

1. Lösungen zu Kapitel 2

Übungsaufgabe 2.1

- a) *Richtig!* Im Idealfall erfüllen Daten die drei zentralen Anforderungen (i) *Objektivität*, (ii) *Zuverlässigkeit* und (iii) *Validität*.
- b) *Falsch!* Es werden - wenn möglich - persönliche Interviews mit allen Haushaltsmitgliedern geführt, die 16 Jahre oder älter sind.
- c) *Falsch!* Aus dem Sozio-oekonomischen Panel lassen sich Querschnittsdatensätze generieren. Ein Querschnittsdatensatz ist eine Stichprobe von Individuen, Haushalten, etc., die zu einem bestimmten Zeitpunkt erhoben wird. Um aus dem Sozio-oekonomischen Panel einen Querschnittsdatensatz zu generieren, ist lediglich die Beschränkung auf ein Befragungsjahr notwendig.

Übungsaufgabe 2.2

Das Schlüsselverzeichnis, welches zusammen mit dem Campus-File des Mikrozensus herunter geladen werden kann, erläutert die im Mikrozensus enthaltenen Variablen. Die Variable „EF539“ beinhaltet die Haushaltsnettoeinkommen. Dem Schlüsselverzeichnis kann man darüber hinaus entnehmen, dass der Wert 50 bedeutet, dass „mindestens ein Haushaltsmitglied ist in seiner Haupttätigkeit selbständiger Landwirt“ und dass der Wert 99 für fehlende Angaben steht. (Für Berechnungen mit Stata sollten diese Werte auf Missing (.) gesetzt werden.)

Die Haushaltsnettoeinkommen sind im Mikrozensus in Kategorien angegeben. Folglich ist die Variable diskret und kann graphisch durch ein Histogramm beschrieben werden. Der Stata-Befehl hierfür lautet „`histogramm`“. Beachtet werden muss, dass das Haushaltsnettoeinkommen für jedes befragte Haushaltsmitglied angegeben ist. Wir möchten das Haushaltseinkommen für unsere Berechnungen jedoch nur einmal pro Haushalt berücksichtigen. Die relativen Häufigkeiten der Nettohaushaltseinkommen sind in Abbildung 1.1 dargestellt.

Zur Berechnung der Armutsquote benötigen wir zunächst das mittlere Haushaltseinkommen (Median). Dieses erhalten wir über den Befehl „`sum ef539, detail`“. Die Armutsquote berechnet sich als Anteil der Haushalte,

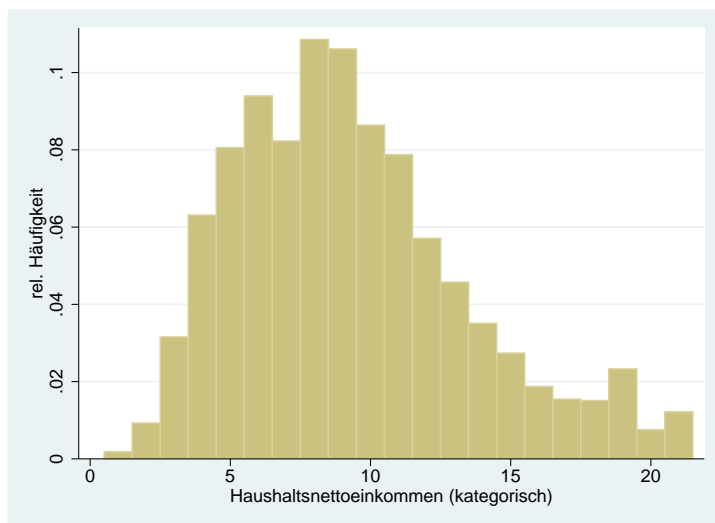


Abbildung 1.1. Wahrscheinlichkeitsfunktion der Nettoeinkommen.

die weniger als die Hälfte des mittleren Haushaltseinkommen zur Verfügung haben. Für das Jahr 1998 erhalten wir eine Armutsquote von 10,6%.

Übungsaufgabe 2.3

$$\begin{aligned}
 E(\hat{\sigma}_X^2) &= E\left(\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2\right) \\
 &= \frac{1}{N-1} E\left((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_N - \bar{X})^2\right) \\
 &= \frac{1}{N-1} [E(X_1 - \bar{X})^2 + E(X_2 - \bar{X})^2 + \dots + E(X_N - \bar{X})^2] \\
 &= \frac{N}{N-1} \cdot \sigma_X^2
 \end{aligned}$$

Übungsaufgabe 2.4

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{N} \sum_{i=1}^N (X_i)\right) = \frac{1}{N^2} \sum_{i=1}^N (\text{Var}(X_i)) = \frac{1}{N^2} \cdot N \cdot \sigma_X^2 = \frac{\sigma_X^2}{N}$$

Übungsaufgabe 2.5

$$\begin{aligned}
 X &= 1, 2, \dots, N \\
 P(x) &= \frac{1}{N} \\
 E(X) &= \frac{1}{N} \sum_{i=1}^N X_i = \frac{N+1}{2} \\
 Var(X) &= \frac{1}{N-1} \sum_{i=1}^N \left(X_i - \frac{N+1}{2} \right)^2
 \end{aligned}$$

Übungsaufgabe 2.6

Zunächst soll gezeigt werden, dass sowohl \bar{X} , das Stichprobenmittel der gesamten Stichprobe mit $N = 100$, als auch \tilde{X} , das Stichprobenmittel das auf Grundlage von 50 Beobachtungen errechnet wurde, unverzerrt sind:

$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{1}{100} \sum_{i=1}^{100} X_i\right) \\
 &= \frac{1}{100} E(X_1 + X_2 + \dots + X_{100}) = \frac{1}{100} (E(X_1) + E(X_2) + \dots + E(X_{100})) \\
 &= \frac{1}{100} \cdot 100 \cdot \mu_X = \mu_X
 \end{aligned}$$

$$\begin{aligned}
 E(\tilde{X}) &= E\left(\frac{1}{50} \sum_{i=1}^{50} X_i\right) \\
 &= \frac{1}{50} E(X_1 + X_2 + \dots + X_{50}) = \frac{1}{50} (E(X_1) + E(X_2) + \dots + E(X_{50})) \\
 &= \frac{1}{50} \cdot 50 \cdot \mu_X = \mu_X
 \end{aligned}$$

Die Varianz der Stichprobenmittel ergibt sich als:

$$Var(\bar{X}) = \frac{1}{100} \sigma_X^2$$

$$Var(\tilde{X}) = \frac{1}{50} \sigma_X^2$$

Beide Schätzer, \bar{X} und \tilde{X} , sind unverzerrte Schätzer, jedoch hat der Schätzer des Stichprobenmittels, der auf der gesamten Stichprobe mit $N = 100$

beruht, eine geringere Varianz. Dies macht \bar{X} zu einem effizienteren Schätzer für μ_x . Es ist folglich nicht lohnenswert auf vorhandene Informationen zu verzichten.

Übungsaufgabe 2.7

Eine Irrtumswahrscheinlichkeit von 0% ist sinnlos, da man ein Konfidenzintervall konstruieren würde, das alle möglichen Werte überdeckt. Dies bedeutet, dass man jede mögliche Nullhypothese als mit der tatsächlichen Ausprägung kompatibel ansehen würde und niemals die Nullhypothese verwerfen würde. Die Wahrscheinlichkeit, die Nullhypothese fälschlicherweise zu verwerfen, beträgt in diesem Fall zwar 0%, allerdings hat der Test keinerlei Aussagekraft und die Trennschärfe geht vollends verloren.

Übungsaufgabe 2.8

a) Der t-Wert eines Zweistichproben-t-Test berechnet sich wie folgt:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}}}$$

mit

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Somit ergibt sich

$$t = \frac{1.275 - 816}{\sqrt{\frac{79.729,73}{100} + \frac{79.729,73}{50}}} = \frac{459}{48,907} = 9,385.$$

Der Tabelle 9.2 können die kritischen Werte der t-Verteilung entnommen werden. $t = 9,385$ ist größer als der kritische Wert der t-Verteilung, unabhängig davon ob man eine Irrtumswahrscheinlichkeit von 1%, 5% oder 10% zugrundelegt. So ist z.B. der kritische Wert der t-Verteilung eines zweiseitigen Tests mit einer Irrtumswahrscheinlichkeit von 5% und 148 Freiheitsgraden in etwa 1,960. Die Differenz in den Gehältern von Männern und Frauen ist signifikant und somit ist die Wahrscheinlichkeit für Frauen ein durchschnittliches Gehalt von 816 € zu beobachten, obwohl es in Wirklichkeit auch 1.275 € beträgt (oder umgekehrt), sehr gering. Wir können folglich die Nullhypothese mit einer Irrtumswahrscheinlichkeit von unter 5% verwerfen.

- b) Auf Grundlage der vorliegenden Daten lassen sich keine Rückschlüsse über die Diskriminierung von Frauen in dem Unternehmen ziehen. Diskriminierung läge vor, wenn Männer und Frauen bei gleicher Leistung unterschiedlich bezahlt würden. Es ist jedoch möglich, dass die Unterschiede in den Löhnen auf unterschiedliche Stellungen mit entsprechend unterschiedlicher Bezahlung in dem Unternehmen zurückzuführen sind.