

**Name:** \_\_\_\_\_

*Nicht bestanden:* ☐

**Vorname:** \_\_\_\_\_

**Matrikelnummer:** \_\_\_\_\_

**Endnote:** \_\_\_\_\_

**M.Sc. Angewandte Nutztier- und Pflanzenwissenschaften**

# **Klausur Biostatistik**

**Hochschule Osnabrück**

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

Klausur vom 07. Juli 2022

## Ergebnis der Klausur

\_\_\_\_\_ von 20 Punkten sind aus dem Multiple Choice Teil erreicht.

\_\_\_\_\_ von 61 Punkten sind aus dem Rechen- und Textteil erreicht.

\_\_\_\_\_ von 101 Punkten in Summe.

Es wird folgender Notenschlüssel angewendet.

<b>Punkte</b>	<b>Note</b>
96 - 101	1,0
91 - 95	1,3
85 - 90	1,7
79 - 84	2,0
73 - 78	2,3
67 - 72	2,7
61 - 66	3,0
55 - 60	3,3
51 - 54	3,7
50	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.

<b>Aufgabe</b>	11	12	13	14	15	16	17
<b>Punkte</b>							

- Es sind \_\_\_\_ von 61 Punkten erreicht worden.

<b>Aufgabe</b>	18	19	20
<b>Punkte</b>			

- Es sind \_\_\_\_ von 20, Punkten erreicht worden.

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

## 2 Aufgabe

(2 Punkte)

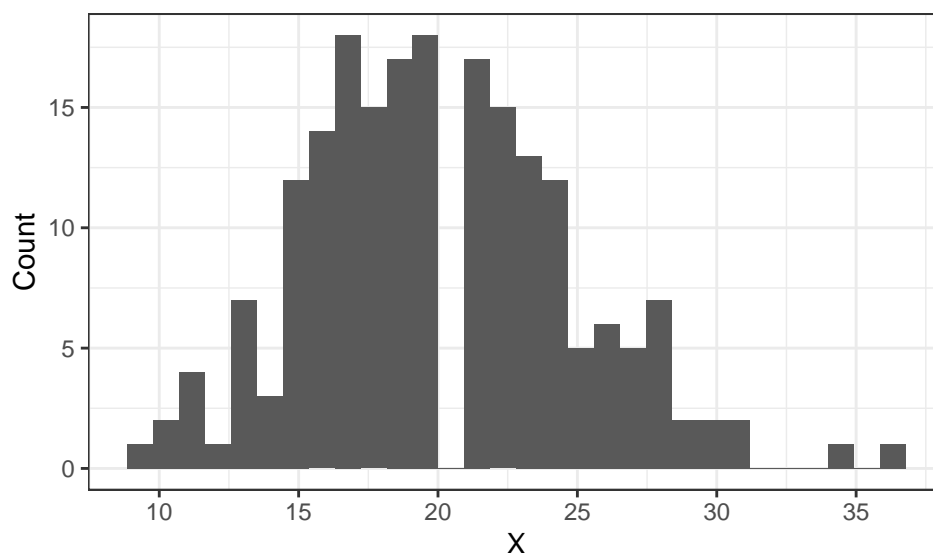
Welche Aussage zum mathematische Ausdruck  $Pr(D|H_0)$  ist richtig?

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

## 3 Aufgabe

(2 Punkte)

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



- A** ☐ Eine rechtsschiefe, multivariate Normalverteilung.
- B** ☐ Es handelt sich um eine Normalverteilung mit  $N(20, 5)$ .
- C** ☐ Eine Standardnormalverteilung mit  $N(0,1)$ .

**D** ☐ Es handelt sich um eine Binomial-Verteilung mit  $\text{Binom}(10)$ .

**E** ☐ Es handelt sich um eine Poisson-Verteilung mit  $\text{Pois}(20)$ .

#### 4 Aufgabe

(2 Punkte)

Das Falsifikationsprinzip besagt...

**A** ☐ ... dass Annahmen an statistische Modelle meist falsch sind.

**B** ☐ ... dass ein schlechtes Modell durch ein weniger schlechtes Modell ersetzt wird. Die Wissenschaft lehnt ab und verifiziert nicht.

**C** ☐ ... dass Modelle meist falsch sind und selten richtig.

**D** ☐ ... dass Fehlerterme in statistischen Modellen nicht verifiziert werden können.

**E** ☐ ... dass in der Wissenschaft immer etwas falsch sein muss. Sonst gebe es keinen Fortschritt.

#### 5 Aufgabe

(2 Punkte)

Sie haben das abstrakte Modell  $Y \sim X$  mit  $X$  als Faktor mit zwei Leveln vorliegen. Welche Aussage über  $n_1 < n_2$  ist richtig?

**A** ☐ Es handelt sich um ein unbalanciertes Design

**B** ☐ Es liegt Varianzhomogenität vor.

**C** ☐ Es liegt Varianzheterogenität vor.

**D** ☐ Es handelt sich um ein balanciertes Design.

**E** ☐ Es handelt sich um abhängige Beobachtungen.

#### 6 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  $\beta$ -Fehler.

**B** ☐ In die Analogie eines Feuerwehrautos: *Car without noise*.

**C** ☐ In die Analogie eines Rauchmelders: *Alarm without fire*, dem  $\alpha$ -Fehler.

**D** ☐ In die Analogie eines Rauchmelders: *Alarm with fire*.

**E** ☐ In die Analogie eines brennenden Hauses ohne Rauchmelder: *House without noise*.

#### 7 Aufgabe

(2 Punkte)

In einer linearen Regression werden die  $\epsilon$  oder Residuen geschätzt. Welcher Verteilung folgen die Residuen bei einer optimalen Modellierung?

**A** ☐ Die Residuen folgen einer Poissonverteilung mit  $\text{Pois}(0)$ .

**B** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(0, 1)$ .

**C** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(0, s^2)$ .

**D** ☐ Die Residuen sind binomialverteilt.

**E** ☐ Die Residuen sind normalverteilt mit  $\mathcal{N}(\bar{y}, s^2)$ .

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

## 9 Aufgabe

(2 Punkte)

Welche Aussage über den t-Test ist richtig?

- A** ☐ Der t-Test vergleicht die Mittelwerte von zwei Gruppen.
- B** ☐ Der t-Test ist ein Vortest der ANOVA und basiert daher auf dem Vergleich von Streuungsparametern
- C** ☐ 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.
- D** ☐ Der t-Test testet generell zu einem erhöhten  $\alpha$ -Niveau von 20%.
- E** ☐ Der t-Test vergleicht die Varianzen von mindestens zwei oder mehr Gruppen

## 10 Aufgabe

(2 Punkte)

Berechnen Sie den Mittelwert und Standardabweichung von y mit 10, 6, 9, 5 und 14.

- A** ☐ Es ergibt sich 8.8 +/- 12.7
- B** ☐ Es ergibt sich 7.8 +/- 6.35
- C** ☐ Es ergibt sich 8.8 +/- 3.56
- D** ☐ Es ergibt sich 9.8 +/- 1.78
- E** ☐ Es ergibt sich 8.8 +/- 1.78

## 11 Aufgabe

(6 Punkte)


1. Skizzieren Sie 3 Normalverteilungen *in einer Abbildung* mit  $\mu_1 \neq \mu_2 \neq \mu_3$  und  $s_1 = s_2 = s_3$ ! (2 Punkte)
2. Beschriften Sie die Normalverteilungen mit den entsprechenden Parametern! (2 Punkte)
3. Liegt Varianzhomogenität oder Varianzheterogenität vor? Begründen Sie Ihre Antwort! (2 Punkte)





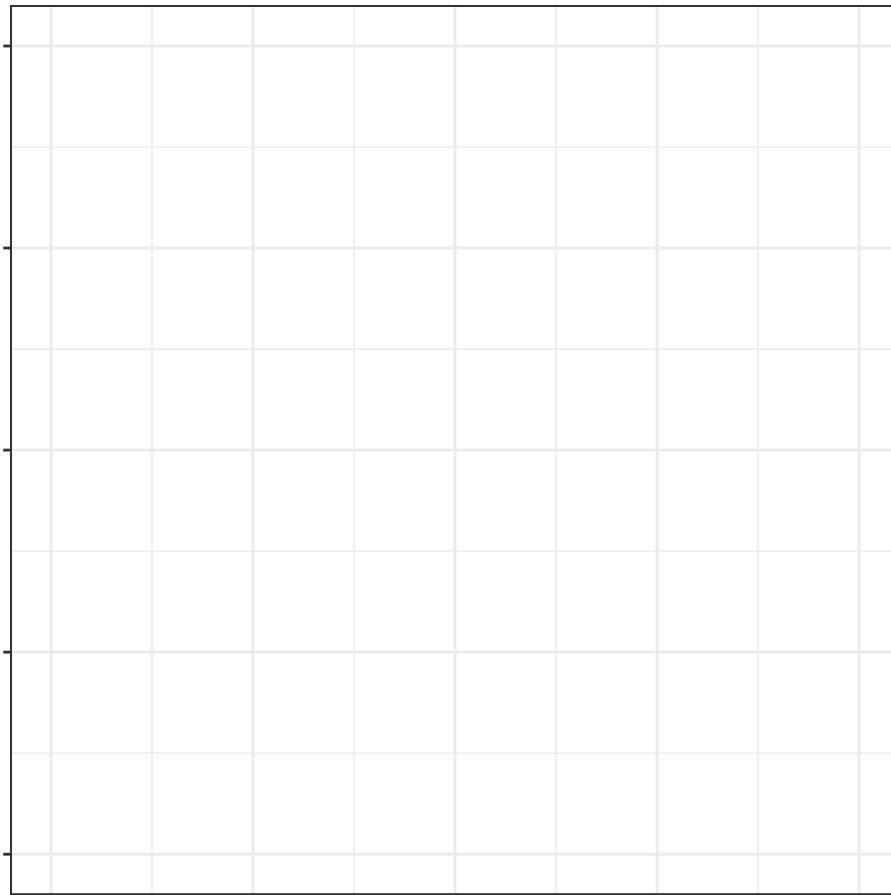
## 12 Aufgabe

(10 Punkte)

In einem Stallexperiment mit  $n = 108$  Ferkeln wurde der Gewichtszuwachs unter bestimmten Lichtverhältnissen 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)	24.71	1.03		
light	1.53	0.10		

1. Berechnen Sie die t Statistik für *(Intercept)* und *light*! **(2 Punkte)**
2. Schätzen Sie den p-Wert für *(Intercept)* und *light* mit  $T_k = 1.96$  ab. Was sagt Ihnen der p-Wert aus? Begründen Sie Ihre Antwort! **(3 Punkte)**
3. Zeichnen Sie die Gerade aus der obigen Tabelle in die untenstehende Abbildung! **(1 Punkt)**
4. Beschriften Sie die Abbildung und die Gerade mit den statistischen Kenngrößen! **(2 Punkte)**
5. Formulieren Sie die Regressionsgleichung! **(2 Punkte)**





### 13 Aufgabe

(14 Punkte)

Der Datensatz PlantGrowth enthält das Gewicht der Pflanzen (*weight*), die unter einer Kontrolle und zwei verschiedenen Behandlungsbedingungen erzielt wurden – dem Faktor *group* mit den Faktorstufen *ctrl*, *trt1*, *trt2*.

1. Füllen Sie die unterstehende einfaktorielle ANOVA Ergebnistabelle aus mit den gegebenen Informationen von Df und Sum Sq! **(4 Punkte)**
2. Schätzen Sie den p-Wert der Tabelle mit der Information von  $F_k = 3.35$  ab. Begründen Sie Ihre Antwort! **(2 Punkte)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	3.96			
Residuals	27	10.3			

3. Was bedeutet ein signifikantes Ergebnis in einer einfaktoriellen ANOVA im Bezug auf die möglichen Unterschiede zwischen den Gruppen? **(2 Punkte)**
4. Berechnen Sie *einen* Student t-Test mit  $T = \frac{\bar{x}_1 - \bar{x}_2}{s_p \cdot \sqrt{2/n_g}}$  für den *vermutlich* signifikantesten Gruppenvergleich anhand der untenstehenden Tabelle mit  $T_k = 2.03$ . Begründen Sie Ihre Auswahl! **(4 Punkte)**

group	n	mean	sd
ctrl	10	5.02	0.58
trt1	10	4.62	0.79
trt2	10	5.51	0.43

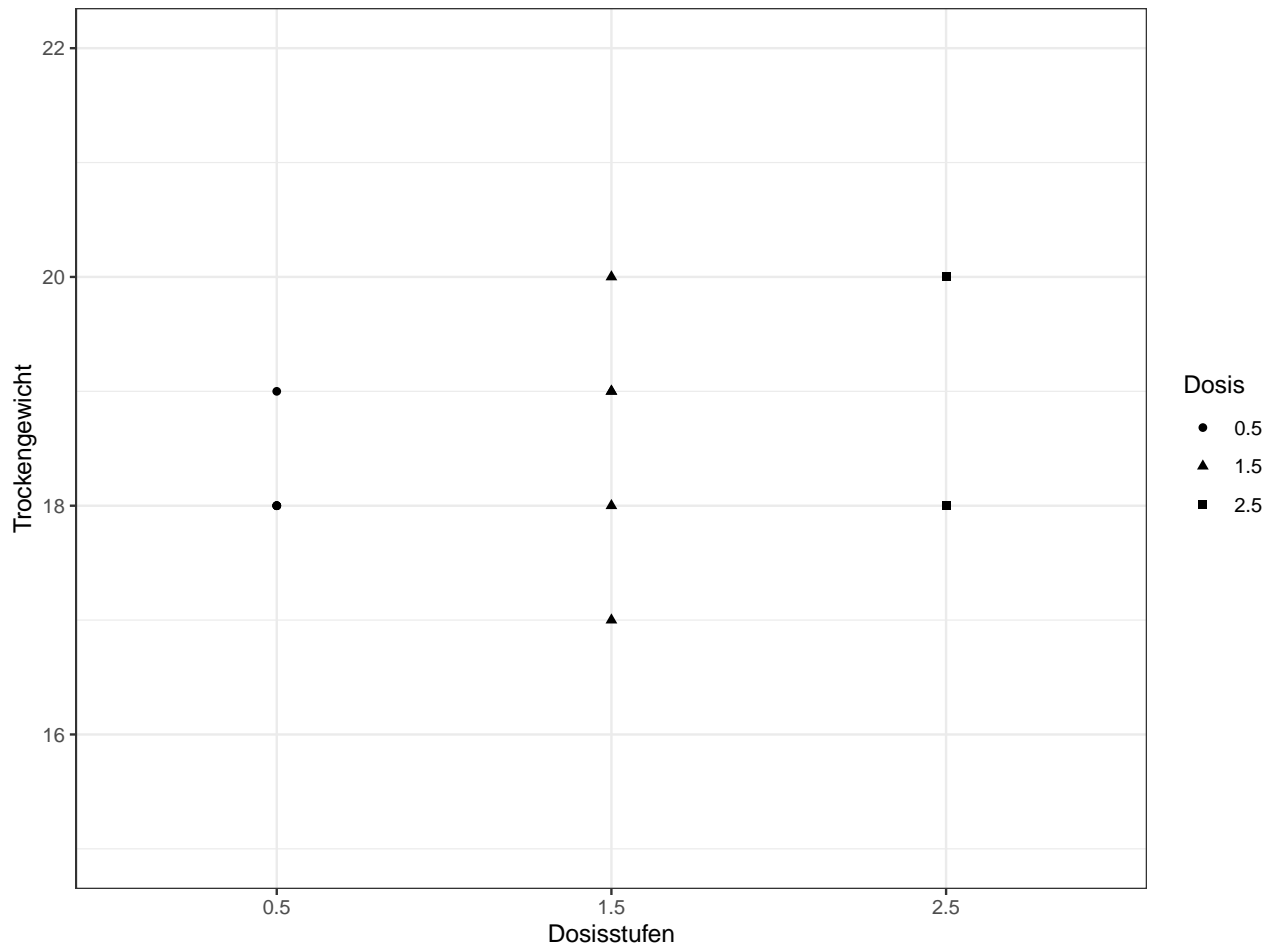
5. Gegebenen der ANOVA Tabelle war das Ergebnis des t-Tests zu erwarten? Begründen Sie Ihre Antwort! **(2 Punkte)**



## 14 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 Abbildung ein! (6 Punkte)

- Total (grand) mean:  $\beta_0$
- Mittelwerte der Dosen:  $\bar{x}_{0.5}$ ,  $\bar{x}_{1.5}$  und  $\bar{x}_{2.5}$
- Effekt der einzelnen Level der Dosen:  $\beta_{0.5}$ ,  $\beta_{1.5}$ , und  $\beta_{2.5}$
- Residuen oder Fehler:  $\epsilon$

2. Schätzen Sie den p-Wert einer einfaktoriellen ANOVA ab. Liegt ein *vermutlicher* signifikanter Unterschied zwischen den Dosisstufen vor? Begründen Sie Ihre Antwort! (2 Punkte)



## 15 Aufgabe

(10 Punkte)

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

water	drymatter
21	24
25	20
22	26
19	28
19	28
16	25
23	25
20	28
20	18

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! **(4 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? **(1 Punkt)**





## 16 Aufgabe

(9 Punkte)

Nach einem Feldexperiment mit zwei Düngestufen (A und B) ergibt sich die folgende Datentabelle mit dem gemessenen Trockengewicht (*drymatter*).

trt	drymatter
A	20
A	28
B	22
A	25
B	30
B	23
B	24
B	30
B	28
B	30
A	24
A	27
A	21
A	24

1. Zeichnen Sie in *einer* Abbildung die beiden Boxplots für die zwei Düngestufen A und B! Beschriften Sie die Achsen entsprechend! **(6 Punkte)**
2. Beschriften Sie *einen* der beiden Boxplots mit den gängigen statistischen Maßzahlen. Beschriften Sie auch die Achsen! **(2 Punkte)**
3. Wenn Sie *keinen Effekt* zwischen de Düngestufen erwarten würden, wie sehen dann die beiden Boxplots aus? **(1 Punkt)**



## 17 Aufgabe

(4 Punkte)

1. Zeichnen Sie über den untenstehenden Boxplot die entsprechende zugehörige Verteilung! (2 Punkte)
2. Zeichnen Sie den entsprechenden Lageparameter in die Abbildung. Beschriften Sie die Abbildung! (2 Punkte)





## Tierzuchtfragen

### 18 Aufgabe

(6 Punkte)

Vergleichen Sie die Standardabweichung mit dem Standardfehler und grenzen Sie die beiden Kennzahlen voneinander ab.

### 19 Aufgabe

(8 Punkte)

Ihnen liegt folgendes Varianzanalysemodell mit der üblichen Beschreibung zur Auswertung des Merkmals fett- und eiweißkorrigierte Milchleistung pro Kuh und Jahr in kg vor:

$$Y_{ijkl} = \mu + Var_i + EKA_j + VarEKA_{ij} + V_k + b(L_{ij} - L) + e_{ijkl}$$

mit

- $Y_{ijkl}$ : l-te Beobachtung
- $\mu$ : Populationsmittel
- $Var_i$ : fixer Effekt der i-ten Variante (i: Kontrolle, Versuchsgruppe 1, Versuchsgruppe 2)
- $EKA_j$ : fixer Effekt der j-ten Erstkalbealtergruppe (j:  $EKA \leq 25$  Monate,  $EKA > 25$  Monate)
- $VarEKA_{ij}$ : fixer Effekt der Interaktion Variante x Erstkalbealtergruppe
- $V_k$ : zufälliger Effekt des Vaters
- $b(L_{ij} - L)$ : lineare Kovariable Laktationsnummer
- $e_{ijkl}$ : zufälliger Restfehler

Erläutern Sie anhand dieses Beispiels die Begriffe fixer Effekt, Interaktion, zufälliger Effekt und Kovariable und grenzen Sie diese Begriffe voneinander ab.

### 20 Aufgabe

(6 Punkte)

Wie bestimmen Sie die richtige Stichprobengröße? Welche Kennzahlen / statistische Maßzahlen benötigen Sie dabei und nennen Sie die Voraussetzungen.



