

Name: _____

Nicht bestanden: ☐

Vorname: _____

Matrikelnummer: _____

Endnote: _____

M.Sc. Angewandte Nutztier- und Pflanzenwissenschaften

Klausur Biostatistik

Hochschule Osnabrück

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

Klausur vom 07. Juli 2022

Ergebnis der Klausur

_____ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.

_____ von 81 Punkten sind aus dem Rechen- und Textteil erreicht.

_____ von 101 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
96 - 101	1,0
91 - 95	1,3
85 - 90	1,7
79 - 84	2,0
73 - 78	2,3
67 - 72	2,7
61 - 66	3,0
55 - 60	3,3
51 - 54	3,7
50	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	18	19
Punkte									

- Es sind ____ von 81 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

Berechnen Sie den Mittelwert und Standardabweichung von y mit 8, 13, 10, 13 und 11.

- A ☐ Es ergibt sich 12 +/- 1.06
- B ☐ Es ergibt sich 11 +/- 1.06
- C ☐ Es ergibt sich 10 +/- 2.25
- D ☐ Es ergibt sich 11 +/- 2.12
- E ☐ Es ergibt sich 11 +/- 4.5

2 Aufgabe

(2 Punkte)

Wenn Sie einen Datensatz erstellen, dann ist es ratsam die Spalten und die Einträge in englischer Sprache zu verfassen, wenn Sie später die Daten in **R** auswerten wollen. Welcher folgende Grund ist richtig?

- A ☐ Alle Funktionen und auch Anwendungen sind in **R** in englischer Sprache. Die Nutzung von deutschen Wörtern ist nicht schick und das ist zu vermeiden.
- B ☐ Die Spracherkennung von **R** ist nicht in der Lage Deutsch zu verstehen.
- C ☐ Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- D ☐ Programmiersprachen können nur englische Begriffe verarbeiten. Zusätzliche Pakete können zwar geladen werden, aber meist funktionieren diese Pakete nicht richtig. Deutsch ist International nicht bedeutend genug.
- E ☐ Im Allgemeinen haben Programmiersprachen Probleme mit Umlauten und Sonderzeichen, die in der deutschen Sprache vorkommen. Eine Nutzung der englischen Sprache umgeht dieses Problem auf einfache Art.

3 Aufgabe

(2 Punkte)

In der Theorie zur statistischen Testentscheidung kann „ H_0 beibehalten obwohl die H_0 falsch ist“ in welche richtige Analogie gesetzt werden?

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

4 Aufgabe

(2 Punkte)

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

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

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

6 Aufgabe

(2 Punkte)

Das Falsifikationsprinzip besagt...

- A** ☐ ... dass Fehlerterme in statistischen Modellen nicht verifiziert werden können.
- B** ☐ ... dass Annahmen an statistische Modelle meist falsch sind.
- C** ☐ ... dass ein schlechtes Modell durch ein weniger schlechtes Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.
- D** ☐ ... dass Modelle meist falsch sind und selten richtig.
- E** ☐ ... dass in der Wissenschaft immer etwas falsch sein muss. Sonst gebe es keinen Fortschritt.

7 Aufgabe

(2 Punkte)

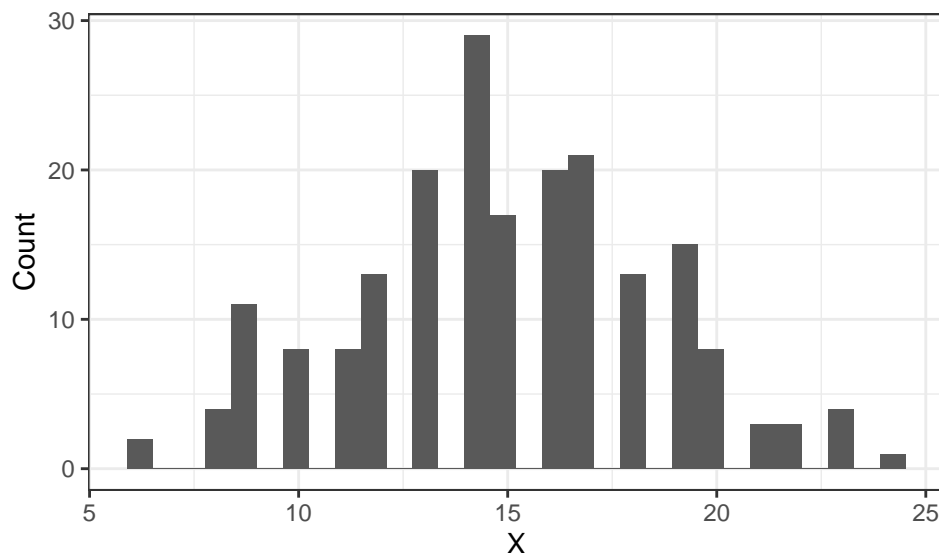
Sie haben das abstrakte Modell $Y \sim X$ mit X als Faktor mit zwei Leveln vorliegen. Welche Aussage über $n_1 = n_2$ ist richtig?

- A** ☐ Es handelt sich um ein balanciertes Design.
- B** ☐ Es handelt sich um ein unbalanciertes Design
- C** ☐ Es handelt sich um abhängige Beobachtungen.
- D** ☐ Es liegt Varianzhomogenität vor.
- E** ☐ Es liegt Varianzheterogenität vor.

8 Aufgabe

(2 Punkte)

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



- A** ☐ Eine rechtsschiefe, multivariate Normalverteilung.
- B** ☐ Es handelt sich um eine Binomial-Verteilung mit $\text{Binom}(10)$.
- C** ☐ Eine Standardnormalverteilung mit $N(0,1)$.
- D** ☐ Es handelt sich um eine Poisson-Verteilung mit $\text{Pois}(15)$.
- E** ☐ Es handelt sich um eine Normalverteilung mit $N(15, 5)$.

9 Aufgabe

(2 Punkte)

Welche Aussage über den t-Test ist richtig?

- A** ☐ Der t-Test testet generell zu einem erhöhten α -Niveau von 20%.
- B** ☐ Der t-Test ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern
- C** ☐ 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.
- D** ☐ Der t-Test vergleicht die Mittelwerte von zwei Gruppen.
- E** ☐ Der t-Test vergleicht die Varianzen von mindestens zwei oder mehr Gruppen

10 Aufgabe

(2 Punkte)

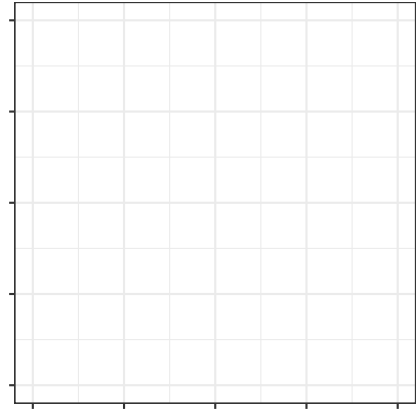
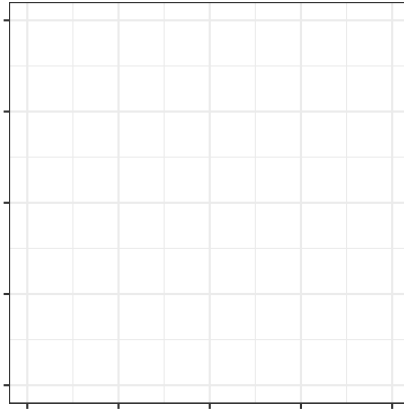
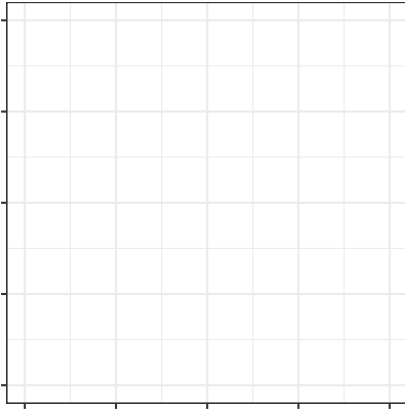
Die Randomisierung von Beobachtungen bzw. Samples zu den Versuchseinheiten ist bedeutend in der Versuchsplanung. Welche der folgenden Aussagen ist richtig?

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

11 Aufgabe


(10 Punkte)

1. Zeichnen Sie in die drei untenstehenden, leeren Abbildungen die Zeile des Regressionskreuzes der Binomialverteilung. Wählen Sie die Beschriftung der y-Achse sowie der x-Achse entsprechend aus! (6 Punkte)
2. Welchen Effektschätzer erhalten Sie aus der entsprechend linearen Regression? Geben Sie ein Beispiel! (2 Punkte)
3. Wenn Sie keinen Effekt erwarten, welchen *Zahlenraum* nimmt dann der Effektschätzer ein? Geben Sie ein Beispiel! (2 Punkte)



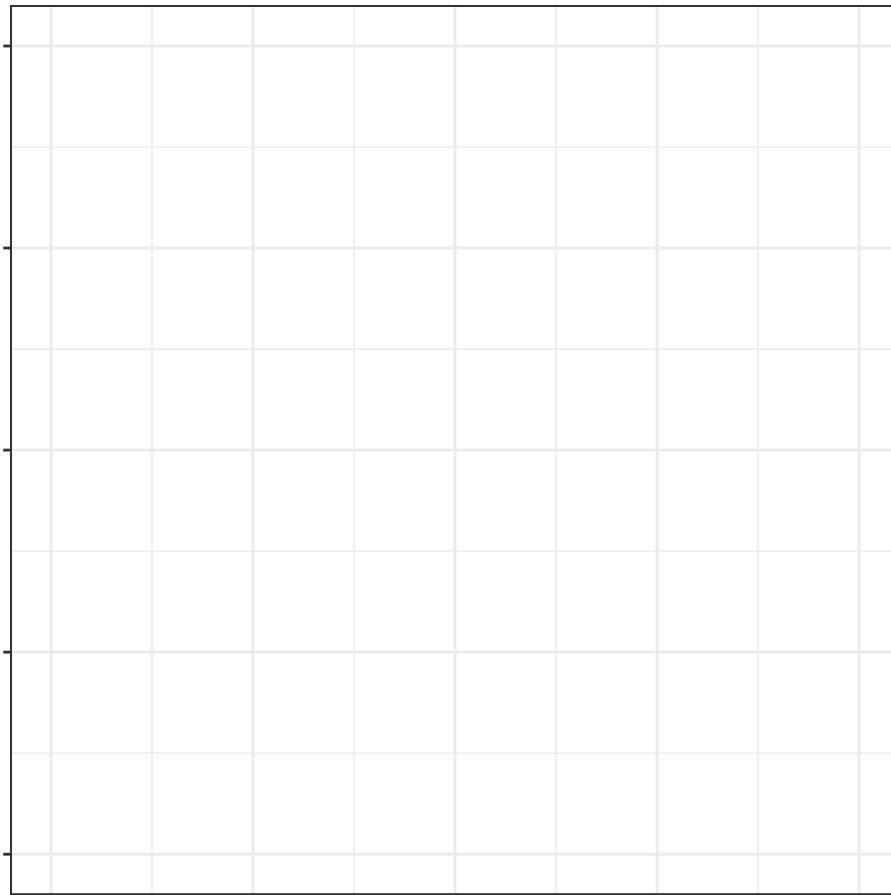
12 Aufgabe

(10 Punkte)

In einem Stallexperiment mit $n = 94$ 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)	25.40	0.96		
light	1.45	0.10		

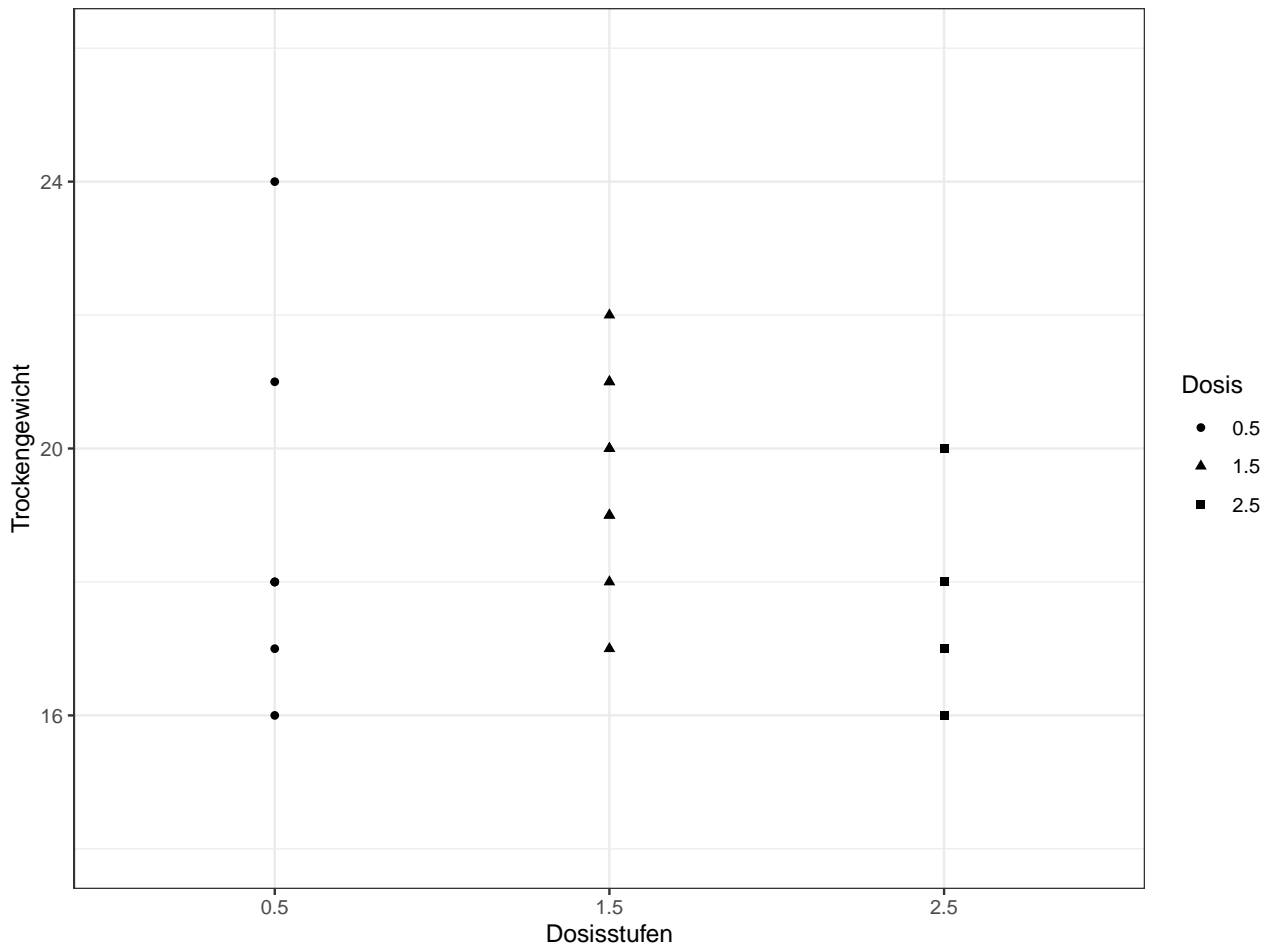
1. Berechnen Sie die t Statistik für *(Intercept)* und *light*! **(2 Punkte)**
2. Schätzen Sie den p-Wert für *(Intercept)* und *light* mit $T_k = 1.96$ ab. Was sagt Ihnen der p-Wert aus? Begründen Sie Ihre Antwort! **(3 Punkte)**
3. Zeichnen Sie die Gerade aus der obigen Tabelle in die untenstehende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



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: β_0
- Mittelwerte der Dosen: $\bar{x}_{0.5}$, $\bar{x}_{1.5}$ und $\bar{x}_{2.5}$
- Effekt der einzelnen Level der Dosen: $\beta_{0.5}$, $\beta_{1.5}$, und $\beta_{2.5}$
- Residuen oder Fehler: ϵ

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

(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	3.29			
Residuals	27	10.05			


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	5.04	0.58
trt1	10	4.67	0.77
trt2	10	5.48	0.42

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

15 Aufgabe

(10 Punkte)

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

```
##  
## Two Sample t-test  
##  
## data:  freshweight by N  
## t = 1.41068, df = 12, p-value = 0.18373  
## alternative hypothesis: true difference in means between group high and group low is not equal to  
## 95 percent confidence interval:  
## -1.5125289  7.0680845  
## sample estimates:  
## mean in group high  mean in group low  
##      18.77778      16.00000
```

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 eine Abbildung in der Sie T_c , $Pr(D|H_0)$, $A = 1$, sowie $T_k = |2.18|$ einzeichnen! **(4 Punkte)**
4. Beschriften Sie die Abbildung und den Graphen entsprechend! **(2 Punkte)**

16 Aufgabe

(6 Punkte)

1. Skizzieren Sie 4 Normalverteilungen *in einer Abbildung* mit $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ und $s_1 = s_2 = s_3 = s_4$! (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)

17 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
22	22
22	20
21	17
19	19
19	24

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)**

18 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	13
A	15
A	18
B	20
A	17
B	23
B	20
A	14
B	28
A	18
A	19
B	28
A	16
A	16
B	20
A	16
A	16
B	26

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)**

19 Aufgabe

(4 Punkte)

1. Zeichnen Sie über den untenstehenden Boxplot die entsprechende zugehörige Verteilung! (2 Punkte)
2. Zeichnen Sie den entsprechenden Lageparameter in die Abbildung. Beschriften Sie die Abbildung! (2 Punkte)



