

Name: _____

Nicht bestanden: ☐

Vorname: _____

Matrikelnummer: _____

Endnote: _____

B.Sc. Bioverfahrenstechnik in Agrar- und Lebensmittelwirtschaft

Klausur Angewandte Statistik für Bioverfahrenstechnik

Hochschule Osnabrück

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

03. Juli 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 65 Punkten sind aus dem Rechen- und Textteil erreicht.
_____ von 85 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
81.5 - 85.0	1,0
77.0 - 81.0	1,3
73.0 - 76.5	1,7
68.5 - 72.5	2,0
64.5 - 68.0	2,3
60.5 - 64.0	2,7
56.0 - 60.0	3,0
52.0 - 55.5	3,3
47.5 - 51.5	3,7
42.5 - 47.0	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	6	10	12	9	10	10

- Es sind ____ von 65 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

Welche Aussage über den Welch t-Test ist richtig?

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

2 Aufgabe

(2 Punkte)

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

- A ☐ 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.
- 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 ☐ 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.
- E ☐ Das OR. Als Chancenverhältnis gibt es das Verhältnis von Relevanz und Signifikanz wieder.

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

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 $\eta^2 = 0.26$. Welche Aussage ist richtig?

- A** ☐ Das η^2 beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen nicht erklärt wird. Somit der Rest an nicht erklärbarer Varianz.
- B** ☐ 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.
- C** ☐ Das η^2 ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.
- D** ☐ 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.
- E** ☐ Die Berechnung von η^2 ist ein Wert für die Interaktion.

5 Aufgabe

(2 Punkte)

Welche Aussage über den Korrelationskoeffizienten nach Pearson ist richtig?

- A** ☐ Der Korrelationskoeffizienten nach Pearson wird genutzt, wenn der Korrelationskoeffizienten zwischen -1 und 1 liegt. Dann sind die Residuen normalverteilt.
- B** ☐ Der Korrelationskoeffizienten nach Pearson wird genutzt, wenn das Outcome Y nicht normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen -1 und 1.
- C** ☐ Der Korrelationskoeffizienten nach Pearson wird genutzt, wenn das Outcome Y nicht normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen 0 und 1.
- D** ☐ Der Korrelationskoeffizienten nach Pearson wird genutzt, wenn das Outcome Y normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen -1 und 1.
- E** ☐ Der Korrelationskoeffizienten nach Pearson wird genutzt, wenn das Outcome Y normalverteilt ist. Der Korrelationskoeffizienten liegt zwischen 0 und 1.

6 Aufgabe

(2 Punkte)

Berechnen Sie den Median, das 1st Quartile sowie das 3rd Quartile von y mit 27, 28, 20, 33, 20, 10, 9, 20, 18, 8 und 51.

- A** ☐ Es ergibt sich 22 [11, 29]
- B** ☐ Es ergibt sich 20 [10, 28]
- C** ☐ Es ergibt sich 20 +/- 28
- D** ☐ Es ergibt sich 22 +/- 10
- E** ☐ Es ergibt sich 20 +/- 10

7 Aufgabe

(2 Punkte)

Das Falsifikationsprinzip besagt...

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

8 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.

9 Aufgabe

(2 Punkte)

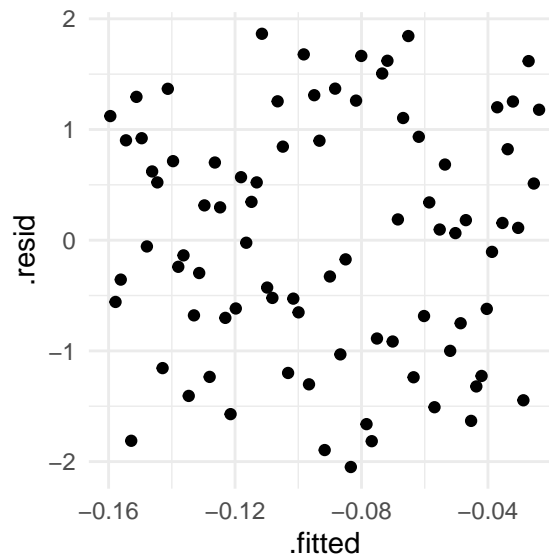
Wenn Sie einen Datensatz erstellen, dann ist es ratsam die Spalten und die Einträge in englischer Sprache zu verfassen, wenn Sie später die Daten in  auswerten wollen. Welcher folgende Grund ist richtig?

- A ☐ 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.
- 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  ist nicht in der Lage Deutsch zu verstehen.
- D ☐ Es gibt keinen Grund nicht auch deutsche Wörter zu verwenden. Es ist ein Stilmittel.
- E ☐ Alle Funktionen und auch Anwendungen sind in  in englischer Sprache. Die Nutzung von deutschen Wörtern ist nicht schick und das ist zu vermeiden.

10 Aufgabe

(2 Punkte)

Sie rechnen eine lineare Regression und erhalten folgenden Residual Plot. Welche Aussage ist richtig?




- A** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Die Punkte liegen zum überwiegenden Teil auf der Diagonalen.
- B** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Kein Muster ist zu erkennen und keine Outlier zu beobachten.
- C** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Es ist kein Muster zu erkennen.
- D** ☐ Die Annahme der normalverteilten Residuen ist nicht erfüllt. Vereinzelt Punkte liegen oberhalb bzw. unterhalb der Geraden um die 0 Linie weiter entfernt. Ein klares Muster ist zu erkennen.
- E** ☐ Die Annahme der normalverteilten Residuen ist erfüllt. Es ist ein Muster zu erkennen.

11 Aufgabe

(9 Punkte)



Der Datensatz *crop_tbl* enthält das Outcome *freshmatter* für ein Experiment mit Maispflanzen, welches unter drei verschiedenen Düngerbedingungen erzielt wurden. Die Düngerbedingungen sind in dem Faktor *trt* mit den Faktorstufen *A*, *trt1* und *C* codiert. Sie erhalten folgenden Output in .

```
## Analysis of Variance Table
##
## Response: freshmatter
##           Df Sum Sq Mean Sq F value    Pr(>F)
## trt         2  67.60   33.800    6.4216 0.007012
## Residuals  20 105.27    5.263
```

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

(6 Punkte)



Nach einer Bonitur von Schnittlauch mit einer Kontrolle und drei Pestiziden (ctrl, pestKill, roundUp, zeroX) ergibt sich die folgende Datentabelle mit den Boniturnoten (*grade*).

pesticide	grade
roundUp	2
pestKill	1
roundUp	5
ctrl	6
pestKill	2
zeroX	1
ctrl	8
roundUp	2
ctrl	9
zeroX	2
zeroX	4
pestKill	1


1. Zeichnen Sie in *einer* Abbildung die Dotplots für die vier Pestizidlevel! Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Ergänzen Sie die Dotplots mit der gängigen statistischen Maßzahl! **(1 Punkt)**
3. Wenn Sie *keinen Effekt* zwischen den Pestizidlevel erwarten würden, wie sehen dann die Dotplots aus? **(1 Punkt)**

13 Aufgabe

(10 Punkte)

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



In einem Stallexperiment mit $n = 60$ Ferkeln wurde der Gewichtszuwachs in kg unter ansteigender Lichteinstrahlung in nm 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)	26.36	1.28		
light	1.83	0.13		

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 Gerade 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

(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.

7.4, 13.6, 8.9, 10.9, 7.3, 9.6, 9.3, 12.4, 15.9, 8.6, 11.4, 16.6, 10.8, 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)**

15 Aufgabe

(10 Punkte)

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



Nach einem Experiment ergibt sich die folgende 2x2 Datentabelle mit einem Pestizid (ja/nein), dargestellt in den Zeilen. Im Weiteren mit dem infizierten Pflanzenstatus (ja/nein) in den Spalten. Insgesamt wurden $n = 124$ Pflanzen untersucht.

	Erkrankt (ja)	Erkrankt (nein)	
Pestizid (ja)	38	21	
Pestizid (nein)	13	52	

1. Ergänzen Sie die Tabelle um die Randsummen! **(1 Punkt)**
2. Formulieren Sie die Fragestellung! **(1 Punkt)**
3. Formulieren Sie das Hypothesenpaar! **(2 Punkte)**
4. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test auf der 2x2 Tafel. Geben Sie Formeln und Rechenweg mit an! **(4 Punkte)**
5. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem $\chi^2_{\alpha=5\%} = 3.841$! **(1 Punkt)**
6. Skizzieren Sie eine 2x2 Tabelle mit $n = 30$ Pflanzen in dem *vermutlich* die Nullhypothese nicht abgelehnt werden kann! **(1 Punkt)**

16 Aufgabe

(10 Punkte)



Sie rechnen einen t-Test für Gruppenvergleiche. Sie schätzen den Unterschied zwischen dem mittleren Befall mit Parasiten zu einer unbehandelten Kontrolle.

1. Beschriften Sie die untenstehende Abbildung mit der Signifikanzschwelle! Begründen Sie Ihre Antwort! **(2 Punkte)**
2. Ergänzen Sie eine *in den Kontext passende* Relevanzschwelle! Begründen Sie Ihre Antwort! **(2 Punkte)**
3. Skizzieren Sie in die untenstehende Abbildung sechs einzelne Konfidenzintervalle (a-f) mit den jeweiligen Eigenschaften! **(6 Punkte)**
 - (a) Ein 95%-Konfidenzintervall mit niedriger Fallzahl n in der Stichprobe als der Rest 95%-der Konfidenzintervalle
 - (b) Ein signifikantes, nicht relevantes 95%-Konfidenzintervall
 - (c) Ein signifikantes, relevantes 95%-Konfidenzintervall
 - (d) Ein 95%-Konfidenzintervall mit höherer Fallzahl n in der Stichprobe als der Rest der 95%-Konfidenzintervalle
 - (e) Ein signifikantes, relevantes 90%-Konfidenzintervall.
 - (f) Ein nicht signifikantes, nicht relevantes 95%-Konfidenzintervall



17 Aufgabe

(12 Punkte)

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



Nach einem Experiment mit zwei Pestiziden (*RoundUp* und *GoneEx*) ergibt sich die folgende Datentabelle mit dem gemessenen Trockengewicht (*drymatter*) von Weizen.

pesticide	drymatter
GoneEx	18
RoundUp	15
GoneEx	19
GoneEx	19
GoneEx	17
RoundUp	12
RoundUp	15
RoundUp	14
RoundUp	16
GoneEx	17
RoundUp	15
GoneEx	18
RoundUp	15

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Bestimmen Sie die Teststatistik T_{calc} eines Student t-Tests für den Vergleich der beiden Pestizide! **(5 Punkte)**
4. Treffen Sie mit $T_{\alpha=5\%} = 2.04$ und dem berechneten T_{calc} eine Aussage zur Nullhypothese! **(2 Punkte)**
5. Wenn Sie *keinen* Unterschied zwischen den beiden Pestiziden erwarten würden, wie groß wäre dann die Teststatistik T_{calc} ? Begründen Sie Ihre Antwort! **(2 Punkte)**