

Wintersemester 2023/2024

Jens Klenke

## R Propädeutikum

### Übungsaufgaben 3

#### 1 Verteilungen und Zufallszahlen

- 1.1 Sei  $X \sim t(5)$ . Berechnen Sie  $P(X < 6)$ ,  $P(3 < X \leq 7)$  und  $P(X > 4)$ .
- 1.2 Berechnen Sie das 0.95-Quantil einer  $F(4, 5)$ -verteilten Zufallsvariable.
- 1.3 Berechnen Sie die Wahrscheinlichkeit dafür, den Jackpot im Lotto zu gewinnen (d.h. 6 Richtige aus 49). Vernachlässigen Sie bei Ihrer Berechnung Zusatz- oder Superzahlen. (Hinweis: Benutzen Sie die hypergeometrische Verteilung.)
- 1.4 Erzeugen Sie 20  $\chi^2(5)$ -verteilte Zufallszahlen ohne (!) dabei die `rchisq()`-Funktion zu benutzen.

*Hinweis:*  $\chi^2(n) = \sum_{i=1}^n Z_i^2$  mit  $Z_i \sim \mathcal{N}(0, 1)$  für alle  $i = 1, \dots, n$ .

- 1.5 Ziehen Sie 10-mal standardnormalverteilte Zufallszahlen vom Umfang  $n = 10000$  und berechnen Sie für jeden Durchlauf das arithmetische Mittel. Schauen Sie sich danach alle 10 Mittelwerte an. Was fällt Ihnen auf? Sind Sie überrascht?

**Zusatzaufgabe:** Führen Sie dieselbe Simulationsstudie mit Cauchy-verteilten Zufallszahlen (`rcauchy()`) durch. Was fällt Ihnen nun auf? Können Sie sich das Ergebnis erklären?

- 1.6 Zeigen Sie, dass das Integral über die Dichtefunktion einer  $\chi^2(15)$ -verteilten Zufallsvariable 1 ist.

*Hinweis:* Zum Integrieren in R können Sie `integrate()` benutzen.

## 2 Grafiken

```
set.seed(385)
results <- rnorm(1000, mean = 100, sd = 15)
```

- 2.1 Kopieren Sie obigen Code und nehmen Sie an, dass dieser eine IQ-Testreihe mit 1000 Probanden simuliert. Zeichnen Sie ein Histogramm der Ergebnisse. Geben Sie Ihrem Plot anschließend eine passende Überschrift sowie passende Achsenbeschriftungen. Spezifizieren Sie darüber hinaus den Bereich von x- und y-Achse auf  $[40, 160]$  bzw.  $[0, 0.03]$ .
- 2.2 Hinterlegen Sie dem Plot die dem IQ zugrundeliegende, theoretische Dichtefunktion, d.h. eine Normalverteilung mit  $\mu = 100$  und  $\sigma = 15$ . Wählen Sie als Zeichenfarbe Rot und machen Sie die einzuzeichnende Linie etwas breiter.
- 2.3 Zeichnen Sie einen Punkt in Form eines Dreiecks an das Maximum der theoretischen Dichte. Wählen Sie als Farbe Blau.
- 2.4 Kennzeichnen Sie sowohl das 0.025- als auch das 0.975-Quantil der theoretischen Dichte, in dem Sie Vertikalen an diesen Punkten einzeichnen. Wählen Sie als Farbe Grün.

## 3 Lineare Regression

- 3.1 Betrachten Sie im Folgenden den Datensatz `faithful`, der Daten zum Old Faithful Geysir im Yellowstone Nationalpark enthält. Sowohl die Dauer einer Eruption in Min. (`eruptions`) als auch die Wartezeit bis zur nächsten Eruption in Min. (`waiting`) sind als Variable im Datensatz verfügbar. Unterstellen Sie nachfolgendes Regressionsmodell und schätzen Sie die entsprechenden Parameter  $\hat{\beta}_0, \hat{\beta}_1$ . Zeichnen Sie anschließend eine geeignete Grafik und interpretieren Sie diese.

$$\text{waiting}_t = \beta_0 + \beta_1 \text{eruptions}_t + u_t$$

- 3.2 Verschaffen Sie sich mit `summary()` einen Überblick über ihr in 3.1 erhaltenes Ergebnis. Interpretieren Sie die geschätzten Koeffizienten und speichern Sie anschließend das  $R^2$  (*Multiple R-squared*) in der Variablen `R2` ab. *Hinweis:* Schauen Sie sich die, beim Ausführen von `summary()`, ausgegebene Datenstruktur genauer an.
- 3.3 Angenommen Sie beobachten einen zusätzlichen Datenpunkt für die Dauer einer Eruption von  $X_{\text{new}} = 4$ . Sagen Sie die entsprechende Wartezeit bis zur nächsten Eruption vorher.