

# Einfache Varianzanalyse

## Übliche Ausgangslage

- Ein Experiment
- Ein primärer Faktor wird gemessen (z.B. Druckfestigkeit in lb)
- Messungen in Gruppen mit verschiedenen Bedingungen gegeben (z.B. Art der Verpackung oder zu verschiedenen Konzentrationen eines Materials)

## Schritt 1: Daten einlesen

- DataFrame verwenden
- ... ersetzen mit Anzahl Messungen pro Gruppe
- Messwerte unbedingt als Float eingeben!

```
import numpy as np
from pandas import DataFrame

df = DataFrame({
    "Gruppe": np.repeat(["G1", "G2", "G3", "G4"], [..., ..., ..., ...]),
    "Messwerte": [5.0, 4.5, ...]
})
```

## Schritt 2: Daten visualisieren

- Stripplot und Boxplot erstellen
- Plots kommentieren
  - Lage
  - Streuung
  - wichtig: offensichtliche Lageunterschiede zwischen Gruppen kommentieren

## Schritt 3: Gruppenmittelmodell wählen

1. Gruppenmittelmodell: meistens  $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ 
  - $i$ :  $i$ -te Gruppe
  - $j$ :  $j$ -te Beobachtung (Messung)
  - $\mu$ : globaler Mittelwert
  - $Y_{ij}$ :  $j$ -te Beobachtung in der  $i$ -ten Gruppe
  - $\tau_i$ : Abweichung des Mittelwertes der  $i$ -ten Gruppen vom globalen Mittelwert
  - $\varepsilon_{ij}$ : Abweichung der  $j$ -ten Beobachtung vom Mittelwert in der  $i$ -ten Gruppe

- Behandlungseffekte:  $\tau_i$  und  $\varepsilon_{ij}$
- 2. Parametrisierung erwähnen:  $\mu = \mu_1$
- 3. Behandlungsspezifische Abweichung erwähnen:  $\tau_1 = 0$
- 4. Parameter berechnen und auflisten

```
fit = ols("Messwerte~Gruppe", data=df).fit()
fit.params
```

```
tau_1 = 0, tau_2 = ..., tau_3 = ..., tau_4 = ...
```