

Übungsaufgaben

Aufgabe 1: (9 Punkte, Praxistransfer)

Geben Sie bitte ein Beispiel für die Ebenen Assoziation, Intervention und Counterfactual auf der kausalen Leiter nach Pearl in Ihrer beruflichen Praxis (*bitte angeben*). Definieren Sie dabei X , Y .

Aufgabe 2: (1 Punkt)

Die Besitzerin einer Eisdiele beobachtet an einem Brückentag (Freitag, 4.10.2019) bei Sonnenschein einen hohen Umsatz. Sie schließt daraus: *Wenn Brückentag ist und die Sonne scheint, kaufen viele Menschen Eis*. Welche Schlussart liegt vor? (max. 1 Antwort ist richtig)

- A. Induktion
- B. Deduktion
- C. Abduktion

Aufgabe 3: (2 Punkte)

Welche Form der Datenerhebung ermöglicht Kausalaussagen (max. 1 Antwort richtig)?

1. Beobachtungsstudie
2. Randomisiertes Experiment
3. Zufällige Stichprobe
4. Eine große Stichprobe

Aufgabe 4: (3 Punkte)

Was verstehen Sie unter der internen Validität eines Forschungsergebnisses?

Aufgabe 5: (2 Punkte)

Bei der Messung einer Eigenschaft A stellt man fest, dass man statt dieser eigentlich die Eigenschaft B misst. Welches Gütekriterium einer Messung ist verletzt? (max. 1 Antwort ist richtig)

- A. Genauigkeit
- B. Objektivität
- C. Reliabilität
- D. Validität

Aufgabe 6: (2 Punkte)

Wie lautet die formal richtige Nullhypothese, wenn die Forschungsthese lautet: *Im Mittelwert essen die Deutschen mehr als 20 Kugeln Eis im Jahr?* (max. 1 Antwort ist richtig)

- A. $\bar{x} = 20$
- B. $\bar{x} \geq 20$
- C. $\bar{x} \leq 20$
- D. $\mu = 20$
- E. $\mu \geq 20$
- F. $\mu \leq 20$

Aufgabe 7: (1 Punkt)

Die Eisdienbesitzerin befragt ihre Kund*innen an dem sonnigen Brückentag (Fr., 4.10.2019), wie viele Kugeln Eis sie im Jahr essen. Um was für eine Stichprobe handelt es sich? (max. 1 Antwort ist richtig)

- A. Eine zufällige Stichprobe.
- B. Eine geschichtete Stichprobe.
- C. Eine Gelegenheitsstichprobe.

Aufgabe 8: (2 Punkte)

Eine zufällige Stichprobe für die Hypothese aus Aufgabe 6 ergibt einen Mittelwert von 25 Kugeln.

Welche Aussage stimmt? (max. 1 Antwort ist richtig)

- A. H_0 wird auf keinen Fall verworfen.
- B. H_0 wird auf jeden Fall verworfen.
- C. Ob H_0 verworfen wird, kann noch nicht entschieden werden.

Aufgabe 9: (2 Punkte)

Stimmt die Aussage: Der Korrelationskoeffizient $r_{x,y}$ ist unabhängig von der Skalierung von x und y ? (max. 1 Antwort ist richtig)

Zur Verdeutlichung ein Beispiel: Wenn die Eisdienbesitzerin anstatt des Korrelationskoeffizienten zwischen der Anzahl Eiskugeln und der Temperatur in °C, den Korrelationskoeffizienten zwischen dem Umsatz (= Anzahl Kugeln $\times 1.10$ Euro) und der Temperatur analysieren würde, so ändert sich das Ergebnis nicht.

$$r_{\text{Eiskugeln, Temperatur}} = r_{\text{Eisumsatz, Temperatur}}$$

- A. Ja.
- B. Nein.

Aufgabe 10: (1 Punkt)

Ändert sich der Korrelationskoeffizient (*ceteris paribus*) substantiell, wenn statt $n = 100$ zufälligen Beobachtungen $n = 1000$ zufällige Beobachtungen zur Berechnung verwendet werden? (max. 1 Antwort ist richtig)

- A. Ja.
- B. Nein.

Aufgabe 11: (4 Punkte)

In einem Skatblatt gibt es 32 Karten: die vier Farben Karo, Herz, Pik und Kreuz mit jeweils den acht Werten 7, 8, 9, 10, Bube, Dame, König, Ass.

Unabhängig von Daten entscheidet sich eine Forscherin, die Nullhypothese zu verwerfen, wenn sie aus einem gut gemischten Blatt eine Karte zieht und dies der Herz-Bube ist.

Was gilt? (max. 2 Antworten sind richtig)

- A. Die Wahrscheinlichkeit für einen Fehler 1. Art ist > 0.05 .
- B. Die Wahrscheinlichkeit für einen Fehler 1. Art ist < 0.05 .
- C. Die Wahrscheinlichkeit für einen Fehler 2. Art ist > 0.05 .
- D. Die Wahrscheinlichkeit für einen Fehler 2. Art ist < 0.05 .

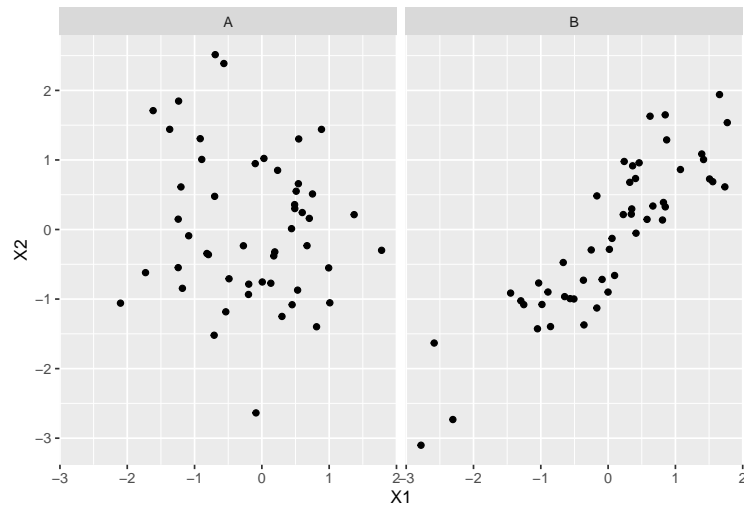
Aufgabe 12: (2 Punkte)

Angenommen bei einer Hypothesenprüfung ergibt sich ein p-Wert von 0.254.

Welche Aussage stimmt? (max. 1 Antwort ist richtig)

- A. Die Wahrscheinlichkeit, dass H_0 stimmt, liegt bei 0.254.
- B. Die Wahrscheinlichkeit, dass H_A stimmt, liegt bei 0.254.
- C. Weder A noch B stimmen.

Aufgabe 13: (1 Punkt)



In welcher Abbildung ist der Korrelationskoeffizient $r = -0.13$? (max. 1 Antwort ist richtig)

- A. In Abbildung A.
- B. In Abbildung B

Aufgabe 14: (2 Punkte)

Mit welcher Methode können Sie den Standardfehler bestimmen?

- A. Bootstrapping (`resample()`).
- B. Permutation (`shuffle()`).

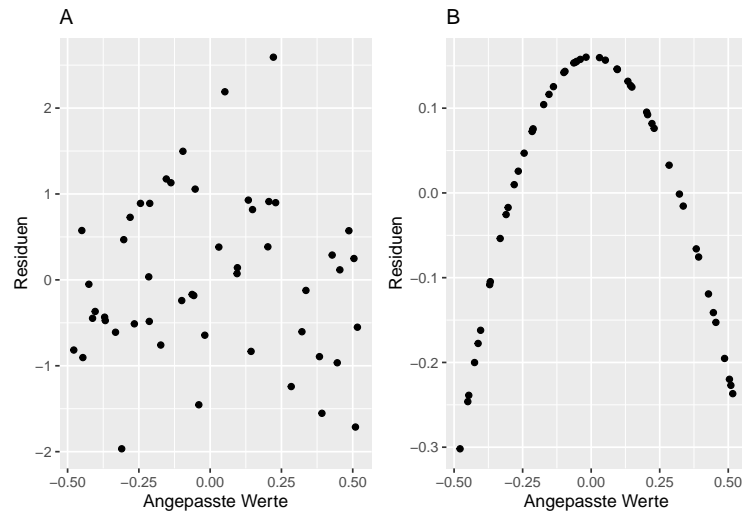
Aufgabe 15: (1 Punkt)

Stimmt die Aussage: Signifikanz bedeutet inhaltliche Relevanz?

- A. Ja.
- B. Nein.

Aufgabe 16: (2 Punkte)

Zwei lineare Regressionsmodelle ergeben folgende Residualplots:



In welchem Fall ist eine Lineare Regression angemessen? (max. 1 Antwort ist richtig)

- A. Nur bei A.
- B. Nur bei B.
- C. Sowohl bei A als auch B.
- D. Weder bei A noch bei B.

Aufgabe 17: (1 Punkt)

Stimmt die Aussage: Das Bestimmtheitsmaß R^2 ist robust gegenüber Ausreißern? (max. 1 Antwort ist richtig)

- A. Ja.
- B. Nein.

Aufgabe 18: (1 Punkt)

Stimmt die Aussage: Das Bestimmtheitsmaß R^2 steigt i. d. R. auch dann, wenn eine Variable x ins Modell aufgenommen wird, die keinen Zusammenhang mit der abhängigen Variable y hat? (max. 1 Antwort ist richtig)

- A. Ja.
- B. Nein.

Aufgabe 19: (4 Punkte)

Eine Forscherin geht von folgenden Zusammenhängen zwischen X , C und Y aus:

- $X \sim C$
- $Y \sim X + C$

Zeichnen Sie den dazugehörigen Graphen (DAG, directed acyclic graph).

Aufgabe 20: (2 Punkte)

Für das Modell in Aufgabe 19 wird angenommen, dass lineare Zusammenhänge gelten. Welches Modell sollte dann gewählt werden, um den kausalen Effekt von X auf Y zu bestimmen? (max. 1 Antwort ist richtig)

- A. $\text{lm}(y \sim 1)$
- B. $\text{lm}(y \sim x)$
- C. $\text{lm}(y \sim c)$
- D. $\text{lm}(y \sim x + c)$

Fallstudie

Es liegt eine Stichprobe einer Erhebung der “Students Attitudes Towards Statistics” vor.

Variable	Beschreibung
Leichtigkeit	Eingeschätzte Leichtigkeit des Faches.
Kompetenz	Eingeschätzte Selbstkompetenz in dem Fach.
Interesse	Interesse am Fach.
Geschlecht	Geschlecht (maennlich, weiblich).
Mathe	Eingeschätztes Leistungsniveau Mathematik (gering, mittel, hoch)
Aufwand	Geplanter Aufwand für das Fach.

Datenstruktur:

```
str(SATS)
```

```
## 'data.frame':    182 obs. of  6 variables:
## $ Leichtigkeit: num  3.6 2.6 3.1 3.4 3.3 4.9 2.7 2.6 2.3 3.7 ...
## $ Kompetenz   : num  3.7 4.7 4.2 3.2 6 6.2 4.3 4.5 3.3 4.2 ...
## $ Interesse    : num  3.2 5.2 6 4 7 6.8 4.8 3.2 3 6.2 ...
## $ Geschlecht   : Factor w/ 2 levels "maennlich","weiblich": 2 2 1 2 2 2 2 1 2 1 ...
## $ Mathe        : Factor w/ 3 levels "gering","mittel",...: 2 3 2 2 3 2 2 2 3 2 ...
## $ Aufwand      : num  6 7 7 7 7 6.5 5 4.2 4.2 6.5 ...
```

Datenbeispiele:

Leichtigkeit	Kompetenz	Interesse	Geschlecht	Mathe	Aufwand
3.6	3.7	3.2	weiblich	mittel	6.0
2.6	4.7	5.2	weiblich	hoch	7.0
3.1	4.2	6.0	maennlich	mittel	7.0
3.4	3.2	4.0	weiblich	mittel	7.0
3.3	6.0	7.0	weiblich	hoch	7.0
4.9	6.2	6.8	weiblich	mittel	6.5

Aufgabe 21: (2 Punkte)

Wie viele Beobachtungen liegen im Datensatz **SATS** vor?

Aufgabe 22: (1 Punkt)

Stimmt die Aussage: Die Daten wurden im Rahmen einer Beobachtungsstudie erhoben? (max. 1 Antwort ist richtig)

- A. Ja
- B. Nein

Aufgabe 23: (1 Punkt)

Mit welcher Abbildung kann die Variable **Mathe** sinnvoll visualisiert werden? (max. 1 Antwort ist richtig)

- A. Säulendiagramm
- B. Histogramm
- C. Boxplot

Aufgabe 24: (1 Punkt)

Stimmt die Aussage: Die Variable **Leichtigkeit** ist diskret? (max. 1 Antwort ist richtig)

- A. Ja
- B. Nein

Aufgabe 25: (1 Punkt)

Für die Analyse von **my.x=Kompetenz** ergeben sich folgende Kennzahlen:

```
favstats( ~ my.x, data = SATS)
```

```
##  min Q1 median  Q3 max    mean      sd  n missing
##  1.2  4    4.6 5.3  7 4.537363 1.189563 182      0
```

Stimmt die Aussage: 75% der Beobachtungen sind größer oder gleich 4.6?

- A. Ja
- B. Nein

Aufgabe 26: (2 Punkte)

Mit welcher Methode können die Verteilungen von Leichtigkeit in Abhängigkeit von Mathe untersucht werden? (max. 1 Antwort ist richtig)

- A. Vergleich eines Anteilswertes
- B. Vergleich eines Mittelwertes
- C. Vergleich zweier Anteilswerte
- D. Vergleich zweier Mittelwerte
- E. Vergleich von mehr als zwei Anteilswerten
- F. Vergleich von mehr als zwei Mittelwerten

Aufgabe 27: (2 Punkte)

Angenommen für eine Hypothesenprüfung von Aufgabe 26 ergibt sich ein p-Wert von 0.0015. Welche Aussage stimmt? (max. 1 Antwort ist richtig)

- A. Der beobachtete Zusammenhang ist unter der Annahme einer Unabhängigkeit unwahrscheinlich.
- B. Der beobachtete Zusammenhang ist unter der Annahme einer Abhängigkeit unwahrscheinlich.
- C. Der beobachtete Zusammenhang ist unter der Annahme einer Unabhängigkeit wahrscheinlich.
- D. Der beobachtete Zusammenhang ist unter der Annahme einer Abhängigkeit wahrscheinlich.

Aufgabe 28: (2 Punkte)

Angenommen eine Hypothesenprüfung von Aufgabe 26 erfolgt zum Niveau $\alpha = 5\%$ mit dem p-Wert von Aufgabe 27. Welche Aussage stimmt? (max. 1 Antwort ist richtig)

- A. Die Nullhypothese wird bestätigt.
- B. Die Nullhypothese wird verworfen.
- C. Weder A noch B stimmen.

Aufgabe 29: (1 Punkt)

Wird das Ergebnis gemäß den obigen Aufgaben (26 – 28) *signifikant* genannt? (max. 1 Antwort ist richtig)

- A. Ja.
- B. Nein.

Aufgabe 30: (2 Punkte)

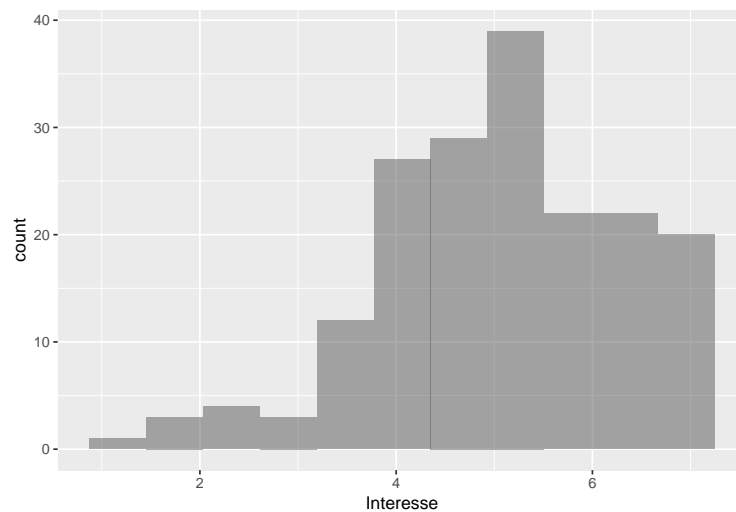
Wie würde sich (ceteris paribus) der p-Wert verändern (alter p-Wert p , neuer p-Wert: \tilde{p}), wenn die Nullhypothese nicht gelten würde, Ihnen aber $\tilde{n} = 728$ Beobachtungen zur Verfügung stehen würden? (max. 1 Antwort ist richtig)

- A. $\tilde{p} \approx p$.
- B. $\tilde{p} < p$.
- C. $\tilde{p} > p$.

Aufgabe 31: (4 Punkte)

Für die Variablen `my.x=Interesse` ergibt sich folgendes Histogramm:

```
gf_histogram( ~ my.x, data = SATS)
```



Beschreiben Sie das Ergebnis.

Aufgabe 32: (4 Punkte)

Für die Variablen `my.x1=Geschlecht` und `my.x2=Mathe` ergibt sich folgender Mosaikplot:

```
mosaicplot(my.x1 ~ my.x2, data = SATS)
```

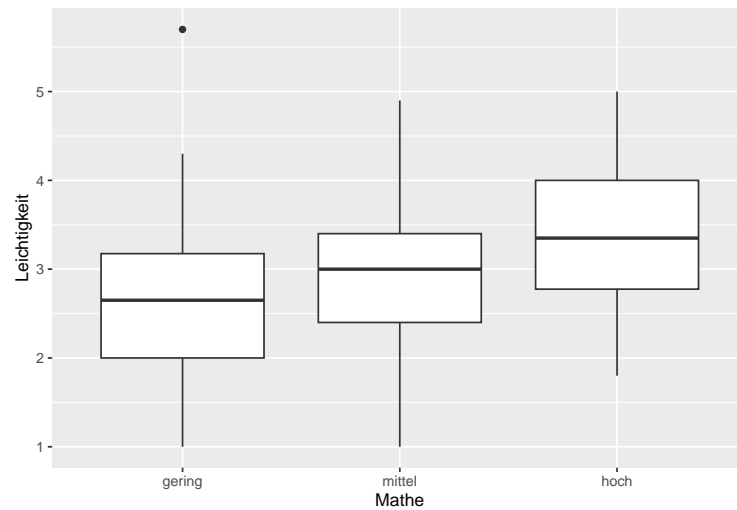


Beschreiben Sie das Ergebnis.

Aufgabe 33: (4 Punkte)

Ein Boxplot mit den Variablen `my.x=Mathe` und `my.y=Leichtigkeit` ergibt folgendes Ergebnis:

```
gf_boxplot(my.y ~ my.x, data = SATS)
```



Beschreiben Sie das Ergebnis.

Aufgabe 34: (4 Punkte)

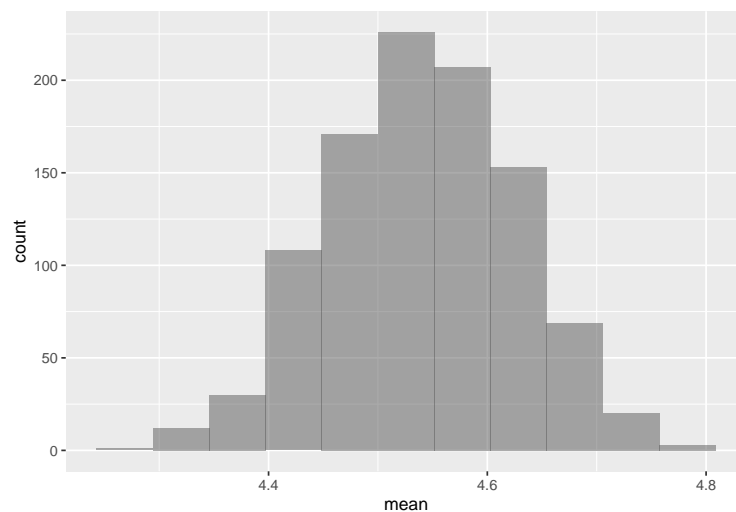
Für die Variable `my.x`=Kompetenz ergibt sich folgende Verteilung in der Stichprobe:

Der Mittelwert in der Stichprobe liegt bei

$$\bar{x} = 4.54.$$

Eine Resamplinganalyse ergibt folgendes Ergebnis:

```
# Bootstrapping
Bootvtlg <- do(1000)* mean(~my.x, data = resample(SATS))
# Bootstrap-Verteilung
gf_histogram( ~ mean, data = Bootvtlg, bins = 11)
# Resampling-Intervall
confint(Bootvtlg)
```



```
##   name   lower  upper level   method estimate
## 1 mean 4.374121 4.70114 0.95 percentile 4.537363
```

Interpretieren Sie das Ergebnis.

Ggf. Fortsetzung Aufgabe 34

Aufgabe 35: (6 Punkte)

Bei der Verteilung der Variable `my.y`=Kompetenz in Abhängigkeit der Variable `my.x`=Geschlecht ergibt sich für die arithmetischen Mittelwerte:

```
mean(my.y ~ my.x, data = SATS)
```

```
## maennlich weiblich  
## 4.77625 4.35000
```

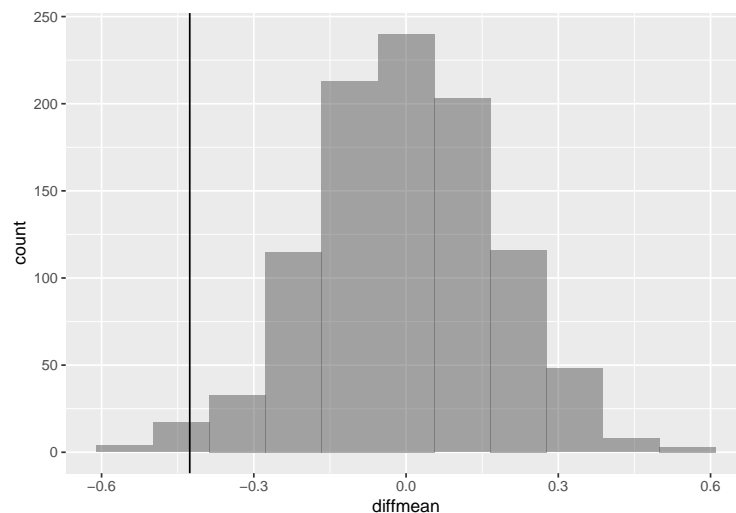
Und damit für die Differenz der Mittelwerte:

```
diff.stipro <- diffmean(my.y ~ my.x, data = SATS)  
diff.stipro
```

```
## diffmean  
## -0.42625
```

Ein Permutationstest ergibt folgendes Ergebnis:

```
# Simulation unter H_0  
Nullvtlg <- do(1000)* diffmean(my.y ~ shuffle(my.x), data = SATS)  
# Permutationsverteilung  
gf_histogram(~diffmean, data = Nullvtlg) %>%  
  gf_vline(xintercept = ~ diff.stipro) # Wert der Stichprobe  
# p-Wert  
prop(~ abs(diffmean)>=abs(diff.stipro), data = Nullvtlg)
```



```
## prop_TRUE  
## 0.024
```

Fassen Sie das Ergebnis der Analyse zusammen. Wie lauten die Forschungsfrage, die Hypothesen und das vorläufige Ergebnis?

Ggf. Fortsetzung Aufgabe 35

Aufgabe 36: (8 Punkte)

Eine Regression von `my.y`=Aufwand auf die Variablen `my.x1`=Interesse und `my.x2`=Geschlecht ergibt folgendes Ergebnis:

```
erglm <- lm(my.y ~ my.x1 + my.x2, data = SATS)
# Zusammenfassung
summary(erglm)

##
## Call:
## lm(formula = get(my.y) ~ get(my.x1) + get(my.x2), data = SATS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5752 -0.2865  0.1635  0.4747  1.3136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.52518    0.25062  18.056 < 2e-16 ***
## (my.x1)         0.25003    0.04671   5.353 2.63e-07 ***
## (my.x2)weiblich 0.51121    0.11592   4.410 1.78e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7761 on 179 degrees of freedom
## Multiple R-squared:  0.2142, Adjusted R-squared:  0.2054
## F-statistic: 24.4 on 2 and 179 DF, p-value: 4.27e-10
```

Fassen Sie das Ergebnis der Analyse zusammen. Wie lauten die Forschungsfrage, die Hypothesen und das vorläufige Ergebnis?

Ggf. Fortsetzung Aufgabe 36

Hinweise

- Die Art und Weise der Aufgaben ist ähnlich denen einer Klausur. Sie können sich vom Anspruch und dem Fragentyp an diesen Übungsaufgaben orientieren, nicht aber was die konkreten Inhalte usw. betrifft.
- Zum Üben besteht diese Probe aus Aufgaben mit insgesamt **90** Punkten.
- Es reicht nicht diese Übungsaufgaben zu lernen: Die Klausur hat andere Aufgaben.
- Tipps zur Klausur:
 - Prüfungsrelevant ist der gesamte Stoff der Vorlesung. Für die Wiederholungsprüfung gelten die gleichen Rahmenbedingungen wie für den regulären Prüfungstermin.
 - Lesen Sie sich erst die ganze Klausur in Ruhe durch und fangen Sie mit den Aufgaben an, die Sie sicher können.
 - Halten Sie sich nicht zu lange mit Aufgaben auf, die wenig Punkte bringen.