



HOCHSCHULE OSNABRÜCK
UNIVERSITY OF APPLIED SCIENCES

Fakultät Agrarwissenschaften und Landschaftsarchitektur

- ☐ Bachelorprüfung
☐ Masterprüfung

Klausur im Modul: _____
Prüfer*in: _____
Name, Vorname: _____
Matrikelnummer _____

Hiermit bestätige ich, dass ich prüfungsfähig bin. Weiter bestätige ich die Kenntnisnahme der umseitigen Klausurbelehrung.

Die für den Prüfungszeitraum ggf. geltenden Hygienebestimmungen habe ich zur Kenntnis genommen.

Ort, Datum, Unterschrift der*des Studierenden

Bewertung - von den Prüfenden auszufüllen -

Prüfer*in: _____
Unterschrift

Punkte: _____ von _____

Note: _____



Klausurbelehrung

Lesen Sie die nachstehende Belehrung bitte sorgfältig durch. Durch Ihre Unterschrift auf dem Klausurendeckblatt bestätigen Sie, die Bestimmungen des Allgemeinen Teils der Prüfungsordnung der Hochschule Osnabrück (ATPO) in der aktuell gültigen Fassung über Rücktritt, Täuschung und Ordnungsverstöße zur Kenntnis genommen zu haben.

Die nachstehenden Hinweise dienen lediglich der Information und verschaffen einen Überblick. Die für Ihre Prüfungen maßgeblichen Regelungen ergeben sich aus den einschlägigen Ordnungen.

Prüfungsfähigkeit aufgrund einer eigenen Krankheit (vgl. § 15 Abs. 2 ATPO)

- Prüflinge, die eine Klausur empfangen, nehmen an der Klausur teil und erklären sich mit Unterschrift auf dem Deckblatt für prüfungsfähig. Prüflinge, die sich gesundheitlich nicht in der Lage fühlen, die Klausur zu bewältigen, müssen vor Beginn der Klausur den Prüfungsraum verlassen. Ein Prüfungsrücktritt nach Beginn der Klausur ist in der Regel ausgeschlossen.

Täuschung und Ordnungsverstöße (vgl. § 15 Abs. 3, 4 ATPO)

- Das Mitsichführen von Unterlagen mit fachlichem Bezug ist untersagt, außer sie sind ausdrücklich zugelassen worden.
- Informations- und kommunikationsfähige Geräte sind abzuschalten, es sei denn, ein Gebrauch ist ausdrücklich zugelassen worden.
- Der Prüfling trägt die Verantwortung dafür, dass sowohl er als Person als auch sein Prüfungsplatz von zur Täuschung geeigneten Materialien frei ist.
- Es gilt ein Sprechverbot für alle Prüflinge untereinander.
- Das Verlassen des Klausurraums ist nur zu dringenden WC-Gängen erlaubt. Es darf jeweils nur ein Prüfling zur gleichen Zeit den Klausurraum verlassen. Das Sprechen mit anderen Studierenden oder dritten Personen im Rahmen eines WC-Gangs ist untersagt.
- Es dürfen nur die Schreibblätter verwendet werden, die von den Prüfungsaufsichten ausgegeben werden, es sei denn, andere Schreibblätter wurden ausdrücklich zugelassen.
- Das Weiterschreiben nach Bekanntgabe des Bearbeitungszeitendes ist untersagt.
- Der Prüfling trägt die Verantwortung für die Abgabe seiner vollständigen Klausur.
- Verstöße gegen diese Verbote können als Täuschungsversuche gewertet werden. Die Prüfungsleistung wird mit „nicht bestanden“ bzw. „nicht ausreichend“ bewertet.
- Auch hier nicht genannte, weitere Formen von Täuschungshandlungen können vergleichbare Konsequenzen nach sich ziehen.
- Prüflinge, die den ordnungsgemäßen Ablauf der Prüfungen stören, können von der aufsichtführenden Person von der Klausur ausgeschlossen werden. Bei festgestellten Ordnungsverstößen wird die Prüfungsleistung mit „nicht bestanden“ bzw. „nicht ausreichend“ bewertet.

Mitwirkungspflichten (allgemeiner Grundsatz)

- Für die aufsichtführende Person nicht ohne Weiteres erkennbare Störungen (störende Geräusche, wackeliger Stuhl/Tisch etc.) sind durch den Prüfling anzuzeigen.

Sonstiges

- Abgelegte Prüfungsleistungen ohne Anmeldung und Zulassung werden nicht bewertet (§ 12 Abs. 3 S. 5 ATPO).

Name: _____

Nicht bestanden: ☐

Vorname: _____

Matrikelnummer: _____

Endnote: _____

**B.Sc. Landwirtschaft; B.Sc. Angewandte Pflanzenbiologie - Gartenbau,
Pflanzentechnologie**

Klausur Spezielle Statistik und Versuchswesen

Hochschule Osnabrück

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

Klausur vom 27. Januar 2023

Erlaubte Hilfsmittel für die Klausur

- Normaler Taschenrechner ohne Möglichkeit der Kommunikation mit anderen Geräten - also ausdrücklich kein Handy!
- Eine DIN A4-Seite als beidseitig, selbstgeschriebene, handschriftliche Formelsammlung - keine digitalen Ausdrucke.

Ergebnis der Klausur

_____ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.
_____ von 61 Punkten sind aus dem Rechen- und Textteil erreicht.
_____ von 81 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
78 - 81	1,0
75 - 77	1,3
70 - 74	1,7
65 - 69	2,0
59 - 64	2,3
54 - 58	2,7
49 - 53	3,0
44 - 48	3,3
41 - 43	3,7
40	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							

- Es sind ____ von 61 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

Der Datensatz PlantGrowth enthält das Gewicht von Pflanzen, die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden. Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein $\eta^2 = 0.37$. Welche Aussage ist richtig?

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

2 Aufgabe

(2 Punkte)

Eine einfaktorielle ANOVA berechnet eine Teststatistik um zu die Nullhypothese abzulehnen. Welche Aussage über die Teststatistik der ANOVA ist richtig?

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

3 Aufgabe

(2 Punkte)

Welche Aussage über die α Adjustierung ist richtig?

- A** ☐ Die α Adjustierung wird durchgeführt um bei multiplen Vergleichen den Fehler 1. Art zu kontrollieren. Es wird die Irrtumswahrscheinlichkeit adjustiert, daher das α -Niveau.
- B** ☐ Die α Adjustierung wird durchgeführt um den Effekt von Interesse, meist die Behandlung, von anderen Effekten zu trennen. Daher eine Adjustierung auf den β -Werten einer Regression.
- C** ☐ Die α ist notwendig um Effekte gegeneinander aufzurechnen. Ohne diese Adjustierung würde der eigentliche Effekt nicht richtig geschätzt. Daher handelt es sich um eine Adjustierung der Fehlerwahrscheinlichkeiten.
- D** ☐ Die α Adjustierung wird durchgeführt um den Fehler 2. Art zu kontrollieren. Ohne diese Adjustierung würde der Fehler 2. Art nicht bei 80% liegen sondern sehr schnell gegen 0 laufen.
- E** ☐ Die α Adjustierung wird meist ignoriert. Wenn die Annahmen an den statistischen Test richtig sind, kann auf eine Adjustierung verzichtet werden.

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% ist richtig?

- A** ☐ 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.
- B** ☐ Die Festlegung von $\alpha = 5\%$ ist eine Kulturkonstante. Wissenschaftler benötigt eine Schwelle für eine statistische Testentscheidung, der Wert von α wurde aber historisch mehr zufällig gewählt.
- C** ☐ Im Rahmen eines langen Disputs zwischen Neyman und Fischer wurde $\alpha = 5\%$ festgelegt. Leider werden die Randbedingungen und Voraussetzungen an statistische Modelle heute immer wieder ignoriert.
- D** ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.
- E** ☐ Der Wert ergab sich aus einer Auswertung von 1042 wissenschaftlichen Veröffentlichungen zwischen 1914 und 1948. Der Wert 5% wurde in 28% der Veröffentlichungen genutzt. Daher legte man sich auf diese Zahl fest.

5 Aufgabe

(2 Punkte)

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Kühen durch. Bei 6 Tieren finden Sie eine Erkrankung der Klauen vor und 12 Tiere sind gesund. Welche Aussage über den Odds ratio Effektschätzer ist richtig?

- A** ☐ Es ergibt sich ein Odds ratio von 0.33, da es sich um eine Chancenverhältnis handelt.
- B** ☐ Es ergibt sich ein Odds ratio von 0.33, da es sich um ein Anteil handelt.
- C** ☐ Es ergibt sich ein Odds ratio von 0.5, da es sich um eine Chancenverhältnis handelt.
- D** ☐ Es ergibt sich ein Odds ratio von 2, da es sich um ein Anteil handelt.
- E** ☐ Es ergibt sich ein Odds ratio von 0.5, da es sich um ein Anteil handelt.

6 Aufgabe

(2 Punkte)

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

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

7 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 sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- B** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.

- C** ☐ Randomisierung erlaubt erst die Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich.
- D** ☐ Randomisierung war bis 1952 bedeutend, wurde dann aber in Folge besserer Rechnerleistung nicht mehr verwendet. Aktuelle Statistik nutzt keine Randomisierung mehr.
- E** ☐ Randomisierung erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte.

8 Aufgabe

(2 Punkte)

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

- A** ☐ Das Konfidenzintervall. Durch die Visualisierung des Konfidenzintervalls kann eine Relevanzschwelle vom Anwender definiert werden. Zusätzlich erlaubt das Konfidenzintervall auch eine Entscheidung über die Signifikanz.
- B** ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- C** ☐ Der p-Wert. Durch den Vergleich mit α lässt sich über die Signifikanz entscheiden und der β -Fehler erlaubt über die Power eine Einschätzung der Relevanz.
- D** ☐ Die Teststatistik. Durch den Vergleich von T_c zu T_k ist es möglich die H_0 abzulehnen. Die Relevanz ergibt sich aus der Fläche rechts vom dem T_c -Wert.
- E** ☐ Das Δ . Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da Δ antiproportional zum p-Wert ist, bedeutet auch ein hohes Δ ein sehr kleinen p-Wert.

9 Aufgabe


(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.34, 0.02, 0.42 und 0.03. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1.36, 0.08, 1.68 und 0.12. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.085, 0.005, 0.105 und 0.0075. Die adjustierten p-Werte werden zu einem α -Niveau von 1.25% verglichen.
- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.085, 0.005, 0.105 und 0.0075. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.08, 1 und 0.12. Die adjustierten p-Werte werden zu einem α -Niveau von 1.25% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.08, 1 und 0.12. Die adjustierten p-Werte werden zu einem α -Niveau von 5% verglichen.

10 Aufgabe

(2 Punkte)

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

- A** ☐ Wir lesen die Daten über eine generische Funktion `read()` ein und müssen dann die Funktion `ggplot()` nur noch installieren. Dann haben wir die Abbildungen als *.png vorliegen.
- B** ☐ 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.
- 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.
- D** ☐ Wir transformieren die Spalten über `mutate()` in ein `tibble` und können dann über `ggplot()` uns die Abbildungen erstellen lassen. Dabei beachten wir das wir keine Faktoren in den Daten haben.

- E** ☐ Wir lesen 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.

11 Aufgabe

(12 Punkte)


Nach einem Experiment mit zwei Pestiziden (*RoundUp* und *GoneEx*) ergibt sich die folgende Datentabelle mit dem gemessenen Trockengewicht (*drymatter*) von Weizen.

pesticide	drymatter
GoneEx	16
GoneEx	12
GoneEx	13
RoundUp	15
RoundUp	13
RoundUp	15
RoundUp	16
GoneEx	17
RoundUp	15
GoneEx	15
GoneEx	13
GoneEx	14

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik T_{calc} eines Student t-Tests für den Vergleich der beiden Pestizide. Geben Sie den Rechenweg und die Formeln mit an! **(5 Punkte)**
4. Treffen Sie mit $T_{\alpha=5\%} = 2.04$ und dem berechneten T_{calc} eine Aussage zur Nullhypothese! **(2 Punkte)**
5. Wenn Sie keinen Unterschied zwischen den beiden Pestiziden erwarten würden, wie groß wäre dann die Teststatistik T_{calc} ? Begründen Sie Ihre Antwort! **(2 Punkte)**

12 Aufgabe

(8 Punkte)

Der Datensatz *PigGain* enthält Daten aus einer Studie zur Bewertung der Wirkung vom Vitamin Selen auf das Wachstum bei Mastschweinen. Der Versuch wurde an 65 Mastschweinen durchgeführt, wobei jedes Tier eine von drei Selen-Dosen (0.5 ng/Tag, 1 ng/Tag und 5 ng/Tag) über eine von zwei Verabreichungsmethoden erhielt (Wasser oder Festnahrung). Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: len
##      Df Sum Sq Mean Sq  F value    Pr(>F)
## supp   1  28.92   28.918    1.80570    0.1841703
## dose    2 3482.38 1741.192   108.72522 < 0.000000000000000222 ***
## supp:dose 2  165.76   82.879    5.17522    0.0084954 **
## Residuals 59  944.86   16.015
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

13 Aufgabe

(6 Punkte)

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

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

14 Aufgabe

(8 Punkte)


Sie rechnen einen t-Test. Sie schätzen einen Mittelwertsunterschied.

1. Beschriften Sie die untenstehende Abbildung mit der Signifikanzschwelle! **(1 Punkt)**
2. Ergänzen Sie eine Relevanzschwelle! **(1 Punkt)**
3. Skizzieren Sie in die untenstehende Abbildung fünf einzelne Konfidenzintervalle (a-e) mit den jeweiligen Eigenschaften! **(6 Punkte)**
 - (a) Ein nicht signifikantes, relevantes Konfidenzintervall
 - (b) Ein signifikantes, nicht relevantes Konfidenzintervall
 - (c) Ein signifikantes, relevantes Konfidenzintervall
 - (d) Ein Konfidenzintervall mit niedriger Varianz s_p in der Stichprobe als der Rest der Konfidenzintervalle
 - (e) Ein Konfidenzintervall mit höherer Varianz s_p in der Stichprobe als der Rest der Konfidenzintervalle



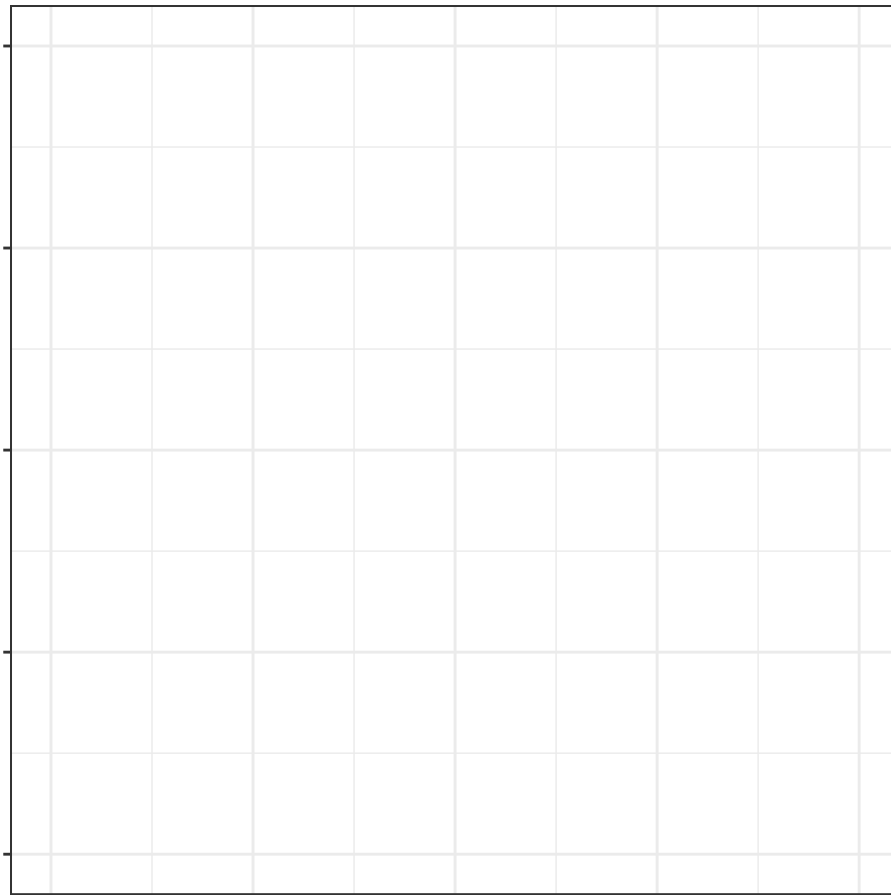
15 Aufgabe

(10 Punkte)

In einem Stallexperiment mit $n = 100$ Ferkeln wurde der Gewichtszuwachs unter bestimmten Lichtverhältnissen gemessen. Sie erhalten den  Output der Funktion `tidy()` einer simplen Gaussian linearen Regression sieben Wochen nach der ersten Messung.

term	estimate	std.error	t statistic	p-value
(Intercept)	25.74	1.01		
light	1.42	0.10		

1. Berechnen Sie die t Statistik für *(Intercept)* und *light*! **(2 Punkte)**
2. Schätzen Sie den p-Wert für *(Intercept)* und *light* mit $T_k = 1.96$ ab. Was sagt Ihnen der p-Wert aus? Begründen Sie Ihre Antwort! **(3 Punkte)**
3. Zeichnen Sie die Gerade aus der obigen Tabelle in die untenstehende Abbildung! **(1 Punkt)**
4. Beschriften Sie die Abbildung und die Gerade mit den statistischen Kenngrößen! **(2 Punkte)**
5. Formulieren Sie die Regressionsgleichung! **(2 Punkte)**



16 Aufgabe

(9 Punkte)

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

trt	drymatter
A	9.1
B	17.4
A	17.9
B	21.9
A	6.4
A	10.9
A	10.1
B	17.7
A	3.6
B	11.2
B	20.2
B	18.4
A	5.1
B	11.8
A	10.4
A	13.5
A	5.6
A	15.8


1. Zeichnen Sie in *einer* Abbildung die beiden Boxplots für die zwei Düngestufen A und B! Beschriften Sie die Achsen entsprechend! **(6 Punkte)**
2. Beschriften Sie *einen* der beiden Boxplots mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen den Düngestufen erwarten würden, wie sehen dann die beiden Boxplots aus? **(1 Punkt)**

17 Aufgabe

(7 Punkte)

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

water	light	P	Fe	drymatter	nematodes
11.74	0	10.05	100.63	67.86	10
8.38	1	9.23	103.07	67.19	5
10.25	1	8.84	102.45	75.68	9
9.87	0	8.93	98.45	68.24	8

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

```
##
## Call:
## lm(formula = reformulate(response = "drymatter", termlabels = wanted_vec),
##     data = data_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.49504 -1.47539  0.00437  1.66276  5.75888
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  5.58029     8.47369   0.6585    0.51177
## Fe           0.65448     0.08358   7.8306 0.0000000000006475 ***
## light       1.03844     0.55354   1.8760    0.06369 .
## water      -0.18008     0.18362  -0.9807    0.32921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.5528 on 96 degrees of freedom
## Multiple R-squared:  0.41186, Adjusted R-squared:  0.39348
## F-statistic: 22.409 on 3 and 96 DF,  p-value: 0.000000000044164
```

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