

**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

25. Juni 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 60 Punkten sind aus dem Rechen- und Textteil erreicht.

\_\_\_\_\_ von 80 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
76.5 - 80.0	1,0
72.5 - 76.0	1,3
68.5 - 72.0	1,7
64.5 - 68.0	2,0
60.5 - 64.0	2,3
56.5 - 60.0	2,7
52.5 - 56.0	3,0
48.5 - 52.0	3,3
44.5 - 48.0	3,7
40.0 - 44.0	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	9	9	8	7	9	8	10

- Es sind \_\_\_\_ von 60 Punkten erreicht worden.

## 1 Aufgabe

(2 Punkte)

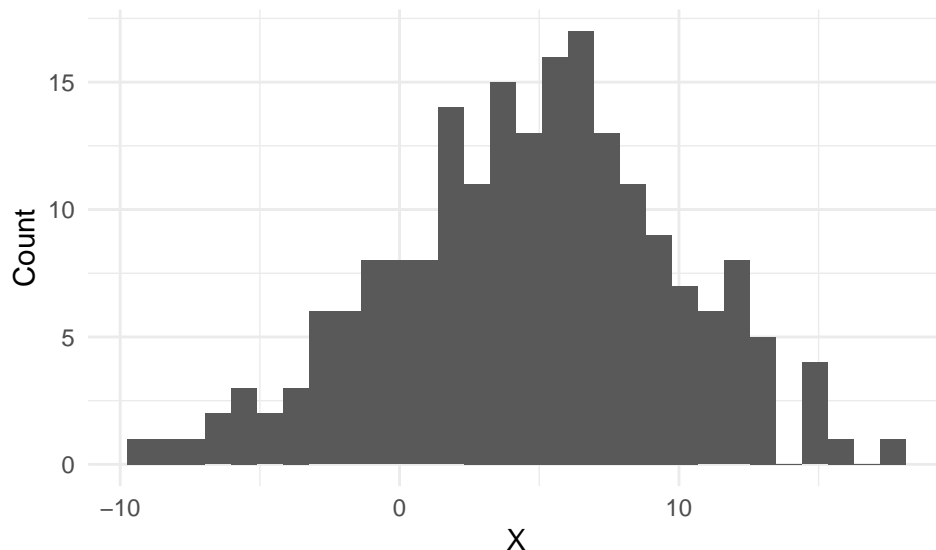
Eine einfaktorielle ANOVA berechnet eine Teststatistik um zu die Nullhypothese abzulehnen. Welche Aussage über die Teststatistik der ANOVA ist richtig?

- A** ☐ Die ANOVA berechnet die 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.
- B** ☐ 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.
- 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 aus den SS Behandlung geteilt durch die SS Fehler.
- E** ☐ 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.

## 2 Aufgabe

(2 Punkte)

In dem folgenden Histogramm von  $n = 187$  Pflanzen ist welche Verteilung mit welchen korrekten Verteilungsparametern dargestellt?



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

## 3 Aufgabe

(2 Punkte)

Sie haben folgende unadjustierten p-Werte gegeben: 0.01, 0.42, 0.89 und 0.21. Sie adjustieren die p-Werte nach Bonferroni. Welche Aussage ist richtig?

- A** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.04, 1, 1 und 0.84. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- B** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0025, 0.105, 0.2225 und 0.0525. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.

- C** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.0025, 0.105, 0.2225 und 0.0525. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.
- D** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.04, 1.68, 3.56 und 0.84. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 5% verglichen.
- E** ☐ Nach der Bonferroni-Adjustierung ergeben sich die adjustierten p-Werte von 0.04, 1, 1 und 0.84. Die adjustierten p-Werte werden zu einem  $\alpha$ -Niveau von 1.25% verglichen.

#### 4 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 erlaubt erst die Varianzen zu schätzen. Ohne eine Randomisierung ist die Berechnung von Mittelwerten und Varianzen nicht möglich.
- C** ☐ Randomisierung bringt starke Unstrukturiertheit in das Experiment 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 sorgt für Strukturgleichheit und erlaubt erst von der Stichprobe auf die Grundgesamtheit zurückzuschliessen.

#### 5 Aufgabe

(2 Punkte)

Die Abkürzung *CLD* steht für welches statistische Verfahren? Welche anschließende Beschreibung der Interpretation ist korrekt?

- A** ☐ Compact line display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Früher wurden keine Buchstaben sondern eine durchgezogene Linie verwendet. Bei mehr als drei Gruppen funktioniert die Linie aber graphisch nicht mehr.
- B** ☐ Compact letter display. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt. Teilweise ist die Interpretation des CLD herausfordernd, da wir ja nach dem Unterschied suchen.
- C** ☐ Compact letter detection. Gleichheit in den Behandlungen wird durch den gleichen Buchstaben oder Symbol dargestellt.
- D** ☐ Compound letter display. Gleichheit in dem Outcomes wird durch den gleichen Buchstaben oder Symbol dargestellt. Teilweise ist die Interpretation des Verbunds (eng. compound) herausfordernd, da wir ja nach dem Unterschied suchen.
- E** ☐ Contrast letter display. Unterschiede in den Behandlungen werden durch den gleichen Buchstaben oder Symbol dargestellt. Die Interpretation des CLD führt häufig in die Irre.

#### 6 Aufgabe

(2 Punkte)

Um zu Überprüfen, ob die Daten die Annahme einer Normalverteilung genügen, können wir folgende Visualisierung nutzen. Dabei kommt dann auch die entsprechende Regel zur Abschätzung der Annahme einer Normalverteilung zur Anwendung.

- A** ☐ Einen Dotplot. Die Punkte müssen sich wie an einer Perlenschnur anreihen. Eine Abweichung führt zur Ablehnung der Annahme.
- B** ☐ Einen Violinplot. Der Bauch der Violine muss hierbei einen höheren Wert annehmen als der Steg der Violine. Dann kann die Annahme angenommen werden.
- C** ☐ Einen Boxplot. Das IQR muss über alle Behandlungen zusammen mit den Whiskers ungefähr gleich sein.

- D** ☐ Einen Barplot. Die Mittelwerte müssen alle auf einer Höhe liegen. Die Fehlerbalken haben hier keine Informationen.
- E** ☐ Einen Boxplot. Der Median als Linie, muss in der Mitte des IQR, dargestellt durch die Box, liegen.

## 7 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  $\alpha$ -Fehler.
- B** ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.
- C** ☐ In die Analogie eines brennenden Hauses ohne Rauchmelder: *House without noise*.
- D** ☐ In die Analogie eines Rauchmelders: *Alarm with fire*.
- E** ☐ In die Analogie eines Rauchmelders: *Fire without alarm*, dem  $\beta$ -Fehler.

## 8 Aufgabe

(2 Punkte)

Welche Aussage über den  $p$ -Wert und dem Signifikanzniveau  $\alpha$  gleich 5% ist richtig?

- A** ☐ Wir vergleichen mit dem  $p$ -Wert und dem Signifikanzniveau  $\alpha$  absolute Werte auf einem Zahlenstrahl und damit den Unterschied der Teststatistiken, wenn die  $H_0$  gilt.
- B** ☐ Wir vergleichen mit dem  $p$ -Wert und dem Signifikanzniveau  $\alpha$  Wahrscheinlichkeiten und damit die Flächen unter der Kurve der Teststatistik, wenn die  $H_0$  gilt.
- C** ☐ Wir vergleichen die Effekte des  $p$ -Wertes mit den Effekten der Signifikanzschwelle unter der Annahme der Nullhypothese.
- D** ☐ Wir vergleichen mit dem  $p$ -Wert und dem Signifikanzniveau  $\alpha$  Wahrscheinlichkeiten und damit die absoluten Werte auf einem Zahlenstrahl, wenn die  $H_0$  gilt.
- E** ☐ Wir machen eine Aussage über die individuelle Wahrscheinlichkeit des Eintretens der Nullhypothese  $H_0$ .

## 9 Aufgabe

(2 Punkte)

Sie führen ein Feldexperiment durch um das Gewicht von Erdbeeren zu steigern. Die Pflanzen wachsen unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen. Nach der Berechnung einer einfaktoriellen ANOVA ergibt sich ein  $\eta^2 = 0.25$ . Welche Aussage ist richtig?

- A** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen nicht erklärt wird. Somit der Rest an nicht erklärbarer Varianz.
- B** ☐ Das  $\eta^2$  ist die Korrelation der ANOVA. Mit der Ausnahme, dass 0 der beste Wert ist.
- C** ☐ Die Berechnung von  $\eta^2$  ist ein Wert für die Interaktion.
- D** ☐ Das  $\eta^2$  ist ein Wert für die Güte der ANOVA. Je kleiner desto besser. Ein  $\eta^2$  von 0 bedeutet ein perfektes Modell mit keiner Abweichung. Die Varianz ist null.
- E** ☐ Das  $\eta^2$  beschreibt den Anteil der Varianz, der von den Behandlungsbedingungen erklärt wird. Das  $\eta^2$  ist damit mit dem  $R^2$  aus der linearen Regression zu vergleichen.

## 10 Aufgabe

(2 Punkte)

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

- A** ☐  $Pr(D|H_0)$  ist die Wahrscheinlichkeit der Alternativhypothese 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 Wahrscheinlichkeit für die Nullhypothese, wenn die Daten wahr sind.
- E** ☐ Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativhypothese überdeckt.

## 11 Aufgabe

(9 Punkte)

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



Der Datensatz `week7_growth_tbl` enthält das Lebendgewicht von Mastschweinen, die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden. Als Behandlung wurden verschiedene Nahrungszusätze in unterschiedlichen Dosen verfüttert. Als Behandlung haben Sie somit den Faktor *group* mit den Faktorstufen *lethal*, *extreme* und *high* vorliegen.

1. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle mit den gegebenen Informationen von **Df** und **Sum Sq** aus! (3 Punkte)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
<b>group</b>	2				
<b>error</b>	20	55.94			
<b>total</b>	22	218.61			

2. Schätzen Sie den p-Wert der Tabelle mit der Information von  $F_{\alpha=5\%} = 3.49$  ab. Begründen Sie Ihre Antwort! (2 Punkte)
3. Berechnen Sie den Effektschätzer  $\eta^2$ . Was sagt Ihnen der Wert von  $\eta^2$  aus? (2 Punkte)
4. Skizzieren Sie Körpergröße von fünf Tieren pro Behandlung für eine nicht signifikante, einfaktorielle ANOVA! (2 Punkte)



## 12 Aufgabe

(9 Punkte)

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



In einer Klimakammer wurde eine neue technische Versuchsanlage getestet. Bei dem Pilotexperiment mit sehr geringer Fallzahl ( $n_1 = n_2 = 3$ ) kamen zwei Behandlungen (*ctrl* und *dose*) zur Wachstumskontrolle zum Einsatz. Es ergibt sich die folgende Datentabelle mit dem gemessenen Gewicht (*weight*) von Kartoffeln.

treatment	weight
ctrl	15.5
dose	11.7
ctrl	22.1
dose	26.2
ctrl	14.0
dose	23.3


1. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
2. Bestimmen Sie die Teststatistik  $T_{calc}$  eines *Welch t-Tests* für den Vergleich der beiden Behandlungen! **(4 Punkte)**
3. Treffen Sie mit  $T_{\alpha=5\%} = 1.64$  und dem berechneten  $T_D$  eine Aussage zur Nullhypothese! Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Berechnen Sie den Effekt der Behandlungen *ctrl* und *dose* auf das Gewicht von Kartoffeln! **(1 Punkt)**

### 13 Aufgabe

(8 Punkte)

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



In einem Experiment für den Zuckergehalt von Erdbeeren in g/kg mit fünf Dosisstufen (ctrl, low, mid, high und pos) erhalten Sie folgendes *Compact letter display (CLD)* als  Ausgabe aus den rohen, unadjustierten *p*-Werten.

```
## ctrl high low mid pos  
## "a" "bc" "a" "b" "c"
```

1. Zeichnen Sie eine Abbildung, der sich ergebenden Barplots! **(2 Punkte)**
2. Ergänzen Sie das *Compact letter display (CLD)* zu der Abbildung! **(1 Punkt)**
3. Erklären Sie *einen* Vorteil und *einen* Nachteil des *Compact letter display (CLD)*! **(2 Punkte)**
4. 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)**

## 14 Aufgabe

(7 Punkte)

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



In einem Gewächshausexperiment mit drei Substrattypen (*torf*, *40p60n* und *70p30n*) als Behandlung (*treatment*) ergibt sich die folgende Datentabelle mit dem gemessenen Frischgewicht (*freshmatter*) von Maiss.

treatment	freshmatter
70p30n	44.8
70p30n	43.6
70p30n	40.9
70p30n	47.2
torf	34.5
40p60n	35.8
70p30n	35.6
40p60n	41.9
torf	56.4
40p60n	37.5
torf	46.7
torf	41.1
40p60n	42.6
40p60n	38.9

1. Zeichnen Sie in *einer* Abbildung die Barplots für die Behandlung von Maiss! Beschriften Sie die Achsen entsprechend! **(4 Punkte)**
2. Beschriften Sie *einen* Barplot mit den gängigen statistischen Maßzahlen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen den Behandlungen von Maiss erwarten würden, wie sehen dann die Barplots aus? Antworten Sie mit einer Skizze der Barplots! **(1 Punkt)**

## 15 Aufgabe

(9 Punkte)



Für ein besseres Verständnis der statistischen Testtheorie, auch Null-Ritual genannt, kann eine Visualisierung als Kreuztabelle genutzt werden.

1. Tragen Sie folgende statistische Fachbegriffe zur statistischen Testtheorie korrekt eine selbst erstellte Kreuztabelle ein! **(3 Punkte)**

$H_0$  falsch    $H_0$  abgelehnt    $\alpha$ -Fehler   Testentscheidung

2. Ergänzen Sie Ihre erstellte Kreuztabelle um vier weitere, passende Fachbegriffe zur statistischen Testtheorie! **(2 Punkte)**

Die Entscheidungsfindung durch einen statistischen Test kann auch durch die Analogie zu einem Feuermelder abgebildet werden. Dabei symbolisiert der Feuermelder den statistischen Test und es soll getestet werden, ob ein Feuer ausgebrochen ist.

3. In der Analogie des Feuermelders, wie lautet der  $\alpha$ -Fehler? **(1 Punkt)**
4. In der Analogie des Feuermelders, wie lautet der  $\beta$ -Fehler? **(1 Punkt)**
5. Wenn der Feuermelder einmal pro Tag messen würde, wie oft würde der Feuermelder mit einem  $\alpha$  von 5% in einem Jahr Alarm schlagen? Begründen Sie Ihre Antwort! **(2 Punkte)**

## 16 Aufgabe

(8 Punkte)

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



Nach einem Gewächshausexperiment zählen Sie folgende Anzahl an Läsionen auf den Blättern von Erdbeeren nach einer durchgestandenen Infektion der Pflanze.

5, 1, 4, 2, 4, 4, 6, 2, 6, 2, 9, 8, 3, 3, 4, 6, 5, 4, 6, 5, 9, 5, 5, 2, 6, 4, 6

1. Zeichnen Sie ein Histogramm um die Verteilung der Daten zu visualisieren! **(3 Punkte)**
2. Beschriften Sie die Achsen der Abbildung! **(2 Punkte)**
3. Ergänzen Sie die absoluten und relativen Häufigkeiten in der Abbildung! **(1 Punkt)**
4. Berechnen Sie aus den Daten die *Wahrscheinlichkeit* mehr als 5 Läsionen zu beobachten! **(1 Punkt)**
5. Berechnen Sie aus den Daten die *Chance* mehr als 5 Läsionen zu beobachten! **(1 Punkt)**

## 17 Aufgabe

(10 Punkte)



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

```
##  
## Two Sample t-test  
##  
## data: drymatter by Fe  
## t = -3.4374, df = 12, p-value = 0.004918  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## -7.469092 -1.673765  
## sample estimates:  
## mean in group ctrl mean in group trt2  
## 14.28571 18.85714
```

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 eine Abbildung in der Sie  $T_D$ ,  $Pr(D|H_0)$ ,  $A = 0.95$ , sowie  $T_{\alpha=5\%} = |2.18|$  einzeichnen! **(4 Punkte)**
4. Beschriften Sie die Abbildung entsprechend! **(1 Punkt)**
5. Berechnen Sie den Effekt der Behandlungen! **(1 Punkt)**