



## Übungsaufgabenblatt No. 3 zur Vorlesung Statistik für WIng

17. Mai 2024

Auf diesem Aufgabenblatt sind Präsenz- und Hausaufgaben gegeben. Die Präsenzaufgaben werden vom Übungsleiter vorgerechnet, die Hausaufgaben sind elektronisch am Tag der nächsten Übung (pünktlich !) abzugeben. Die Ergebnisse der Berechnungen sind als Excel-, die Dokumentation und Interpretation als PDF-Datei zuzusenden.

Anmerkung zur Bearbeitung:

Für die hier im folgenden zu bearbeitenden Aufgaben steht Ihnen die Datei 'Aufg3.xlsm' zur Verfügung. Bitte nehmen Sie vor Beginn der Bearbeitung die Individualisierung der Daten über Ihre Emailadresse vor.

**Abgabe der H-Aufgaben dieses Aufgabenblattes: 23.05.2024**

### **Präsenzaufgaben:**

P1: Regressionsanalyse: Berechnung der notwendigen Größen

In der Datei 'Aufg3P.xlsm', Tabelle 'P1' finden Sie einen Datensatz mit zwei stetigen Merkmalen. Führen Sie folgende Untersuchungen durch:

- Berechnen Sie die deskriptive Statistik für beide Datenspalten (Mittelwerte  $(\bar{x}, \bar{y})$ , Standardabweichungen  $(s_x, s_y)$ , Varianzen  $(s_x^2, s_y^2)$ )
- Berechnen Sie die Kovarianz  $s_{xy}$  zwischen beiden Merkmalen
- Korrelationskoeffizient zwischen  $x$  und  $y$

- Berechnen Sie die Regressionsgerade  $y = a * x + b$
- Wie lautet die Prognose für  $y$ , wenn  $x = 5$

## P2: Regressionsanalyse der Umfragedaten

Welche (wie viele ...) Zusammenhänge zwischen quantitativen Merkmalen sind in den Daten der 'Umfrage' (siehe dazu Übung 1) vermutbar (unter ignorieren der z.B. Klassifizierungen, z.B. 'männlich / weiblich') ? Seien Sie insb. vorsichtig bzgl. der Zuordnung von  $x$  (Unabhängiges Merkmal) und  $y$  ((von  $x$ ) Abhängiges Merkmal) ...

Analysieren Sie drei dieser möglichen Fragestellungen und interpretieren Sie das Ergebnis.

## Hausaufgaben:

### H1: Regressionsanalyse: Berechnung der notwendigen Größen

In der Datei 'Aufg3P.xlsm', Tabelle 'P1' finden Sie einen Datensatz mit zwei stetigen Merkmalen. Führen Sie folgende Untersuchungen durch:

- Berechnen Sie die deskriptive Statistik für beide Datenspalten (Mittelwerte  $(\bar{x}, \bar{y})$ , Standardabweichungen  $(s_x, s_y)$ , Varianzen  $(s_x^2, s_y^2)$ )
- Berechnen Sie die Kovarianz  $s_{xy}$  zwischen beiden Merkmalen
- Korrelationskoeffizient zwischen  $x$  und  $y$
- Berechnen Sie die Regressionsgerade  $y = a * x + b$
- Wie lautet die Prognose für  $y$ , wenn  $x = 5$

### H2: Regressionsanalyse für eine Studie

Ein Forscherteam hat einen Datensatz im Rahmen einer Studie durch eine Stichprobe gewonnen, aufgrund derer sie behaupten, dass 'Männer mit wenig Haaren mehr Geld verdienen'.

In der Tabelle 'P2' der o.g. Excel-Datei sind Daten dieser Studie gegeben. Hierbei ist  $x =$  'Anzahl Haare pro  $cm^2$  Kopfhaut' und  $y =$  'Monatliches Nettogehalt in Euro'.

Analysieren Sie diese Daten (Grafiken, Deskriptive Statistik, Regressionsanalyse). Schliessen Sie sich der Meinung der Studienwissenschaftler an ?

Wie viele Haare hatten die Studienteilnehmer durchschnittlich pro  $cm^2$  ?

Die Zeitschrift empfiehlt aufgrund der Studie sich eine Glatze rasieren zu lassen. Sollten Sie diesem Vorschlag - z.B. vor einem Bewerbungsgespräch - folgen ?