

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

Version vom 27. Juni 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.
- **You can answer the questions in English without any consequences.**

Ergebnis der Klausur

_____ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.
_____ von 64 Punkten sind aus dem Rechen- und Textteil erreicht.
_____ von 84 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
80.5 - 84.0	1,0
76.0 - 80.0	1,3
72.0 - 75.5	1,7
67.5 - 71.5	2,0
63.5 - 67.0	2,3
59.5 - 63.0	2,7
55.0 - 59.0	3,0
51.0 - 54.5	3,3
46.5 - 50.5	3,7
42.0 - 46.0	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	8	10	8	12	9	9	8

- Es sind ____ von 64 Punkten erreicht worden.

1 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 aus den SS Behandlung geteilt durch die SS Fehler.
- 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 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.
- 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 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.

2 Aufgabe

(2 Punkte)

Das Falsifikationsprinzip besagt...

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

3 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 erlaubt erst die Mittelwerte zu schätzen. Ohne Randomisierung keine Mittelwerte.
- C ☐ Randomisierung sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.
- 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.

4 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** ☐ 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.
- C** ☐ Die Spracherkennung von **R** ist nicht in der Lage Deutsch zu verstehen.
- D** ☐ Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- E** ☐ 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.

5 Aufgabe

(2 Punkte)

Nachdem Sie in einem Experiment die Daten D erhoben haben, berechnen Sie den Mittelwert und den Median. Der Mittelwert \bar{y} und der Median \tilde{y} unterscheiden sich. Welche Aussage ist richtig?

- A** ☐ Da sich der Mittelwert und der Median unterscheiden, ist der Datensatz nicht zu verwenden. Mittelwert und Median müssen gleich sein.
- B** ☐ Da sich der Mittelwert und der Median unterscheiden, liegen vermutlich keine Outlier in den Daten vor.
- C** ☐ Da sich der Mittelwert und der Median nicht unterscheiden, liegen vermutlich keine Outlier in den Daten vor. Wir verwenden den Datensatz so wie er ist.
- D** ☐ Da sich der Mittelwert und der Median nicht unterscheiden, liegen vermutlich Outlier in den Daten vor.
- E** ☐ Da sich der Mittelwert und der Median unterscheiden, liegen vermutlich Outlier in den Daten vor. Wir untersuchen den Datensatz nach auffälligen Beobachtungen.

6 Aufgabe

(2 Punkte)

Berechnen Sie den Mittelwert und Standardabweichung von y mit 12, 7, 16, 10 und 13.

- A** ☐ Es ergibt sich 10.6 ± 5.65
- B** ☐ Es ergibt sich 11.6 ± 3.36
- C** ☐ Es ergibt sich 11.6 ± 11.3
- D** ☐ Es ergibt sich 11.6 ± 1.68
- E** ☐ Es ergibt sich 12.6 ± 1.68

7 Aufgabe

(2 Punkte)

Welche Aussage über den Welch t-Test ist richtig?

- A ☐ Der Welch t-Test ist ein Post-hoc Test der ANOVA und basiert daher auf dem Vergleich der Varianz.
- B ☐ Der Welch t-Test vergleicht die Varianz von zwei Gruppen.
- C ☐ Der Welch t-Test wird angewendet, wenn Varianzheterogenität zwischen den beiden zu vergleichenden Gruppen vorliegt.
- D ☐ Der Welch t-Test vergleicht die Mittelwerte von zwei Gruppen unter der strikten Annahme von Varianzhomogenität.
- E ☐ Der Welch t-Test ist die veraltete Form des Student t-Test und wird somit nicht mehr verwendet.

8 Aufgabe

(2 Punkte)

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

- A ☐ 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.
- B ☐ 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.
- C ☐ Das Konfidenzintervall. Durch die Visualisierung des Konfidenzintervalls kann eine Relevanzschwelle vom Anwender definiert werden. Zusätzlich erlaubt das Konfidenzintervall auch eine Entscheidung über die Signifikanz.
- 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 OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.

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: *Fire without alarm*, dem β -Fehler.
- B ☐ In die Analogie eines Rauchmelders: *Alarm without fire*, dem α -Fehler.
- 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 Feuerwehrautos: *Car without noise*.

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

11 Aufgabe

(9 Punkte)



Der Datensatz *crop_tbl* enthält das Trockengewicht der Maispflanzen (*drymatter*), die unter drei verschiedenen Düngerbedingungen erzielt wurden. Die Düngerbedingungen sind in dem Faktor *trt* mit den Faktorstufen *A*, *B* und *C* codiert. Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: drymatter
##           Df Sum Sq Mean Sq F value    Pr(>F)
## trt         2 202.681  101.340   63.556 7.212e-10
## Residuals  22  35.079    1.595
```

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

12 Aufgabe

(8 Punkte)

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



Nach einem Experiment zählen Sie folgende Trockengewichte von Sonnenblumen nach einer durchgestandenen Infektion.

8.7, 10.2, 9.1, 6.9, 9, 11.6, 7.4, 10.2, 7.1, 7.5, 11.5, 7.3, 5.9, 13.9, 12.9

1. Zeichnen 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)**

13 Aufgabe

(10 Punkte)

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



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

water	drymatter
13	14
17	16
17	21
19	18
19	14
25	13
27	17
25	15
21	15

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! Geben Sie die numerischen Zahlenwerte mit an! **(3 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? **(2 Punkt)**

14 Aufgabe

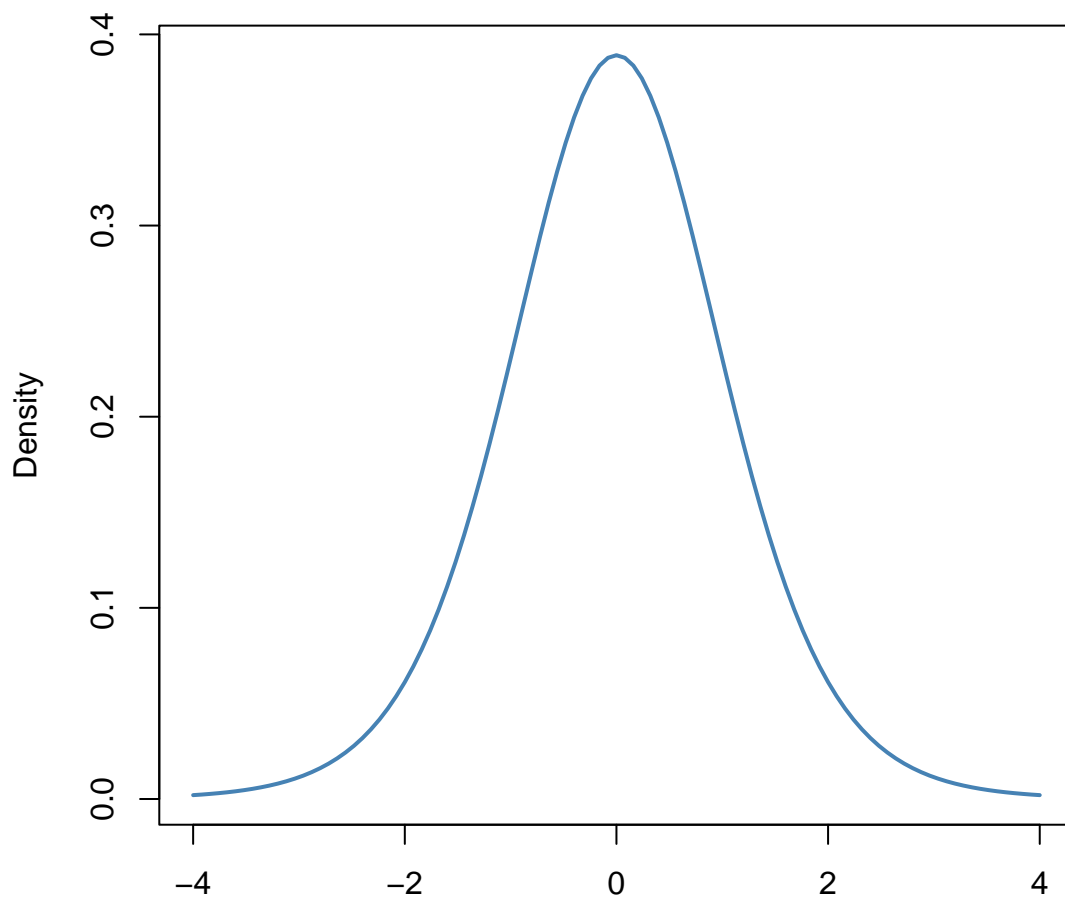
(8 Punkte)



Im folgenden ist eine t-Verteilung 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)**

t Distribution




15 Aufgabe

(8 Punkte)



Sie erhalten folgende R Ausgabe der Funktion `lm()` nach einem Experiment mit zwei Behandlungen (A und B) sowie dem Ertragsgewicht von Weizen.


```
##
## Call:
## lm(formula = weight ~ trt, data = data_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.00  -2.95   0.10   2.95   5.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    15.800      1.836   8.607 2.57e-05
## trtB           13.200      2.596   5.084 0.000948
##
## Residual standard error: 4.105 on 8 degrees of freedom
## Multiple R-squared:  0.7637, Adjusted R-squared:  0.7341
## F-statistic: 25.85 on 1 and 8 DF,  p-value: 0.0009479
```

1. Ist die Annahme der Normalverteilung an das Outcome *rsp* erfüllt? Begründen Sie die Antwort! **(2 Punkte)**
2. Wie groß ist der Effekt des *Trt*? Liegt ein signifikanter Effekt vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Erklären Sie *kurz* den Begriff Multiple R-squared! Was sagt Ihnen der Wert 0.7637 aus? **(2 Punkte)**
4. Schreiben Sie das Ergebnis der  Ausgabe in einen Satz nieder, der die Information zum Effekt und der Signifikanz enthält! **(2 Punkte)**

16 Aufgabe

(9 Punkte)



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

```
##  
## Two Sample t-test  
##  
## data: waterintake by infusion  
## t = -4.8654, df = 16, p-value = 0.0001719  
## alternative hypothesis: true difference in means between group high and group low is not equal to 0  
## 95 percent confidence interval:  
## -9.092832 -3.573835  
## sample estimates:  
## mean in group ctrl mean in group trt1  
## 12.66667 19.00000
```

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 die sich ergebenden Boxplot! Welche Annahmen an die Daten haben Sie getroffen? Begründen Sie Ihre Antwort! **(3 Punkte)**
4. Skizzieren Sie die sich ergebenden Barplots! **(2 Punkte)**

17 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
ProGain	12
FatDown	19
FatDown	15
ProGain	13
FatDown	18
ProGain	19
ProGain	17
ProGain	18
FatDown	17
FatDown	15
FatDown	19
FatDown	23

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 $T_{\alpha=10\%} = 1.36$ 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 Punkt)**