

# Programmieren mit statistischer Software

Eva Endres, M.Sc.

Institut für Statistik

Ludwig-Maximilians-Universität München

*Simulationen*



Grundlagen

Aufbau von Simulationen

Durchführung von Simulationen

Auswertung von Simulationen

Grundlagen

Aufbau von Simulationen

Durchführung von Simulationen

Auswertung von Simulationen

# Problemstellung

---

- Ausgangspunkt: statistisches Verfahren zur Beschreibung und/oder Modellierung von Daten
- Untersuchung von Eigenschaften bzw. Bestimmung der Güte des zugrundeliegenden Verfahrens
- Durchführung anhand von Daten, die ein reales Szenario darstellen, deren wahre Struktur aber bekannt ist

- Wie sehen meine Daten aus?
- Welches statistische Modell liegt zugrunde?
- Welchen Aspekt möchte ich untersuchen?
- Welche Modellkomponenten bzw. Parameter sind fixiert?
- Welche Modellkomponenten bzw. Parameter sollten variiert werden?
- Welches Kriterium verwende ich zur Beurteilung der Ergebnisse?

# Simulation von Zufallszahlen in R

---

## 1. Simulation aus eindimensionalen Verteilungen

- Normalverteilung: `rnorm()`
- Binomialverteilung: `rbinom()`
- Poissionverteilung: `rpois()`

...

## 2. Simulation aus mehrdimensionaler Normalverteilung

- `library(MASS)`
- `mvrnorm()`

Grundlagen

Aufbau von Simulationen

Durchführung von Simulationen

Auswertung von Simulationen

# Bestandteile einer Simulation

---

Prinzipiell besteht eine Simulation aus den drei folgenden Bestandteilen:

- Modell zur Generierung der Daten (Datengenerierender Prozess)
- Zu untersuchendes statistisches Verfahren
- Beurteilungskriterien der resultierenden Modellparameter

Jeder Bestandteil enthält Parameter, die variiert werden können (einfachstes Beispiel: Stichprobenumfang  $n$ ). Eine konkrete Kombination dieser Parameter ergibt ein Simulationsszenario.

# Beachtenswertes

---

1. Bestimmte Anzahl an Wiederholungen eines Simulationsszenarios notwendig, falls Komponenten der Simulation zufällig sind.
2. Systematische Variation der Parameter der Simulation notwendig. Anzahl an Simulationsszenarien sollte wohl überlegt sein (Laufzeit!)
3. Welche Teile der Simulation sollten (zwischen-)gespeichert werden? (`save()`, `pdf()`)
4. Welche Teile der Simulation sollten replizierbar sein? Oftmals Verwendung von `set.seed()` notwendig.

Grundlagen

Aufbau von Simulationen

Durchführung von Simulationen

Auswertung von Simulationen

# Simulationsausführung

---

Simulationsaufbau sollte optimalerweise so gestaltet sein, dass durch einen Funktionsaufruf

```
results <- simulation()
```

alle Berechnungen vollständig durchgeführt werden.

# Beachtenswertes

---

1. Simulation immer erst in sehr kleinem Umfang ausführen, um Fehler frühzeitig zu finden und eine Zeitabschätzung vorzunehmen (`system.time()`, `microbenchmark()`)
2. Sollen Statusmeldungen ausgegeben werden, die den Fortschritt der Simulation anzeigen? (`cat()`, `trace()`)
3. Oftmals ist eine Parallelisierung von Teilen der Simulation sinnvoll/erforderlich. (Funktion `foreach()` aus package 'foreach' bzw. package 'batchjobs')
4. Längere Simulationen im Hintergrund auf einem Server ausführen

Grundlagen

Aufbau von Simulationen

Durchführung von Simulationen

Auswertung von Simulationen

- Auswertung der Simulationsergebnisse mithilfe von statistischen Kennzahlen
- Darstellung der Ergebnisse in Graphiken und Tabellen
- Die Auswertung sollte von der Simulationsausführung getrennt sein und als nachfolgender Teil stattfinden.