



HOCHSCHULE OSNABRÜCK
UNIVERSITY OF APPLIED SCIENCES

Fakultät Agrarwissenschaften und Landschaftsarchitektur

- ☐ Bachelorprüfung
☐ Masterprüfung

Klausur im Modul: _____
Prüfer*in: _____
Name, Vorname: _____
Matrikelnr.: _____ Fachsemester.: _____
Tisch Nr.: _____ Zugel. Hilfsmittel: _____

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 -

1. Prüfer*in: _____ Unterschrift	2. Prüfer*in: _____ Unterschrift
Punkte: _____ von _____	Punkte: _____ von _____
Note: _____	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.Eng. Wirtschaftsingenieurwesen im Agri- und Hortibusiness,
B.Sc. Angewandte Pflanzenbiologie - Gartenbau, Pflanzentechnologie**

Klausur Angewandte Statistik und Versuchswesen

Hochschule Osnabrück

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

Klausur vom 06. Juli 2022

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)

Welche Aussage zum mathematische Ausdruck $Pr(D|H_0)$ ist richtig?

- A ☐ Die Wahrscheinlichkeit für die Nullhypothese, wenn die Daten wahr sind.
- B ☐ Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativhypothese überdeckt.
- C ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit die Daten D zu beobachten wenn die Nullhypothese wahr ist.
- D ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit der Alternativhypothese und somit $1 - Pr(H_A)$
- E ☐ Die Wahrscheinlichkeit der Daten unter der Nullhypothese in der Grundgesamtheit.

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 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.
- C ☐ 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.
- D ☐ 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.
- 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)

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

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

4 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 $\eta^2 = 0.27$. Welche Aussage ist richtig?

- A ☐ 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.
- 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 ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.
- E ☐ Die Berechnung von η^2 ist ein Wert für die Interaktion.

5 Aufgabe

(2 Punkte)

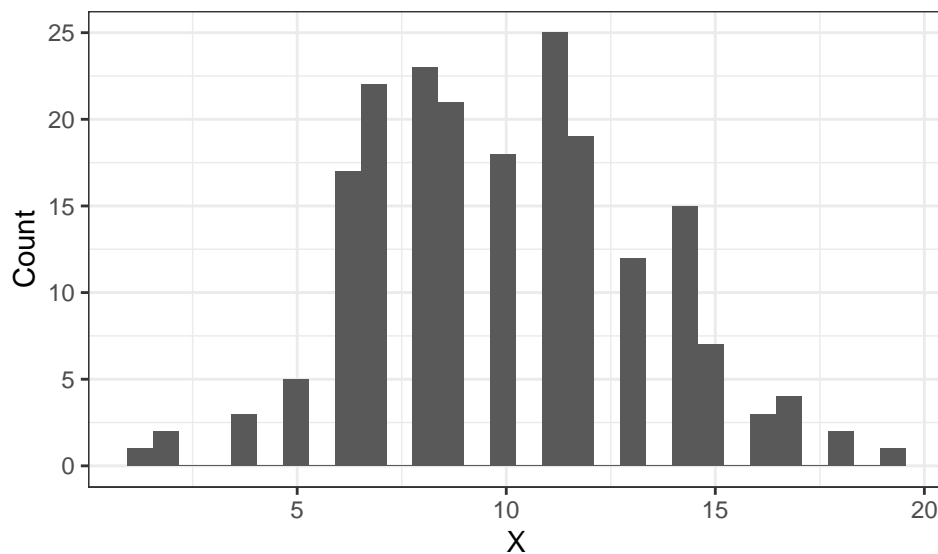
Sie haben das abstrakte Modell $Y \sim X$ vorliegen. Welche Aussage über X ist richtig?

- ☐ A X beinhaltet eine Spalte. Die Spalte gibt die Verteilungsfamilie vor.
- ☐ B X beinhaltet mehrere Spalten. Die Spalten geben die Verteilungsfamilie vor.
- ☐ C X beinhaltet eine Spalte. Die Spalte gibt nicht die Verteilungsfamilie vor.
- ☐ D X beinhaltet mehrere Spalten. Die Spalten enthalten die Behandlung und weitere potenzielle Einflussvariablen
- ☐ E X beinhaltet die Zeilen. Die Zeilen geben die Verteilungsfamilie vor.

6 Aufgabe

(2 Punkte)

In dem folgenden Histogramm von $n = 200$ Pflanzen ist welche Verteilung mit welchen korrekten Verteilungsparametern dargestellt?



- ☐ A Es handelt sich um eine Poisson-Verteilung mit $\text{Pois}(10)$.
- ☐ B Es handelt sich um eine Binomial-Verteilung mit $\text{Binom}(10)$.
- ☐ C Es handelt sich um eine Normalverteilung mit $N(10, 5)$.
- ☐ D Eine Standardnormalverteilung mit $N(0,1)$.
- ☐ E Eine rechtsschiefe, multivariate Normalverteilung.

7 Aufgabe

(2 Punkte)

Sie führen ein Experiment zur Behandlung von Klaueninfektionen bei Kühen durch. Bei 3 Tieren finden Sie eine Erkrankung der Klauen vor und 8 Tiere sind gesund. Welche Aussage über den Risk ratio Effektschätzer ist richtig?

- ☐ A Es ergibt sich ein Risk ratio von 0.38, da es sich um ein Anteil handelt.
- ☐ B Es ergibt sich ein Risk ratio von 0.27, da es sich um eine Chancenverhältnis handelt.
- ☐ C Es ergibt sich ein Risk ratio von 0.38, da es sich um eine Chancenverhältnis handelt.
- ☐ D Es ergibt sich ein Risk ratio von 2.67, da es sich um ein Anteil handelt.
- ☐ E Es ergibt sich ein Risk ratio von 0.27, da es sich um ein Anteil handelt.

8 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 des absoluten Abstands.
- C ☐ ... das Produkt der kleinsten Quadrate.
- D ☐ ... die Methode des absoluten, quadrierten Abstands.
- E ☐ ... die Methode der aufaddierten, absoluten Abstände.

9 Aufgabe

(2 Punkte)

In der Theorie zur statistischen Testentscheidung kann „ H_0 ablehnen obwohl die H_0 gilt“ in welche richtige Analogie gesetzt werden?

- A ☐ In die Analogie eines Rauchmelders: *Alarm without fire*, dem α -Fehler.
- B ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.
- C ☐ In die Analogie eines Rauchmelders: *Alarm with fire*.
- D ☐ In die Analogie eines brennenden Hauses ohne Rauchmelder: *House without noise*.
- E ☐ In die Analogie eines Rauchmelders: *Fire without alarm*, dem β -Fehler.

10 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 ☐ 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.
- B ☐ 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.
- C ☐ 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.
- D ☐ 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.
- E ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.

11 Aufgabe

(6 Punkte)


1. Skizzieren Sie 2 Normalverteilungen *in einer Abbildung* mit $\mu_1 \neq \mu_2$ und $s_1 = s_2$! (2 Punkte)
2. Beschriften Sie die Normalverteilungen mit den entsprechenden Parametern! (2 Punkte)
3. Liegt Varianzhomogenität oder Varianzheterogenität vor? Begründen Sie Ihre Antwort! (2 Punkte)

12 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
10.89	0	9.95	98.18	65.78	7
9.06	0	8.58	102.29	78.00	9
9.76	0	10.76	98.87	64.10	12
11.02	0	11.52	101.14	68.39	6

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
## -5.05427 -1.74085 -0.12722  1.46784  8.02216
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -14.860829   7.934943  -1.8728    0.06417 .
## light         0.553279   0.537136   1.0301    0.30560
## water        -0.292197   0.178048  -1.6411    0.10408
## Fe           0.869687   0.076051  11.4355 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.4072 on 95 degrees of freedom
## Multiple R-squared:  0.58867, Adjusted R-squared:  0.57568
## F-statistic: 45.319 on 3 and 95 DF,  p-value: < 0.00000000000000222
```

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)**

13 Aufgabe

(7 Punkte)

Gegeben sind folgende Randsummen in einer 2x2 Kreuztabelle aus einem Experiment mit $n = 133$ Sauen. In dem Experiment wurde gemessen, ob eine Sau nach einer Behandlung mit einem Medikament (ja/nein) mehr als 30 Ferkel pro Jahr bekommen konnte (ja/nein).

	>30 Ferkel (ja)	≤30 Ferkel (nein)	
Medikament (ja)			72
Medikament (nein)			61
	76	57	133

1. Ergänzen Sie die Felder innerhalb der 2x2 Kreuztabelle in dem Sinne, dass *kein* signifikanter Effekt zu erwarten wäre! (2 Punkte)
2. Erklären und Begründen Sie Ihr Vorgehen an der Formel des Chi-Quadrat-Tests mit

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

Sie können dies an einem Beispiel erklären! (2 Punkte)

3. Was ist die Mindestanzahl an Beobachtungen je Zelle? Wenn in einer der Zellen weniger Beobachtungen auftreten, welchen Test können Sie anstatt des „normalen“ Chi-Quadrat-Tests anwenden? (2 Punkte)
4. Warum hat die obige Vierfeldertafel einen Freiheitsgrad von $df = 1$? (1 Punkt)

14 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	22
B	22
B	25
A	22
A	20
B	30
A	19
B	27
B	26
A	21
A	16
A	18
A	22
B	26
A	22
B	24
A	21
A	16

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. Beschriften Sie auch die Achsen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen de Düngestufen erwarten würden, wie sehen dann die beiden Boxplots aus? **(1 Punkt)**

15 Aufgabe

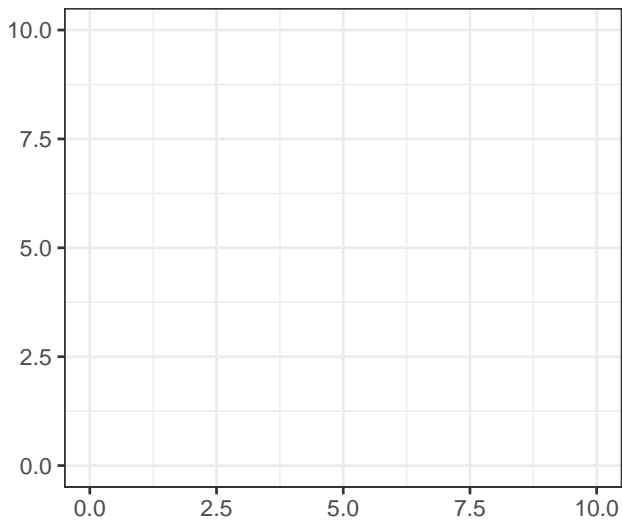
(9 Punkte)

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

1. Zeichnen Sie für die angegebene ρ -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)**

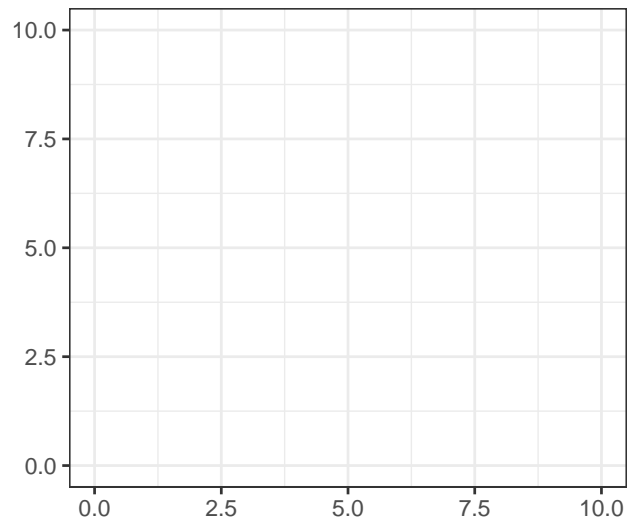
Pearsons $\rho = 0.75$

$R^2 = 0.75$



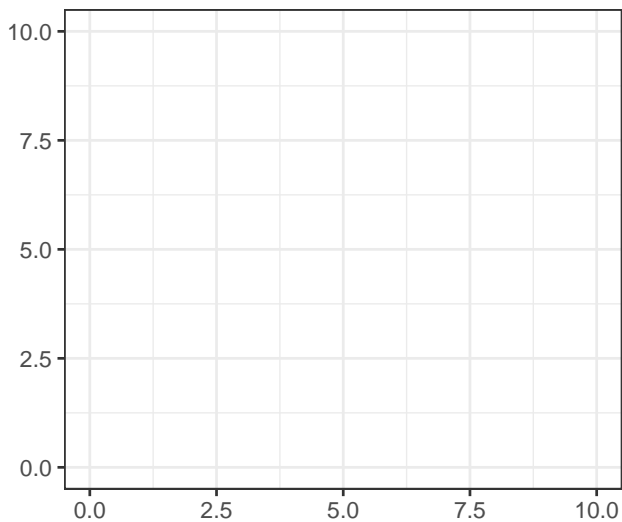
Pearsons $\rho = -0.5$

$R^2 = 0.5$



Pearsons $\rho = 0.5$

$R^2 = 1$



16 Aufgabe

(10 Punkte)

In einem Feldexperiment für die Bodendurchlässigkeit wurde der Niederschlag pro Parzelle sowie der durchschnittliche Ertrag gemessen. Es ergibt sich folgende Datentabelle.

water	drymatter
14	24
17	16
19	27
14	23
16	27
11	12
14	30

1. Erstellen Sie den Scatter-Plot für die Datentabelle. Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Zeichnen Sie eine Gerade durch die Punkte! **(1 Punkt)**
3. Beschriften Sie die Gerade mit den gängigen statistischen Maßzahlen! **(4 Punkte)**
4. Wenn kein Effekt von dem Niederschlag auf das Trockengewicht vorhanden wäre, wie würde die Gerade verlaufen und welche Werte würden die statistischen Maßzahlen annehmen? **(1 Punkt)**

17 Aufgabe

(14 Punkte)

Der Datensatz PlantGrowth enthält das Gewicht der Pflanzen (*weight*), die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden – dem Faktor *group* mit den Faktorstufen *ctrl*, *trt1*, *trt2*.

1. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus mit den gegebenen Informationen von Df und Sum Sq! **(4 Punkte)**
2. Schätzen Sie den p-Wert der Tabelle mit der Information von $F_k = 3.35$ ab. Begründen Sie Ihre Antwort! **(2 Punkte)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	3.64			
Residuals	27	11.97			

3. Was bedeutet ein signifikantes Ergebnis in einer einfaktoriellen ANOVA im Bezug auf die möglichen Unterschiede zwischen den Gruppen? **(2 Punkte)**
4. Berechnen Sie *einen* Student t-Test mit $T = \frac{\bar{x}_1 - \bar{x}_2}{s_p \cdot \sqrt{2/n_g}}$ für den *vermutlich* signifikantesten Gruppenvergleich anhand der untenstehenden Tabelle mit $T_k = 2.03$. Begründen Sie Ihre Auswahl! **(4 Punkte)**

group	n	mean	sd
ctrl	10	5.10	0.58
trt1	10	4.70	0.85
trt2	10	5.55	0.52

5. Gegebenen der ANOVA Tabelle war das Ergebnis des t-Tests zu erwarten? Begründen Sie Ihre Antwort! **(2 Punkte)**

