Name: ____________________ *Nicht bestanden:* ☐

Vorname: ____________________

Matrikelnummer: ____________________ **Endnote:** ____________

B.Eng. Wirtschaftsingenieurwesen Agrar/Lebensmittel

Klausur Statistik

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

23. Januar 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 63 Punkten sind aus dem Rechen- und Textteil erreicht.

______ von 83 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	**Note**
79.5 - 83.0	1,0
75.5 - 79.0	1,3
71.0 - 75.0	1,7
67.0 - 70.5	2,0
63.0 - 66.5	2,3
59.0 - 62.5	2,7
55.0 - 58.5	3,0
50.5 - 54.5	3,3
46.5 - 50.0	3,7
41.5 - 46.0	4,0

Es ergibt sich eine Endnote von ________.

Multiple Choice Aufgaben

- Pro Multipe 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	8	8	10	12	8	8	9

- Es sind ____ von 63 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 $\eta^2 = 0.32$. 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 ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.

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 ☐ Die Berechnung von η^2 ist ein Wert für die Interaktion.

E ☐ Das η^2 beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen nicht erklärt wird. Somit der Rest an nicht erklärbarer Varianz.

2 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 Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte.

B ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.

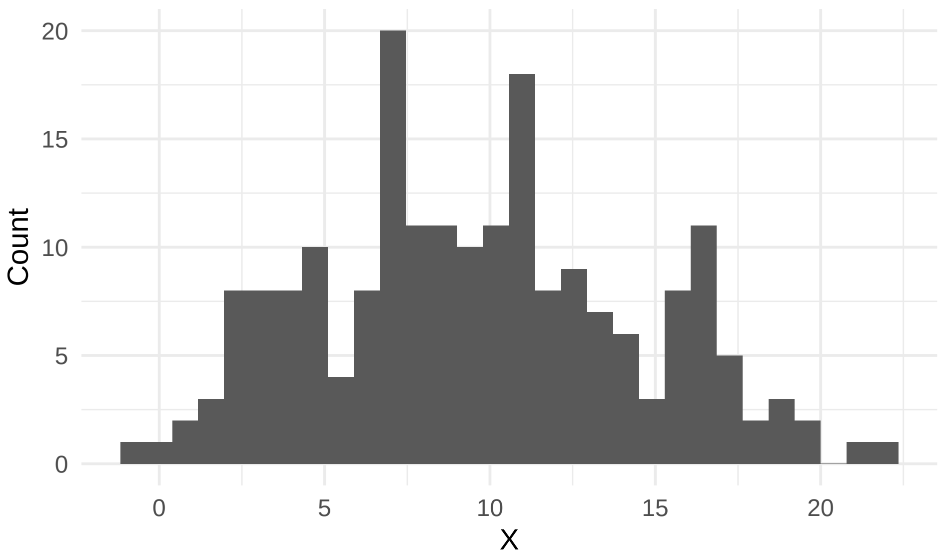
C ☐ Randomisierung sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.

D ☐ Randomisierung 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 Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich.

3 Aufgabe **(2 Punkte)**

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



A ☐ Eine Standardnormalverteilung mit N(0,1).

B ☐ Es handelt sich um eine Poisson-Verteilung mit Pois(10).

C ☐ Es handelt sich um eine Binomial-Verteilung mit Binom(10).

D ☐ Eine rechtsschiefe, multivariate Normalverteilung.

E ☐ Es handelt sich um eine Normalverteilung mit N(10, 5).

4 Aufgabe (2 Punkte)

Berechnen Sie den Mittelwert und Standardabweichung von y mit 3, 11, 16, 7 und 10.

A ☐ Es ergibt sich 9.4 +/- 23.3

B ☐ Es ergibt sich 10.4 +/- 2.415

C ☐ Es ergibt sich 9.4 +/- 4.83

D ☐ Es ergibt sich 9.4 +/- 2.415

E ☐ Es ergibt sich 8.4 +/- 11.65

5 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 T-Statistik aus der Multiplikation der MS Behandlung mit der MS der Fehler. Wenn die F-Statistik genau 0 ist, kann die Nullhypothese abgelehnt werden.

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 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.

E ☐ Die ANOVA berechnt die F-Statistik aus den SS Behandlung geteilt durch die SS Fehler.

6 Aufgabe (2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.34, 0.001, 0.21, 0.03 und 0.01. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

A ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.068, 2e-04, 0.042, 0.006 und 0.002. Die adjustierten p-Werte werden zu einem α-Niveau von 5% verglichen.

B ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.068, 2e-04, 0.042, 0.006 und 0.002. Die adjustierten p-Werte werden zu einem α-Niveau von 1% verglichen.

C ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1.7, 0.005, 1.05, 0.15 und 0.05. 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, 0.005, 1, 0.15 und 0.05. Die adjustierten p-Werte werden zu einem α-Niveau von 5% verglichen.

E ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 1, 0.005, 1, 0.15 und 0.05. Die adjustierten p-Werte werden zu einem α-Niveau von 1% verglichen.

7 Aufgabe (2 Punkte)

Beim statistischen Testen wird `signal` mit `noise` zur Teststatistik T verrechnet. Welche der Formel berechnet korrekt die Teststatistik T?

A ☐ Es gilt $T = \frac{signal}{noise^2}$

B ☐ Es gilt $T = \frac{signal}{noise}$

C ☐ Es gilt $T = signal \cdot noise$

D ☐ Es gilt $T = (signal \cdot noise)^2$

E ☐ Es gilt $T = \frac{noise}{signal}$

8 Aufgabe (2 Punkte)

Die empfohlene Mindestanzahl an Beobachtungen für ein Histogramm sind...

A ☐ 1 Beobachtung.

B ☐ 10 Beobachtungen.

C ☐ 5 und mehr Beobachtungen.

D ☐ 2-5 Beobachtungen.

E ☐ mindestens 20 Beobachtungen.

9 Aufgabe (2 Punkte)

Der Fehler 1. Art oder auch Signifikanzniveau α genannt, liegt bei 5%. Welcher der folgenden Gründe für diese Festlegeung 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 ☐ Der Begründer der modernen Statistik, R. Fischer, hat die Grenze simuliert und berechnet. Dadurch ergibt sich dieser optimale Cut-Off.

D ☐ 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.

E ☐ Im Rahmen eines langen Disputs zwischen Neyman und Fischer wurde $\alpha = 5\%$ festgelegt. Leider werden die Randbedingungen und Voraussetzungen an statistsiche Modelle heute immer wieder ignoriert.

10 Aufgabe (2 Punkte)

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

A ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit der Alternativehypothese und somit $1 - Pr(H_A)$

B ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit die Daten D zu beobachten wenn die Nullhypothese wahr ist.

C ☐ Die Wahrscheinlichkeit der Daten unter der Nullhypothese in der Grundgesamtheit.

D ☐ Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativehypothese überdeckt.

E ☐ Die Wahrscheinlichkeit für die Nullhypothese, wenn die Daten wahr sind.

11 Aufgabe (8 Punkte)

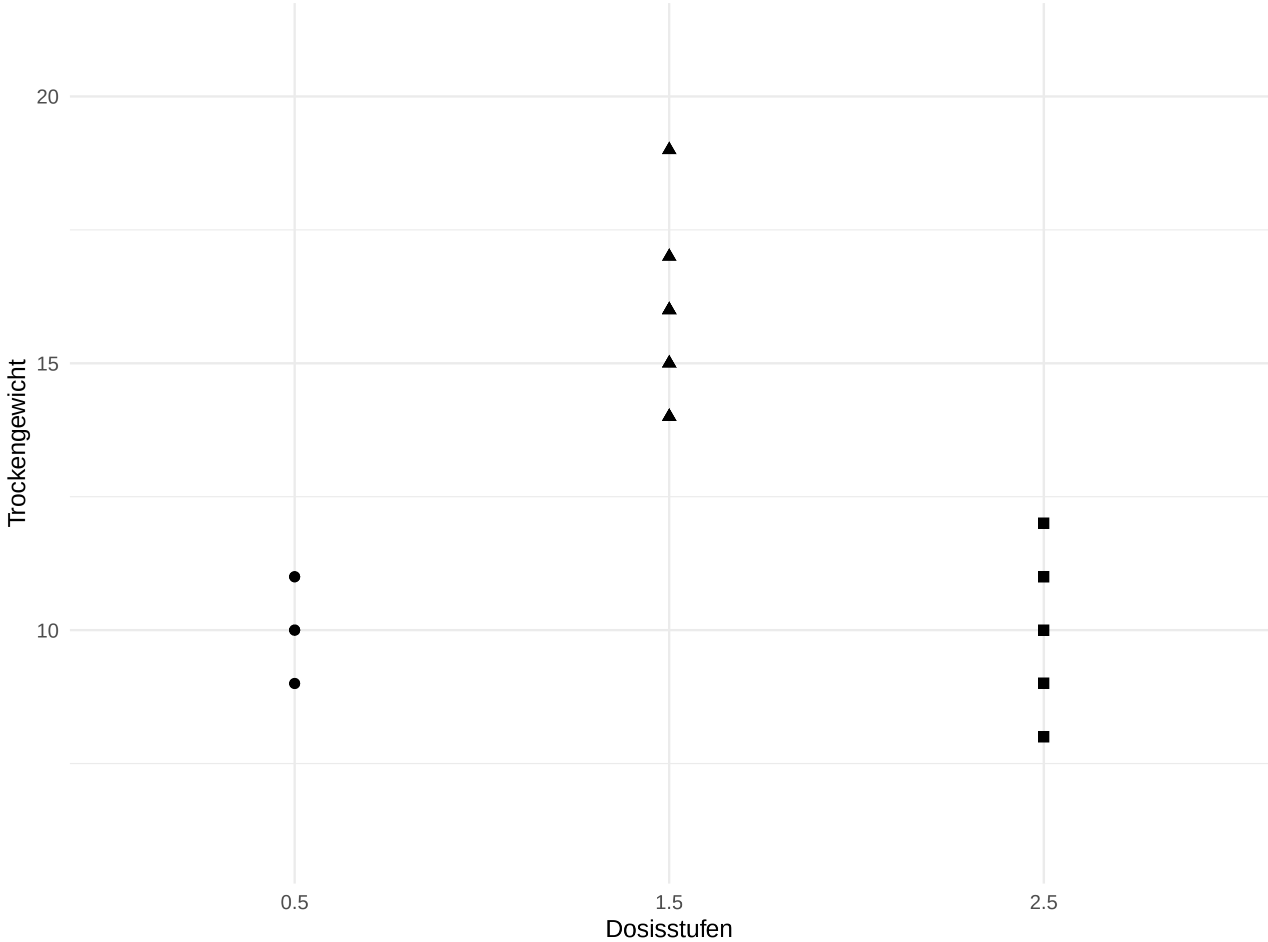
In einem Experiment für den Proteingehalt von Wasserlinsen in g/l mit fünf Dosisstufen (A, B, C, D und E) erhalten Sie folgendes *Compact letter display (CLD)* als Ausgabe aus den rohen, unadjustierten p-Werten.

```
##   A   B   C   D   E
## "a" "a" "b" "a" "a"
```

1. Erstellen Sie eine Matrix mit den paarweisen p-Werten, die sich näherungsweise aus dem *Compact letter display (CLD)* ergeben würde! Begründen Sie Ihre Antwort! **(3 Punkte)**
2. Zeichnen Sie eine Abbildung, der sich ergebenden Barplots! **(2 Punkte)**
3. Ergänzen Sie das *Compact letter display (CLD)* zu der Abbildung! **(1 Punkt)**
4. Erklären Sie *einen* Vorteil und *einen* Nachteil des *Compact letter display (CLD)*! **(2 Punkte)**

12 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 Abildung ein! Beschriften Sie die statistischen Maßzahlen! **(6 Punkte)**
   - Total (grand) mean: β_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: ϵ
2. Liegt ein *vermutlicher* signifikanter Unterschied zwischen den Dosisstufen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**

13 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 R Output einer simplen Gaussian linearen Regression sieben Wochen nach der ersten Messung.

term	estimate	std.error	t statistic	p-value
(Intercept)	1.918159	1.0944357		
light	0.118555	0.1091051		

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

14 Aufgabe (12 Punkte)

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

Nach einem Experiment mit zwei Futtermitteln (*FatDown* und *ProGain*) an Puten ergibt sich die folgende Datentabelle mit dem gemessenen Gewichtszunahmen nach fünf Wochen Mast.

feed	weight
FatDown	17
ProGain	13
ProGain	16
FatDown	21
FatDown	20
ProGain	11
ProGain	14
FatDown	18
ProGain	15
FatDown	16
FatDown	16
ProGain	20
FatDown	19

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik T_{calc} eines Welch t-Tests für den Vergleich der beiden Futtermittel! **(4 Punkte)**
4. Treffen Sie mit $T_{\alpha=10\%} = 1.35$ und dem berechneten T_{calc} eine Aussage zur Nullhypothese! **(1 Punkt)**
5. Berechnen Sie das 90% Konfidenzintervall unter der Verwendung von s_p und der gemittelten Fallzahl über die beiden Gruppen! **(3 Punkte)**
6. Nennen Sie den statistischen Grund, warum Sie sich zwischen einem Student t-Test und einem Welch t-Test entscheiden müssen! **(1 Punk)**

15 Aufgabe (8 Punkte)

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

YouTube

Nach einem Freilandexperiment bestimmen Sie folgende Trockengewichte von Erbsen nach einer durchgestandenen Infektion der Pflanzen.

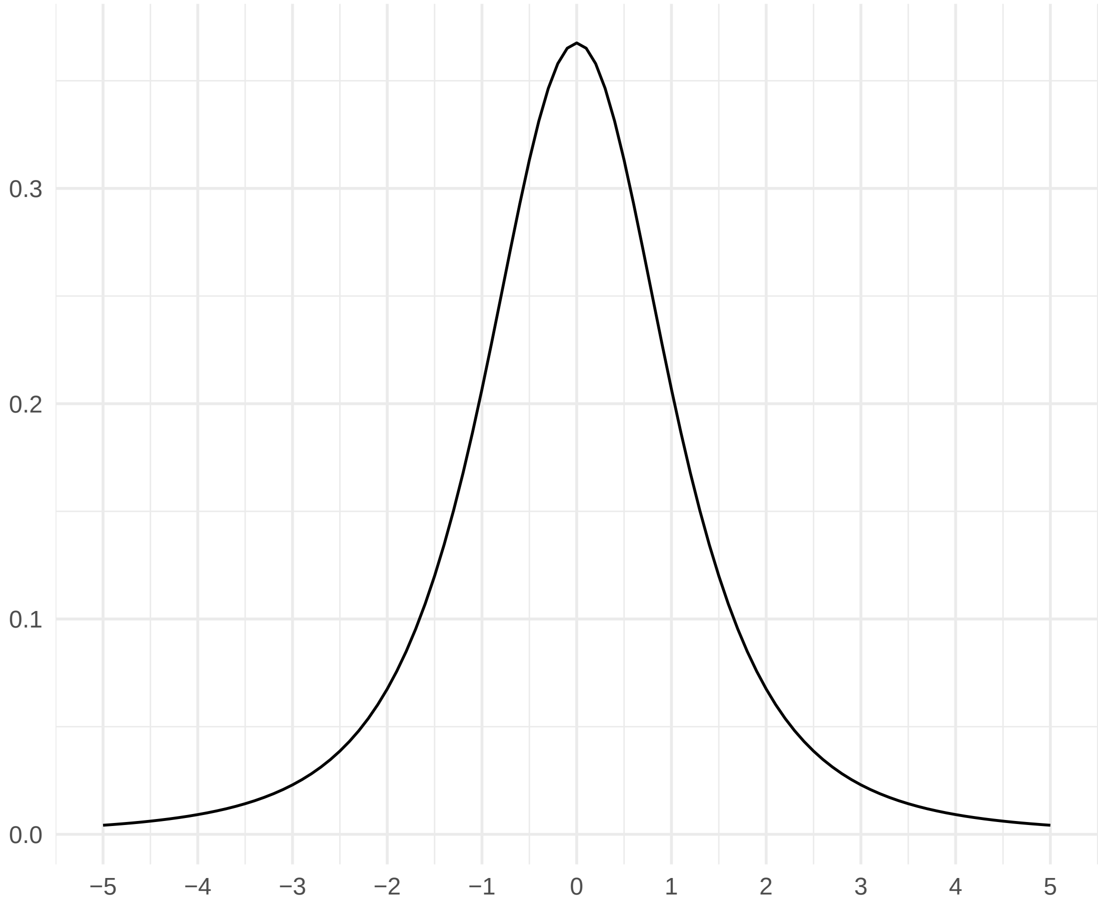
$$11.9, 9.7, 12.6, 6.7, 13.5, 7.3, 10.2, 8.5, 12.3, 10.6, 11.7, 14$$

1. Zeichen Sie ein Histogramm um die Verteilung der Daten zu visualisieren! (**3 Punkte**)
2. Erläutern Sie Ihr Vorgehen um ein Histogramm für kontinuierliche Daten zu zeichnen! (**2 Punkte**)
3. Beschriften Sie die Achsen der Abbildung! (**2 Punkte**)
4. Ergänzen Sie die relativen Häufigkeiten in der Abbildung! **(1 Punkt)**

16 Aufgabe (8 Punkte)

Im Folgenden ist die t-Verteilung unter der Anahme der Gültigkeit der Nullhypothese abgebildet. Ergänzen Sie die Abbildung wie folgt.

1. Zeichnen Sie das Signifikanzniveau α in die Abbildung! **(2 Punkte)**
2. Zeichnen Sie einen signifikant p-Wert in die Abbildung! **(2 Punkte)**
3. Ergänzen Sie „$\bar{y}_1 = \bar{y}_2$"! **(1 Punkt)**
4. Ergänzen Sie „$A = 0.95$"! **(1 Punkt)**
5. Zeichnen Sie $T_{\alpha=5\%}$ in die Abbildung! **(1 Punkt)**
6. Zeichnen Sie $-T_{calc}$ in die Abbildung! **(1 Punkt)**



17 Aufgabe (9 Punkte)

Der Datensatz *crop_tbl* enthält das Outcome *drymatter* für ein Experiment mit Maispflanzen, welches unter drei verschiedenen Düngerbedingungen erzielt wurden. Die Düngerbedingungen sind in dem Faktor *trt* mit den Faktorstufen *ctrl*, *mid* und *C* codiert. Sie erhalten folgenden Output in R.

```
## Analysis of Variance Table
##
## Response: drymatter
##           Df Sum Sq Mean Sq F value Pr(>F)
## trt        2  5.294  2.6470  0.6489 0.5323
## Residuals 22 89.746  4.0794
```

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. Berechen Sie den Effektschätzer η^2. Was sagt Ihnen der Wert von η^2 aus? **(2 Punkte)**
4. Skizieren Sie eine Abbildung, der dem obigen Ergebnis der einfaktoriellen ANOVA näherungsweise entspricht! **(3 Punkte)**