Übungsaufgaben 1

1. **Aufgabe**

   Ob wohl die PS-Zahl (A) und der Spritverbrauch (B) voneinander abhängig sind? Was meinen Sie? Was ist Ihre Einschätzung dazu? Vermutlich haben Sie ein (wenn vielleicht auch implizites) Vorab-Wissen zu dieser Frage. Lassen wir dieses Vorab-Wissen aber einmal außen vor und schauen uns rein Daten dazu an.

   Um es konkret zu machen, nutzen wir den Datensatz `mtcars`:

```
library(tidyverse)
data(mtcars)
glimpse(mtcars)
```

```
## Rows: 32
## Columns: 11
## $ mpg  <dbl> 21.0, 21.0, 22.…
## $ cyl  <dbl> 6, 6, 4, 6, 8, …
## $ disp <dbl> 160.0, 160.0, 1…
## $ hp   <dbl> 110, 110, 93, 1…
## $ drat <dbl> 3.90, 3.90, 3.8…
## $ wt   <dbl> 2.620, 2.875, 2…
## $ qsec <dbl> 16.46, 17.02, 1…
## $ vs   <dbl> 0, 0, 1, 1, 0, …
## $ am   <dbl> 1, 1, 1, 0, 0, …
## $ gear <dbl> 4, 4, 4, 3, 3, …
## $ carb <dbl> 4, 4, 1, 1, 2, …
```

   Weitere Infos zum Datensatz bekommen Sie mit `help(mtcars)` in R.

   Definieren wir uns das Ereignis "hohe PS-Zahl" (und nennen wir es `hp_high`, klingt cooler). Sagen wir, wenn die PS-Zahl größer ist als der Median, dann trifft `hp_high` zu, ansonsten nicht:

```
mtcars %>%
  summarise(median(hp))
```

```
##   median(hp)
## 1        123
```

   Mit dieser "Wenn-Dann-Abfrage" können wir die Variable `hp_high` mit den Stufen `TRUE` und `FALSE` definieren:

```
mtcars <-
  mtcars %>%
  mutate(hp_high = case_when(
    hp > 123 ~ TRUE,
    hp <= 123 ~ FALSE
  ))
```

   Genauso gehen wir mit dem Spritverbrauch vor (`mpg_high`):

```
mtcars <-
  mtcars %>%
  mutate(mpg_high = case_when(
    mpg > median(mpg) ~ TRUE,
    mpg <= median(mpg) ~ FALSE
  ))
```

   a. Schauen Sie im Datensatz nach, ob unser Vorgehen (Erstellung von `hp_high` und `mpg_high`) überhaupt funktioniert hat. Probieren geht über Studieren.

b. Visualisieren Sie in geeignter Form den Zusammenhang.

c. Berechnen Sie $Pr\,(\text{mpg\_high}|\text{hp\_high})$ und $Pr\,(\text{mpg\_high}|\neg\text{hp\_high})$!

Lösung

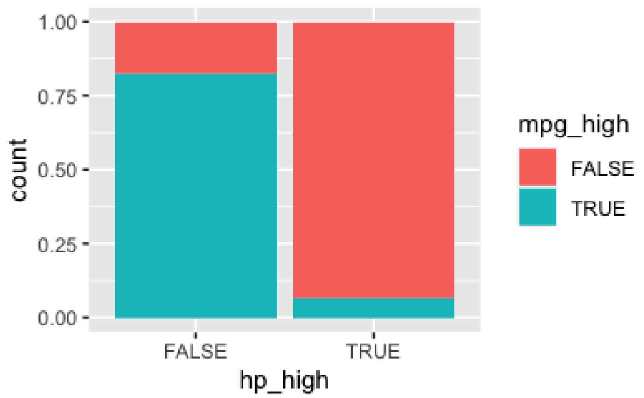
a. Schauen wir mal in den Datensatz:

```
mtcars %>%
  select(hp, hp_high, mpg, mpg_high) %>%
  slice_head(n = 5)
```

```
##                    hp hp_high
## Mazda RX4         110   FALSE
## Mazda RX4 Wag     110   FALSE
## Datsun 710         93   FALSE
## Hornet 4 Drive    110   FALSE
## Hornet Sportabout 175    TRUE
##                    mpg
## Mazda RX4         21.0
## Mazda RX4 Wag     21.0
## Datsun 710        22.8
## Hornet 4 Drive    21.4
## Hornet Sportabout 18.7
##                   mpg_high
## Mazda RX4             TRUE
## Mazda RX4 Wag         TRUE
## Datsun 710            TRUE
## Hornet 4 Drive        TRUE
## Hornet Sportabout    FALSE
```

b.

```
mtcars %>%
  select(hp_high, mpg_high) %>%
  ggplot() +
  aes(x = hp_high, fill = mpg_high) +
  geom_bar(position = "fill")
```



Hey, sowas von abhängig voneinander, die zwei Variablen, `mpg_high` und `hp_high`!

Der linke Balken zeigt $Pr\,(\text{mpg\_high}|\text{hp\_high})$ und $Pr\,(\neg\text{mpg\_high}|\text{hp\_high})$.Der rechte Balken zeigt $Pr\,(\text{mpg\_high}|\neg\text{hp\_high})$ und $Pr\,(\neg\text{mpg\_high}|\neg\text{hp\_high})$.

c. Berechnen wir die relevanten Anteile:

```
mtcars %>%
  select(hp_high, mpg_high) %>%
  count(hp_high, mpg_high) %>%  # Anzahl pro Zelle der Kontingenztabelle
```

```
  group_by(hp_high) %>%  # die Anteile pro "Balken" s. Diagramm
  mutate(prop = n / sum(n))
```

```
## # A tibble: 4 × 4
## # Groups:   hp_high [2]
##   hp_high mpg_high     n
##   <lgl>   <lgl>    <int>
## 1 FALSE   FALSE        3
## 2 FALSE   TRUE        14
## 3 TRUE    FALSE       14
## 4 TRUE    TRUE         1
## # … with 1 more variable:
## #   prop <dbl>
```

Am besten, Sie führen den letzten Code Schritt für Schritt aus und schauen sich jeweils das Ergebnis an, das hilft beim Verstehen.

Alternativ kann man sich die Häufigkeiten auch schön bequem ausgeben lassen:

```
library(mosaic)
tally(mpg_high ~ hp_high,
      data = mtcars,
      format = "proportion")
```

```
##          hp_high
## mpg_high       TRUE
##    TRUE  0.06666667
##    FALSE 0.93333333
##          hp_high
## mpg_high      FALSE
##    TRUE  0.82352941
##    FALSE 0.17647059
```

2. **Aufgabe**

In der klassischen Statistik (Frequentismus) spielt der *p-Wert* eine zentrale Rolle. Der p-Wert ist (oft) das Entscheidungskriterium, um zu entscheiden, ob man eine Aussage (d.h. Hypothese) beibehält oder zurückweist, sozusagen ob man auf “ja, stimmt” wettet oder auf “nein, stimmt nicht”.

Der p-Wert ist etwas unintuitiv und muss daher aufmerksam betrachtet werden.

Ein Beispiel zur Verdeutlichung des p-Werts: Sagen wir, wir möchten wissen, ob eine Münze fair ist, also ob die Hypothese H_0 gilt: $H_0 : Pr(K) = Pr(Z) = 1/2$ gilt. Dazu führen wir folgenden Versuch (einmal aus): Wir werfen die Münze $n = 10$ mal und zählen den Anteil von “Kopf”. Wie gesagt: Wir wissen nicht, ob die Münze fair ist!

Wir bekommen 8 Treffer (von 10 Würfen), also einen Anteil von 80% ($p = 0.8$). Das sind unsere Daten (unsere Stichprobe) bzw. unsere Statistik. Was meinen Sie, geht das mit rechten Dingen zu? Sind 8 von 10 Treffern “erwartbar”, “plausibel”, “wahrscheinlich” bzw. “häufig” wenn man faire Münzen wirft? Hm!

Ok, probieren wir es aus! Wir nehmen jetzt eine *unserer* Münzen. Eine Münze, vor der wir (sicher) wissen, dass sie fair ist, dass also die zu überprüfende Hypothese $H_0 : Pr(K) = Pr(Z) = 1/2$ gilt.

Wir führen also den Versuch sozusagen unter “kontrollierten” Bedingungen mit unserer fairen Münze durch. Ergebnis: 7 Treffer (von 10), nennen wir es den “empirischen Anteil” (`prop_emp`).

Dann denken wir uns, hm, eine einzige Durchführung des Versuchs ist zu sehr vom Zufall abhängig. Besser wir wiederholen den Versuch oft, sagen wir 1000 Mal. Nach 1000 Würfen mit einer fairen Münze werden wir ja sehr genau wissen, ob 8 von 10 Treffern ein häufiges oder seltenes Ereignis ist.

Wir vereinbaren folgende Entscheidungsregel: Wenn unser kontrolliertes Experiment zeigen wird, dass 8 von 10 Treffern ein *seltenes* Ereignis ist, dann glauben wir nicht mehr die Hypothese der fairen Münze. Wenn unser Experiment aber zeigen wird, hey, 8 von 10 Treffern kommt gar nicht so selten vor (ist also ein häufiges Ereignis), dann haben wir keinen Grund, die Hypothese der fairen Münze zu verwerfen, bleiben also bei der Annahme, dass die Münze wohl fair ist – oder, etwas spitzfindiger formuliert, schließen nicht aus, dass die Münze fair ist.

Wenn man Zeit hat, kann man das Experiment mit den 1000 Versuchen ausprobieren … Aber um eine Sehnenscheidenentzündung zu vermeiden, lohnt es sich, diese Aufgabe an einen Golem (bzw. eine Maschine) zu delegieren (meckert nicht). Hey R, komm mal her!

Das R-Paket `mosaic` hält ein paar praktische Funktionen dafür bereit.

```
library(mosaic)
library(tidyverse)
```

So können wir den Versuch mit 10 zufälligen (random) Münzwürfen (coin flips) mit R simulieren:

```
rflip(n = 10)
```

```
##
## Flipping 10 coins [ Prob(Heads) = 0.5 ] ...
##
## T H H T T H H H T H
##
## Number of Heads: 6 [Proportion Heads: 0.6]
```

Oder gleich die Anzahl der Treffer (Kopf) zählen:

```
rflip(n = 10) %>% sum()
```

```
## [1] 4
```

Jetzt wiederholen wir den Versuch 1000 Mal. Hey R, *tue 1000 Mal den 10-fachen Münzwurf*:

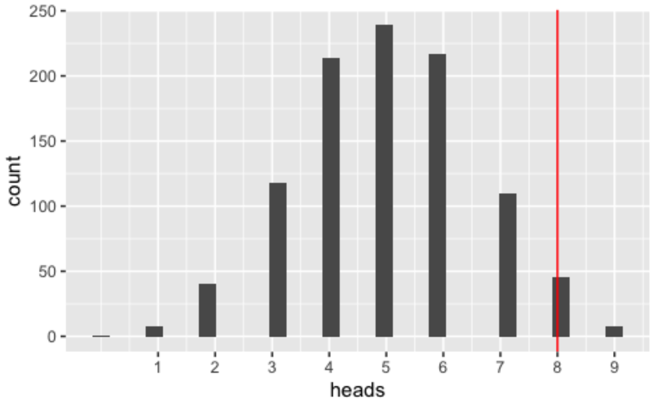
```
muenzversuch <-
  do(1000) * rflip(n = 10)  #*tue 1000 Mal den 10-fachen Münzwurf*

head(muenzversuch) %>%
  gt()
```

n	heads	tails	prop
10	5	5	0.5
10	6	4	0.6
10	6	4	0.6
10	6	4	0.6
10	4	6	0.4
10	7	3	0.7

OK, jetzt visualisieren wir die 1000 Versuche bzw. die Tabelle `muenzversuch` (es ist egal, ob wir die Spalte `prop` verwenden oder `heads`):

```
muenzversuch %>%
  ggplot() +
  aes(x = heads) +
  geom_histogram() +
  scale_x_continuous(breaks = 1:10) +
  geom_vline(xintercept = 8,  # prop_emp ist 8
             color = "red")  # vertikale Linie, um unser emp. Ergebnis anzuzeigen
```



Lange Rede, kurze Fragen:

a. Sind (mind.) 8 von 10 Treffer (unser empirisches Ergebnis) ein häufiges oder ein seltenes Ereignis? Genauer gefragt: Wie häufig kommt dieses Ergebnis in unseren (Ihren) Daten vor?

b. Finden Sie (ja, Sie persönlich!), dass das Ereignis *zu* selten ist, als dass Sie der Hypothese H_0 ("faire Münze") glauben würden? Wo ziehen *Sie* Ihre "rote Linie". Die Grenze, wo Sie sagen, dieses Ereignis tritt so selten auf (wenn man faire Münzen wirft), dass Sie *nicht* glauben, dass eine faire Münze geworfen worden ist. Sie also sagen: "Wäre eine faire Münze geworfen, so wäre das Ereignis (8 von 10 Treffern) sehr selten, daher glaube ich nicht an die Hypothese der fairen Münze".

Hinweis: Wenn Sie selber (oder Ihr Golem, R) die Münzen wirft, kann Ihr Ergebnis etwas von dem hier gezeigten abweichen, schließlich ist ein Münzwurf ein Zufallsexperiment.

Lösung

a. Wie häufig kamen jetzt 8 Treffer in den Daten vor? Im Diagramm kann man es grob sehen. Zählen wir in der Tabelle nach:

```
muenzversuch %>%
 # filter(heads == 8) %>%
  count(heads)
```

```
##    heads   n
## 1      0   1
## 2      1   8
## 3      2  40
## 4      3 118
## 5      4 214
## 6      5 239
## 7      6 217
## 8      7 110
## 9      8  45
## 10     9   8
```

Das ist ein Anteil von 45 von 1000, also 0.045.

Und zählen wir noch, wie oft 8 *oder mehr* Treffer:

```
muenzversuch %>%
  filter(heads >= 8) %>%
  count(heads)
```

```
##   heads  n
## 1     8 45
## 2     9  8
```

Das ist ein Anteil von 53 von 1000, also 0.053. Diesen Anteil nennt man den *p-Wert*. Ja, ist so.

b. Ob dieser Anteil “selten” oder “häufig” ist, ist eine subjektive Frage! Darauf gibt es keine objektive Antwort. Allerdings ist die Konvention, 5% als Grenze zu nehmen. Ein Ergebnis von kleiner als 5% ist nach dieser Konvention “selten”. Man nennt es auch *statistisch signifikant*. Als Konsequenz glaubt man dann nicht mehr an die getestete Hypothese. Das ist das Vorgehen der klassischen Statistik. Ansonsten schätzt man es als nicht selten ein und nennt es *nicht statistisch signifikant*.

Achtung, jetzt kommt eine Definition:

Den Anteil von 8 oder mehr Treffern (also ein Ergebnis, das mindestens so extrem ist wie unsere Daten) nennt man den p-Wert.

3. **Aufgabe**

Nehmen wir an, $n = 10$ voneinander unabhängige Eigenschaften $E_1, E_2, \ldots, E_{10}$ bestimmen, ob eine Person als “normal” angesehen wird. Jede dieser Eigenschaften kann entweder mit “normal” (*n*) oder aber “nichtnormal” (*nn*) ausgeprägt sein, wobei wir nicht genau vorhersagen können, wie diese Eigenschaften bei einer Person bestellt sein werden.

Als Zufallsexperiment ausgedrückt: $\Omega_E := \{n, nn\}$ mit den zwei Ergebnissen n und nn.

Mit der Wahrscheinlichkeit $Pr_{E_i} = 0.9$ treffe das Ereignis $N_i := E_i = n$ (für alle $i = 1, \ldots, n$) zu.

Nehmen wir weiter an, als “voll normal” (VN) wird eine Person genau dann angesehen, wenn sie in allen n Eigenschaften “normal” ausgeprägt ist, das Ereignis N also für alle n Eigenschaften auftritt.

a. Nennen Sie Beispiele für mögliche Eigenschaften E!
b. Wie groß ist die Wahrscheinlichkeit - unter den hier geschilderten Annahmen -, dass eine Person “voll normal” ist?
c. Diskutieren Sie die Plausibilität der Annahmen!

Lösung

a. Intelligenz, Aussehen, Gesundheit, Herkunft, Hautfarbe, sexuelle Identität oder Neigung, …

b. Für unabhängige Ereignisse ist die Wahrscheinlichkeit, dass sie alle eintreten, gleich dem Produkt ihrer Einzelwahrscheinlichkeiten:

$$VN = Pr(E_i)^{10} = 0.9^{10} \approx 0.3486784$$

Die Wahrscheinlichkeit, dass VN nicht eintritt (Nicht-Voll-Normal, NVN), ist dann die Gegenwahrscheinlichkeit: $NVN = 1 - VN$.

c. Mehrere der Annahmen sind diskutabel. So könnten die Eigenschaften nicht unabhängig sein, dann wäre der hier gezeigte Rechenweg nicht anwendbar. Die Wahrscheinlichkeit für “normal” könnte höher oder niedriger sein, wobei 90% nicht ganz unplausibel ist. Schließlich unterliegt n sozialpsychologischen bzw. soziologischen Einflüssen und kann variieren.

4. **Aufgabe**

Betrachten wir das Ereignis “Schwerer Coronaverlauf” (S); ferner betrachten wir das Ereignis “Blutgruppe ist A” (A) und das Gegenereignis von A: “Blutgruppe ist nicht A”. Ein Gegenereignis wird auch als *Komplementärereignis* oder auch als *Komplement* (complement) mit dem Term A^C bezeichnet.

Sei $Pr\left(S|A\right) = 0.01$ und sei $Pr\left(S|A^C\right) = 0.01$.

Was kann man auf dieser Basis zur Abhängigkeit der Ereignisse S und A sagen?

Lösung

S und A sind unabhängig: Offenbar ist die Wahrscheinlichkeit eines schweren Verlaufs gleich groß unabhängig davon, ob die Blutgruppe A ist oder nicht. In diesem Fall spricht man von *stochastischer Unabhängigkeit*.

$$Pr\left(S|A\right) = Pr\left(S|A^C\right) = Pr\left(S\right)$$

5. **Aufgabe**

Prof. Salzig untersucht eine seiner Lieblingsfragen: Wie viel bringt das Lernen auf eine Klausur? Dabei konzentriert er sich auf das Fach Statistik (es gefällt ihm gut). In einer aktuellen Untersuchung hat er $n = 80$ Studierende untersucht (s. Tabelle und Diagramm) und jeweils erfasst, ob die Person die Klausur bestanden (b) hat oder durchgefallen (d) ist. Dabei hat er zwei Gruppen unterschieden: Die “Viel-Lerner” (VL) und die “Wenig-Lerner” (WL).

Berechnen Sie die folgende *bedingte Wahrscheinlichkeit*: p(Bestehen|Viellerner).

Beispiel: Wenn Sie ausrechnen, dass die Wahrscheinlichkeit bei 42 Prozentpunkten liegt, so geben Sie ein: `0,42` bzw. `0.42` (das Dezimalzeichen ist abhängig von Ihren Spracheinstellungen).

Hinweise:

- Geben Sie *nur eine Zahl* ein (ohne Prozentzeichen o.Ä.), z.B. `0,42`.
- Andere Angaben können u.U. nicht gewertet werden.
- Runden Sie auf zwei Dezimalstellen.
- Achten Sie darauf, das *korrekte Dezimaltrennzeichen* einzugeben; auf Geräten mit deutscher Spracheinstellung ist dies oft ein Komma.

Ergebnisse der Studie

	Viellerner	Weniglerner
Bestehen	27	41
Durchfallen	5	7

Lösung

Der gesuchte Wert liegt bei 0.84.

Lerntyp	Klausurergebnis	n	n_group	prop_conditional_group	N_gesamt
Viellerner	Bestehen	27	32	0.84375	80

6. **Aufgabe**

Als Bildungsforscher(in) untersuchen Sie den Lernerfolg in einem Statistikkurs.

Eine Gruppe von Studierenden absolviert einen Statistikkurs. Ein Teil lernt gut mit (Ereignis A), ein Teil nicht (Ereignis A^C). Ein Teil besteht die Prüfung (Ereignis B); ein Teil nicht (B^C).

Hinweis: Das Gegenereignis zum Ereignis A wird oft das Komplementärereignis oder kurz Komplement von A genannt und mit A^C bezeichnet.

Wir ziehen zufällig eine/n Studierende/n: Siehe da – Die Person hat bestanden. Yeah!

Gesucht ist die Wahrscheinlichkeit, dass *diese Person* gut mitgelernt hat.

Die Anteile der Gruppen (bzw. Wahrscheinlichkeit des Ereignisses) lassen sich unten stehender Tabelle entnehmen.

row_ids	B	Bneg
A	0.68	0.07
Aneg	0.17	0.08

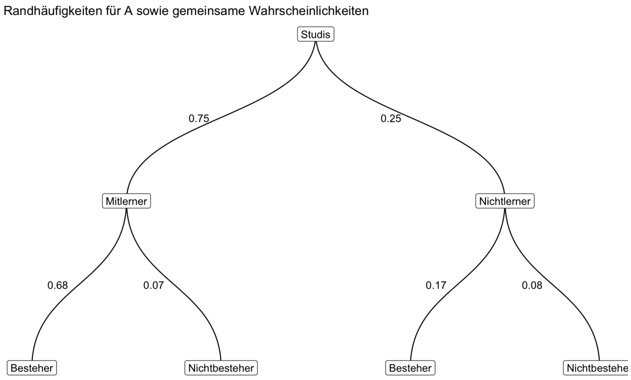
Hinweise:

- Runden Sie auf 2 Dezimalstellen.
- Geben Sie Anteile stets in der Form `0.42` an (mit führender Null und Dezimalzeichen).
- “Aneg” bezieht sich auf das Komplementärereignis zu A.

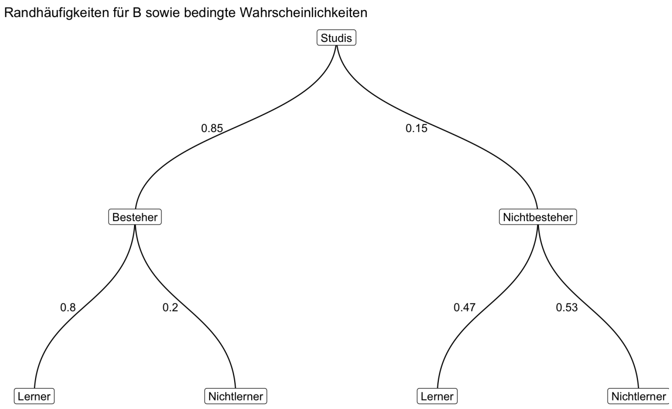
a. Zeichnen Sie (per Hand) ein Baumdiagramm, um die gemeinsamen Wahrscheinlichkeiten darzustellen. Weiterhin sollen die Randwahrscheinlichkeiten für A dargestellt sein.
b. Zeichnen Sie (per Hand) ein Baumdiagramm, um diesen Sachverhalt darzustellen.
c. Geben Sie die Wahrscheinlichkeit des gesuchten Ereignisses an.

Lösung

A.



B.



C.

0.8

```
A_cond_B <- AandB / B_marg %>% round(2)
Aneg_cond_B <- AnegandB / B_marg %>% round(2)
A_cond_Bneg <- AandBneg / Bneg_marg %>% round(2)
Aneg_cond_Bneg <- AnegandBneg / Bneg_marg %>% round(2)
```

$$Pr(A) = 0.75.$$

$$Pr(B) = 0.8525.$$

$$Pr(AB) = 0.6825.$$

$$Pr(A|B) = 0.8029412.$$

$$Pr(\neg A|B) = 0.2.$$

$$Pr(A|\neg B) = 0.4666667.$$

$$Pr(\neg A|\neg B) = 0.5333333.$$

7. **Aufgabe**

Prof. Süß-Sauer untersucht eine seiner Lieblingsfragen: Wie viel bringt das Lernen auf eine Klausur? Dabei konzentriert er sich auf das Fach Statistik (es gefällt ihm gut). In einer aktuellen Untersuchung hat er $n = 80$ Studierende untersucht (s. Tabelle und Diagramm) und jeweils erfasst, ob die Person die Klausur bestanden (b) hat oder durchgefallen (d) ist. Dabei hat er zwei Gruppen unterschieden: Die “Viel-Lerner” (VL) und die “Wenig-Lerner” (WL).

Berechnen Sie die folgende: *gemeinsame Wahrscheinlichkeit*: p(Bestehen UND Viellerner).

Beispiel: Wenn Sie ausrechnen, dass die Wahrscheinlichkeit bei 42 Prozentpunkten liegt, so geben Sie ein: `0,42` bzw. `0.42` (das Dezimalzeichen ist abhängig von Ihren Spracheinstellungen).

- Geben Sie *nur eine Zahl* ein (ohne Prozentzeichen o.Ä.), z.B. `0,42`.
- Andere Angaben können u.U. nicht gewertet werden.
- Runden Sie auf zwei Dezimalstellen.
- Achten Sie darauf, das *korrekte Dezimaltrennzeichen* einzugeben; auf Geräten mit deutscher Spracheinstellung ist dies oft ein Komma.

Lerntyp	Bestehen	Durchfallen
Viellerner	27	5
Weniglerner	41	7

Lösung

Die gemeinsame Wahrscheinlichkeit beträgt 0.34.

Lerntyp	Klausurergebnis	n	n_group	prop_conditional_group	joint_prob
Viellerner	Bestehen	27	32	0.84375	0.3375

Die gemeinsame Wahrscheinlichkeit berechnet sich hier als der Quotient der Zellenhäufigkeit und der Gesamthäufigkeit.

8. **Aufgabe**

Ein renommiertes Unternehmen sucht einen Kandidaten für eine (hoch dotierte) Führungsposition. Ein Managementberatungsunternehmung führt ein Assessmentcenter durch, welches pro Kandidat/in eine positive bzw. negative Empfehlung ergibt. Aus früheren Erfahrungen heraus wissen die Berater, dass die tatsächlich geeigneten Kandidaten (Ereignis E wie *eligible*) mit 70% eine positive Empfehlung für die Stelle ausgesprochen bekommen (Ereignis R wie *recommendation*). Weiterhin bekommen von den *nicht* geeigneten Kandidaten 73% eine negative Empfehlung. Insgesamt wissen die Berater, dass 9% der Bewerber/innen tatsächlich geeignet sind.

Was ist die entsprechende Häufigkeitstabelle? Geben Sie alle vier Einträge in Prozent an!

Hinweis: Das Gegenereignis vom Ereignis A wird als Komplementärereignis oder kurz als Komplement bezeichnet und mit A^C oder $\overline{A}$ abgekürzt. Im vorliegenden Fall meint $\overline{R} = R^C$ das Ereignis, dass ein Kandidat *keine* Empfehlung ausgesprochen bekommt.

a. $P(E \cap R)$
b. $P\left(\overline{E} \cap R\right)$
c. $P\left(E \cap \overline{R}\right)$
d. $P\left(\overline{E} \cap \overline{R}\right)$

Lösung

Einige Wahrscheinlichkeiten lassen sich direkt aus dem Text errechnen:

$$\begin{aligned} P(E \cap R) &= P(R|E) \cdot P(E) = 0.7 \cdot 0.09 = 0.063 = 6.3\% \\ P\left(\overline{E} \cap \overline{R}\right) &= P\left(\overline{R}|\overline{E}\right) \cdot P\left(\overline{E}\right) = 0.73 \cdot 0.91 = 0.6643 = 66.43\%. \end{aligned}$$

Die restlichen gemeinsamen Wahrscheinlichkeiten lassen sich durch Addieren und Subtrahieren in der Kontingenztabelle errechnen:

	R	$\overline{R}$	**Summe**
E	**6.30**	*2.70*	**9.00**
$\overline{E}$	*24.57*	**66.43**	*91.00*
Summe	*30.87*	*69.13*	**100.00**

a. $P(E \cap R) = 6.30\%$
b. $P\left(\overline{E} \cap R\right) = 24.57\%$
c. $P\left(E \cap \overline{R}\right) = 2.70\%$
d. $P\left(\overline{E} \cap \overline{R}\right) = 66.43\%$

9. **Aufgabe**

Erstellen Sie ein Meme, das sich auf den Stoff der aktuellen Stunde im Unterricht bezieht.

Hier finden Sie einige Beispiele zur Anregung.

Wer eine kulturphilosophische Abhandlung sucht zur Frage “Was ist eine Meme?”, der wird hier glücklich.

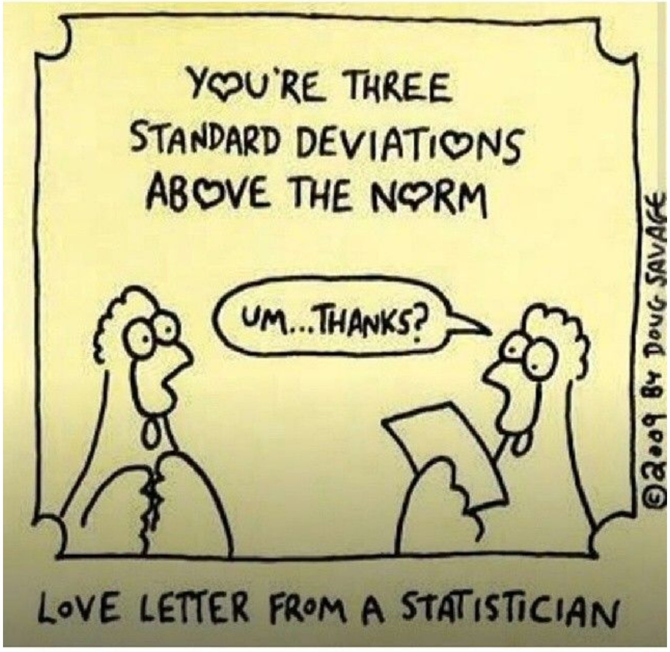
Ok, ich höre gleich auf, aber hier ist eine große Sammlung an Memes.

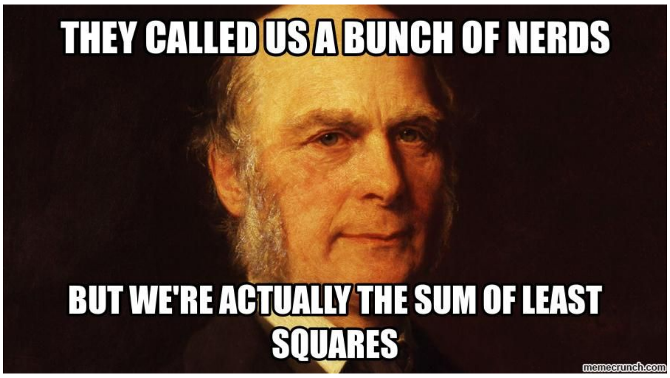
NUR wer krass ist und auf Memes abfährt, sollte sich diese ausführliche Analyse hier anschauen.

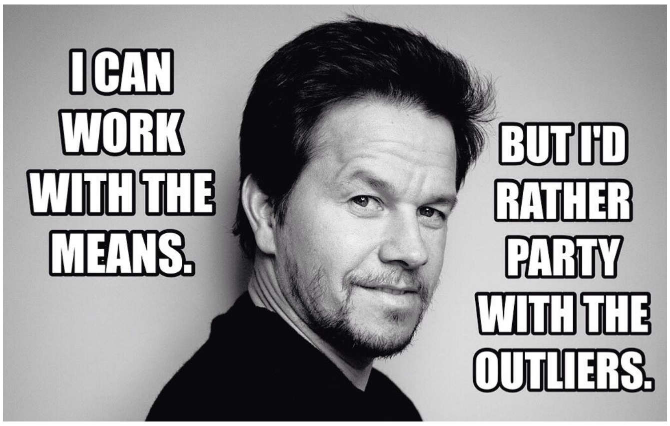
Memes erstellen kann man z.B. hier.

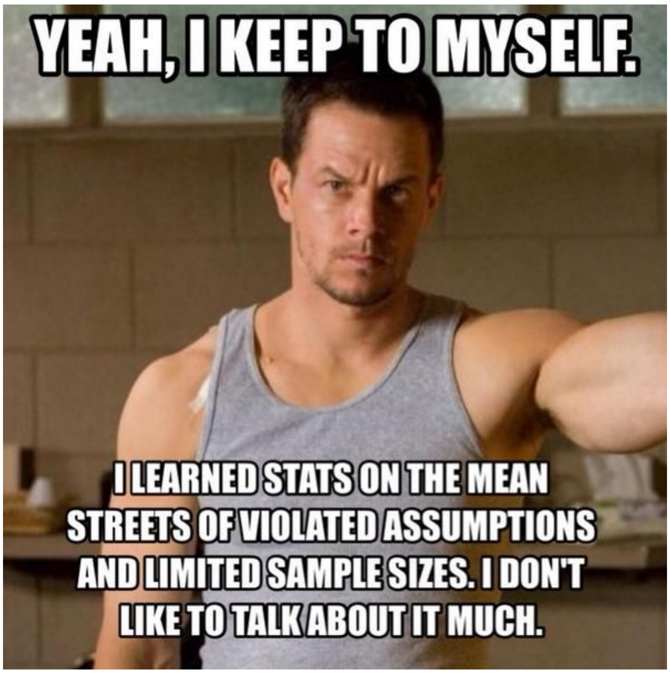
Lösung

Eine Auswahl an allgemeinen Memes zum Thema Statistik:









10. **Aufgabe**

Alkohol ist ein weit verbreites Genussmittel in vielen Gesellschaften. Insgesamt sind die negativen (kausalen) Konsequenzen für die Gesundheit unstrittig. So findet man etwa in dieser Studie:

> This meta-analysis found that alcohol most strongly increased the risks for cancers of the oral cavity, pharynx, esophagus, and larynx. Statistically significant increases in risk also existed for cancers of the stomach, colon, rectum, liver, female breast, and ovaries.

Allerdings gibt es auch Stimmen, die Alkohol mit gesundheitlich wünschenswerten Effekten in Verbindung bringen. Dabei wird in einigen Fällen die "mediterrane Ernährung" als Erkärungsnarrativ ins Spiel gebracht. So kann man etwa hier lesen:

> Adhering to a Mediterranean diet (…) were associated with a lower risk of all-cause mortality (…).

Solche Befunde wurden von der Breiten- oder Boulevardpresse dankbar aufgenommen, wie man z.B. hier nachlesen kann:

> Small Amounts of Alcohol in Mediterranean Diet Could Boost Brain Health, Claims Study

Man beachte, dass "boost your healht" eine kausale Aussage ist, die über einen reinen Zusammenhang hinausgeht. Nach dieser Lesart heißt es: Trink etwas Alkohol (A), das macht dich gesünder (G).

Zeigen Sie ein alternatives Kausalmodell auf, das erklärt, warum ein Zusammenhang (wie eine Korrelation) zwischen A und G zu beobachten ist, aber ohne dass es einen (kausalen) Effekt zwischen beiden Größen gäbe!

Lösung

Eine Erklärung lautet - frei erfunden! -, dass die *Lebenszufriedenheit* (L) jeweils einen (positiven, kausalen) Effekt auf Alkoholkonsum (A) und auf die Gesundheit (G) ausübt.

```
confounder_triangle(x = "A", y = "G", z = "L") %>%  # DAG definieren
  ggdag_dconnected(text = FALSE, use_labels = "label")  # DAG zeichnen
```

```
## Error: object of type 'closure' is not subsettable
```

```
## Error: object of type 'closure' is not subsettable
```

Übrigens: Eine Art von Diagramm, das Kausalbeziehungen zwischen Variablen aufzeigt, ist ein sog. *Directed Acyclic Graph*, oder kurz ein DAG. Hier ist so ein DAG gezeichnet.