

Serie 4

1. Der Geysir Old Faithful im Yellowstone National Park ist eine der bekanntesten heissen Quellen. Für die Zuschauer und den Nationalparkdienst ist die Zeitspanne zwischen zwei Ausbrüchen und die Eruptionsdauer von grossem Interesse.

Im File <http://stat.ethz.ch/Teaching/Datasets/geysir.dat> sind die Messungen vom 1.8.1978–8.8.1978 in 3 Spalten abgelegt: "Tag", "Zeitspanne" und "Eruptionsdauer".

- a) Zeichnen Sie Histogramme von der Zeitspanne zwischen zwei Ausbrüchen:

```
geysir <- read.table("http://stat.ethz.ch/Teaching/Datasets/geysir.dat",
                    header = TRUE) ## Datensatz einlesen
par(mfrow = c(2,2)) ## 4 Grafiken im Grafikfenster
## Histogramme zeichnen
hist(geysir[, "Zeitspanne"])
hist(geysir[, "Zeitspanne"], breaks = 20)
hist(geysir[, "Zeitspanne"], breaks = seq(41, 96, by = 11))
```

Was fällt auf? Was ist der Unterschied zwischen diesen drei Histogrammen?

Bemerkung:

Wenn man die Anzahl Klassen mit `breaks = 20` vorgibt, so wird dies nur als "Vorschlag" interpretiert und intern unter Umständen noch abgeändert.

- b) Zeichnen Sie Histogramme (Anzahl Klassen variieren) von der Eruptionsdauer.

```
hist(geysir[, "Eruptionsdauer"], ...)
```

Was fällt auf? Vergleichen Sie mit der ersten Teilaufgabe.

2. Bei einem Zufallsexperiment werden 3 Würfel gleichzeitig geworfen. Statt dieses Experiment analog zu Aufgabe 1 der Übung 1 theoretisch zu analysieren, wollen wir hier seine Eigenschaften durch Simulation näher kennenlernen.

- a) Simuliere mit *R* 100 Realisierungen von je 3 Würfeln und berechne jeweils die Summe. Speichere die Resultate in die Vektoren `wuerfel1`, `wuerfel2`, `wuerfel3` und `wuerfelSumme`.

```
> n <- 100
> wuerfel1 <- sample(1:6, size=n, replace=TRUE)
# Vektor wuerfel1 erzeugen, die anderen analog
> wuerfelSumme <- wuerfel1+wuerfel2+wuerfel3
```

- b) Zeichne ein Histogramm der erzeugten Augensummen und berechne Mittelwert und Standardabweichung (diese entsprechen den theoretischen Grössen Erwartungswert und Standardabweichung).

```
> hist(wuerfelSumme, breaks=2.5:18.5, freq=FALSE, ylab="rel. Haeufigkeit")
> mean(wuerfelSumme); sd(wuerfelSumme)
```

- c) Erhöhe die Anzahl Realisierungen auf 10000 und wiederhole a) und b). Was fällt dabei auf?

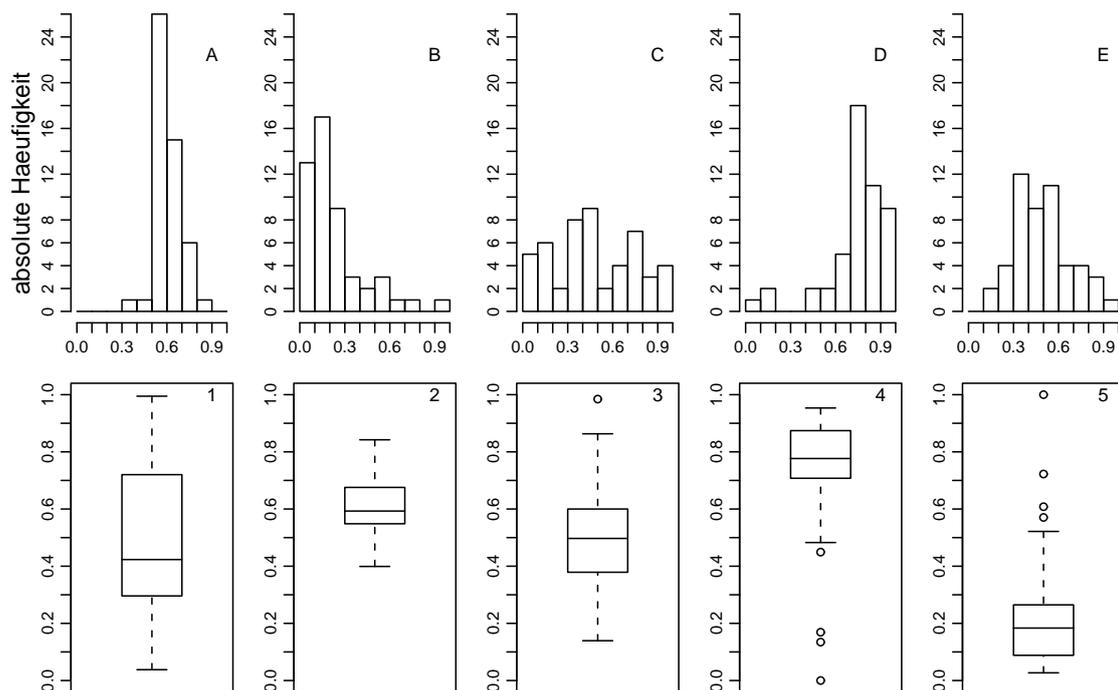
- d) Ein Casino bietet das folgende Spiel an: Drei Würfel werden gleichzeitig geworfen. Ist die Summe der Augenzahlen grösser als 12 gewinnt der Spieler 2.-, sonst verliert er 1.-. Simuliere 100 Wiederholungen dieses Spiels und berechne den mittleren Gewinn.

```
> gewinn <- ifelse(wuerfelSumme>12, 2, -1)
> mean(gewinn)
```

- e) Stelle einen möglichen Gewinnverlauf eines Spielers dar, der 100 mal in Folge das oben erklärte Spiel spielt. Kommentar?

```
> kumGewinn <- cumsum(gewinn) # Kumulativer Gewinn
> plot(1:100, kumGewinn, xlab="Spielnummer", ylab="Gesamtgewinn")
```

3. Für fünf Stichproben vom Umfang $n = 100$ wurden je ein Histogramm und ein Boxplot gezeichnet. Ordnen Sie die fünf Boxplots den entsprechenden Histogrammen zu. Geben Sie für jede Zuordnung eine kurze Begründung!



Besprechung: 20., 21. März.