

Lösungen zu den Aufgaben

1. Aufgabe

Ob wohl die *PS-Zahl* (Ereignis A) und der *Spritverbrauch* (Ereignis 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. Vereinfachen wir die Frage etwas, indem wir fragen, ob die Ereignisse “hoher Spritverbrauch” (A) und “hohe PS-Zahl” voneinander abhängig sind.

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

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

## Rows: 32
## Columns: 11
## $ mpg   <dbl> 21, 21, 23, 21, 19, 18, 14, 24, 23, 19, 18, 16, 17, 15, 10, 10, 1...
## $ cyl   <dbl> 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 4, 4, 4, 4, 8,...
## $ disp  <dbl> 160, 160, 108, 258, 360, 225, 360, 147, 141, 168, 168, 276, 276, ...
## $ hp    <dbl> 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, 180, 180...
## $ drat  <dbl> 3.9, 3.9, 3.9, 3.1, 3.1, 2.8, 3.2, 3.7, 3.9, 3.9, 3.9, 3.1, 3.1, ...
## $ wt    <dbl> 2.6, 2.9, 2.3, 3.2, 3.4, 3.5, 3.6, 3.2, 3.1, 3.4, 3.4, 4.1, 3.7, ...
## $ qsec  <dbl> 16, 17, 19, 19, 17, 20, 16, 20, 23, 18, 19, 17, 18, 18, 18, 1...
## $ vs    <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0,...
## $ am    <dbl> 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0,...
## $ gear  <dbl> 4, 4, 4, 3, 3, 3, 4, 4, 4, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3,...
## $ carb  <dbl> 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, 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)

#> # A tibble: 1 x 1
#>   median_hp
#>   <dbl>
#> 1 123
```

median(hp)
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
  ))
```

- Schauen Sie im Datensatz nach, ob unser Vorgehen (Erstellung von `hp_high` und `mpg_high`) überhaupt funktioniert hat. Probieren geht über Studieren.
- Visualisieren Sie in geeigneter Form den Zusammenhang.
- Berechnen Sie $Pr(\text{mpg_high}|\text{hp_high})$ und $Pr(\text{mpg_high}|\neg\text{hp_high})$!

Lösung

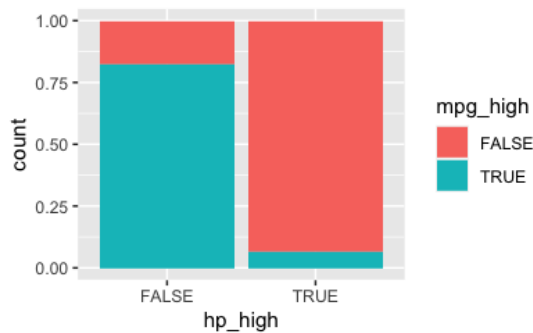
- Schauen wir mal in den Datensatz:

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

	hp	hp_high	mpg	mpg_high
Mazda RX4	110	FALSE	21	TRUE
Mazda RX4 Wag	110	FALSE	21	TRUE
Datsun 710	93	FALSE	23	TRUE
Hornet 4 Drive	110	FALSE	21	TRUE
Hornet Sportabout	175	TRUE	19	FALSE

-

```
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 *rechte* Balken zeigt $Pr(\text{mpg_high} | \text{hp_high})$ und $Pr(\neg \text{mpg_high} | \text{hp_high})$. Der *linke* 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 Kontingenztafel
  group_by(hp_high) %>% # die Anteile pro "Balken" s. Diagramm
  mutate(prop = n / sum(n))

## # A tibble: 4 x 4
## # Groups:   hp_high [2]
##   hp_high mpg_high     n prop
##   <lg1>   <lg1>   <int> <dbl>
## 1 FALSE  FALSE     3 0.176
## 2 FALSE  TRUE     14 0.824
## 3 TRUE   FALSE    14 0.933
## 4 TRUE   TRUE      1 0.0667
```

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")
```

	TRUE	FALSE
TRUE	0.07	0.82
FALSE	0.93	0.18

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 Würfeln), 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 Würfeln 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!

So können wir $n = 1$ Versuch mit 10 (`size=10`) zufälligen (`random`) Münzwürfen (`fair`, d.h. `prob = 0.5`) mit R simulieren:

```
rbinom(n = 1, size = 10, prob = 0.5)
## [1] 4
```

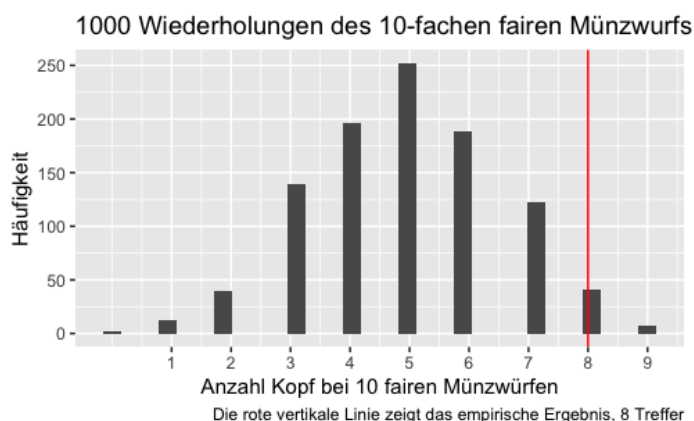
Jetzt wiederholen wir den Versuch 1000 Mal. Hey R, *tue 1000 Mal den 10-fachen Münzwurf*.

```
library(tidyverse) # für "tibble"
muenzversuch <-
  tibble(
    id = 1:1000,
    anzahl_kopf = rbinom(1000, size = 10, prob = 1/2))
head(muenzversuch) %>%
  gt()
```

id	anzahl_kopf
1	7
2	7
3	4
4	7
5	6
6	5

OK, jetzt visualisieren wir die 1000 Versuche bzw. die Tabelle `muenzversuch`:

```
muenzversuch %>%
  ggplot() +
    aes(x = anzahl_kopf) +
    geom_histogram() +
    scale_x_continuous(breaks = 1:10) +
    labs(x = "Anzahl Kopf bei 10 fairen Münzwürfen",
         y = "Häufigkeit",
         title = "1000 Wiederholungen des 10-fachen fairen Münzwurfs",
         caption = "Die rote vertikale Linie zeigt das empirische Ergebnis, 8 Treffer") +
    geom_vline(xintercept = 8, # prop_emp ist 8
               color = "red") # vertikale Linie, um unser emp. Ergebnis anzuzeigen
```



Lange Rede, kurze Fragen:

- 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?
- 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 *mindestens* 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 %>%  
  count(anzahl_kopf)
```

anzahl_kopf	n
0	2
1	13
2	40
3	139
4	196
5	252
6	188
7	122
8	41
9	7

Das ist ein Anteil von 41 von 1000, also 0.04.

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

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

anzahl_kopf	n
8	41
9	7

Das ist ein Anteil von 48 von 1000, also 0.05. 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 (ein *p*-Wert) 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, man *verwirft die H_0* . Ansonsten schätzt man es als nicht selten ein, verwirft die H_0 nicht und nennt es *nicht statistisch signifikant*. Das ist das Vorgehen der klassischen Statistik.

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, $k = 10$ voneinander unabhängige Eigenschaften E_1, E_2, \dots, 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 $i = 1, \dots, k$) zu.

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

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

Lösung

- Intelligenz, Aussehen, Gesundheit, Herkunft, Hautfarbe, sexuelle Identität oder Neigung, ...
- Für unabhängige Ereignisse ist die Wahrscheinlichkeit, dass sie alle eintreten, gleich dem Produkt ihrer Einzelwahrscheinlichkeiten:

$$VN = Pr(E_i)^{10} = 0.9^{10} \approx 0.35$$

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

- 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 das Ereignis E_N mit den Ergebnissen n bzw. nn 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 *Komplement* (complement) mit dem Term A^C bezeichnet.

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

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

Geben Sie ein *Adjektiv* an, dass diesen Sachverhalt kennzeichnet!

Lösung

Die Lösung lautet: unabhängig.

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*.

$$Pr(S|A) = Pr(S|A^C) = Pr(S)$$

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 $n = 60$ 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*: $p(\text{Bestehen}|\text{Viellerner})$.

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	25	13
Durchgefallen	17	5

Lösung

Der gesuchte Wert liegt bei 0.6.

Lerntyp	Klausurergebnis	n	n_group	prop_conditional_group	N_gesamt
Viellerner	Bestehen	25	42	0.6	60

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!

Aufgabe: Gesucht ist die Wahrscheinlichkeit, dass *diese Person* gut mitgelernt hat, gegeben der Tatsache, dass dieser Person bestanden hat.

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

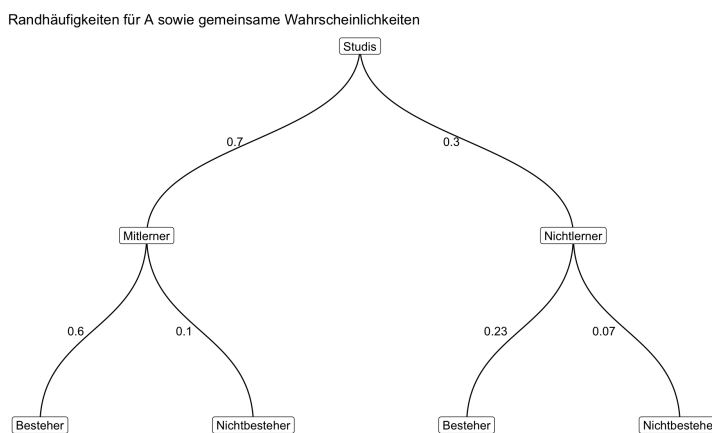
row_ids	B	Bneg
A	0.59	0.10
Aneg	0.23	0.07

Hinweise:

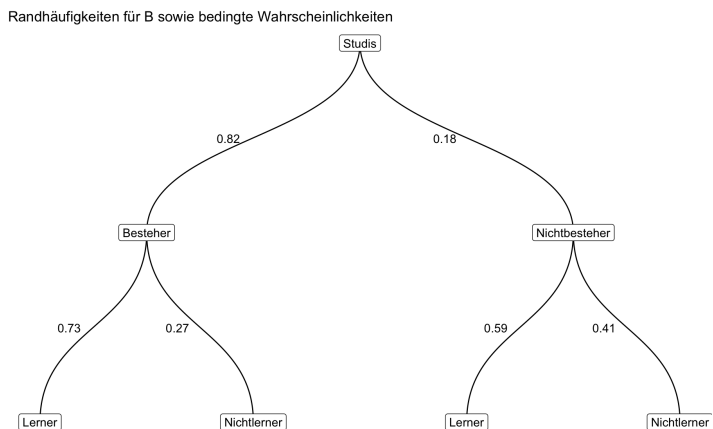
- Runden Sie auf 2 Dezimalstellen.
 - Geben Sie Anteile stets in der Form 0.42 an (mit führender Null und Dezimalzeichen).
 - "Aneg" bezieht sich auf das Komplementärereignis zu A .
- a. Zeichnen Sie (per Hand) ein Baumdiagramm, um die gemeinsamen Wahrscheinlichkeiten darzustellen. Weiterhin sollen die 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.

0.73

```
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)
```

$Pr(A) = 0.7.$

$Pr(B) = 0.82.$

$Pr(AB) = 0.6.$

$Pr(A|B) = 0.73.$

$Pr(\neg A|B) = 0.27.$

$Pr(A|\neg B) = 0.59.$

$Pr(\neg A|\neg B) = 0.41.$

7. Aufgabe

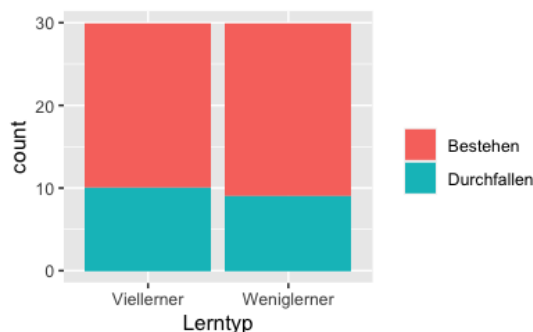
Prof. Salzig untersucht eine seiner Lieblingsfragen: Wie viel bringt das Lernen auf eine Klausur? Dabei konzentriert er sich auf das Fach Statistik (es gefällt ihm gut). In einer aktuellen Untersuchung hat er $n = 60$ Studierende untersucht (s. Tabelle und Diagramm) und jeweils erfasst, ob die Person 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*: $p(\text{Durchgefallen UND Weniglerner})$.

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.

Das folgende Diagramm zeigt die Häufigkeiten pro Gruppe:



Hier ist die Kontingenztafel mit den Häufigkeiten pro Gruppe:

Lerntyp	Bestehen	Durchgefallen
Viellerner	20	10
Weniglerner	21	9

Lösung

Die gemeinsame Wahrscheinlichkeit beträgt 0.15.

Lerntyp	Klausurergebnis	n	n_group	prop_conditional_group	joint_prob
Weniglerner	Durchgefallen	9	30	0.3	0.15

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 Managementberatungsunternehmen 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 64% eine positive Empfehlung für die Stelle ausgesprochen bekommen (Ereignis R wie *recommendation*). Weiterhin bekommen von den *nicht* geeigneten Kandidaten 76% eine negative Empfehlung. Insgesamt wissen die Berater, dass 10% 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 \bar{A} abgekürzt. Im vorliegenden Fall meint $\bar{R} = R^C$ das Ereignis, dass ein Kandidat *keine* Empfehlung ausgesprochen bekommt.

- a. $P(E \cap R)$
- b. $P(\bar{E} \cap R)$
- c. $P(E \cap \bar{R})$
- d. $P(\bar{E} \cap \bar{R})$

Lösung

Einige Wahrscheinlichkeiten lassen sich direkt aus dem Text errechnen:

$$\begin{aligned} P(E \cap R) &= P(R|E) \cdot P(E) = 0.64 \cdot 0.1 = 0.06 = 6.4\% \\ P(\bar{E} \cap \bar{R}) &= P(\bar{R}|\bar{E}) \cdot P(\bar{E}) = 0.76 \cdot 0.9 = 0.68 = 68.4\%. \end{aligned}$$

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

	R	\bar{R}	Summe
E	6.40	3.60	10.00
\bar{E}	21.60	68.40	90.00
Summe	28.00	72.00	100.00

- a. $P(E \cap R) = 6.4\%$
- b. $P(\bar{E} \cap R) = 21.6\%$
- c. $P(E \cap \bar{R}) = 3.6\%$
- d. $P(\bar{E} \cap \bar{R}) = 68.4\%$

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.

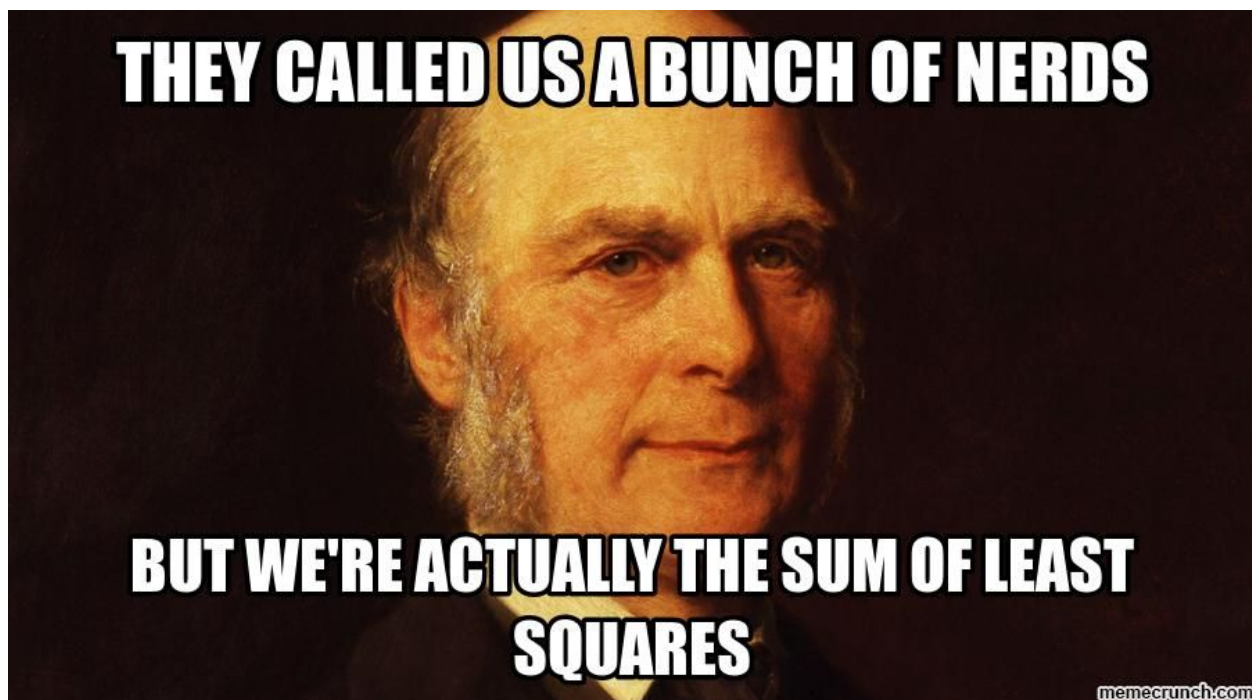
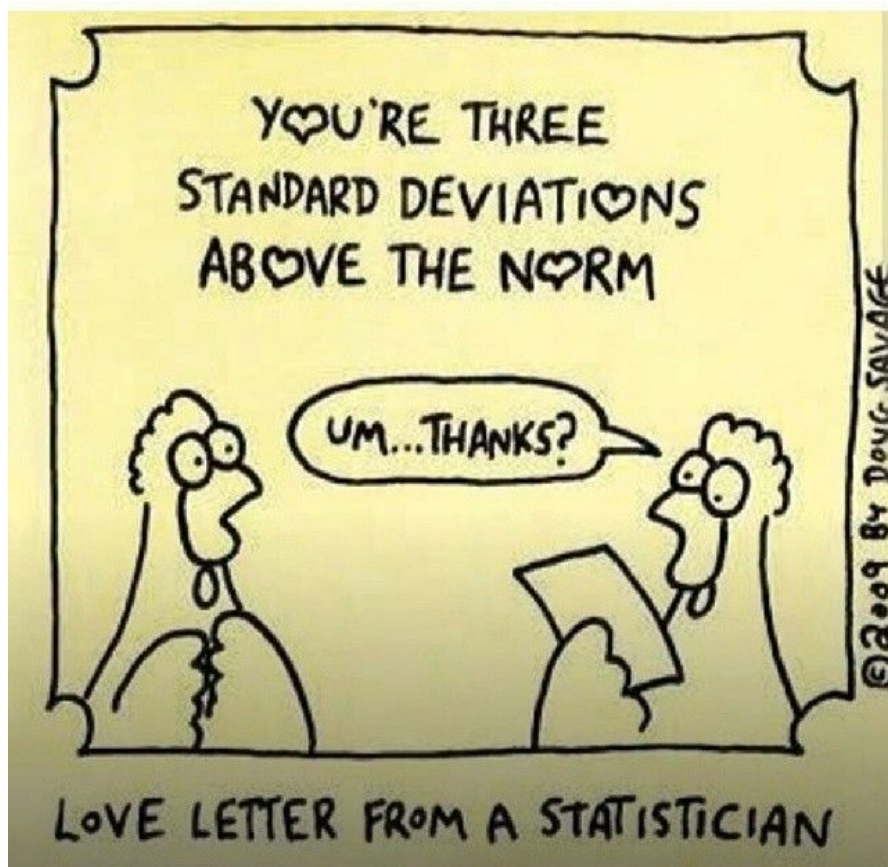
Ok, ich höre gleich auf, aber hier ist eine [große Sammlung an Memes](#).

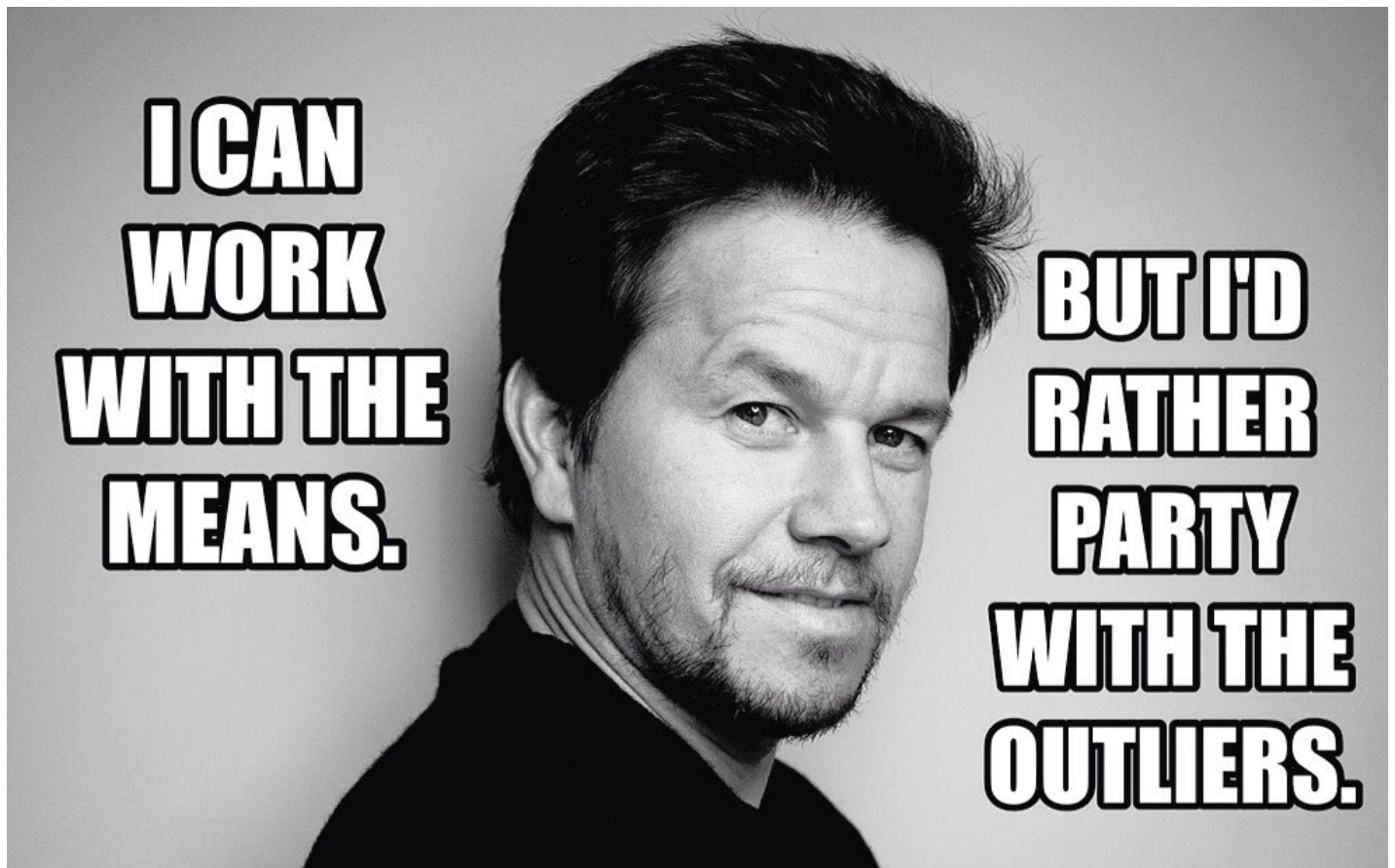
NUR wer krass ist und auf Memes abfährt, sollte sich [diese ausführliche Analyse hier](#) anschauen.

Memes erstellen kann man z.B. [hier](#).

Lösung

Eine Auswahl an allgemeinen Memes zum Thema Statistik:





10. Aufgabe

Alkohol ist ein weit verbreitetes 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 Erklärungsnarrativ 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 health" eine kausale Aussage ist, die über einen reinen Zusammenhang hinausgeht. Nach dieser Lesart heißt es: Trink etwas Alkohol (A), das macht dich gesünder (G).

Ihre Aufgabe: 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.

```
detach("package:mosaic", unload = TRUE)

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

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

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.