

Übungsaufgaben

zur Klausurvorbereitung (QM-Eval)

Aufgabe 1: (2 Punkte)

1 Pkt $\hat{=}$ 1 Min. Bearbeit.-zeit

Alle Schwäne die ich bisher gesehen habe waren weiß. Ich schließe daraus: alle Schwäne sind weiß.

Um welche Schlussart handelt es sich (max. 1 Antwort richtig)?

- ☒ 1. Induktion
- ☐ 2. Deduktion
- ☐ 3. Abduktion

Aufgabe 2: (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 3: (2 Punkte)

Welches Skalenniveau hat das Merkmal Postleitzahl (max. 1 Antwort richtig)?

- ☐ 1. Numerisch diskret
- ☐ 2. Numerisch stetig
- ☐ 3. Kategorial ordinal
- ☒ 4. Kategorial nominal

↑ Kochen! ✓

Wenn es numerisch wäre, dann müsste
45130+44386 einen SINN haben!

Die reine lexikographische Ordnung der
Zahlen ist kein Ordinalitätskriterium!

Aufgabe 4: (5 Punkte) (Praxistransferaufgabe!)

Benennen Sie die fünf Punkte des quantitativen Forschungsprozesses (PPDAC).

P (Problem)

Problemdefinition, Forschungsfrage: Was soll untersucht werden?

P (Plan)

Planung der Analyse, z. B.

Operationalisierung / Variablenauswahl. Wahl des Stichprobenverfahren und / oder Versuchsplanung.

D (Data)

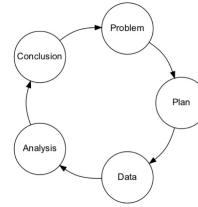
Datenerhebung, Datenmanagement, Datenbereinigung.

A (Analysis)

Explorative Datenanalyse, Modellierung, Hypothesenprüfung.

C (Conclusion)

Schlussfolgerungen, Interpretation, neue Ideen, Kommunikation.



Aufgabe 5: (1+3=4 Punkte)

Es liegen folgende Daten einer zufälligen Bevölkerungssstichprobe aus dem Jahre 1985 vor (USA).

```
inspect(einkommen)
```

```
##
## categorical variables:
##      name  class levels   n missing
## 1  Geschlecht factor      2  534      0
## 2  Gewerkschaft factor      2  534      0
##                                     distribution
## 1  männlich (54.1%), weiblich (45.9%)
## 2  kein Mitglied (82%) ...
##
## quantitative variables:
##      name  class min   Q1 median   Q3  max    mean      sd
## ...1  Stundenlohn numeric   1  5.25  7.78 11.25 44.5  9.024064  5.139097
## ...2  Ausbildungsjahre integer  2 12.00 12.00 15.00 18.0 13.018727  2.615373
## ...3  Alter integer  18 28.00 35.00 44.00 64.0 36.833333 11.726573
##      n missing
## ...1  534      0
## ...2  534      0
## ...3  534      0
```

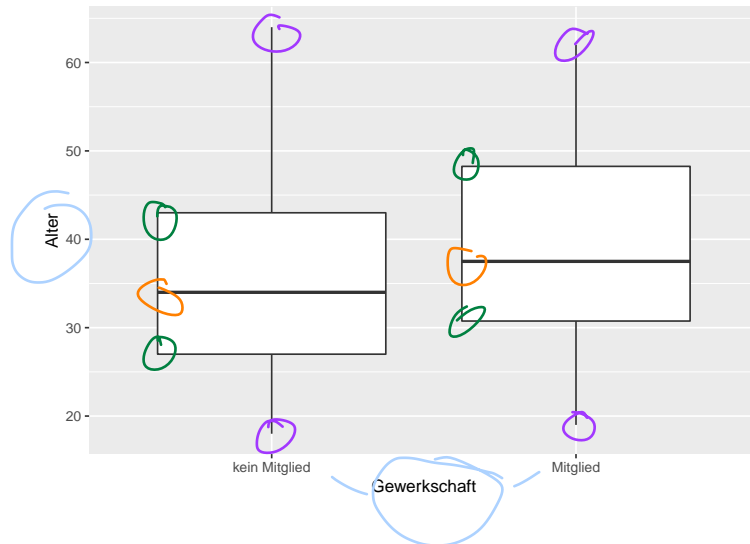
1. Wie groß ist der Stichprobenumfang n ? 534
2. Welche numerischen Merkmale liegen vor? Sind diese Verhältnis- oder Intervallskaliert?

Stundenlohn, Ausbildungsjahre und Alter sind verhältnisskaliert!

Aufgabe 6: (5 Punkte)

Was können Sie der folgenden Abbildung entnehmen?

```
gf_boxplot(Alter ~ Gewerkschaft, data = einkommen)
```



① Alter / Gewerkschaft

② Min / Max der beiden Gruppen

③ $Q_1 - Q_3$ als IQR der beiden Gruppen

④ Median

⑤ Es gibt keine Ausreißer

} jeweils im Vergleich!

Aufgabe 7: (2 Punkte)

Mit welchen Diagrammtyp könnte die Verteilung des Merkmals **Stundenlohn** sinnvoll visualisiert werden (max. 2 Antworten richtig)?

- 1. Histogramm ✓
- 2. Boxplot ✓
- 3. Liniendiagramm (zwei Merkmale, 2. „Zeit“)
- 4. Streudiagramm/ Scatterplot (zwei Merkmale)
- 5. Säulendiagramm (diskret / kategorisch)

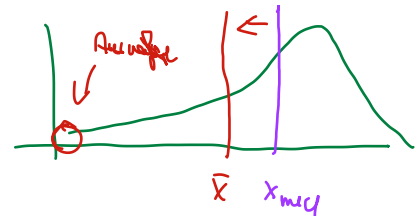
das \Rightarrow EIN Merkmal!

\hookrightarrow numerisch \Rightarrow stetig

Aufgabe 8: (2 Punkte)

Was gilt in der Regel für linksschiefe Verteilungen (max. 1 Antwort richtig)

- 1. Der Mittelwert ist gleich dem Median
- 2. Der Mittelwert ist kleiner als der Median
- 3. Der Mittelwert ist größer als der Median
- 4. keine Aussage möglich



Aufgabe 9: (4 Punkte)

Ausgehend von der "Gleichung": $\text{Daten} = \text{Modell} + \text{Rest}$: Was verstehen Sie unter Modellierung?

Unter Modellierung verstehen wir die Erstellung eines Modells, das, bis auf einen Rest/Fehler die Daten, die gegeben sind, erklärt und/oder neue Daten vorhersagt.

Aufgabe 10: (3 Punkte)

Benennen Sie drei Kommunikationsformen bei Fragebögen

Kommunikationsformen

- ▶ **standardisiert:** nicht
gestalbarer, starrer
Ablauf
- ▶ **teilstandardisiert:**
teilweise gestaltbarer
Ablauf
- ▶ **nicht standardisiert:**
gestaltbarer, flexibler
Ablauf

Aufgabe 11: (2+4=6 Punkte)

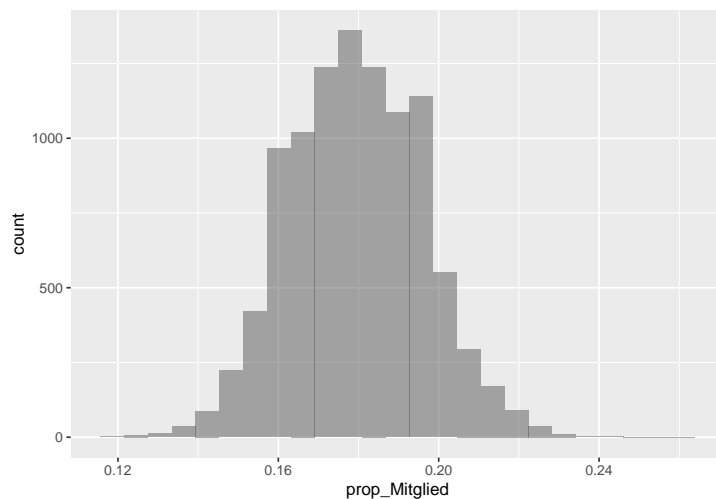
Der Anteil der Personen in der Stichprobe, die Gewerkschaftsmitglied sind,

```
prop( ~ Gewerkschaft, success="Mitglied", data = einkommen)
```

```
## prop_Mitglied
##      0.1797753
```

liegt bei $\approx 18\%$.

```
gewerkschaftsim <- do(10000) * prop( ~ Gewerkschaft, success="Mitglied",
                                     data = resample(einkommen))
gf_histogram( ~ prop_Mitglied, data = gewerkschaftsim)
```



```
qdata( ~ prop_Mitglied, data = gewerkschaftsim, p = c(0.025, 0.975))
```

```
## [ 2.5% 97.5% ] ← 20% = 0,20 sind enthalten ⇒ H₀ kann nicht
## [ 0.1479401 0.2116105 ] ← abgelehnt werden!
                                ↗ π
```

1. Formulieren Sie die formalen Hypothesen, für die **Forschungsthese**, dass der **Anteil** der **Gewerkschaftsmitglieder** in der Population (Bevölkerung) **nicht 20%** beträgt.
2. Können Sie anhand des Bootstrap Konfidenzintervalls den Hypothesentest entscheiden? ~~Warum?~~ Und wie lautet ggfs. die Testentscheidung?

1. $H_0: \pi_{\text{Gew.}} = 20\% \quad \text{vs.} \quad H_A: \pi_{\text{Gew.}} \neq 20\%$

Teilt in Gruppen auf

7

Mittelwert $\Rightarrow \mu$

2) Ja. Da 20% im 95%-KI liegen, kann die H_0 nicht abgelehnt werden!

Aufgabe 12: (3 Punkte)

Welche der folgenden Aussagen über den p-Wert stimmt (max. 1 Antwort richtig)?

- ~~1. Der p-Wert weist auf die ökonomische Relevanz der Ergebnisse hin.~~
- ~~2. Der p-Wert weist auf die wissenschaftliche Relevanz der Ergebnisse hin.~~
- ~~3. Der p-Wert gibt die Wahrscheinlichkeit an, dass die Nullhypothese stimmt.~~
- ~~4. Der p-Wert gibt die Wahrscheinlichkeit an, dass die Alternativhypothese stimmt.~~
- ~~5. 1.) und 3.) stimmen.~~
- ~~6. 2.) und 4.) stimmen.~~
- ☒ 7. Keine der genannten Antworten stimmt.

Aufgabe 13: (2 Punkte)

Ein Test für $H_0 : \mu \leq 100$ gegen $H_A : \mu > 100$ ergibt die Teststatistik $\bar{x} = 95$. Was können Sie über das Testergebnis aussagen? Begründen Sie kurz.

Die H_0 wird nicht verworfen, weil $\bar{x} \leq \mu \leq 100$ gilt! ✓

Aufgabe 14: (1 Punkte)

Welches der genannten ist das richtige Testverfahren um einen Zusammenhang zwischen Gewerkschaft und Alter zu analysieren (max. 1 Antwort richtig)?

~~1. Vergleich zweier Anteilswerte.~~

2. Vergleich zweier Mittelwerte.

num. \rightarrow 2-seitig \rightarrow Mittelwert

2. 2 Faktoren / Mittelwert (2in Mitglied)

Aufgabe 15: (2 Punkte)

Was reduziert die Wahrscheinlichkeit einen Fehler 2. Art zu begehen (max. 2 Antworten richtig)?

~~1. Ein geringerer Stichprobenumfang.~~

~~2. Ein geringeres Signifikanzniveau.~~

☒ 3. Ein höherer Stichprobenumfang.

☒ 4. Ein höheres Signifikanzniveau.

Aufgabe 16: (6 Punkte)

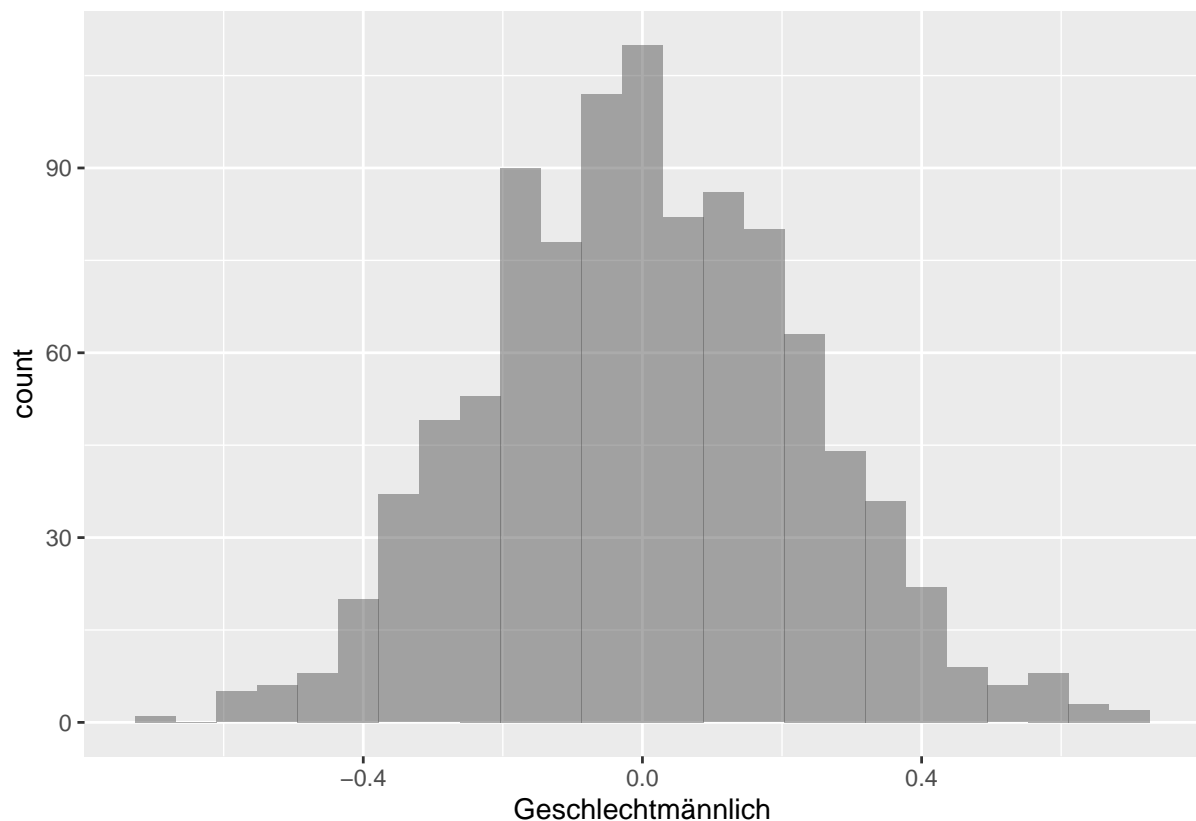
Ein lineares Modell der Ausbildungsjahre je Geschlecht ergibt folgendes Ergebnis:

```
linmod <- lm(Ausbildungsjahre ~ Geschlecht, data = einkommen)
coef(linmod)
```

```
##      (Intercept) Geschlechtmännlich
##      13.02448980 -0.01064897
```

← gemessene Wert ∇

```
vertlg <- do(1000) * lm(Ausbildungsjahre ~ sample(Geschlecht), data = einkommen)
gf_histogram( ~ Geschlechtmännlich, data = vertlg)
```



```
qdata( ~ Geschlechtmännlich, p = c(0.025, 0.975), data = vertlg)
```

```
##      2.5%      97.5%
## (-0.4254502 0.4795707)
```

Schwankung unter H_0

Fassen Sie kurz Forschungsfrage, Hypothese und Ergebnis zusammen.

gemessene Wert liegt in diesem Intervall

=> kein unüblicher Wert => H_0 kann nicht abgelehnt werden ∇

Forschungsfrage:

Gibt es einen Unterschied der Ausbildungsdauer zwischen den Geschlechtern?

Hypothesen:

$$H_0: \mu_{\sigma} = \mu_{\varphi} \quad \text{vs.} \quad H_A: \mu_{\sigma} \neq \mu_{\varphi}$$

Ergebnis:

Da der gemessene Wert (-0.01...) im Bereich der üblichen Werte unter der H_0 liegt (-0.43 bis 0.48) kann die H_0 nicht abgelehnt werden!

Aufgabe 17: (9 Punkte)

Eine lineare Regression des Stundenlohns auf die anderen Variablen ergibt folgendes Ergebnis:

```
linmod <- lm(Stundenlohn ~ Ausbildungsjahre + Geschlecht + Alter + Gewerkschaft,  
             data = einkommen)
```

```
coef(linmod)
```

```
##           (Intercept)      Ausbildungsjahre  Geschlechtmännlich  
##           -7.115950           0.8282483           2.1455241  
##           Alter GewerkschaftMitglied  
##           0.1067311           1.4713063
```

```
vertlg <- do(1000) *  
  lm(Stundenlohn ~ Ausbildungsjahre + Geschlecht + Alter + Gewerkschaft,  
     data = resample(einkommen))
```

```
confint(vertlg) %>%  
  filter(!(name %in% c("sigma", "F")))
```

| ## | name | lower | upper | level | method | estimate |
|------|----------------------|-------------|------------|-------|------------|------------------|
| ## 1 | Intercept | -9.94939316 | -4.6020139 | 0.95 | percentile | -7.115950 |
| ## 2 | Ausbildungsjahre | 0.66060008 | 0.9948884 | 0.95 | percentile | 0.8282483 |
| ## 3 | Geschlechtmännlich | 1.36933520 | 2.9222921 | 0.95 | percentile | 2.1455241 |
| ## 4 | Alter | 0.07180131 | 0.1420811 | 0.95 | percentile | 0.1067311 |
| ## 5 | GewerkschaftMitglied | 0.44570388 | 2.5871915 | 0.95 | percentile | 1.4713063 |
| ## 6 | r.squared | 0.19016837 | 0.3627599 | 0.95 | percentile | <u>0.2646481</u> |

Beschreiben und diskutieren Sie das Ergebnis.

① Stundenlohn =

② Übertragbarkeit des Modells in die Population ohne weiter möglich?

| |
|-------------|
| (Intercept) |
| -7.115950 |

| |
|------------------|
| Ausbildungsjahre |
| 0.8282483 |

| |
|--------------------|
| Geschlechtmännlich |
| 2.1455241 |

| |
|-----------|
| Alter |
| 0.1067311 |

| |
|----------------------|
| GewerkschaftMitglied |
| 1.4713063 |

① $\text{Stundenlohn}_i = -7,116 + 0,828 \cdot \text{Ausbildungsjahre} + 0,107 \cdot \text{Alter}$
 $+ 2,146 \cdot \begin{cases} 1 : \text{Geschlecht} = \text{männlich} \\ 0 : \text{sonst} \end{cases}$
 $+ 1,411 \cdot \begin{cases} 1 : \text{Gewerkschaft} = \text{Mitglied} \\ 0 : \text{sonst} \end{cases}$

② Da alle Nullhypothese bzgl. der β_i ~~verworfen~~ werden können, da keine 0 in der 95%-KI liegt, läßt sich das Modell in die Population übertragen

③ R^2 liegt bei 26,46%

Übersicht

Aufgaben: 17

Punkte: 60