



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.Eng. Wirtschaftsingenieurwesen Agrar/Lebensmittel**

# **Klausur Statistik**

**Hochschule Osnabrück**

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

Klausur vom 24. 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 60 Punkten sind aus dem Rechen- und Textteil erreicht.  
\_\_\_\_\_ von 80 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
78 - 80	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 60 Punkten erreicht worden.

## 1 Aufgabe

(2 Punkte)

Berechnen Sie den Median und das IQR von  $x$  mit 18, 13, 24, 14, 17, 31, 13, 23, 21, 17 und 63.

- A ☐ Es ergibt sich 23 +/- 14
- B ☐ Es ergibt sich 23 [14, 24]
- C ☐ Es ergibt sich 18 +/- 24
- D ☐ Es ergibt sich 18 +/- 14
- E ☐ Es ergibt sich 18 [14, 24]

## 2 Aufgabe

(2 Punkte)

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

- A ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.
- B ☐ Das  $\Delta$ . Durch die Effektstärke haben wir einen Wert für die Relevanz, die vom Anwender bewertet werden muss. Da  $\Delta$  antiproportional zum p-Wert ist, bedeutet auch ein hohes  $\Delta$  ein sehr kleinen p-Wert.
- C ☐ Der p-Wert. Durch den Vergleich mit  $\alpha$  lässt sich über die Signifikanz entscheiden und der  $\beta$ -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 Konfidenzintervall. Durch die Visualisierung des Konfidenzintervalls kann eine Relevanzschwelle vom Anwender definiert werden. Zusätzlich erlaubt das Konfidenzintervall auch eine Entscheidung über die Signifikanz.

## 3 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 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.
- B ☐ Die ANOVA berechnet die F-Statistik aus den SS Behandlung geteilt durch die SS Fehler.
- C ☐ 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.
- D ☐ 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.
- 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.

## 4 Aufgabe

(2 Punkte)

Welche Aussage über den t-Test ist richtig?

- A ☐ Der t-Test vergleicht die Varianzen von mindestens zwei oder mehr Gruppen
- B ☐ Der t-Test vergleicht die Mittelwerte von zwei Gruppen unter der strikten Annahme von Varianzhomogenität. Sollte keine Varianzhomogenität vorliegen, so gibt es keine Möglichkeit den t-Test in einer Variante anzuwenden.
- C ☐ Der t-Test vergleicht die Mittelwerte von zwei Gruppen.
- D ☐ Der t-Test testet generell zu einem erhöhten  $\alpha$ -Niveau von 20%.
- E ☐ Der t-Test ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern

## 5 Aufgabe

(2 Punkte)

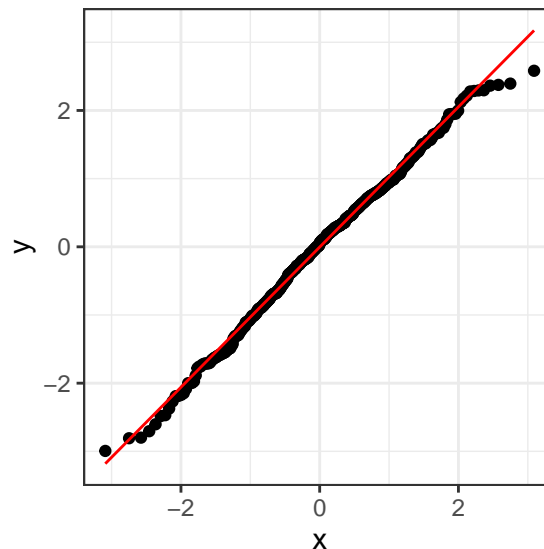
Eine der gängigsten Methode der Statistik um einen Fehler zu bestimmen ist...

- A** ☐ ... die kleinste Quadrate Methode oder auch least square method genannt.
- B** ☐ ... die Methode der aufaddierten, absoluten Abstände.
- C** ☐ ... die Methode des absoluten Abstands.
- D** ☐ ... die Methode des absoluten, quadrierten Abstands.
- E** ☐ ... das Produkt der kleinsten Quadrate.

## 6 Aufgabe

(2 Punkte)

Sie rechnen in eine linearen Regression und erhalten folgenden QQ Plot. Welche Aussage ist richtig?



- A** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden.
- B** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- C** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden und Korrelation ist negativ.
- D** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Geraden.
- E** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil nicht auf der Geraden.

## 7 Aufgabe

(2 Punkte)

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

- A** ☐ 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.
- B** ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.

- 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** ☐ Die Festlegung von  $\alpha = 5\%$  ist eine Kulturkonstante. Wissenschaftler benötigt eine Schwelle für eine statistische Testentscheidung, der Wert von  $\alpha$  wurde aber historisch mehr zufällig gewählt.
- E** ☐ 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.

## 8 Aufgabe

(2 Punkte)

In einer linearen Regression werden die  $\epsilon$  oder Residuen geschätzt. Welcher Verteilung folgen die Residuen bei einer optimalen Modellierung?

- A** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(0, s^2)$ .
- B** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(\bar{y}, s^2)$ .
- C** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(0, 1)$ .
- D** ☐ Die Residuen sind binomialverteilt.
- E** ☐ Die Residuen folgen einer Poissonverteilung mit  $\text{Pois}(0)$ .

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

## 10 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  $\eta^2 = 0.21$ . Welche Aussage ist richtig?

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



## 11 Aufgabe

(8 Punkte)

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

```
##  
## Two Sample t-test  
##  
## data: waterintake by infusion  
## t = -1.80637, df = 12, p-value = 0.095986  
## alternative hypothesis: true difference in means between group high and group low is not equal to  
## 95 percent confidence interval:  
## -7.9422662 0.7422662  
## sample estimates:  
## mean in group high mean in group low  
## 15.0 18.6
```

1. Formulieren Sie die wissenschaftliche Fragestellung! **(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)**

## 12 Aufgabe

(8 Punkte)

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

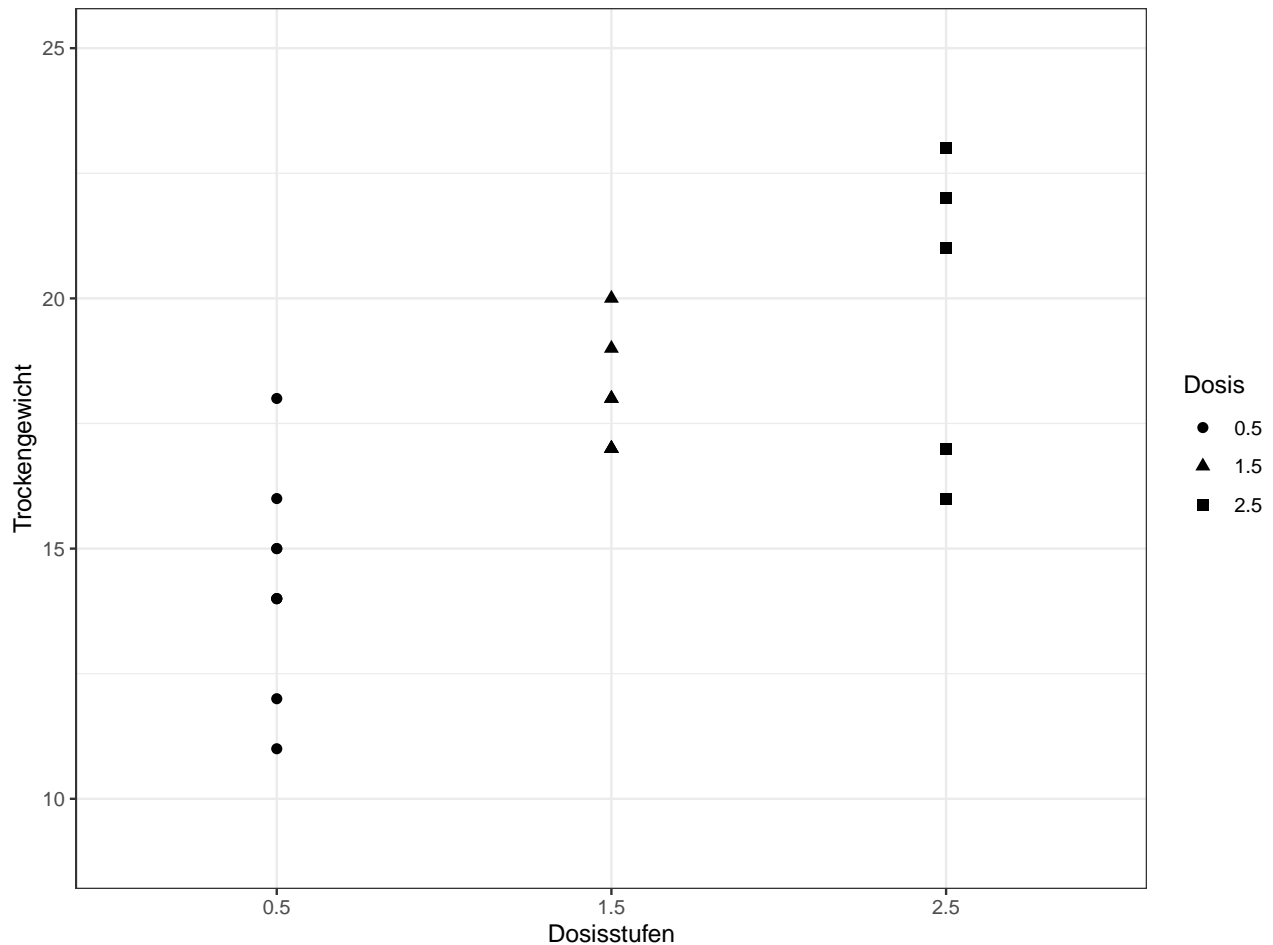
1. Beschriften Sie die untenstehende Abbildung mit der Signifikanzschwelle! **(1 Punkt)**
2. Ergänzen Sie eine Relevanzschwelle! **(1 Punkt)**
3. Skizzieren Sie in die untenstehende Abbildung fünf einzelne Konfidenzintervalle (a-e) mit den jeweiligen Eigenschaften! **(6 Punkte)**
  - (a) Ein signifikantes,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



### 13 Aufgabe

(8 Punkte)

In einem Experiment wurde der Ertrag von Erbsen unter drei verschiedenen Pestizid-Dosen 0.5 g/l, 1.5 g/l und 2.5 g/l gemessen. Unten stehenden sehen Sie die Visualisierung des Datensatzes.



1. Zeichnen Sie folgende statistischen Masszahlen in die Abbildung ein! (6 Punkte)

- Total (grand) mean:  $\beta_0$
- Mittelwerte der Dosen:  $\bar{y}_{0.5}$ ,  $\bar{y}_{1.5}$  und  $\bar{y}_{2.5}$
- Effekt der einzelnen Level der Dosen:  $\beta_{0.5}$ ,  $\beta_{1.5}$ , und  $\beta_{2.5}$
- Residuen oder Fehler:  $\epsilon$

2. Schätzen Sie den p-Wert einer einfaktoriellen ANOVA ab. Liegt ein *vermutlicher* signifikanter Unterschied zwischen den Dosisstufen vor? Begründen Sie Ihre Antwort! (2 Punkte)

## 14 Aufgabe

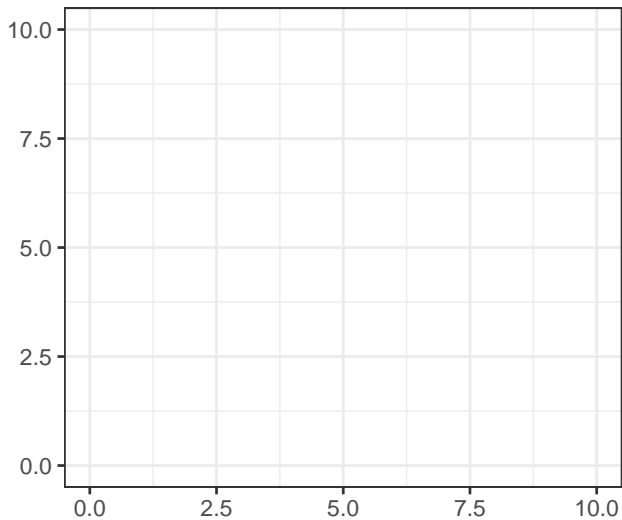
(9 Punkte)

Im folgenden sehen Sie drei leere Scatterplots. Füllen Sie diese Scatterplots nach folgenden Anweisungen.

1. Zeichnen Sie für die angegebene  $\rho$ -Werte eine Gerade in die entsprechende Abbildung! **(3 Punkte)**
2. Zeichnen Sie für die angegebenen  $R^2$ -Werte die entsprechende Punktwolke um die Gerade. **(3 Punkte)**
3. Sie rechnen ein statistisches Modell. Was sagen Ihnen die  $R^2$ -Werte über das jeweilige Modell? **(3 Punkte)**

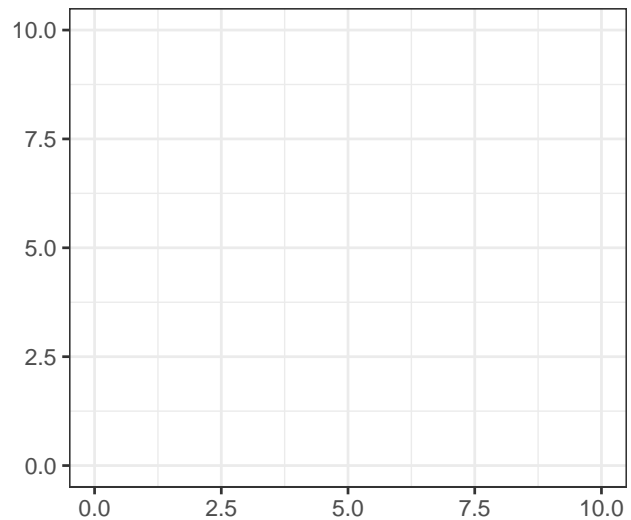
Pearsons  $\rho = 1$

$R^2 = 1$



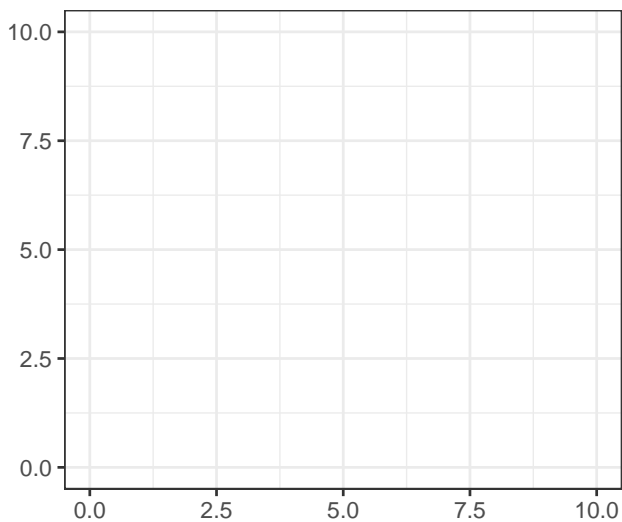
Pearsons  $\rho = 0$

$R^2 = 0.75$



Pearsons  $\rho = -1$

$R^2 = 0.25$



## 15 Aufgabe

(7 Punkte)


Nach einem Gewächshausexperiment mit drei Bewässerungstypen (*low*, *mid* und *high*) ergibt sich die folgende Datentabelle mit dem gemessenen Frischgewicht (*freshmatter*).

water_type	freshmatter
high	19
mid	24
high	16
high	20
high	19
mid	25
mid	27
low	12
low	16
low	14
mid	27
low	17

1. Zeichnen Sie in *einer* Abbildung die Barplots für die Bewässerungstypen! Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Beschriften Sie *einen* Barplot mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen der Bewässerungstypen erwarten würden, wie sehen dann die Barplots aus? **(1 Punkt)**

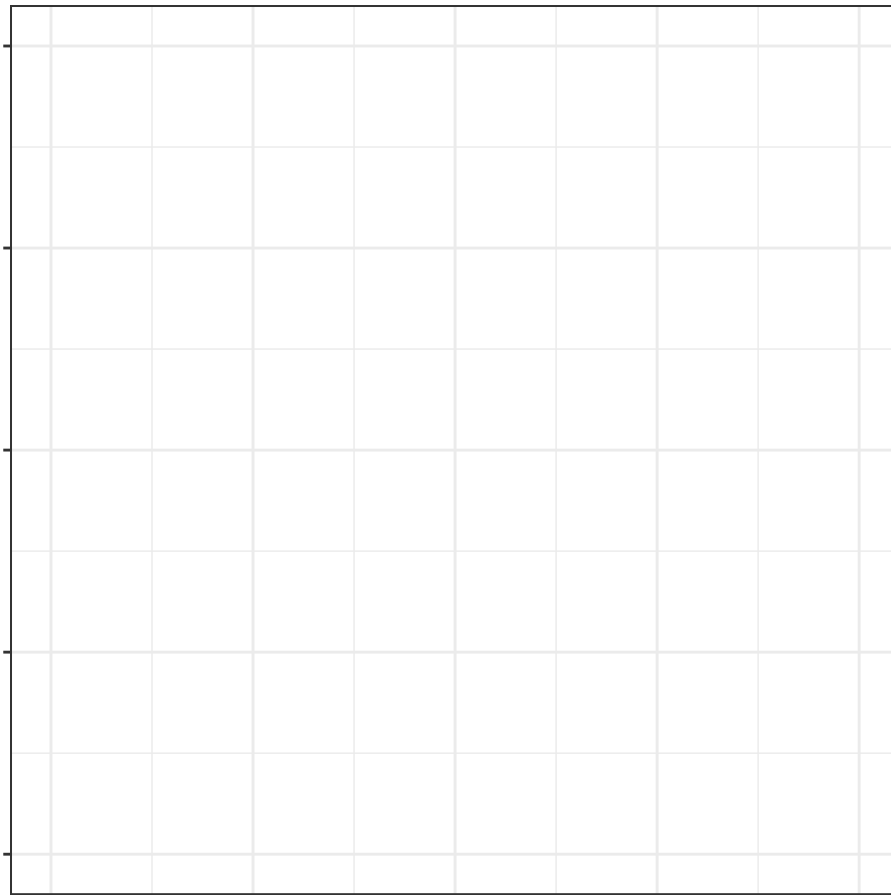
## 16 Aufgabe

(10 Punkte)

In einem Stallexperiment mit  $n = 98$  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)	25.76	1.11		
light	1.44	0.11		

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)**



## 17 Aufgabe

(8 Punkte)

Der Datensatz *Crop* enthält das Trockengewicht der Maispflanzen (*drymatter*), die unter drei verschiedenen Düngerbedingungen erzielt wurden – dem Faktor *trt* mit den Faktorstufen *low*, *mid*, *high*. Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: drymatter
##           Df    Sum Sq Mean Sq F value    Pr(>F)
## trt         2    3.05157  1.52578  3.22242 0.055586 .
## Residuals  27   12.78423  0.47349
## ---
## 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 einfaktorielle ANOVA auf! **(2 Punkte)**
2. Interpretieren Sie das Ergebnis der einfaktoriellen ANOVA! **(2 Punkt)**
3. Berechnen Sie den Effektschätzer  $\eta^2$ . Was sagt Ihnen der Wert von  $\eta^2$  aus? **(2 Punkte)**
4. Skizzieren Sie eine Abbildung, der dem obigen Ergebnis der einfaktoriellen ANOVA näherungsweise entspricht! **(2 Punkte)**