

Übungsblatt 5

Einführung in R
Institut der deutschen Wirtschaft

Pekka Sagner

25. Oktober & 11. November 2022

Deskriptive Statistiken und Häufigkeitstabellen

Aufgabe 1: »Mehr als nur fünf Zahlen« - deskriptive Statistiken

- a) Laden Sie den Datensatz `soep_dummy.dta` von der Kurswebseite herunter und lesen Sie diesen mit der Funktion `read_dta` aus dem Paket `haven` ein.
- b) Verschaffen Sie sich einen Überblick über den Datensatz.
- c) Bestimmen Sie das durchschnittliche Einkommen, das durchschnittliche Einkommen pro Jahr, sowie das durchschnittliche Einkommen pro Jahr und Regionstyp (`region`).
- d) Um auf Basis von Befragungsdaten Aussagen über die Gesamtpopulation zu treffen, müssen diese in der Regel gewichtet werden. Wiederholen Sie die Schritte aus c), dieses Mal für den gewichteten Durchschnitt.

Hinweis: Das Paket `spatstat.geom` enthält die Funktionen `weighted.median()`, `weighted.quantile()`, `weighted.var()`, welche für das Bestimmen weiterer gewichteter Summary Statistics hilfreich sind.

- e) Bestimmen Sie, in einem Schritt, die durchschnittliche Zahl der Kinder, den Durchschnitt des Einkommens und der Bildungsjahre.
- f) Bestimmen Sie den Median und die Standardabweichung für alle numerischen Variablen.

Aufgabe 2: »Wie oft ist zu oft?« - Häufigkeitstabellen

- a) Wie groß ist der Anteil der befragten Männer und Frauen im Jahr 2002?
- b) Wie hat sich der Anteil der befragten Männer und Frauen über die Zeit entwickelt?
- c) Wiederholen Sie a) und b) und treffen Sie Aussagen über die Gesamtpopulation.
- d) Wie groß ist der Anteil derjenigen mit einem Einkommen über 10.000 Euro in den verschiedenen Jahren? Überführen Sie das Ergebnis in eine klassische Kreuztabelle.
- e) Speichern Sie die Tabelle als `xlsx`-Datei.