirken, aber auf einer anderen Einheit sehr viel größer sein. Der p-Wert wird auf einer einheitslosen Teststatistik bestimmt.
- C** ☐ Wenn der Effekt β_{PO_2} winzig ist, dann kann es an einer falsch gewählten Einheit liegen. Der Anstieg von einer Einheit in X führt ja zu einer Änderung von β_{PO_2} in x . Wir müssen daher die Einheit von y entsprechend anpassen.

- D** ☐ Das Gewicht und die PO_2 -Konzentration korrelieren sehr stark, deshalb wird der β_{PO_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 PO_2 -Konzentration ist zu klein gewählt. Dadurch sehen wir den sehr kleinen p-Wert. Der p-Wert und die Einheit von der PO_2 -Konzentration hängen antiproportional zusammen.

7 Aufgabe

(2 Punkte)

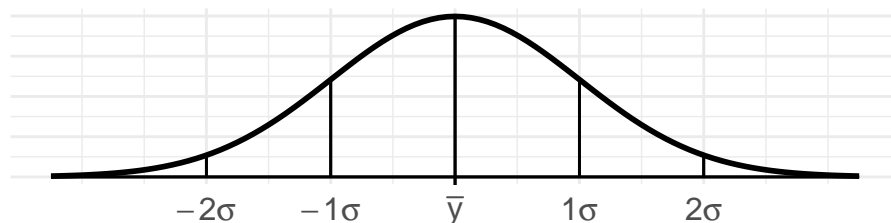
In Ihrer Abschlußarbeit wollen Sie Ihre Daten für den Ertrag in einem Barplot darstellen. Sie nutzen den Barplot auch, da der Barplot zu den meist genutzten Visualisierungen von Daten gehört. Welche statistischen Maßzahlen stellt der Barplot dar?

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

8 Aufgabe

(2 Punkte)

Im Folgenden sehen Sie ein Normalverteilung dargestellt. Welche Aussage zu der Normalverteilung und der Standardabweichung σ 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 unter der Kurve entspricht dem Signifikanzniveau α von 5%. Damit ist die Standardabweichung σ gleich 1 in der obigen Abbildung.
- C** ☐ Die Fläche zwischen -1σ und 1σ ist 0.95 und 95% der Beobachtungen liegen somit zwischen $\bar{y} \pm \sigma$ in der obigen Verteilung.
- D** ☐ Die Fläche rechts von 2σ ist der p-Wert mit $Pr(D|H_0)$ in der obigen Abbildung.
- E** ☐ Die Fläche zwischen -2σ und 2σ ist 0.95 und 95% der Beobachtungen liegen somit zwischen $\bar{y} \pm \sigma$ in der obigen Verteilung.

9 Aufgabe

(2 Punkte)

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

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

10 Aufgabe

(2 Punkte)

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

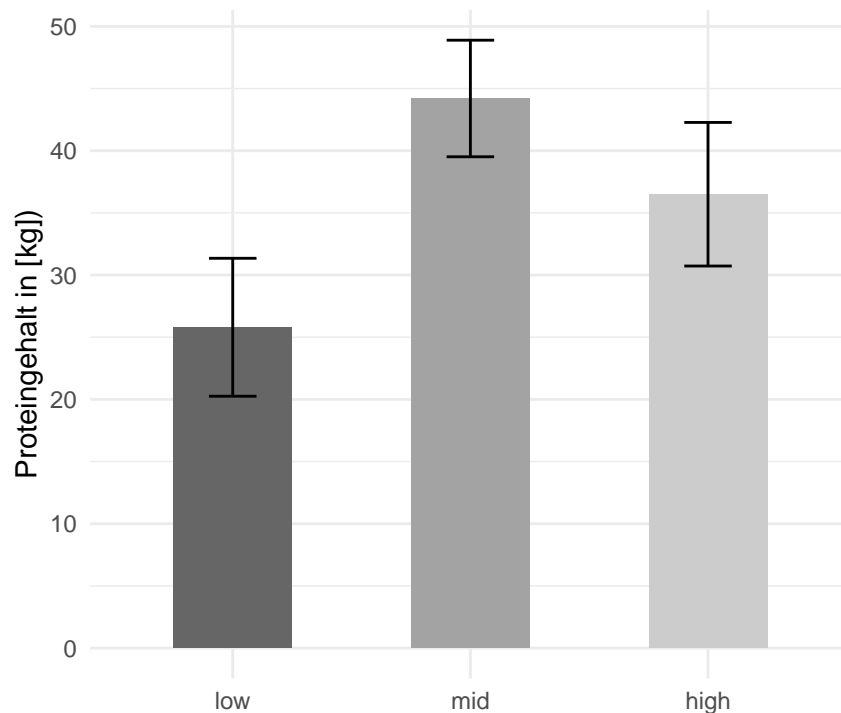
- A** ☐ ... dass ein minderwertiges Modell durch ein minderwertiges Modell ersetzt wird. Es gilt das Verifikationsprinzip nach Karl Popper.
- B** ☐ ... dass ein schlechtes Modell durch ein schlechteres Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.
- C** ☐ ... dass ein schlechtes Modell durch das Falsifikationsprinzip durch ein noch schlechteres Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.
- D** ☐ ... dass in der Wissenschaft immer etwas falsch sein muss. Sonst gebe es keinen Fortschritt.
- E** ☐ ... dass ein schlechtes Modell durch ein weniger schlechtes Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.

11 Aufgabe

(8 Punkte)

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

Barplots sind bedeutend in der Darstellung von wissenschaftlichen Ergebnissen. Leider hat sich Paula nicht gemerkt, welche statistischen Maßzahlen für einen Barplot erhoben werden müssen. Besser wäre was anderes gewesen. Paula liebt Harry Potter. Darin kann sie sich wirklich verlieren und immer wieder neu begeistern. Das ist in soweit doof, da nach ihrer Betreuer erstmal ein Barplot nachgebaut werden soll, bevor es mit ihrer Hausarbeit losgeht. Dann hat sie schonmal den R Code vorliegen und nachher geht dann alles schneller. Na dann mal los. Paula schafft sich die nötige Stimmung. Paula nickt im Takt von White Lies und bemerkt dabei gar nicht was die Ratte schon wieder anstellt. In der Behandlung für Lauch werden verschiedene Bewässerungstypen (*low*, *mid* und *high*) sein. Erfasst wird als Outcome (Y) *Proteingehalt*. Paula soll dann *protein* in ihrer Exceldatei eintragen.



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

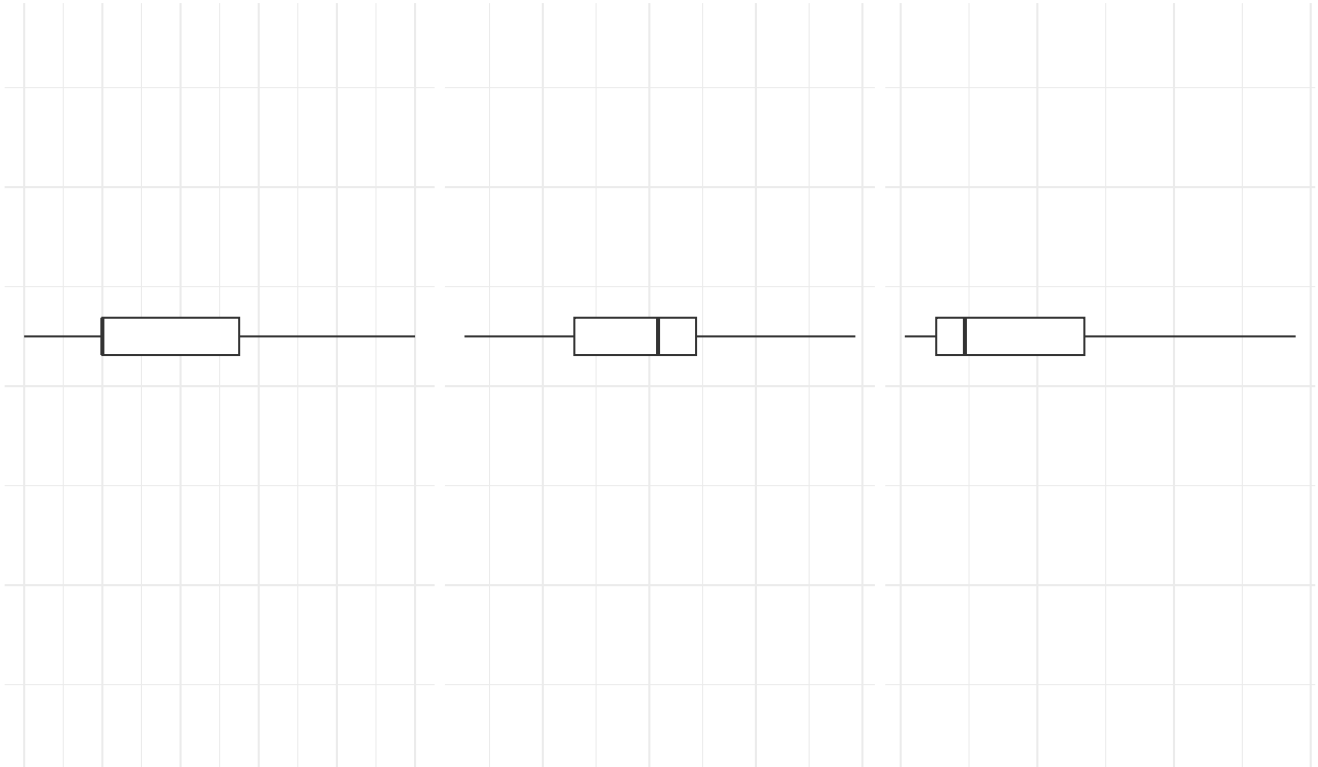
1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Erstellen Sie eine Tabelle mit den statistischen Maßzahlen der drei Barplots! *Beachten Sie die korrekte Darstellungsform der statistischen Maßzahlen!* **(3 Punkte)**
3. Erstellen Sie einen beispielhaften Datensatz im R üblichen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

12 Aufgabe

(10 Punkte)

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

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



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

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

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

13 Aufgabe

(9 Punkte)

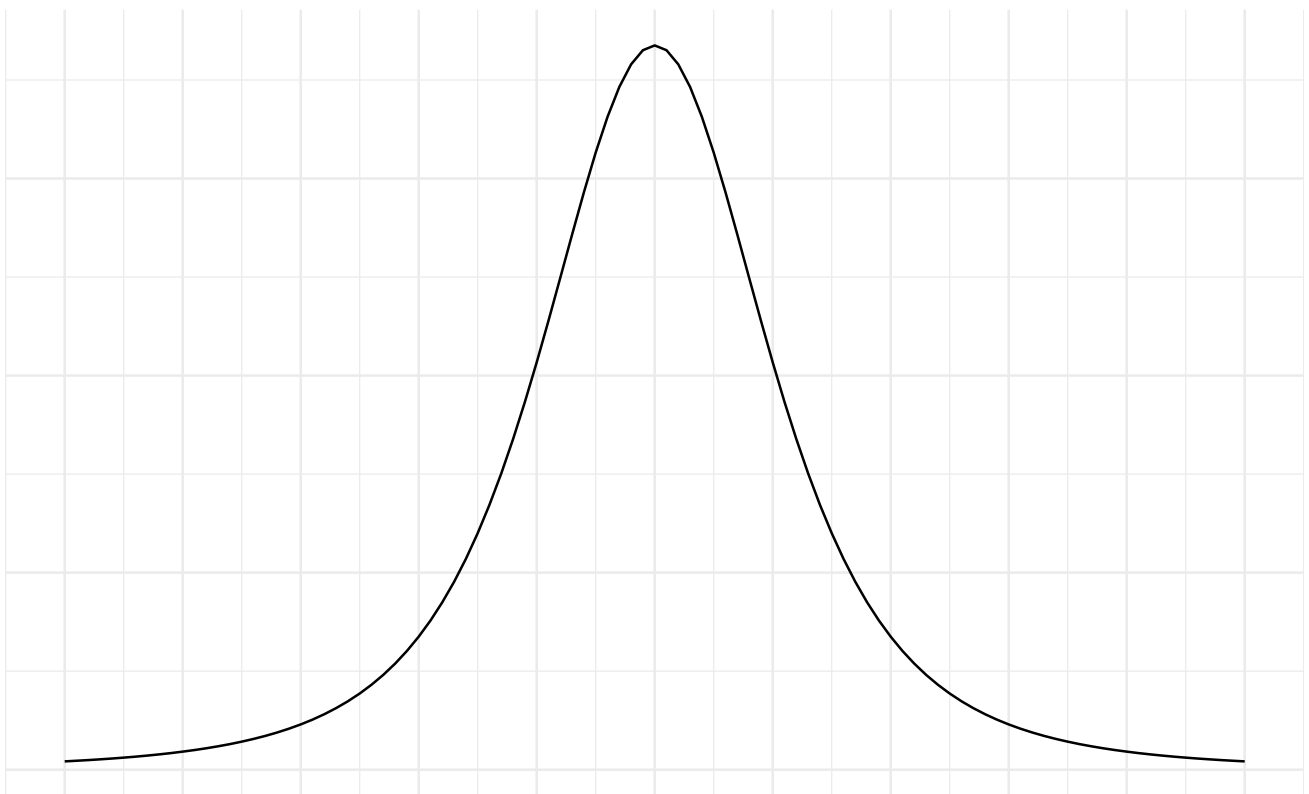
Geben Sie grundsätzlich Formeln und Rechenweg zur Lösung der Teilaufgaben mit an!

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

Leider kennen sich Jessica und Paula mit der Visualisierung der Teststatistik T_D und dem p-Wert überhaupt nicht aus und brauchen daher Ihre Hilfe!

Beachten Sie, dass im Folgenden keine numerisch korrekte Darstellung verlangt wird! Es gilt Erkennbarkeit vor Genauigkeit!


1. Ergänzen Sie eine beschriftete x-Achse! **(1 Punkt)**
2. Ergänzen Sie „ $\bar{y}_1 = \bar{y}_2$ “! **(1 Punkt)**
3. Ergänzen Sie „ $A = 0.95$ “! **(1 Punkt)**
4. Zeichnen Sie $T_{\alpha=5\%}$ in die Abbildung! **(1 Punkt)**
5. Zeichnen Sie das Signifikanzniveau α in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Zeichnen Sie $+T_D$ in die Abbildung! **(1 Punkt)**
7. Zeichnen Sie einen nicht signifikant p-Wert in die Abbildung! Begründen Sie Ihre Antwort! **(2 Punkte)**



14 Aufgabe

(10 Punkte)

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

'Uff... die einfaktorielle ANOVA und . Nicht so einfach... Was sagt mir jetzt die Ausgabe der ANOVA und wo sehe ich, ob da was signifikant ist?', denkt Jonas und hebt die Augenbraue. Jonas hatte sich einen Leistungssteigerungsversuch mit Fleischrindern angeschaut. Als wäre das nicht alles schon schwer genug. Wenn die Erschöpfung nicht wäre, ja dann wäre wohl vieles möglich für Jonas! Aber so.. Dabei ging es beim Experiment herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Elterlinie (*ctrl*, *Standard*, *Yray* und *Xray*) und dem Messwert Schlachtgewicht [kg] gibt. Nun möchte sein Betreuer seiner Hausarbeit erstmal eine ANOVA sehen und die Ergebnisse präsentiert bekommen. Und eigentlich will er ja was anderes... Schon dutzende Male gesehen: Mission Impossible. Aber immer noch großartig zusammen mit Snickers.

```
## Analysis of Variance Table
##
## Response: Schlachtgewicht
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Elterlinie  3  297.81   99.271    4.6464 0.01157
## Residuals  22  470.03   21.365
```

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

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

15 Aufgabe

(12 Punkte)

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

Es ist schon kurz nach fünf und Yuki wird langsam nervös. Yuki wollte heute Abend noch seine E-Sport Qualifikation schauen. Stattdessen versucht seine Betreuerin die Ausgabe der zweifaktoriellen ANOVA zu visualisieren und zu überprüfen, ob es mit der Visualisierung der Daten als Boxplots zusammenpasst. Yuki hatte im Teuteburgerwald ein Kreuzungsexperiment mit Puten durchgeführt. Es gab dabei zwei Behandlungen. Einmal Lüftungssystem (*keins*, *storm*, *tornado* und *thunder*) sowie als zweite Behandlung Flüssignahrung (*ctrl* und *flOw*). Gemessen wurde der Messwert (Y) Gewichtszuwachs in der 1LW. So kompliziert kann das jetzt doch nicht sein! Eigentlich wollte Yuki nachher noch zum Sport. Um zu Boldern geht Yuki dann später nochmal raus. Echte Entspannung.

```
## Analysis of Variance Table
##
## Response: Gewichtszuwachs
##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## Lüftungssystem	2	477.23	238.615	13.7342	0.0002388
## Flüssignahrung	1	5.30	5.299	0.3050	0.5875432
## Lüftungssystem:Flüssignahrung	2	194.43	97.213	5.5954	0.0128898
## Residuals	18	312.73	17.374		

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

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

16 Aufgabe

(10 Punkte)

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

Nilufar betrachtet in sich gekehrt die Poster vor dem Büro von ihrer Betreuerin. Viele der explorativen Abbildungen sagen ihr etwas. Die Barplots und die Boxplots könnte sie dann schon nachbauen. Das macht sie dann zuversichtlich die Abschlussarbeit auch hinzukriegen. Etwas komischer sind die seltsamen Buchstaben über den Barplots. Nilufar betrachtet ein Poster das sich mit Maiss beschäftigt. Substrattypen (*kompost*, *torf*, *40p60n* und *70p30n*) und Chlorophyllgehalt (SPAD-502Plus) [SPAD] wurden dort bestimmt. So richtig schlau, wird sie daraus nicht.

Behandlung	Compact letter display
kompost	B
torf	AB
40p60n	A
70p30n	AB

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

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

17 Aufgabe

(20 Punkte)

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

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

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
\vdots	\vdots	\vdots	\vdots	\vdots
1	2	1.9	9.6	0

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

Beantworten Sie die folgenden Fragen anhand eines selbst gewählten Beispiels!

Allgemeiner Aufgabenteil

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

Spezieller Aufgabenteil für die Variablen f_1 , f_2 und y_1

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