Universität Duisburg-Essen
Lehrstuhl für Ökonometrie
Dr. Yannick Hoga
M.Sc. Martin Arnold

Wintersemester 2022/23

Übungsblatt 1 — Grundlagen in Statistik mit R

Zu diesem Übungsblatt empfehlen wir neben der Lektüre der Kapitel 2 und 3 des Lehrbuches *Introduction to Econometrics* von *Stock & Watson* eine Aufarbeitung mithilfe der Kapitel 2 und 3 in unserem Online-Companion *Introduction to Econometrics with R*.

Aufgabe 1 – Rechnen mit der Normalverteilung

Dem *Sozio-ökonomischen Panel* (SOEP) zufolge beträgt die Körpergröße von volljährigen, männlichen Bundesbürgern im Mittel 179cm bei einer Standardabweichung von etwa 7cm. Gehen Sie davon aus, dass die Körpergröße ein normalverteiltes Merkmal ist. Für diese Population gilt also:

$$\text{Körpergröße} \sim \mathcal{N}(\mu = 179, \sigma^2 = 49)$$

(a) Wie viel Prozent der Männer oben genannter Gruppe haben eine Körpergröße von *nicht mehr* als 170cm?

Lösung:
Im folgenden sei X die Körpergröße, d.h. $X \sim N(179, 49)$ und $Z \sim \mathcal{N}(0, 1)$.
Gesucht ist also $P(X \leq 170)$.

$$\begin{aligned} P(X \leq 170) &= P\left(Z \leq \frac{170 - 179}{\sqrt{49}}\right) \\ &= P(Z \leq -1.2857) \\ &= \Phi(-1.2857) \\ &= 1 - \Phi(1.2857) \\ &\approx 0.1 \end{aligned}$$

(b) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig aus dieser Population ausgewählter Mann größer als 190.5cm ist?

Lösung: Wir suchen $P(X \geq 190.5)$.

$$\begin{aligned} P(X \geq 190.5) &= P\left(Z \geq \frac{190.5 - 179}{\sqrt{49}}\right) \\ &= P(Z \geq 1.6429) \\ &= 1 - \Phi(1.6429) \\ &\approx 0.05 \end{aligned}$$

(c) Die Verteilung der Körpergröße für volljährige Basketballspieler weicht von o.g. Verteilung ab. Sie vermuten, dass hier die Verteilung

$$\text{Körpergröße} \sim \mathcal{N}(\mu = 199, \sigma^2 = 56.25)$$

gilt. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Basketballspieler zwischen 190cm und 205cm groß ist?

Lösung: Gesucht ist $P(190 \leq X \leq 205)$.

$$\begin{aligned} P(190 \leq X \leq 205) &= P\left(\frac{190-199}{\sqrt{56.25}} \leq Z \leq \frac{205-199}{\sqrt{56.25}}\right) \\ &= P(-1.2 \leq Z \leq 0.8) \\ &= \Phi(0.8) - \Phi(-1.2) \\ &= \Phi(0.8) - [1 - \Phi(1.2)] \\ &\approx 0.67 \end{aligned}$$

(d) Nutzen Sie `R`, um die in den Teilaufgaben (a) bis (c) gesuchten Ergebnisse zu berechnen. Hierfür ist die Funktion `pnorm()` hilfreich.

```
# (a)
pnorm(170, mean = 179, sd = 7)

# (b)
pnorm(190.5, mean = 179, sd = 7, lower.tail = F)

# (c)
pnorm(205, mean = 199, sd = 7.5) - pnorm(190, mean = 199, sd = 7.5)
```

Aufgabe 2 – Wahrscheinlichkeitsverteilungen in R

(a) Plotten Sie die $\mathcal{N}(0,1)$-Dichte mit den Funktionen `dnorm()` und `curve()`.

```
# Wir plotten die N(0,1)-Dichte über dem Intervall [-4,4]
curve(dnorm(x),
      from = -4,
      to = 4,
      main = "N(0,1)-Dichtefunktion")
```

(b) Berechnen Sie die Dichte einer $\mathcal{N}(3, 4^2)$-verteilten Zufallsvariable an der Stelle 4 mit `dnorm()`.

```
# Beachte, dass sd die Standardabweichung ist
dnorm(x = 4, mean = 3, sd = 4)
```

(c) Berechnen Sie $P(X \leq 4)$ für $X \sim N(3, 4^2)$ mit `pnorm()`.

```
# Gesucht ist P(X<=4) für X ~ N(3,16)
pnorm(q = 4, mean = 3, sd = 4)
```

(d) Berechnen Sie das 0.95-Quantil einer $\mathcal{N}(3, 4^2)$-verteilten Zufallsvariable mit `qnorm()`.

```
# Quantile können mit qnorm() berechnet werden
qnorm(p=0.95, mean = 3, sd = 4)
```

(e) Erstellen Sie eine 100-elementige Zufallsstichprobe der χ^2_k-Verteilung, wobei die Anzahl der Freiheitsgerade $k = 10$ beträgt. Nutzen Sie die Funktion `rchisq()`. Berechnen Sie das Stichprobenmittel.

```r
# Wir nutzen rchisq() um 100 Beobachtungen von X ~ Chi^2_10 zu erzeugen
# df ist die Anzahl der Freiheitsgrade
X <- rchisq(n = 100, df = 10)

# Stichprobenmittel berechnen mit mean()
mean(X)
```

(f) Sei X binomialverteilt mit $n = 50$ und $p = 1/3$. Berechnen Sie $P(10 \leq X \leq 30)$.

```r
# Siehe '?Binomial'. Wir nutzen pbinom() umd Wahrscheinlichkeiten für
# binomialverteilte Zufallsvariablen zu berechnen
pbinom(q = 30, size = 100, prob = 1/3) - pbinom(q = 9, size = 100, prob = 1/3)
```

Aufgabe 3 – Daten einlesen und t-Test mit R

(a) Die Datei "`Daten.csv`" enthält 1000 Beobachtungen einer Variable X. Lesen Sie den Datensatz in `R` ein. *Hinweis:* Nutzen Sie hierfür die Funktion `read.csv2()`.

```r
# Arbeitsverzeichnis abfragen
getwd()

# Arbeitsverzeichnis setzen
# (Beispielpfad)
setwd("Z:/RUebung")

# Daten anzeigen/einlassen
read.csv2("Daten.csv")

# Daten einlesen
Daten <- read.csv2("Daten.csv")
```

(b) Berechnen Sie mit `R` deskriptive Statistiken für den Datensatz und stellen Sie die Verteilung der Daten graphisch dar. Was fällt Ihnen auf?

```r
# Übersicht über die Struktur mit str(), siehe ?str
str(Daten)

# Deskriptive Statistiken
summary(Daten$X)

### Grafische Darstellung(en)

# Plot trägt X gegen den Index ab
plot(Daten$X,
     main = "Grafische Darstellung der Daten")

# Besser: Histogramm. Mit freq=TRUE werden absolute Häufigkeiten an der
# Y-Achse abgetragen, bei freq=FALSE die (geschätzte) Dichte.
hist(Daten$X,
     main = "Darstellung der Daten als Histogramm",
     freq = FALSE)

# Das Histogramm lässt vermuten, dass X ~ Chi^2 verteilt ist.
```

```r
# Wir sehen, dass eine Chi^2_10 Verteilung gut passt.
curve(dchisq(x, df = 10),
      from = 0,
      to = 30,
      add = T)
```

(c) Nutzen Sie die Funktion `t.test()` für den Hypothesentest von

$$H_0 : \mu_X = 10 \text{ vs. } H_1 : \mu_X \neq 10.$$

Welche weiteren Informationen finden Sie im Output von `t.test()`? Erläutern Sie kurz.

```r
# Aufgrund der Plots vermuten wir, dass mu = 10 ist.
# Wir nutzen die Funktion t.test() um zu testen, ob mu = 10 ist.
?t.test

# t.test berechnet standardmäßig den p-Wert für einen Test von
# H_0 gegen eine beidseitige Alternativhypothese.
t.test(Daten$X, mu = 10)

# Wir vergleichen den p-Wert mit dem 5\% Signifikanzniveau:
#
# Da P > 0.05 wird H_0: mu = 10 zum 5%-Niveau beibehalten.
# Der Output enthält außerdem ein 95%-Konfidenzintervall
# für mu.
```

Aufgabe 4 – Normalverteilungsdichte als R-Funktion definierem

Die Normalverteilung besitzt die Dichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

(a) Implementieren Sie diese Funktion als `R`-Funktion mit dem Namen `fnorm`.

(b) Überprüfen Sie, ob Ihre Funktion für $\mu = 0$ und $\sigma = 1$ die Dichte an den Stellen $-1.96, 0$ und 1.96 korrekt berechnet. Benutzen Sie hierzu eine geeignete Funktion der Familie `Normal`, siehe `?Normal`. Welche Quantile sind $-1.96, 0$ und 1.96?

(c) Ziehen Sie 1000 Zufallszahlen aus der Standardnormalverteilung und berechnen Sie geläufige deskriptive Statistiken für Ihre Stichprobe.

Hinweis zu (a): Der folgende Code definiert in `R` die Dichtefunktion der *Standardnormalverteilung* als Funktion `f`.

```r
f <- function(x) {
  1/(sqrt(2 * pi)) * exp(-0.5 * x^2)
}
```