

Aufgabe 4: Daten modellieren

FOM R-Training

HINWEIS: Für die Beantwortung der Aufgaben und für weitere Informationen kann das Cheat Sheet *C4-Datenanalyse-mit-Mosaic* zu Hilfe genommen werden.

Aufgabe a) Hypothesentest auf Mittelwertdifferenz

Die Funktion `t.test` aus dem Paket `mosaic` führt einen statistischen Hypothesentest des Mittelwertes durch. Je nach Spezifikation führt die Funktion einen einfachen Mittelwerttest gegen einen festen Wert durch

```
t.test( ~ variable_of_interest, data=mydata,
        mu=fixed_value, alternative="greater")
```

oder einen Test auf Differenz des Mittelwerts zwischen zwei Gruppen

```
t.test( variable_of_interest ~ grouping_variable, data=mydata)
```

Laden Sie den Datensatz `tips.csv` ein und testen Sie die folgenden Nullhypothesen zum Niveau $\alpha = 0.05$:

- H_0 : Die mittlere Rechnungshöhe ist nicht größer als 15\$.
- H_0 : Es gibt keinen Unterschied in der Trinkgeldhöhe zwischen Rauchern und Nichtrauchern.

Aufgabe b) Lineare Regression

Die Funktion `lm` passt ein lineares Modell an die Daten an und gibt u.a. die geschätzten Parameter und die Modellgüte (R^2) zurück. Der Befehl zum Anpassen des Modells lautet:

```
reg_fit <- lm(dependent_variable ~ independent_variable, data = mydata)
```

und der Befehl zum Ausgeben der Ergebnisse

```
summary(reg_fit)
```

Passen Sie ein Modell an, was die Trinkgeldhöhe durch die Höhe des Rechnungsbetrages erklärt. Wie hoch ist das R^2 ?