

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

19. November 2024

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 64 Punkten sind aus dem Rechen- und Textteil erreicht.

_____ von 84 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
80.5 - 84.0	1,0
76.0 - 80.0	1,3
72.0 - 75.5	1,7
67.5 - 71.5	2,0
63.5 - 67.0	2,3
59.5 - 63.0	2,7
55.0 - 59.0	3,0
51.0 - 54.5	3,3
46.5 - 50.5	3,7
42.0 - 46.0	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	9	8	10	9	9	9

- Es sind ____ von 64 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

Sie wollen in Ihrer Abschlussarbeit über eine explorative Datenanalyse überprüfen, ob Ihr gemessener Endpunkt einer Normalverteilung folgt. Welche drei Abbildungen eignen sich insbesondere für die Überprüfung?

- A ☐ Violinplot, Scatterplot, Barplot
- B ☐ Violinplot, Boxplot, Densityplot
- C ☐ Scatterplot, Densityplot, Barplot
- D ☐ Histogramm, Scatterplot, Boxplot
- E ☐ Scatterplot, Mosaicplot, Boxplot

2 Aufgabe

(2 Punkte)

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

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

3 Aufgabe

(2 Punkte)

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

- A ☐ ... dass in der Wissenschaft immer etwas falsch sein muss. Sonst gebe es keinen Fortschritt.
- B ☐ ... dass Annahmen an statistische Modelle meist falsch sind.
- C ☐ ... dass ein minderwertes Modell durch ein minderwertiges Modell ersetzt wird. Es gilt das Verifikationsprinzip nach Karl Popper.
- D ☐ ... dass ein schlechtes Modell durch das Falsifikationsprinzip durch ein weniger schlechtes Modell ersetzt wird.
- E ☐ ... dass Modelle meist falsch sind und selten richtig.

4 Aufgabe

(2 Punkte)

Gegeben ist y mit 8, 12, 14, 6 und 12. Berechnen Sie den Mittelwert und Standardabweichung.

- A ☐ Es ergibt sich 11.4 +/- 1.645
- B ☐ Es ergibt sich 9.4 +/- 5.4
- C ☐ Es berechnet sich 11.4 +/- 10.8
- D ☐ Sie erhalten 10.4 +/- 1.645
- E ☐ Es ergibt sich 10.4 +/- 3.29

5 Aufgabe

(2 Punkte)

In einem Feldversuch haben Sie einen Behandlungsfaktor mit mehreren Levels vorliegen. Sie rechnen einen multiplen Vergleich. Vorher hatten Sie eine einfaktorielle ANOVA mit einem signifikanten Ergebnis vorliegen. Welche Aussage ist richtig?

- A** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt Δ nach Bonferroni adjustiert werden. Dafür wird der Effekt mit der Anzahl an Vergleichen k multipliziert. Dies geschieht analog zu den p-Werten.
- B** ☐ Beim multiplen Testen kann es zu einer Effektüberschätzung (Δ -Inflation) kommen. Daher müssen die Effekte angepasst werden. Dies geschieht nicht händisch sondern intern in den angewendeten Algorithmen.
- C** ☐ Wenn ein multipler Test gerechnet wird, dann muss der Effekt Δ adjustiert werden im Gegensatz zu den p-Werten.
- D** ☐ Beim multiplen Testen werden die Effekte der paarweisen Vergleiche ignoriert. Der Nachteil des multiplen Testens ist ja auch, dass wir am Ende keine Effekte mehr vorliegen haben. Eine ANOVA liefert hier bessere Informationen.
- E** ☐ Beim multiplen Testen muss der Effekt, hier der Mittelwertsunterschied Δ aus den paarweisen t-Tests, nicht adjustiert werden.

6 Aufgabe

(2 Punkte)

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

- A** ☐ Mit der Nullhypothese H_A und der Alternativhypothese H_0 gibt es zwei Hypothesen, die aber selten genutzt werden.
- B** ☐ Es gibt - bedingt durch das das Falsifikationsprinzip - ein Set von k Nullhypothesen, die iterative gegen $k - 1$ Alternativhypothesen getestet werden.
- C** ☐ Die Hypothesen H_0 und H_A sind rein prosarischer Natur und bilden keinen mathematischen Hintergrund ab. In der Statistik wird die wissenschaftliche Fragestellung getestet. Daher stehen auch die verständlichen Hypothesen im Mittelpunkt der biologischen Interpretation.
- D** ☐ Ein statistisches Hypothesenpaare gibt es. Zum einen die Nullhypothese H_0 und zum anderen die Alternativhypothese H_A oder H_1
- E** ☐ Es gibt ein statistisches Hypothesenpaar mit der Hypothese für und gegen die wissenschaftliche Fragestellung. Die Hypothesen werden H_{pro} und H_{contra} bezeichnet.

7 Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.89, 0.001, 0.03 und 0.34. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.004, 0.12 und 1. Die adjustierten p-Werte werden zu einem α -Niveau von 1.25% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.004, 0.12 und 1. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.2225, 3e-04, 0.0075 und 0.085. Die adjustierten p-Werte werden zu einem α -Niveau von 1.25% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.2225, 3e-04, 0.0075 und 0.085. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 3.56, 0.004, 0.12 und 1.36. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.

8 Aufgabe

(2 Punkte)

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

- A** ☐ Durch eine Randomisierung können wir von Strukturgleichheit zwischen der Stichprobe und der Grundgesamtheit ausgehen.
- B** ☐ Randomisierung war bis 1952 bedeutend, wurde dann aber in Folge besserer Rechnerleistung nicht mehr verwendet. Aktuelle Statistik nutzt keine Randomisierung mehr.
- C** ☐ Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte. Ohne Mittelwerte keine Varianz und somit auch kein statistischer Test.
- D** ☐ Durch eine Randomisierung können wir nicht von Strukturgleichheit zwischen der Stichprobe und der Grundgesamtheit ausgehen.
- E** ☐ Randomisierung erlaubt erst die Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich. Dadurch lässt sich erst ein Experiment auswerten.

9 Aufgabe

(2 Punkte)

Sie versuchen folgende Aussage richtig in die Analogie der statistischen Testtheorie zu setzen. Welche Analogie ist richtig?

H_0 ablehnen obwohl die H_0 gilt

- A** ☐ *Fire without alarm*, dem β -Fehler als Analogie von Rauch im Haus.
- B** ☐ In die Analogie eines Rauchmelders: *Alarm without fire*, dem α -Fehler.
- C** ☐ In die Analogie eines Rauchmelders: *Fire without alarm*, dem β -Fehler.
- D** ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.
- E** ☐ *Alarm with fire*, dem α -Fehler in der Analogie von Feuer.

10 Aufgabe

(2 Punkte)

Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein $\eta^2 = 0.78$. Welche Aussage ist richtig?

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

11 Aufgabe

(10 Punkte)

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



Visualisierung der linearen Regression 'Hä? Hatten wir das als Aufgabe nicht schon mal, das wir aus kontinuierlichen Daten eine Abbildung bauen sollten?', fragt Paula. Mark schaut fragend zurück. 'Kann mich wie immer an nichts erinnern. Können wir trotzdem jetzt erstmal die Daten auswerten? Columbo?', antwortet Mark leicht angespannt. Die beiden hatten ein Freilandversuch im Emsland mit Lauch durchgeführt. Dabei wurden die beiden folgenden Variablen gemessen: durchschnittlicher Regenwurmdichte [Anzahl/l] und Proteingehalt [g/kg]. Jetzt haben die beiden eigentlich alles zusammen. *Eigentlich...*

Durchschnittlicher Regenwurmdichte [Anzahl/l]	Proteingehalt [g/kg]
13.0	16.5
12.0	16.1
22.5	30.8
10.9	14.5
10.1	17.8
11.5	20.4
15.1	20.6
6.2	12.4
11.0	15.4
18.7	27.9
11.2	19.7

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Erstellen Sie eine Visualisierung für die Datentabelle. Beschriften Sie die Achsen! **(2 Punkte)**
3. Schätzen Sie die Regressionsgleichung aus der obigen Abbildung ab! **(2 Punkte)**
4. Beschriften Sie die Grade mit den statistischen Maßzahlen der linearen Regressionsgleichung! **(2 Punkte)**
5. Liegt ein Zusammenhang zwischen x und y vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
6. Wenn kein Zusammenhang zu beobachten wäre, wie würde die Grade aussehen? *Antworten Sie mit einer Skizze der Geraden!* **(1 Punkt)**

12 Aufgabe

(9 Punkte)

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



Visualisierung des Boxplots Jonas steht vor einem ersten Problem, denn wenn es nach seinem Betreuer geht, soll er in einem Versuch in einer Klimakammer Erbsen auswerten. Soweit eigentlich alles passend. Besser wäre was anderes gewesen. Jonas liebt Stricken. Darin kann er sich wirklich verlieren und immer wieder neu begeistern. Die Behandlung waren verschiedene Genotypen (AA und BB). In seiner Exceldatei hat er den Outcome (Y) *Proteingehalt* als *protein* aufgenommen. Nun soll Jonas die Daten einmal als Boxplots in einer Präsentation visualisieren, damit seinem Betreuer wieder klar wird, was er eigentlich nochmal gemacht hat und was für ein Ergebnis in einem statistischen Test zu erwarten wäre. Anhand von Boxplots lässt sich eine Aussage über die Normalverteilung von Y treffen. Wäre da nicht noch etwas. Wenn die Erschöpfung nicht wäre, ja dann wäre wohl vieles möglich für Jonas! Aber so.. Aber egal. Jonas will später nochmal raus um zu Schwimmen. Druck ablassen, dass muss er auch.

treatment	drymatter
BB	40.8
AA	27.9
BB	40.3
BB	43.4
AA	32.1
AA	34.4
AA	29.0
AA	35.9
BB	35.5
BB	43.6
AA	24.3
BB	33.8
AA	36.9
BB	39.1
AA	23.1
AA	25.5

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


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

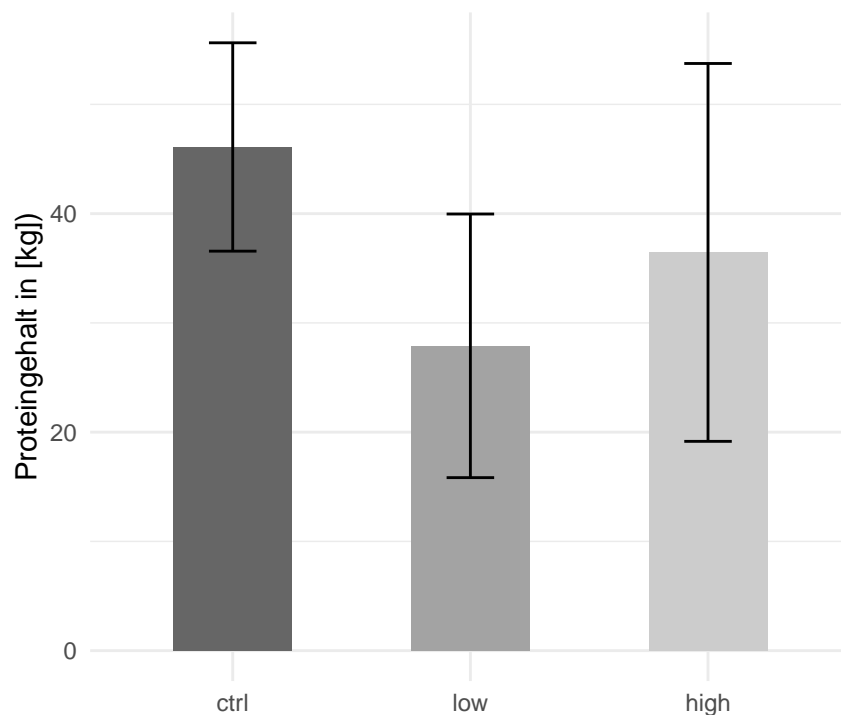
13 Aufgabe


(8 Punkte)


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



Zerforschen des Barplots 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. Harry Potter. Ein wunderbares Hobby um sich drin zu verlieren und Abstand zu bekommen. Paula denkt gerne über Harry Potter nach. 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  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 Düngestufen (*ctrl*, *low* 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  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  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)**

14 Aufgabe

(10 Punkte)

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



Die zweifaktorielle ANOVA in R Es ist schon kurz nach fünf und Tina wird langsam nervös. Tina wollte heute Abend noch ihre E-Sport Qualifikation schauen. Stattdessen versucht ihre Betreuerin die Ausgabe der zweifaktoriellen ANOVA zu visualisieren und zu überprüfen, ob es mit der Visualisierung der Daten als Boxplots zusammenpasst. Tina hatte im Emsland ein Kreuzungsexperiment mit Hühnern durchgeführt. Es gab dabei zwei Behandlungen. Einmal Ernährungszusatz (*ctrl*, *fedX* und *getIt*) sowie als zweite Behandlung Flüssignahrung (*ctrl* und *fLOW*). Gemessen wurde der Messwert (Y) Protein/Fettrate [%/kg]. So kompliziert kann das jetzt doch nicht sein! Eigentlich wollte Tina nachher noch zum Sport. Tina will später nochmal raus um zu Boxen. Druck ablassen, dass muss sie auch.

```
## Analysis of Variance Table
##
## Response: Protein/Fettrate
##
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Ernährungszusatz      2    1.95    0.974   0.0633 0.9388808
## Flüssignahrung        1 110.50 110.505   7.1783 0.0153118
## Ernährungszusatz:Flüssignahrung  2 338.49 169.243 10.9938 0.0007588
## Residuals            18 277.10   15.394
```

Leider kennt sich Tina 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)**

15 Aufgabe

(9 Punkte)



Grundgesamtheit und experimentelle Stichprobe 'Grundlage des statistischen Testen ist das Verständnis von der Grundgesamtheit (eng. *population* oder *ground truth*) und der experimentellen Stichprobe (eng. *sample*).', liest Nilufar laut aus dem Skript vor. Tina war kurz eingenickt und wird mit einem Stoß geweckt. 'Reiz dich zusammen und iss noch ein paar Takis Blue Heat das hilft mir immer. Alleine komme ich hier nicht weiter.', tadelt Nilufar Tina etwas zu forsch. 'War ne lange Nacht', mault Tina. Beide sollen in ihrer Abschlussarbeit einen statistischen Test interpretieren und versuchen die Grundlagen zu wiederholen. Tina war auf einem Konzert von Tocotronic.

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

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

16 Aufgabe

(9 Punkte)

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



```
## Error in eval(expr, envir, enclos): object 'attr_name_sec_vec' not found
```

Interpretation des t-Tests in R - die Visualisierung 'Grundlage des statistischen Testen ist das Verständnis von der Grundgesamtheit (eng. *population* oder *ground truth*) und der experimentellen Stichprobe (eng. *sample*).', liest Nilufar laut aus dem Skript vor. Tina war kurz eingenickt und wird mit einem Stoß geweckt. 'Reiz dich zusammen und iss noch ein paar Takis Blue Heat das hilft mir immer. Alleine komme ich hier nicht weiter.', tadelt Nilufar Tina etwas zu forsch. 'War ne lange Nacht', mault Tina. Beide sollen in ihrer Abschlussarbeit einen statistischen Test interpretieren und versuchen die Grundlagen zu wiederholen. Tina war auf einem Konzert von Tocotronic.

```
##
## Two Sample t-test
##
## data: Schlachtgewicht by Ernährungszusatz
## t = 0.24562, df = 16, p-value = 0.8091
## alternative hypothesis: true is not equal to [condensed]
## 95 percent confidence interval:
## -8.284795 10.456223
## sample estimates:
## mean in group ctrl mean in group fedX
## 43.28571 42.20000
```

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

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

17 Aufgabe

(12 Punkte)

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



Berechnung des Student t-Test Der t-Test. Yuki erschauert. Yuki und die Faulheit, eine unendliche Geschichte mit kniffligen Wendungen. Ein mächtiges Werkzeug ist der t-Test in den Händen desjenigen, der einen normalverteilten Endpunkt (Y) hat. Aber erstmal überhaupt den t-Test rechnen können. Wie sah das Experiment von Yuki überhaupt aus? Aus den Boxen wummert London Grammar und sein Mund ist verklebt von Reese's Peanut Butter Cups. 'Herrlich', denkt Yuki. Yuki hat einen Leistungssteigerungsversuch mit Milchvieh durchgeführt. Dabei wurde die Behandlung Genotypen (AA und BB) an den Milchvieh getestet. Gemessen hat Yuki dann als Messwert Schlachtgewicht [kg]. Warum der Versuch im Emsland für seinen Projektbericht stattfinden musste, ist ihm bis heute ein Rätsel. Egal. Gibt es jetzt einen Zusammenhang zwischen der Behandlung und Schlachtgewicht [kg]?

Genotypen	Schlachtgewicht
AA	27.4
AA	40.8
AA	30.9
BB	32.0
BB	20.5
BB	24.4
BB	22.3
AA	29.9
AA	25.0
AA	28.6
BB	32.6
BB	28.5
AA	26.8
BB	32.6
BB	28.9

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

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Bestimmen Sie die Teststatistik T_D eines Student t-Tests! **(3 Punkte)**
4. Treffen Sie mit $T_{\alpha=5\%} = 1.64$ eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effekt des Student t-Tests! **(1 Punkt)**
6. Wenn Sie *einen* Unterschied zwischen den Behandlungsgruppen erwarten würden, wie groß wäre dann der Effekt? Begründen Sie Ihre Antwort! **(2 Punkte)**
7. Formulieren Sie eine Antwort an Yuki über das Ergebnis Ihrer statistischen Analyse! **(2 Punkte)**