

Fallstudie zu quantitativen Berufsfeldern der Wirtschaftspsychologie

2019-05-21

1 Vorbereitung

```
library(mosaic)
```

2 Aufgabe OE

Eine Organisationsentwicklungsabteilung möchte herausfinden, was die Ursachen für Unterschiede in der Motivation ihrer Mitarbeiter sind. Dazu berechnet ein Mitarbeiter den Zusammenhang von der Zufriedenheit mit der direkten Führungskraft (Variable: **zuf**) und der Arbeitsmotivation (**engagement**); der Datensatz mit den Daten heie **d** (keine echte Daten; es ist eine theoretische Aufgabe).

1. Schreiben Sie die R-Syntax, um ein 95%-Konfidenzintervall zu berechnen!
2. Der Abteilungsleiter sthnt, "So kompliziert! Inferenzstatistik braucht doch kein Mensch!" Erlutern Sie, in welchen Fllen bzw. wozu Inferenzstatistik ntzlich ist.

LSUNG

1. Syntax (hier nur theoretisch, also ohne echte Daten)

```
# DO NOT RUN THIS CODE
h0_vert1 <- do(1000) * cor(zuf ~ engagement, data = resample(d))

gf_histogram(~ cor, data = h0_vert1)

quantile(~ cor, probs = c(0.025, 0.975), data = h0_vert1)
```

2. Argumentation zum Chef

"Mchte man nur Aussagen treffen ber eine Stichprobe, braucht man keine Inferenzstatistik. Mchte man von einer Stichprobe auf eine Grundgesamtheit schließen, so quantifiziert die Inferenzstatistik die Ungewissheit dieses Schlusses".

3 Aufgabe Vertrieb

Ein Business-Development-Abteilung untersucht, ob die neue Version des Webshops ihrer Firma mehr Vertriebs­erfolg bringt als die alte Version (**version**). Vertriebs­erfolg wird gemessen als der Anteil der Besucher der Seite, die etwas kaufen beim aktuellen Besuch (**buyer**). (Der Datensatz heie **d**; keine echte Daten; es ist eine theoretische Aufgabe).

1. Ist folgende Syntax korrekt, um herauszufinden, ob die neue Seite signifikant mehr Umsatz erzielt als die alte Seite?

```
# DO NOT RUN THIS CODE
h0_vert2 <- do(1000) * diffmean(buyer ~ shuffle(version),
                                data = d)
```

2. Korrigieren Sie die Syntax ggf.
3. Was würde ein Konfidenzintervall hier aussagen?

LÖSUNG

1. Nein.
2. Korrekte Syntax:

```
# DO NOT RUN THIS CODE
h0_vert2 <- do(1000) * diffprop(buyer ~ shuffle(version),
                                data = d)

gf_histogram(~ diffprop, data = h0_vert2)
```

3. KI: Das KI gibt einen Schätzbereich für den gesuchten Parameter an, hier also die Größe des Unterschieds in den Kaufquoten.

4 Aufgabe Eignungstest

Die Personalabteilung einer technisch orientierten Firma möchte einen Wissenstest zur Diagnose der Berufseignung implementieren. Der zuständige Facharbeiter entwirft 20 Wissensitems (Aussagen) zu einem IT-technischen Wissensgebiet, die jeweils mit *Richtig* bzw. *Falsch* zu beantworten sind (d.h. jede Aussage ist entweder richtig oder falsch). Jetzt überlegt sie: “Hm, wie viele Items muss eine Bewerberin richtig beantworten, damit wir sicher genug sind, dass sie nicht bloß geraten hat?”.

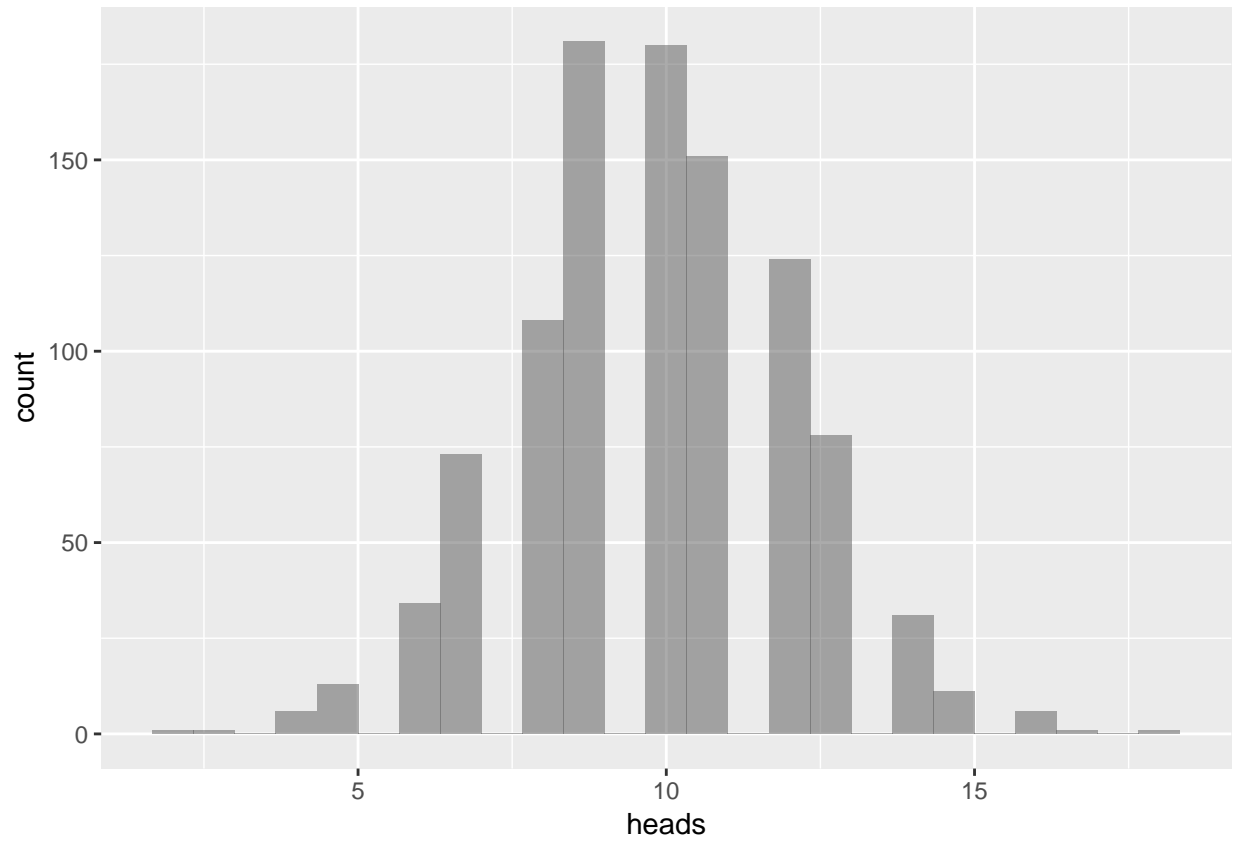
1. Berechnen Sie die Anzahl für eine Sicherheit von 95%!
2. Wie lautet der Fachbegriff für die Wahrscheinlichkeit, die die Facharbeiterin oben anspricht?
3. Die erste Bewerberin hat 13 Aufgaben richtig; angenommen sie hat geraten, wie hoch ist die Wahrscheinlichkeit für dieses (oder ein noch extremes) Ergebnis?
4. Wie 2) nur mit 15 Treffern.
5. Angenommen, eine Bewerberin, Frau Sch. Ummel, mit 15 Treffern wird eingestellt. Tatsächlich hat sich aber nur geraten. Die Personalabteilung denkt als fälschlich, sie hätte einen “Treffer” (Effekt) gefunden. Nennen Sie den statistischen Fachbegriff für diese Art von Fehler.

LÖSUNG

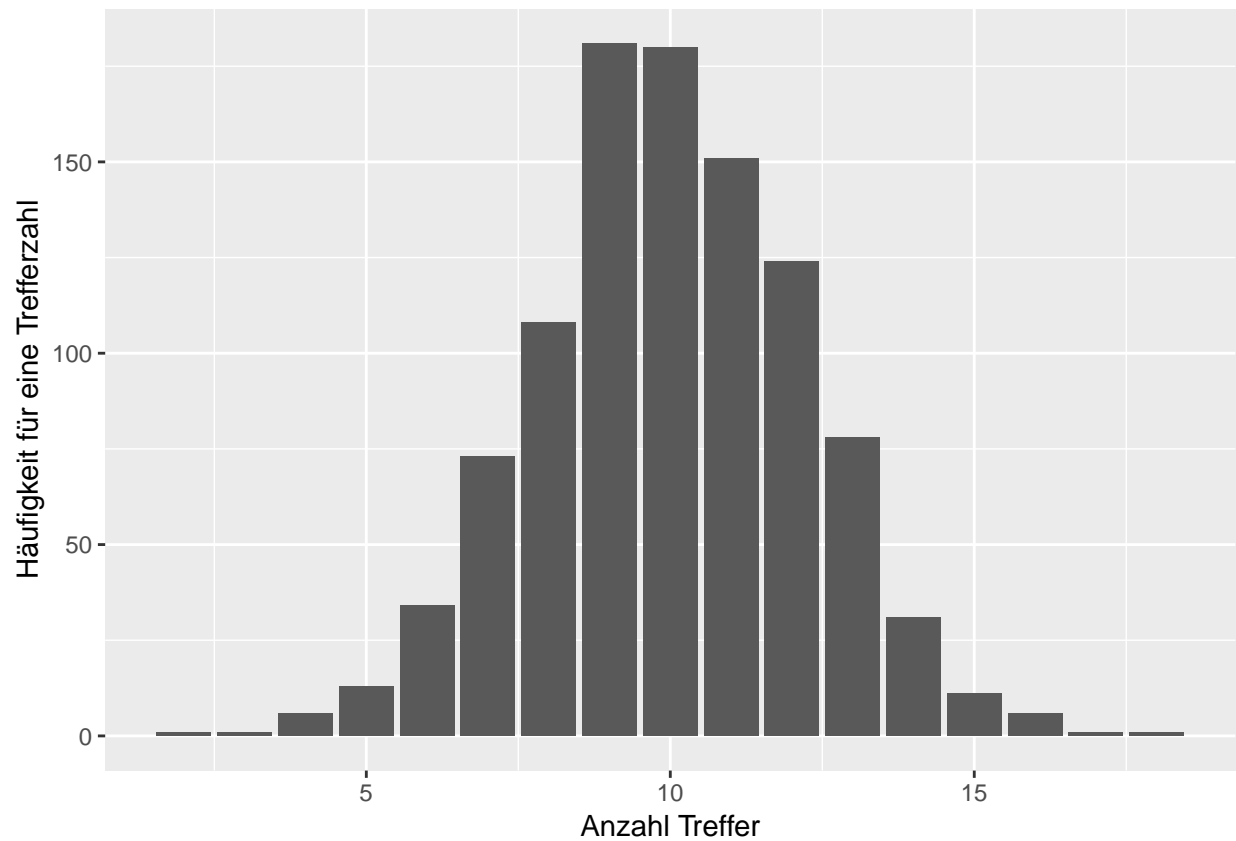
1. Anzahl nötiger Treffer:

```
set.seed(42)
h0_vert3 <- do(1000) * rflip(n = 20)

gf_histogram(~ heads, data = h0_vert3)
```



```
gf_bar(~ heads, data = h0_vert3) %>% # schaut hier schöner aus als das Histogramm
gf_labs(x = "Anzahl Treffer", y = "Häufigkeit für eine Trefferzahl")
```



Jetzt ausprobieren:

```
prop( ~ heads>=13, data = h0_vert3) # 13%
```

```
## prop_TRUE
##      0.128
```

```
prop( ~ heads>=14, data = h0_vert3) # 5%
```

```
## prop_TRUE
##      0.05
```

```
prop( ~ heads>=15, data = h0_vert3) # 2%
```

```
## prop_TRUE
##      0.019
```

Antwort: 15 Treffer. Bei 15 Treffern ist die Wahrscheinlichkeit $< 5\%$ (für so ein oder noch extremeres Ereignis), wenn man bloß rät. MaW: Der p-Wert ist kleiner als 5%.

Eleganter:

```
quantile(~ heads, prob = .95,  
        data = h0_vert3)
```

```
## 95%  
## 13.05
```

Man könnte die Münzwurfverteilung (sog. Binominalverteilung) durch die Normalverteilung approximieren. Berechnen wir dazu zunächst die Streuung (sd) und den Mittelwert (M) unserer Stichprobenverteilung der Münzwürfe:

```
sd_h0_vert3 <- sd(~ heads, data = h0_vert3)  
mean(~ heads, data = h0_vert3)
```

```
## [1] 9.995
```

Dann können wir uns das Quantil (d.h. den Cutoff-Wert, X-Achse) bequem ausgeben lassen (unter Annahme einer Normalverteilung, was hier ausreichend gegeben ist).

```
xqnorm(p = .95, mean = 10, sd= sd_h0_vert3)
```

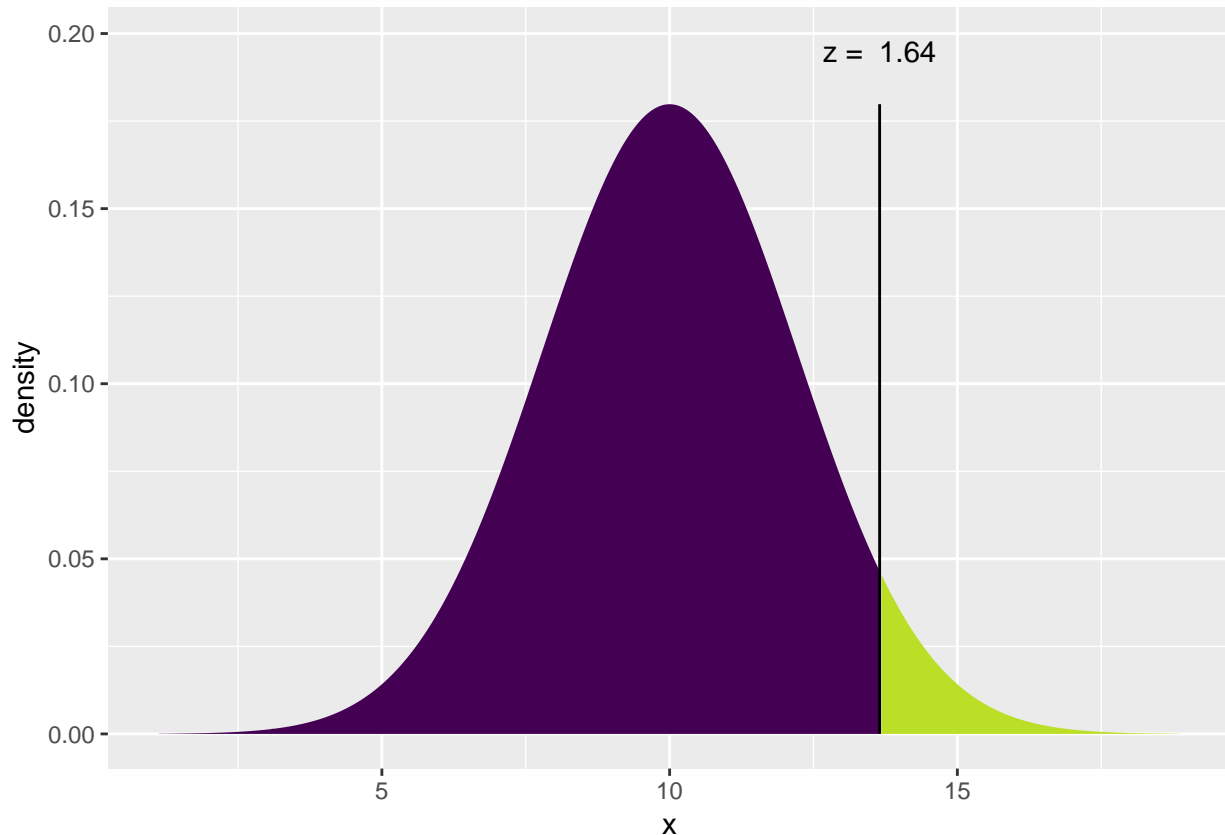
```
##
```

```
## If  $X \sim N(10, 2.218535)$ , then
```

```
##  $P(X \leq 13.64917) = 0.95$ 
```

```
##  $P(X > 13.64917) = 0.05$ 
```

```
##
```



```
## [1] 13.64917
```

Antwort: Etwa 14 Treffer sind nötig, um einen Fehlalarm mit mind. 95% Wahrscheinlichkeit auszuschließen.

2. Die Facharbeiterin spricht den p-Wert an. Der p-Wert ist der Anteil der Stichprobenergebnisse in der H_0 -Verteilung, die mind. so extrem sind wie das echte (empirische) Ergebnis. Das ist *nicht* genau das, was die Facharbeiterin gesagt hat, aber es geht in eine ähnliche Richtung und der p-Wert ist das beste Maß zur Entscheidung über eine Hypothese, das wir in diesem Kurs kennenlernen.
3. $p = 13\%$
4. $p = 2\%$
5. Es handelt sich um einen Alphafehler oder Fehler 1. Art

5 Aufgabe “Freies Spiel der Kräfte”

Recherchieren Sie einen Datensatz, der zu ihrem (angestrebten) Berufsbild passt und importieren Sie ihn in Ihr RStudio. Treffen Sie ggf. Annahmen für den weiteren Verlauf der Aufgabe wo nötig.

1. Nennen Sie eine Forschungsfrage für einen ungerichteten Zusammenhang (zu diesem Datensatz)!
2. Berechnen Sie die relevanten deskriptiven Statistiken!
3. Visualisieren Sie die Datenlage zur Ihrer Forschungsfrage!
4. Berechnen Sie ein Konfidenzintervall für die Forschungsfrage!

5. Formulieren Sie die passenden formalen statistischen Hypothesen dazu.
6. Berechnen Sie den p-Wert!
7. Nennen Sie die statistische Entscheidung und begründen Sie sie.