

Nicht bestanden: ☐

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

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

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

**M.Sc. Angewandte Nutztier- und Pflanzenwissenschaften**

# Klausur Biostatistik

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

27. Juni 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 62 Punkten sind aus dem Rechen- und Textteil erreicht.

\_\_\_\_\_ von 82 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
78.5 - 82.0	1,0
74.5 - 78.0	1,3
70.0 - 74.0	1,7
66.0 - 69.5	2,0
62.0 - 65.5	2,3
58.0 - 61.5	2,7
54.0 - 57.5	3,0
49.5 - 53.5	3,3
45.5 - 49.0	3,7
41.0 - 45.0	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	9	10	10	7	8	8	10

- Es sind \_\_\_\_ von 62 Punkten erreicht worden.

## 1 Aufgabe

(2 Punkte)

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

- A ☐ ... dass Modelle meist falsch sind und selten richtig.
- B ☐ ... dass Annahmen an statistische Modelle meist falsch sind.
- C ☐ ... dass Fehlerterme in statistischen Modellen nicht verifiziert werden können.
- 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.

## 2 Aufgabe

(2 Punkte)

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

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

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

## 4 Aufgabe

(2 Punkte)

Die ANOVA ist ein statistisches Verfahren welches häufig in den Auswertungen von Experimenten in den Agrarwissenschaften angewendet wird. Dabei wird die ANOVA als ein erstes statistischen Werkzeug für die Übersicht über die Daten benutzt. Eine ANOVA testet dabei ...

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

## 5 Aufgabe

(2 Punkte)

Nach einer simplen linearen Regression zur Untersuchung vom Einfluss der  $\text{CO}_2$ -Konzentration [ $\mu\text{g}$ ] im Wasser auf das Wachstum von Wasserlinsen [ $\text{kg}$ ] erhalten Sie einen  $\beta_1$  Koeffizienten von 0.00001 und einen hoch signifikanten  $p$ -Wert mit  $2.3 \cdot 10^{-9}$ . Warum sehen Sie so einen kleinen Effekt bei einer so deutlichen Signifikanz?

- A** ☐ Die 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.
- 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 Einheit der  $\text{CO}_2$ -Konzentration ist zu klein gewählt. Dadurch sehen wir den sehr kleinen  $p$ -Wert. Der  $p$ -Wert und die Einheit von der  $\text{CO}_2$ -Konzentration hängen zusammen.
- D** ☐ Die Einheit der  $\text{CO}_2$ -Konzentration ist zu klein gewählt. Die Erhöhung der  $\text{CO}_2$ -Konzentration um 1 führt nur zu einem sehr winzigen Anstieg im Gewicht der Wasserlinsen. Die Einheit muss besser gewählt werden.
- E** ☐ Das Gewicht und die  $\text{CO}_2$ -Konzentration korrelieren sehr stark, deshalb wird der  $\beta_1$  Koeffizient sehr klein.

## 6 Aufgabe

(2 Punkte)

Die Abkürzung CLD steht für welches statistische Verfahren? Welche anschließende Beschreibung der Interpretation ist korrekt?

- A** ☐ Contrast letter display. Unterschiede in den Behandlungen werden durch den gleichen Buchstaben oder Symbol dargestellt. Die Interpretation des CLD führt häufig in die Irre.
- B** ☐ Compact letter detection. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt.
- C** ☐ Compact letter display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Teilweise ist die Interpretation des CLD herausfordernd, da wir ja nach dem Unterschied suchen.
- D** ☐ Compound letter display. Gleichheit in den Outcomes wird durch den gleichen Buchstaben oder Symbol dargestellt. Teilweise ist die Interpretation des Verbunds (eng. compound) herausfordernd, da wir ja nach dem Unterschied suchen.
- E** ☐ Compact line display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Früher wurden keine Buchstaben sondern eine durchgezogene Linie verwendet. Bei mehr als drei Gruppen funktioniert die Linie aber graphisch nicht mehr.

## 7 Aufgabe

(2 Punkte)

Bei einem multiplen Vergleich oder Posthoc Test kann es zu einer Besonderheit beim statistischen Testen kommen. Wie nennt man diese Besonderheit beim statistischen Testen und wie kann man mit ihr umgehen?

- A** ☐ Beim multiplen Testen kann es zu einer  $\beta$ -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 20%. Daher müssen die  $p$ -Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der  $p$ -Werte nach Bonferroni das bekannteste Verfahren ist.
- B** ☐ Beim multiplen Testen kann es zu einer  $\alpha$ -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern weit darunter. Daher müssen die  $p$ -Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der  $p$ -Werte nach Welch das bekannteste Verfahren ist.
- C** ☐ Beim multiplen Testen kann es zu einer  $\alpha$ -Deflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern weit darunter. Daher müssen die  $p$ -Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der  $p$ -Werte nach Bonferroni das bekannteste Verfahren ist. Die  $p$ -Werte werden durch die Anzahl an Vergleichen geteilt

- D** ☐ Beim multiplen Testen kann es zu Varianzheterogenität kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5%. Daher müssen die p-Werte entsprechend adjustiert werden. Das Verfahren nach Welch, bekannt aus dem t-Test, ist hier häufig anzuwenden.
- E** ☐ Beim multiplen Testen kann es zu einer  $\alpha$ -Inflation kommen. Das globale Signifikanzniveau liegt nicht mehr bei 5% sondern sehr viel höher. Daher müssen die p-Werte entsprechend adjustiert werden. Hierfür gibt es verschiedene Verfahren, wobei das Verfahren zur Adjustierung der p-Werte nach Bonferroni das bekannteste Verfahren ist.

## 8 Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.21, 0.001, 0.01, 0.42 und 0.89. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.005, 0.05, 1 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.005, 0.05, 1 und 1. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.042, 2e-04, 0.002, 0.084 und 0.178. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.042, 2e-04, 0.002, 0.084 und 0.178. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1.05, 0.005, 0.05, 2.1 und 4.45. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.

## 9 Aufgabe

(2 Punkte)

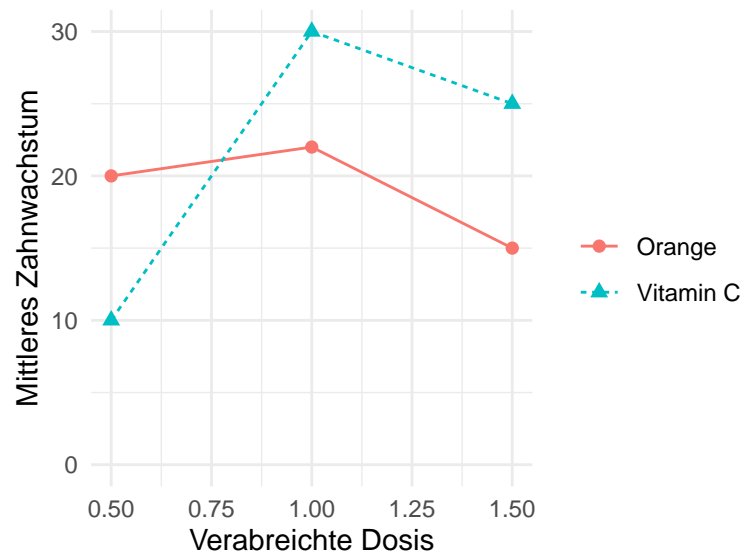
Nachdem Sie in einem Feldexperiment zu Leistungssteigerung von Erdbeeren durchgeführt haben, berechnen Sie den Mittelwert und den Median. Der Mittelwert  $\bar{y}$  und der Median  $\tilde{y}$  unterscheiden sich nicht. Welche Aussage ist richtig?

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

## 10 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 48 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 (Orange) oder Ascorbinsäure (Vitamin C) erhielt.



Welche Aussage ist richtig im Bezug auf eine zweifaktorielle ANOVA?


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

## 11 Aufgabe

(9 Punkte)

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



In einem Experiment für den Zuckergehalt von Erdbeeren in g/kg 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 0.0320553 0.0647177 0.7807677
## high 0.0320553 1.0000000 0.0002264 0.0586476
## low  0.0647177 0.0002264 1.0000000 0.0356299
## mid  0.7807677 0.0586476 0.0356299 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	9.90	0.99
high	9	5.28	6.76
low	9	13.84	3.68
mid	9	9.32	4.01

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




## 12 Aufgabe

(10 Punkte)

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



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


term	estimate	std.error	t statistic	p-value
(Intercept)	1.9291273	1.0835963		
light	0.9871926	0.1096176		

1. Zeichnen Sie die Gerade aus der obigen Tabelle in ein Koordinatenkreuz! **(1 Punkt)**
2. Beschriften Sie die Abbildung und die Gerade mit den statistischen Kenngrößen! **(2 Punkte)**
3. Formulieren Sie die Regressionsgleichung! **(2 Punkte)**
4. Berechnen Sie die t Statistik für *(Intercept)* und *light*! **(2 Punkte)**
5. 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)**


### 13 Aufgabe

(10 Punkte)



Der Datensatz *gain\_weight\_tbl* enthält Daten aus einer Studie zur Bewertung der Wirkung vom Vitamin Selen auf das Wachstum von Hühnern. Der Versuch wurde an 67 Hühnern durchgeführt, wobei jedes Tier eine von drei Selen-Dosen *dose* (0.2 ng/Tag, 1 ng/Tag und 10 ng/Tag) über eine von zwei Verabreichungsmethoden *form* erhielt (Wasser oder Festnahrung). Sie erhalten folgende Ausgabe in .

```
## Analysis of Variance Table
##
## Response: weight
##      Df  Sum Sq Mean Sq F value Pr(>F)
## dose      2   48.731    24.365   1.5723 0.23478
## form      1  107.350   107.350   6.9272 0.01692
## dose:form  2  164.558    82.279   5.3094 0.01540
## Residuals 18  278.944    15.497
```

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

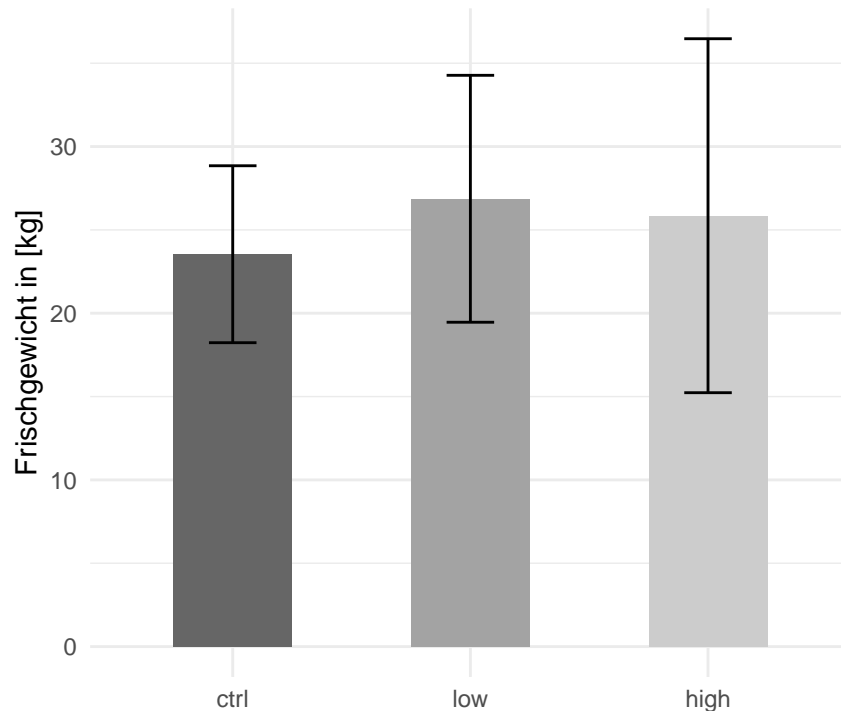
## 14 Aufgabe


(7 Punkte)

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



In einer Klimakammer mit drei Düngestufen (*ctrl*, *low* und *high*) als Behandlung (*treatment*) ergeben sich die folgenden Barplots mit dem gemessenen Frischgewicht (*freshmatter*) von Kartoffeln.



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

## 15 Aufgabe

(8 Punkte)



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

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

## 16 Aufgabe

(8 Punkte)



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

```
##  
## Two Sample t-test  
##  
## data: drymatter by Fe  
## t = -0.24574, df = 10, p-value = 0.8109  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -2.876338 2.304910  
## sample estimates:  
## mean in group high mean in group trt2  
## 19.00000 19.28571
```

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)**

## 17 Aufgabe

(10 Punkte)



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

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

