

Übungen – Einheit 3:

Für die Übungen der Einheit 3 nehmen wir an, dass der SOEP-Datensatz der Grundgesamtheit entspricht. Eine einfache Möglichkeit, aus der Grundgesamtheit eine Zufallsstichprobe zu ziehen besteht darin, eine Zufallszahl zu erzeugen, die Daten nach der Zufallszahl zu sortieren, und dann die $n < N$ Erhebungseinheiten mit den kleinsten Zufallszahlen auszuwählen.

Verwenden Sie den Datensatz „SOEP_2009_alle_Variablen_mit_Zufallszahl.xlsx“. Ich habe hier eine „Zufallszahl“ (Spalte A) erzeugt. Wählen Sie immer die Personen mit den 200 bzw. 800 kleinsten Zufallszahlen aus, um eine Zufallsstichprobe zu ziehen!

(In EXCEL besteht auch die Möglichkeit, über „Daten -> Analyse -> Datenanalyse -> Stichprobenziehung -> zufällig“ eine Zufallsstichprobe zu ziehen. Ich würde aber davon abraten, das für diese Aufgabe zu tun. Zum einen zieht dann jeder Studierende eine andere Stichprobe, was den Vergleich der Ergebnisse erschwert. Zum anderen ist diese Zufallsstichprobe ein „Ziehen mit zurücklegen“, d.h. es kann vorkommen, dass eine Person mehrmals in der Stichprobe vorkommt.)

Aufgabe 3.1:

Betrachten Sie die Merkmale „ymove“ und „income“. „ymove“ gibt das Kalenderjahr an, in dem die Person in die derzeitige Wohnung gezogen ist. „income“ bezeichnet die individuellen Arbeitseinkommen. Gehen Sie von der Grundgesamtheit aus und berechnen Sie das Intervall, sodass der Mittelwert einer Zufallsstichprobe vom Umfang (a) $n=200$ bzw. (b) $n=800$ mit einer 95%-igen Wahrscheinlichkeit innerhalb des Intervalls liegt.

Berechnungen: siehe Excel.

Aufgabe 3.2:

Ziehen Sie (nach der oben beschriebenen Vorgangsweise) eine Stichprobe vom Umfang (a) 200 bzw. (b) 800 Personen. Berechnen Sie für die Merkmale „ymove“ und „income“ die Stichprobenmittelwerte! Beachten Sie, dass bei den ersten 200 bzw. 800 Beobachtungen die entsprechenden Merkmale für einige Personen fehlen. Der Stichprobenumfang reduziert sich daher etwas (d.h. $n < 200$ bzw. $n < 800$).

Nehmen Sie an, dass die Varianz der Grundgesamtheit (die Sie in Aufgabe 3.1 berechnet haben) bekannt ist (und der Varianz aus Aufgabe 3.1 entspricht). Berechnen Sie ein Konfidenzintervall für die Stichprobenmittelwerte der beiden Variablen (für die beiden Stichprobenumfänge), sodass das Intervall den tatsächlichen Mittelwert in 95% der Fälle überdeckt!

Berechnungen: siehe Excel.

Aufgabe 3.3:

Starten Sie mit den gleichen Stichproben wie in Aufgaben 3.2. Nehmen Sie nun an, dass die Varianz der Grundgesamtheit unbekannt ist. Berechnen Sie nun das 95%-Konfidenzintervall für die beiden Variablen für die beiden Stichprobenumfänge! Was ändert sich an der Vorgehensweise im Vergleich zu Aufgabe 3.2?

Berechnungen: siehe Excel.

Änderung Vorgehensweise: Da die Varianz der Grundgesamtheit nicht beobachtet wird, muss diese geschätzt werden. D.h. die unbekannte Varianz der Grundgesamtheit wird durch die Varianz der Stichprobe ersetzt. Um die damit verbundene Zunahme der Unsicherheit zu berücksichtigen, werden zur Berechnung der Intervalle die entsprechenden Perzentile der t-Verteilung (anstelle der Standard-Normalverteilung) verwendet. (Dadurch werden die Konfidenzintervalle etwas breiter.)

Aufgaben 3.4:

Berechnen Sie anstelle des 95%-Konfidenzintervalls (wie in Aufgabe 3.3) nun das 90%- und das 99%-Konfidenzintervall für die beiden Variablen (hier ist es ausreichend, nur die Stichprobe mit 200 Personen zu verwenden).

Aufgabe 3.5:

Sie sind (bei der Stichprobe von 200 Personen) daran interessiert, ob sich die durchschnittlichen Einkommen (income) signifikant von 15,800 Euro unterscheiden. Formulieren Sie eine entsprechende Null- und Alternativhypothese! Testen Sie die Hypothesen, wobei Sie eine Irrtumswahrscheinlichkeit (α) von 5% annehmen.

Berechnungen: siehe Excel.

Nullhypothese H_0 : income = 15,800

Alternativhypothese H_1 : income \neq 15800

Aufgabe 3.6:

Sie sind (bei der Stichprobe vom Umfang $n=200$) daran interessiert, ob die durchschnittlichen Einkommen (income) signifikant größer sind als 15,800 Euro. Formulieren Sie eine entsprechende Null- und Alternativhypothese! Testen Sie die Hypothesen, wobei Sie eine Irrtumswahrscheinlichkeit (α) von 5% annehmen.

Berechnungen: siehe Excel.

Nullhypothese H_0 : income \leq 15800

Alternativhypothese H_1 : income $>$ 15800