

Übungsaufgaben

Aufgabe 1: (6 Punkte) *Praxistransfer*

Geben Sie exemplarisch ein Beispiel aus Ihrer Berufspraxis¹ (angeben!) für die Analyse eines Anteilswertes (*oder*: eines Mittelwerts, Vergleich zweier Anteilswerte, Vergleich zweier Mittelwerte, lineare Regression) an. Gehen Sie auch kurz auf die Forschungsfrage, die Datenerhebung, die Variable(n) und die Verfahren der Datenanalyse ein.

¹d. h. keine Fallstudie o. ä. aus der Vorlesung

(Fortsetzung Aufgabe 1)

Aufgabe 2: (1 Punkte)

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 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: (2 Punkte)

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

1. Numerisch diskret
2. Numerisch stetig
3. Kategorial ordinal
4. Kategorial nominal

Aufgabe 5: (3 Punkte)

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

Aufgabe 6: (1+3=4 Punkte)

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

```
inspect(einkommen)
```

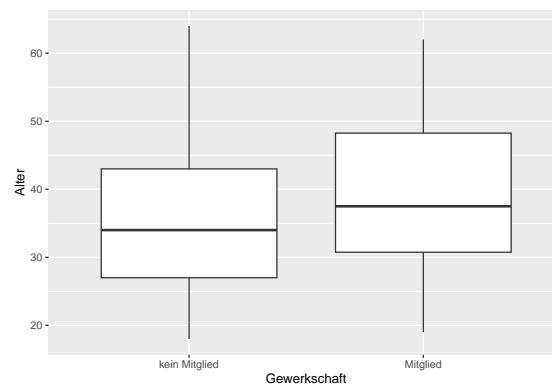
```
##
## categorical variables:
##      name  class levels  n missing
## 1  Geschlecht factor    2  534      0
## 2  Gewerkschaft factor    2  534      0
##                                     distribution
## 1  männlich (54.1%), weiblich (45.9%)
## 2  kein Mitglied (82%) ...
##
## quantitative variables: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

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

Aufgabe 7: (5 Punkte)

Was können Sie der folgenden Abbildung entnehmen?

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



Aufgabe 8: (2 Punkte)

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

1. Histogramm
2. Boxplot
3. Liniendiagramm
4. Streudiagramm/ Scatterplot
5. Balkendiagramm

Aufgabe 9: (2 Punkte)

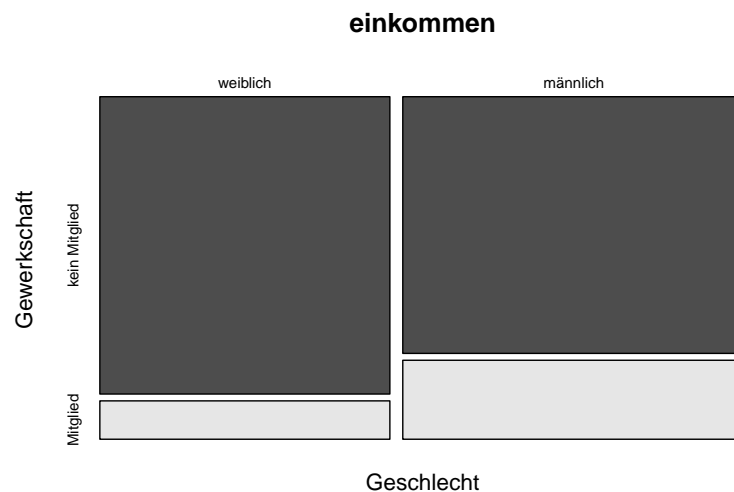
Was gilt in der Regel für linksschiefe Verteilungen (max. 1 Antwort richtig)

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 10: (4 Punkte)

Für die Variablen **Geschlecht** und **Gewerkschaft** ergibt sich folgender Mosaikplot:

```
mosaicplot(Geschlecht ~ Gewerkschaft, data = einkommen)
```



Beschreiben Sie das Ergebnis.

Aufgabe 11: (2+4=6 Punkte)

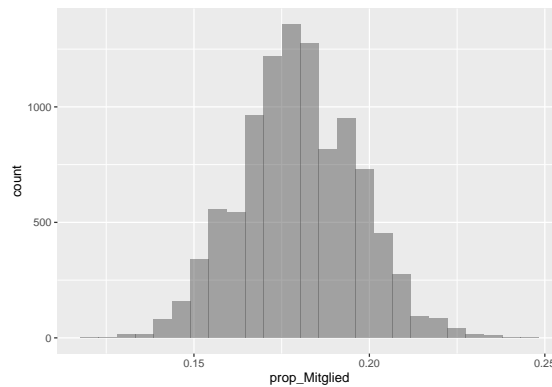
Der Anteil der Personen in der Stichprobe, die Gewerkschaftsmitglied sind,

```
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)
```



```
quantile( ~ prop_Mitglied, data = gewerkschaftsim, probs=c(0.025, 0.975))
```

```
##      2.5%      97.5%  
## 0.1479401 0.2116105
```

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?

(Fortsetzung Aufgabe 11)

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.

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.

Aufgabe 15: (2 Punkte)

Angenommen die Jahre der Arbeitserfahrung sei normalverteilt mit $\mu = 25$ und $\sigma = 10$. Ein Freund erzählt Ihnen, dass er 30 Jahre Arbeitserfahrung hat.

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

- A. 30 Jahre ist ein außergewöhnlicher (d. h. unwahrscheinlicher) Wert.
- B. Der dazugehörige z-Wert ist 2.
- C. Der dazugehörige z-Wert ist -0.5 .
- D. Keine der Aussagen A–C stimmt.

Aufgabe 16: (6 Punkte)

Ein t-Test der Ausbildungsjahre je Geschlecht ergibt folgendes Ergebnis:

```
t.test(Ausbildungsjahre ~ Geschlecht, data = einkommen)

##
##  Welch Two Sample t-test
##
## data:  Ausbildungsjahre by Geschlecht
## t = 0.047345, df = 531.35, p-value = 0.9623
## alternative hypothesis: true difference in means between group weiblich and group männlich
## 95 percent confidence interval:
##  -0.4311944  0.4524924
## sample estimates:
## mean in group weiblich mean in group männlich
##                13.02449                13.01384
```

Fassen Sie kurz Forschungsfrage, Hypothese und Ergebnis zusammen.

(Fortsetzung Aufgabe 16)

Aufgabe 17: (9 Punkte)

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

```
summary(lm(Stundenlohn ~ Ausbildungsjahre + Geschlecht + Alter + Gewerkschaft,
           data=einkommen))
```

```
##
## Call:
## lm(formula = Stundenlohn ~ Ausbildungsjahre + Geschlecht + Alter +
##     Gewerkschaft, data = einkommen)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.493 -2.710 -0.712  1.925 37.779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -7.11560     1.26018   -5.646 2.68e-08 ***
## Ausbildungsjahre    0.82825     0.07411  11.176 < 2e-16 ***
## Geschlechtmännlich  2.14552     0.39097   5.488 6.33e-08 ***
## Alter           0.10673     0.01673   6.381 3.85e-10 ***
## GewerkschaftMitglied 1.47131     0.50936   2.889 0.00403 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.424 on 529 degrees of freedom
## Multiple R-squared:  0.2646, Adjusted R-squared:  0.2591
## F-statistic: 47.6 on 4 and 529 DF, p-value: < 2.2e-16
```

Beschreiben und diskutieren Sie das Ergebnis.

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 **60** 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.