Lösungen zu den Aufgaben

1. **Aufgabe**

   Eine Klettererin verwendet ein Seil, dass eine Sicherheit von $r = .99$ hat: mit einer Wahrscheinlichkeit von 1% reißt das Seil. Jetzt knüpft sie mehrere dieser Seile zusammen zu einem “Gesamtseil”. Wie groß ist die Gefahr, dass das „Gesamtseil“ reist?

   a. Sie knüpft 2 zusammen
   b. Sie knüpft 5 zusammen
   c. Sie knüpft 10 zusammen
   d. Sie knüpft 20 zusammen

   Lösung

   Sei R die Wahrscheinlichkeit, dass das Gesamtseil hält (nicht reißt). $1 - R$ ist dann die Wahrscheinlichkeit des Gegenereignisses: das Gesamtseil reißt.

   Allgemein ist R bei k Tests gleich r hoch k: $R = r^k$.

   Das Aufaddieren der Fehlalarm-Wahrscheinlichkeit bezeichnet man als Alphafehler-Inflation.

```
r <- .99
R2 <- r^2 %>% round(2)  # Auf 2 Dezimalen runden
R5  <- r^5  %>% round(2)
R10 <- r^10  %>% round(2)
R20 <- r^20  %>% round(2)
```

   Die Antworten (Gesamtsicherheiten) lauten also:

```
R2
```

```
## [1] 0.98
```

```
R5
```

```
## [1] 0.95
```

```
R10
```

```
## [1] 0.9
```

```
R20
```

```
## [1] 0.82
```

Vertiefung

Betrachten wir abschließend aus Neugier die Wahrscheinlichkeit, dass die Klettererin abstürzt ($1 - R$) als Funktion der Anzahl der Seie.

Diese Überlegung ist etwas weiterführender und nicht ganz so zentral, aber ziemlich interessant.

Definieren wir die Parameter:

```
anz_seile <- 1:20  # von 1 bis max 20 Seile
r <- c(.9, .95, .99, .999)  # verschiedene Seil-Sicherheiten
```

Jetzt erstellen wir einen Tabelle, die alle `anz_seile * r` Werte kombiniert:

```
d <-
  expand_grid(anz_seile, r)

head(d)
```

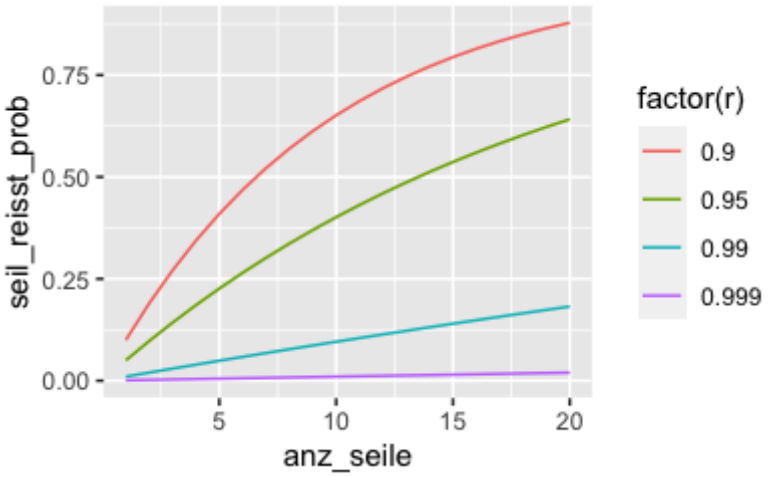
```
## # A tibble: 6 × 2
##   anz_seile     r
##       <int> <dbl>
## 1         1 0.9
## 2         1 0.95
## 3         1 0.99
## 4         1 0.999
## 5         2 0.9
## 6         2 0.95
```

Jetzt berechnen wir für jede Kombination die Gesamtsicherheit `R` sowie die Wahrscheinlichkeit, dass das Seil reißt, $1 - R$:

```
d <-
  d %>%
  mutate(R = r^anz_seile,
         seil_reisst_prob = 1 - R)
```

Und plotten das Ganze:

```
d %>%
  ggplot(aes(x = anz_seile,
             y = seil_reisst_prob,
             color = factor(r))) +
  geom_line()
```



Hat ein Seil eine Sicherheit von 90%, dann will man nicht dranhängen, wenn 20 Seile zusammengeknotet sind!

a. $R_2 = r \cdot r = r^2 = 0.98$
b. $R_5 = r^5 = 0.95$
c. $R_{10} = r^{10} = 0.9$
d. $R_{20} = r^{20} = 0.82$

2. **Aufgabe**

Das “Maschinendisaster” sei als folgendes Szenario beschrieben:

Eine Maschine bestehe aus einer Menge Teile, die alle recht zuverlässig arbeiten. Außerdem arbeiten alle Teile unabhängig voneinander (vermutlich keine ganz realistische Annahme). Die Zuverlässigkeit eines Teils sei $r = .9999$ für ein bestimmtes Zeitintervall t. Mit $1 - r$ fällt also ein Teil innerhalb von t aus.

Ein interessanter Schnörkel ist, dass man “Maschine” auch als “Computerprogramm” oder “biologisches System” lesen kann (ohne Verlust der Allgemeinheit).

Ei Forschi fragt sich, aus wie vielen k Teilen die Maschine höchstens bestehen darf, damit es mit einer Wahrscheinlichkeit von 99% zu keinem Ausfall innerhalb von $t = 1$ kommt.

Lösung

```
r <- .9999
loesung <- log(.99, base = r) %>% trunc()
loesung
```

```
## [1] 100
```

`trunc()` schneidet die Dezimalstellen ab, rundet also ab.

$$\begin{aligned} r^k &= .99 \qquad |log_r \\ log_r\left(r^k\right) &= log_r(.99) \\ k &\approx 100 \end{aligned}$$

Die Lösung lautet also 100.

3. **Aufgabe**

Ein Glücksspieler will ausprobieren, ob er einer Münze befehlen kann mit Kopf nach oben zu laden (wäre ja praktisch).

Er überlegt sich als Experiment, eine Münze 50 Mal zu werfen und jedes Mal konzentriert “Kopf!” zu befehlen.

Da er aber auch Realist ist, überlegt er sich, wie hoch die Wahrscheinlichkeit ist, per Zufall so oft “Kopf” zu bekommen.

Er hat gehört, dass bei vielen Wissenschaftlern ein Ereignis, dass eine Wahrscheinlichkeit von weniger als 5% hat, als “signifikant” gilt und als Beleg für einen Effekt gilt, in dem Fall wäre der Effekt seine Münz-Befehl-Zauberkraft.

Als Erstes kommt der langweilige Teil: Er wirft 100 Mal 50 Münzen, um sich ein “Gefühl” für “typische” Ergebnisse zu bekommen. Puh! Sehnenscheidenentzündung, und überhaupt, diese

ganzen unhygienischen Münzen…

Einen Münzwurf kann man auch als “Bernoulli-Experiment” bezeichnen.

Werfen wir also $n = 50$ Münzen mit R. “Kopf” sei das Ereignis, das wir suchen:

```
rbernoulli(n = 50)
```

```
##  [1] FALSE  TRUE  TRUE FALSE FALSE  TRUE
##  [7]  TRUE  TRUE  TRUE FALSE FALSE FALSE
## [13] FALSE  TRUE FALSE FALSE FALSE  TRUE
## [19]  TRUE  TRUE  TRUE  TRUE FALSE FALSE
## [25] FALSE FALSE FALSE  TRUE FALSE  TRUE
## [31] FALSE FALSE FALSE FALSE FALSE FALSE
## [37]  TRUE  TRUE FALSE FALSE  TRUE FALSE
## [43]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## [49]  TRUE  TRUE
```

Zählen wir die Treffer (Anzahl Kopf):

```
rbernoulli(n = 50) %>% sum()
```

```
## [1] 25
```

Die obige Zeile könnten wir jetzt 100 Mal tippen (Sehnenscheidenentzündung)…

Oder … wir probieren es cleverer:

```
replicate(n = 100, expr = rbernoulli(n = 50) %>% sum())
```

```
##   [1] 23 26 23 20 17 28 25 22 23 31 28 19 25
##  [14] 29 25 32 24 22 25 30 26 24 27 28 25 24
##  [27] 25 22 25 24 22 23 28 25 26 26 23 24 29
##  [40] 27 29 23 24 22 31 28 29 26 25 23 29 29
##  [53] 32 25 21 25 18 21 25 30 28 26 27 23 21
##  [66] 21 33 23 30 21 26 27 25 26 29 29 24 23
##  [79] 26 27 24 22 26 23 30 23 32 25 30 23 27
##  [92] 19 13 21 22 23 27 21 26 28
```

`replicate()` wiederholt (repliziert) einen Ausdruck (`expr`ession) `n` Mal.

Tabellieren wir dieses Ergebnis, und spaßeshalber, lassen wir unseren Glücksspieler den Versuch jetzt 1000 Mal durchführen (zum Glück mit Hilfe des Computers):

```
set.seed(42)
oft_50_muenzen <- replicate(n = 1000, expr = rbernoulli(n = 50) %>% sum())

tibble(oft_50_muenzen) %>%
  count(oft_50_muenzen)
```

```
## # A tibble: 21 × 2
##    oft_50_muenzen     n
##             <int> <int>
##  1             15     1
##  2             16     2
##  3             17     6
##  4             18    18
##  5             19    20
##  6             20    37
##  7             21    62
##  8             22    98
##  9             23   107
```

```
## 10             24    93
## # … with 11 more rows
```

Mit `set.seed(42)` legen wir die Generierung von Zufallszahlen auf eine bestimmte Auswahl fest. Diese Zeile ist nur dazu dar, damit Sie die exakt gleichen Ergebnisse bekommen wie in diesem Text. Ansonsten würden Ihre Ergebnisse leicht abweichen - schließlich sind es ja zufällige Ereignisse.

Unser Glücksspieler resümiert: “Ah, so alles über 30 Treffer, also 60% Trefferquote, scheint etwas besonderes zu sein, kommt zumindest selten vor!”.

Aber kann man den Weg zum Glück nicht abkürzen?

Er nimmt sich seine 50 Münzen und wirft sie hintereinander. Dabei notiert er sich den kumulativen Anteil von Kopf (s. Diagramm).

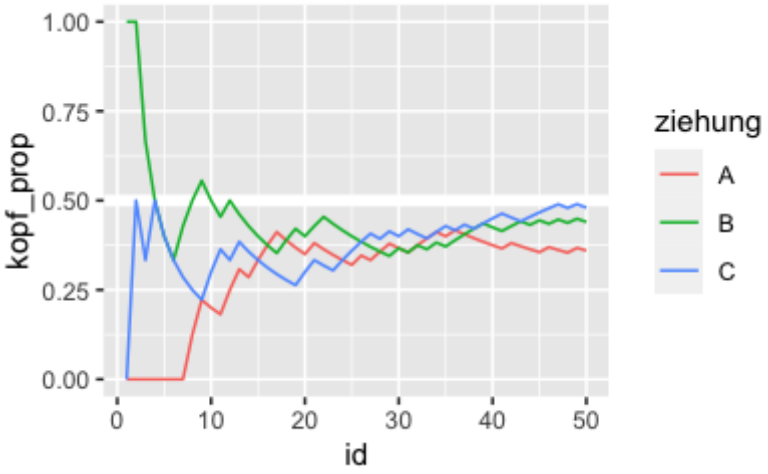
“Falls ich vorzeitig einen Anteil von über 60% Kopf habe, breche ich die Ziehung einfach ab, werfe also nicht 50 Mal, sondern weniger, und berichte einfach nur dieses Experiment!”, denkt sich unser Glücksspieler.

Mal sehen, ob seine Rechnung aufgeht:

```
d2 <-
  tibble(
    kopf = rbernoulli(150),
    ziehung = c(rep("A", 50), rep("B", 50), rep("C", 50)),
    id = c(1:50, 1:50, 1:50))

d3 <-
  d2 %>%
  group_by(ziehung) %>%
  mutate(
    kopf_sum = cumsum(kopf),
    kopf_prop = cummean(kopf))

d3 %>%
  ggplot(aes(x = id, y = kopf_prop, color = ziehung)) +
  geom_hline(yintercept = .5, color = "white", size = 2) +
  geom_line()
```



Frage

Würde es dem Glücksspieler etwas bringen, vorzeitig den Versuch zu stoppen, also *weniger* als 50 Mal die Münze zu werfen?

Zur Erinnerung: Sein Ziel ist es, einen hohen Anteil von “Kopf” zu werfen.

a. Nein, der Anteil von Kopf ist in konstant über die 50 Münzwürfe (für jede der drei Ziehungen).
b. Nein, der Anteil von Kopf sinkt im Laufe der 50 Münzwürfe (für jede der drei Ziehungen).
c. Ja, es gibt Würfe, bei denen der Anteil der (kumulierten) “Kopf-Ergebnisse” höher ist als der sich stabilisierende Mittelwert.

Lösung

Tatsächlich: Seine Rechnung geht auf. Betrügen ist möglich: Zu einigen Zeiten ist der Mittelwert oberhalb des langristig zu erwartenden Mittelwert (hier ist der zu erwartende Mittelwert 50%).

Daher sollte ein Experiment *vorab* angeben, wie groß die Stichprobe ist.

Dieser vorab geplante Stichprobenwert sollte dann an ein Gremium eingereicht werden, bevor die Datenerhebung beginnt. Das nennt man auch “preregistration”.

- Falsch
- Falsch
- Richtig