

Name: _____

Nicht bestanden: ☐

Vorname: _____

Matrikelnummer: _____

Endnote: _____

Studierende der Fakultät Agrarwissenschaften und Landschaftsarchitektur (AuL)

Probeklausur Bio Data Science

für Pflichtmodule

im 1. & 2. Semester B.Sc./M.Sc.

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

9. Dezember 2025

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 70 Punkten sind aus dem Rechen- und Textteil erreicht.

_____ von 90 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

Punkte	Note
86.0 - 90.0	1,0
81.5 - 85.5	1,3
77.0 - 81.0	1,7
72.5 - 76.5	2,0
68.0 - 72.0	2,3
63.5 - 67.5	2,7
59.0 - 63.0	3,0
54.5 - 58.5	3,3
50.0 - 54.0	3,7
45.0 - 49.5	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	Hausarbeit
Punkte	8	10	11	11	10	10	10

- Es sind ____ von 70 Punkten erreicht worden.

1 Aufgabe

(2 Punkte)

In Ihrer Forschungsarbeit wollen Sie eine Aussage über ein untersuchtes Individuum treffen. Dazu nutzen Sie einen statistischen Test. Erhalten Sie eine valide Aussage aus einem statistischen Test?

- A** ☐ Ja, es ist möglich ein untersuchtes Individuum mit einem statistischen Test wie einem t-Test auszuwerten. Wir erhalten dann eine Aussage zum Individuum.
- B** ☐ Ja, wir erhalten nur eine Aussage zu zwei Individuen. Ein statistischer Test liefert Informationen zu einem Individuum im Vergleich zu einem anderen Individuum.
- C** ☐ Nein, wir erhalten eine *einfache* Aussage. Wir müssen daher das Individuum im Kontext der Population adjustieren.
- D** ☐ Nein, ein untersuchtes Individuum können wir mit einem statistischen Test nicht auswerten. Wir erhalten keine Aussage zum Individuum.
- E** ☐ Nein, wir können ein untersuchtes Individuum nicht mit einem t-Test auswerten. Wir erhalten keine Aussage zum Individuum. Wir können aber den Effekt als Quelle der Relevanz nutzen.

2 Aufgabe

(2 Punkte)

In Ihrer Abschlussarbeit rechnen Sie einen Student t-Test. Welche Aussage ist auch für den Welch t-Test richtig?

- A** ☐ Der t-Test ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern
- B** ☐ Der t-Test testet generell zu einem erhöhten α -Niveau von 20%.
- C** ☐ Der t-Test vergleicht zwei Gruppen indem die Mittelwerte miteinander verglichen werden.
- D** ☐ 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.
- E** ☐ Der t-Test berechnet die Differenz von zwei Mittelwerten als Effekt und gibt eine Entscheidung, ob sich die beiden Mittelwerte *jeweils* von Null unterscheiden.

3 Aufgabe

(2 Punkte)

Gegeben ist y mit 10, 23, 21, 32, 26, 23, 21, 11, 16, 28 und 63. Berechnen Sie den Median, das 1st Quartile sowie das 3rd Quartile.

- A** ☐ Sie erhalten 23 [14; 26]
- B** ☐ Es berechnet sich 24 [17; 27]
- C** ☐ Es ergibt sich 23 [16; 28]
- D** ☐ Sie erhalten 23 +/- 28
- E** ☐ Es ergibt sich 25 +/- 16

4 Aufgabe

(2 Punkte)

Der Boxplot stellt folgende statistische Maßzahlen in einer Abbildung dar. Damit gehört der Boxplot zu einem der am meisten genutzten statistischen Verfahren zur Visualisierung von Daten.

- A** ☐ Durch die Abbildung des Boxplot erhalten wir die Informationen über die Mittelwerte und die Standardabweichung.
- B** ☐ Den Median und die Quartile.
- C** ☐ Den Mittelwert und die Varianz.
- D** ☐ Der Boxplot stellt den Median und die Streuung dar.
- E** ☐ Den Median und die Standardabweichung.

5 Aufgabe

(2 Punkte)

Viele statistische Verfahren nutzen eine Teststatistik um eine Aussage über den Zusammenhang zwischen der Grundgesamtheit und der Stichprobe abzubilden. Ein statistisches Testwerkzeug ist hierbei die ANOVA. Die ANOVA rechnet dabei...

- A** ☐ ... den Unterschied zwischen der Mittelwerte und der Varianz aus verschiedenen Behandlungsguppen. Wenn die ANOVA signifikant ist, ist bekannt welcher Vergleich konkret unterschiedlich ist.
- B** ☐ ... den Unterschied zwischen der F-Statistik anhand der Varianz der Gruppen. Wenn die F-Statistik exakt 0 ist, kann die Nullhypothese abgelehnt werden.
- C** ☐ ... den Unterschied zwischen der globalen Varianz und der Varianz aus verschiedenen Behandlungsguppen. Wenn die ANOVA signifikant ist, ist nicht bekannt welcher Vergleich konkret unterschiedlich ist.
- D** ☐ ... den Unterschied zwischen der Varianz ausgelöst durch alle Behandlungsgruppen und der Varianz aus globalen Behandlungsgruppen der Kontrollen. Wenn die ANOVA nicht signifikant ist, muss ein Posthoc-Test ausgeschlossen werden.
- E** ☐ ... den Unterschied zwischen der Varianz über alle Behandlungsgruppen oder der Varianz aus verschiedenen Behandlungsgruppen. Wenn die ANOVA signifikant ist, muss sich zwischen einem der beiden Varianzquellen entschieden werden.

6 Aufgabe

(2 Punkte)

Sie haben den mathematischen Ausdruck $Pr(D|H_0)$ vorliegen, welche Aussage ist richtig?

- A** ☐ Die Inverse der Wahrscheinlichkeit unter der die Nullhypothese nicht mehr die Alternativhypothese überdeckt.
- B** ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit der Alternativhypothese und somit $1 - Pr(H_A)$
- C** ☐ $Pr(D|H_0)$ beschreibt die Wahrscheinlichkeit die Teststatistik T_D aus den Daten D zu beobachten, wenn die Nullhypothese wahr ist.
- D** ☐ $Pr(D|H_0)$ ist die Wahrscheinlichkeit nicht die Daten D zu beobachten sondern die Nullhypothese, wenn diese wahr ist.
- E** ☐ $Pr(D|H_0)$ stellt die Wahrscheinlichkeit die Teststatistik T zu beobachten dar, wenn die Nullhypothese falsch ist.

7 Aufgabe

(2 Punkte)

Die Testtheorie hat einen philosophischen Unterbau. Eins der Prinzipien ist das Falsifikationsprinzip. Das Falsifikationsprinzip besagt,

- A** ☐ ... dass Annahmen an statistische Modelle meist falsch sind.
- B** ☐ ... dass ein schlechtes Modell durch das Falsifikationsprinzip durch ein noch schlechteres Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.
- C** ☐ ... dass ein minderwertiges Modell durch ein minderwertiges Modell ersetzt wird. Es gilt das Verifikationsprinzip nach Karl Popper.
- 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)

Aus einem Feldversuch ergibt sich die Notwendigkeit der Berechnung einer einfaktoriellen ANOVA. Es ergibt sich ein $\eta^2 = 0.52$. 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.
- C** ☐ Die Berechnung von η^2 ist ein Wert für die Interaktion in der einfaktoriellen ANOVA.
- D** ☐ Das η^2 ist die Korrelation der ANOVA. Mit der Ausnahme, dass $\eta^2 = 0$ der beste Wert ist.
- E** ☐ 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.

9 Aufgabe


(2 Punkte)

Ein statistischer Test benötigt für die richtige Durchführung Hypothesen H , sonst ist der Test nicht zu interpretieren. Welche Aussage ist richtig?

- A** ☐ Mit der Nullhypothese H_A und der Alternativhypothese H_0 gibt es zwei Hypothesen, die aber selten genutzt werden.
- B** ☐ Es gibt - bedingt durch das das Falsifikationsprinzip - ein Set von k Nullhypothesen, die iterative gegen $k - 1$ Alternativhypothesen getestet werden.
- C** ☐ Es gibt ein Hypothesenset bestehend aus k Hypothesen. Meistens wird die Nullhypothese H_0 und die Alternativhypothese H_A verwendet. Wegen des Falsifikationsprinzips ist es wichtig, die bekannte falsche und unbekannte richtige Hypothese mit in das Set zu nehmen.
- D** ☐ Es gibt ein statistisches Hypothesenpaar mit der Nullhypothese H_0 und der Alternativhypothese H_A oder H_1 .
- E** ☐ Ein statistisches Hypothesenpaare gibt es. Zum einen die Nullhypothese und zum anderen die Alternativhypothese. Es ist aber nur notwendig die Alternative anzugeben, da die Nullhypothese nicht beim Testen benötigt wird.

10 Aufgabe

(2 Punkte)

Wenn wir ein Experiment in  auswerten wollen, dann müssen die Daten in Excel in einem bestimmten Format angelegt werden. Wir sprechen auch vom Long-Format. Welche Aussage zum Long-Format ist richtig?

- A** ☐ Das Long-Format beschreibt in den Spalten die Beobachtungen sowie in den Zeilen die *unabhängigen* Beobachtungen.
- B** ☐ In den Spalten finden sich die einzelnen Beobachtungen und in den Zeilen die gemessenen Variablen (Y) sowie die experimentellen Faktoren (X).
- C** ☐ Wichtig ist, dass sich in den Zeilen die Beobachtungen finden und in den Spalten die Variablen, wie die Messwerte (Y) sowie die experimentellen Faktoren (X).
- D** ☐ Wichtig ist, dass sich in den Zeilen die Variablen finden, wie die Messwerte (Y) sowie die experimentellen Faktoren (X). In den Spalten stehen die Variablen.
- E** ☐ Das Long-Format hat die gleiche Anzahl an Spalten wie auch an Zeilen.

11 Aufgabe

(8 Punkte)

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

Eine echte Herausforderung für ihn war schon immer die Romantik gewesen. Ein leidiges Lied. Deshalb gilt anschauen, was andere vor einem gemacht haben. Für Steffen ist es eine Möglichkeit schneller ans Ziel zu gelangen. Deshalb hat sich Steffen viele Poster in der Fakultät angeschaut und ist zum Schluß gekommen, dass Barplots eine häufig genutzte Abbildung sind. Steffen soll nun in seiner Abschlussarbeit Maiss untersuchen. Die Behandlung in seiner Abschlussarbeit sind verschiedene Substrattypen (*torf*, *40p60n* und *70p30n*). Erhoben wurden von Steffen als Endpunkt (Y) *Frischegewicht* benannt als *freshmatter* in seiner Exceldatei. Erwartungsgemäß erhält er von seinem Betreuer den Auftrag die erhobenen Daten als Barplots darzustellen. Dann kann Steffen auch schonmal abschätzen, was bei einem statistischen Test rauskommen könnte. Na dann mal los. Steffen schafft sich die nötige Stimmung. Steffen streichelt liebevoll die Schlange. Der Kopf ist in seinem Schloß vergraben um den Klang von Taylor Swift zu dämpfen.

treatment	freshmatter
40p60n	25.9
70p30n	51.3
40p60n	31.7
torf	38.2
70p30n	38.2
40p60n	44.0
70p30n	34.6
70p30n	49.7
70p30n	34.1
torf	37.0
torf	21.1
torf	18.6
torf	36.6

Leider kennt sich Steffen mit der Erstellung von Barplots nicht aus. Deshalb braucht er bei der Visualisierung Ihre Hilfe!

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

12 Aufgabe

(10 Punkte)

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

Das Verrückte ist, dass die Schlange Harry Potter wirklich liebt. Das ist Steffen sehr recht, denn er braucht Entspannung. Aber Ablenkung hilft nur begrenzt. 'Uff!', denkt sich Steffen. Jetzt hat er doch tatsächlich zwei kategoriale Variablen in seiner Hausarbeit gemessen. Zum einen die Behandlung Pestizideinsatz [ja/nein] und zum anderen die Messung Proteingehalt im Zielbereich [ja/nein] im Kontext von Brokkoli. Hierfür hat er ein Feldexperiment im Teuteburgerwald durchgeführt. Jetzt möchte Steffen die Daten einmal in einer explorativen Datenanalyse darstellen. Danach kann er dann über den passenden statistischen Test nachdenken. Dabei unterstützt sein Betreuer diesen Ansatz bevor es in der Datenanalyse weiter geht. So schön wie so gut. Wenn die Romantik nicht wäre, ja dann wäre wohl vieles möglich für Steffen! Aber so..

Proteingehalt im Zielbereich	Pestizideinsatz	Proteingehalt im Zielbereich	Pestizideinsatz
ja	ja	nein	ja
nein	ja	nein	ja
nein	ja	nein	nein
nein	nein	nein	ja
nein	nein	ja	ja
ja	ja	ja	ja
ja	ja	ja	nein
ja	ja	ja	ja
nein	ja	ja	ja
nein	ja	nein	ja
ja	ja	nein	ja
ja	ja	ja	ja
ja	ja	ja	ja
ja	ja	nein	ja
nein	ja	ja	ja
nein	ja	nein	ja
ja	ja	ja	ja

Leider kennt sich Steffen mit der Erstellung einer explorativen Datenanalyse für kategoriale Daten überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Stellen Sie den Zusammenhang zwischen den beiden kategorialen Variablen in einer zusammenfassenden Tabelle dar! **(3 Punkte)**
2. Berechnen Sie die Verhältnisse in der zusammenfassenden Tabelle! Welche Annahme haben Sie getroffen? **(2 Punkte)**
3. Visualisieren Sie den Zusammenhang zwischen den beiden kategorialen Variablen! **(3 Punkte)**
4. Wenn *ein* Effekt von der Behandlung vorliegen würde, wie würde die Tabelle und die Visualisierung aussehen? **(2 Punkt)**

13 Aufgabe

(10 Punkte)

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

'Wir waren im Wendland um Lauch in einem Versuch in einer Klimakammer zu messen.', Paula legt das Dokument auf den Tisch und schaut Alex und Steffen fragend an. Beide schauen fragend zurück. Gäbe es die Romantik nicht, dann wäre es für Steffen irgendwie einfacher hier zu helfen. Echt unangenehm. Die beiden sind zu Paula gekommen, da sie sich nicht mit **R** auskennen und daher Hilfe bei der Interpretation des t-Tests brauchen. Im Hintergrund wummert White Lies und leere Smarties Packungen stappeln sich auf dem Boden. 'Kein Problem', sagt Paula und streichelt langsam die Ratte. 'Aber worum es in dem Versuch geht, lässt sich nur aus dem Text in seiner Hand erraten.' merkt sie an. Vielleicht hilft da ja die Ausgabe des t-Tests in R weiter. Draußen geht blutrot die Sonne unter.

```
##  
## Two Sample t-test  
##  
## data: Chlorophyllgehalt by Genotypen  
## t = 10.698, df = 14, p-value = 4.025e-08  
## alternative hypothesis: true is not equal to [condensed]  
## 95 percent confidence interval:  
## 11.61292 17.43708  
## sample estimates:  
## mean in group AA mean in group BB  
## 43.700 29.175
```

Helfen Sie Paula bei der Interpretation des t-Tests! Sonst geht es auch für Alex und Steffen nicht weiter.

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(1 Punkt)**
3. Liegt ein signifikanter Unterschied zwischen den Gruppen vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
4. Berechnen Sie den Effekt des t-Tests! **(1 Punkt)**
5. Skizzieren Sie eine Abbildung in der Sie T_D , $Pr(D|H_0)$, $A = 0.95$, sowie $T_{\alpha=5\%} = |2.14|$ einzeichnen! **(4 Punkte)**
6. Beschriften Sie die Abbildung! **(1 Punkt)**

14 Aufgabe

(11 Punkte)

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

Nilufar und Paula schauen sich etwas entnervt an. Gemeinsam schreiben die beiden ihre Abschlussarbeit und sollen nun als erstes einmal die Daten visualisieren damit abgeschätzt werden kann, ob überhaupt signifikante Ergebnisse zu erwarten sind. Die beiden waren im Emsland um ein Kreuzungsexperiment mit Zandern durchzuführen. Dabei haben Nilufar und Paula den Messwert Fettgehalt [%/kg] unter der Behandlung Genotypen (AA, AB und BB) ermittelt. Kennengelernt haben sich die beiden auf einem Konzert von White Lies. Später wird noch Jagd auf roter Oktober geguckt. Paula befürwortet das!

Genotypen	Fettgehalt
BB	28
AB	49
AA	46
AA	46
AA	44
AA	45
BB	24
AA	45
BB	24
AB	45
AB	46
BB	28
BB	27
AB	45
AB	44
BB	24


Leider kennen sich Nilufar und Paula mit Darstellung einer einfaktoriellen ANOVA überhaupt nicht aus.

1. Erstellen Sie eine Visualisierung der Datentabelle! Beschriften Sie die Abbildung! **(2 Punkte)**
2. Benennen Sie die Visualisierung mit dem korrekten, statistischen Fachbegriff! **(1 Punkt)**
3. Zeichnen Sie folgende statistischen Maßzahlen passend ein!
 - Den globalen Mittelwert β_0 **(1 Punkt)**
 - Die Mittelwerte der einzelnen Behandlungsstufen **(1 Punkt)**
 - Die Mittelwertsdifferenz der einzelnen Behandlungsstufen mit β_{AA} , β_{AB} und β_{BB} **(1 Punkt)**
 - Die Residuen oder Fehler mit ϵ **(1 Punkt)**
4. Liegt ein *vermutlicher* signifikanter Unterschied vor? Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Schätzen Sie die Effekte der Behandlungsstufen! **(2 Punkte)**

15 Aufgabe

(10 Punkte)

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

Alex schaut sich fragend in der Bibliothek um. Alex hatte gehofft, dass jemand hier sein würde, den er kennt und sich mit  auskennt. Wird aber enttäuscht. Alex war im Emsland um ein Kreuzungsexperiment mit Hühnern durchzuführen. Nun möchte seine Betreuerin seiner Abschlussarbeit erstmal eine ANOVA sehen und die Ergebnisse präsentiert bekommen. Dabei ging es herauszufinden, ob es einen Zusammenhang zwischen der Behandlung Flüssignahrung (*ctrl*, *superIn* und *fLOW*) und dem Messwert Fettgehalt [%/kg] gibt. Und eigentlich wollte Alex doch noch zum Sport! Um zu Laufen geht Alex dann später nochmal raus. Echte Entspannung.

```
## Analysis of Variance Table
##
## Response: Fettgehalt
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Flüssignahrung  2 3045.53 1522.77  34.715 4.523e-07
## Residuals      19  833.42   43.86
```

Leider kennen sich Alex mit Berechnung einer einfaktoriellen ANOVA überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Formulieren Sie das statistische Hypothesenpaar! **(2 Punkte)**
3. Interpretieren Sie das Ergebnis der einfaktoriellen ANOVA! **(2 Punkte)**
4. Berechnen Sie den Effektschätzer η^2 . Was sagt Ihnen der Wert von η^2 aus? **(2 Punkte)**
5. Skizzieren Sie eine Abbildung, der dem obigen Ergebnis der einfaktoriellen ANOVA näherungsweise entspricht! **(3 Punkte)**

16 Aufgabe

(11 Punkte)

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

Am Ende war es für Steffen in seiner Abschlussarbeit dann doch kein normalverteiltes Outcome. Das was jetzt etwas doof, da er sich auf eine ANOVA gefreut hatte. Dann noch schnell Taylor Swift auf das Ohr und los gehts. Prinzipiell ginge das auch irgendwie, aber nun möchte sein Betreuer gerne einen χ^2 -Test auf einer 2x2-Kreuztabelle berechnet bekommen. Steffen hatte sich in ein Gewächshausexperiment $n = 202$ Beobachtungen von Kartoffeln angeschaut. Dabei hat er als Behandlung *Herbizideinsatz* [ja/nein] bestimmt und zum anderen die Variable *Frischegewicht über Zielwert* [ja/nein] ermittelt. Jetzt muss Steffen mal schauen, wie er das jetzt rechnet. Eigentlich wollte Steffen nachher noch zum Sport. Um zu Ringen geht Steffen dann später nochmal raus. Echte Entspannung.

	56	51	
	43	52	

Leider kennt sich Steffen mit der Berechnung eines χ^2 -Test für kategoriale Daten überhaupt nicht aus. Deshalb braucht er bei der Erstellung Ihre Hilfe!

1. Formulieren Sie die wissenschaftliche Fragestellung! **(1 Punkt)**
2. Ergänzen Sie die Tabelle um die fehlenden Informationen! **(1 Punkt)**
3. Berechnen Sie die Teststatistik eines Chi-Quadrat-Test! **(3 Punkte)**
4. Treffen Sie eine Entscheidung im Bezug zu der Nullhypothese gegeben einem $\chi^2_{\alpha=5\%} = 5.61$! Begründen Sie Ihre Antwort! **(2 Punkte)**
5. Berechnen Sie den Effektschätzer *Cramers V* auf der 2x2 Kreuztabelle! **(1 Punkt)**
6. Welchen Wertebereich kann der Effektschätzer *Cramers V* annehmen? Wann liegt kein Effekt und wann ein starker Effekt vor? **(2 Punkte)**
7. Interpretieren Sie den berechneten Effektschätzer *Cramers V* unter Berücksichtigung der Fragestellung! **(1 Punkt)**

17 Hausarbeit

(10 Punkte)

Teil der Prüfungsleistung in dem Modul ist eine Hausarbeit zur Datenanalyse in R.

Diese Aufgabe dient in der Klausur als Platzhalter für die erreichten Punkte in der Hausarbeit.

Zum Bestehen des Moduls ist die Abgabe des Berichts nicht notwendig.

Alle Informationen zu der Hausarbeit sowie dem Stichtag der Abgabe finden Sie auf ILIAS.