

## Exercise 1 - Berksons und Simpson Paradox

Bearbeite entweder das Berksons Paradox oder das Simpson Paradox - oder bearbeite als **Zusatz** beide:

1. Berksons Paradox Neben vielen anderen medial verbreiteten Falschmeldungen und Fehlschlüssen zum Thema Covid-19 wurde auch zwischenzeitlich behauptet, Rauchen schütze vor Covid-19. In einigen Ländern waren nämlich unter den wegen Covid-19 hospitalisierten Personen weniger Raucher als in der allgemeinen Bevölkerung. Der hier vorliegende Trugschluss ist in der Statistik bekannt u.a. als Berksons Paradox.  
(a) Was ist das Berksons Paradox? Beschreiben Sie das Berksons Paradox in allgemeiner verständlicher Form.
2. Simpson Paradox Ein weiteres spannendes Paradoxon aus der Statistik ist das nach Simpson benannte. Beschreiben Sie es in allgemeiner Form, unter Zuhilfenahme geeigneter Tabellen, und erstellen Sie selbst wieder ein eigenes illustratives und überzeugendes Beispiel

## Exercise 2 - R: EEG Data

In diesem Beispiel wollen wir uns einen EEG Datensatz anschauen (*library(HRM)*). Der Datensatz enthält 40 Messungen von 160 Versuchsperson. Diese Messungen sind pro Versuchsperson abhängig, wir betrachten daher immer nur die erste Messung. Die Versuchspersonen sind in 4 Gruppen eingeteilt: Subjective Cognitive Complaint (SCC+, SCC-), Mild Cognitive Impairment (MCI), und Alzheimer's Disease (AD).

- (a) Erstellen Sie Kastendiagramme (box plots) für die vier Behandlungsgruppen. Benutzen Sie dazu die Funktion `boxplot` in R.
- (b) Interpretieren Sie anschließend das Ergebnis hinsichtlich Unterschiede zwischen den Gruppen. Klären Sie in dem Zusammenhang auch, was in einem Kastendiagramm zu sehen ist (Stichwort: 5-Zahlen).
- (c) Erstellen Sie dann auch Kastendiagramme für alle zehn Regionen (die vier Gruppen werfen wir hier alle zusammen). Gibt es hier Unterschiede je nach Region?

## Exercise 3 - R: EEG Data Fortsetzung

Die Funktion `ggplot` benötigt einen "tidy" Datensatz (bzw. long table Format). Der Vorteil von `ggplot2` gegenüber den Standard R Funktionen ist, dass Sie viel mehr Möglichkeiten haben (und es ist vor allem leichter), Ihre Grafik so zu gestalten, wie Sie es haben möchten. Betrachten Sie folgenden R Code:

```
library(HRM)
library(ggplot2)
ggplot(data = EEG, aes(x=dimension, y=value))
+geom_line(aes(group=subject, col=group))
+theme_bw()
```

- (a) Finden Sie zuerst heraus, was dieser Code macht. Die Grafik ist etwas unübersichtlich, ändern Sie deshalb den Code ab, sodass nur mehr die Daten für eine Gruppe angezeigt werden und dass jede Versuchsperson eine andere Farbe bekommt.
- (b) Ersetzen Sie die Zeile mit `geom_line(...)+` durch `geom_boxplot()+`. Auf der x-Achse sollen die vier Behandlungsgruppen aufgetragen werden. Wie müssen Sie die Argumente innerhalb von `aes()` der Funktion `ggplot` abändern? Optional können Sie die vier Kastendiagramme auch unterschiedlich einfärben, dies geht ähnlich wie beim Beispielcode.

## Exercise 4 - Operationen

In einer Studie wurde untersucht, ob es einen Unterschied in der Überlebenschance zwischen operativen Eingriffen am Tag bzw. in der Nacht gibt. Dabei ergab sich folgende Kontingenztafel:

	Überlebte	Überlebte nicht
Operation am Tag	700	100
Operation in der Nacht	300	900

1. Berechnen und interpretieren Sie das *Odds Ratio* und den *Risikoquotienten*. Wie ist der oben beschriebene Zusammenhang einzuschätzen?
2. Berechnen Sie die Risikodifferenz für die Untergruppe der Nicht-Überlebenden und interpretieren Sie das Ergebnis.
3. Ist es bei diesem Studiendesign sinnvoll, ein *Odds Ratio* zu berechnen? Geben Sie eine kurze Begründung an.