



## Projektarbeiten

---

### Statistics for WING 2020

Liebe Studierende

Wie Sie bereits wissen, werden Sie im Verlaufe dieses Semesters ~~eine grössere~~ **zwei** Projektarbeiten machen

- Projektarbeit I (statistische Analyse) bestehend aus Teil I und II
- Projektarbeit II (Spamfilter)

Sehr gute Arbeiten können mit zusätzliche Bonuspunkte belohnt werden. Hier finden Sie alle nötigen Informationen.

## Projektarbeit I

---

### 1 Allgemeine Kriterien für die Projektarbeit

Sie können einen beliebigen Datensatz wählen, wobei die folgenden Kriterien erfüllt sein müssen.

- ❖ Niemand darf den gleichen oder einen ähnlichen Datensatz wählen (Liste der Datensätze wird auf Moodle ersichtlich sein).
- ❖ Der Datensatz muss mehr als 3 Spalten und mehr als 100 Zeilen erhalten (Daten nicht manipulieren z.B. nicht mit NA-Werte füllen!).
- ❖ Es müssen mindestens 3 verschiedene Variablen-Typen vorkommen (z.B. character, integer, numeric).
- ❖ Sie müssen mit Python arbeiten.

Es gibt **keine Punkte** falls diese Kriterien nicht eingehalten werden.

### 2 Teil I – beschreibende Statistik

- ❖ Sie müssen (gibt je ~~0.5~~ **1** Punkt)
  - eine klare Fragestellung formulieren, die Sie mit Ihrer Visualisierung beantworten.
  - mind. eine Schlussfolgerung ziehen (Fazit).
  - alle Quellen sauber angeben.
  - die gesamte Arbeit in einer Sprache verfassen (Deutsch, Englisch **xor** Französisch).
- ❖ Ihre Arbeit soll mind. die folgenden Plot-Formen enthalten (gibt je ~~0.5~~ **1** Punkt):
  - ein Histogramm
  - ein Boxplot
  - ein Scatterplot
  - ein Plot mit *facets*
  - und eine Animation

Alle Plots müssen richtig beschriftet und klar lesbar sein (**-2 Punkte** sonst).

- ❖ Die Visualisierung muss klar, schön und sinnvoll sein (2 Punkte).

- ❖ Sie müssen Ihre Story gut erzählen/aufschreiben (1 Punkt).
- ❖ Falls Sie z.B. einen nicht präparierten Datensatz wählen oder Datensätze aus unterschiedlichen Quellen zusammenfügen, können erhebliche Bereinigungsaufwände entstehen. Wenn Sie den Code sauber präsentieren und Ihre *data wrangling* Künste zeigen, können Sie dafür 3 Punkte gewinnen.

### 3 Teil II – schliessende Statistik

Damit wir die Kriterien für Teil II gut verstehen können, müssen wir zuerst ein Paar wichtige Begriffe sauber definieren bzw. kennenlernen.

- ❖ Einen statistischen Test oder eine Regressionsanalyse durchführen (5 Punkte)
  - alle Schritte durchführen
  - Nullhypothese ablehnen (oder nicht) mit einem geeigneten Signifikanzniveau
  - Konfidenzintervalle berechnen
- ❖ Alles sauber aufschreiben und dokumentieren inklusive einer geeigneten Visualisierung (3 Punkte)
- ❖ Fazit (2 Punkte)

**Bemerkung:** Sie müssen mit mir Kontaktaufnehmen, falls es nicht möglich ist eine Regression oder einen statistischen Test mit Ihren Daten durchzuführen. Wir vereinbaren in diesem Fall eine andere Lösung.

## Projektarbeit II (Spamfilter)

---

Ziel Ihrer Projektarbeit ist es, den bestmöglichen Spamfilter zu programmieren. Die Person mit dem besten bzw. zweit besten Ergebnis erhält 10 respektive 9 Punkte (bei Gleichstand gibt es 9 bzw. 8 Punkte). Falls Sie einen funktionsfähigen Spamfilter liefern (und nicht einfach eine Kopie meines Codes), erhalten Sie mindestens 5 Punkte. Ihr Code inkl. Dokumentation geben Sie auf Moodle ab.