
Datenanalyse für Naturwissenschaftler*Innen

Statistische Methoden in Theorie und Praxis

Übung 1

Aufgabe 1: SARS-COV-2 Antigenschnelltests (6P)

Die aktuellen Mindestkriterien für SARS-COV-2 Antigenschnelltests in Deutschland lauten

- Sensitivität: $> 82 \%$
- Spezifität: $> 98 \%$

Nehmen wir an, dass ein Schnelltest eine Sensitivität von 82% und eine Spezifität von 98% erfüllt. Und nehmen wir zusätzlich an, dass 3% aller Menschen in Deutschland derzeit mit SARS-COV-2 infiziert sind und damit weitere Menschen anstecken könnten. Nehmen Sie der Einfachheit halber an, Deutschland hätte eine Bevölkerung von 80 Millionen.

- Wie groß ist die Wahrscheinlichkeit, das Virus zu haben, wenn der Schnelltest negativ ist (man sagt, das Testresultat wäre dann “falsch negativ”)? (2P)
- Wie groß ist die Wahrscheinlichkeit, das Virus nicht zu haben, wenn der Schnelltest positiv ist (“falsch positiv”)? (2P)
- Stellen Sie sich vor, es nähmen 5000 Menschen an einem Konzert teil, die alle negativ getestet sind. Wie viele Personen sind trotzdem erkrankt und können andere Personen bei dem Konzert anstecken? (2P)

Aufgabe 2: Varianz und Kovarianz (9P)

- Die Varianz einer Stichprobe x vom Umfang N , mit den Elementen $\{x_1, x_2, \dots, x_N\}$, ist gegeben durch

$$V(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2,$$

wobei

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

der Mittelwert ist. Zeigen Sie, dass dies zu

$$V(x) = \overline{x^2} - \bar{x}^2$$

äquivalent ist. Mit anderen Worten: Die Varianz ergibt sich aus dem Mittelwert der Quadrate minus dem Quadrat des Mittelwertes. (3P)

- (b) Die Kovarianz der N Paare von Messobservablen $(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)$ ist gegeben durch

$$\text{cov}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

- (i) Zeigen Sie, dass dies äquivalent zu

$$\text{cov}(x, y) = \overline{xy} - \bar{x} \bar{y}$$

ist. (3P)

- (ii) Zeigen Sie, dass unter der Variablentransformation $x \rightarrow x + C$, wobei C eine beliebige Konstante ist, die Kovarianz $\text{cov}(x, y)$ erhalten ist. (3P)

Aufgabe 3: Autounfälle (11P)

Man hört oft, dass starker Regen Autounfälle verursachen kann. Unten steht die Anzahl der Autounfälle in Freiburg an einem spezifischen Tag zusammen mit der Niederschlagsmenge für diesen Tag.

Autounfälle	Niederschlagsmenge [mm]
7	0
16	10,2
4	0
12	0,4
11	2,8
9	3,0
18	7,5
14	7,6
20	15,8
24	25,5
23	24,1
35	30,4

- (a) Berechnen Sie den Mittelwert und die Standardabweichung für die Anzahl der Autounfälle und der Niederschlagsmenge pro Tag. (2.5P)
- (b) Was ist die Korrelation zwischen den beiden Größen in diesem Datensatz? (2.5P)
- (c) Zeichnen Sie die Anzahl der Autounfälle in ein Histogramm. Wählen Sie ein sinnvolles Binning. Berechnen Sie Mittelwert und Standardabweichung erneut aus den Binmittelpunkten und der Binhöhe. Was fällt Ihnen auf? (6P)

Aufgabe 4: Wechseln oder nicht? (4P)

In einer Spielshow gibt es drei Türen. Hinter einer Tür wartet ein Gewinn, hinter den beiden anderen Türen befinden sich Nieten. Nachdem eine der Kandidat*Innen eine Tür ausgewählt hat, wird eine der anderen beiden Türen geöffnet hinter der sich eine Niete befindet. Nun darf erneut zwischen den beiden übrigen Türen gewählt werden.

Sollte man bei der ursprünglichen Tür bleiben oder besser die andere Tür wählen, um die Gewinnchance zu erhöhen? Benutzen Sie explizit den Satz von Bayes, um Ihre Antwort zu begründen.