

☒ Folgende Stichprobe beinhaltet 10 Beobachtungen² über (Haushalts-)Einkommen von Personen in Österreich und deren Stichprobengewichte³:

	1	2	3	4	5	6	7	8	9	10
eqIncome	22656	6649	16691	19494	9006	17314	8367	17192	29408	26314
weights	868	616	493	642	556	459	493	655	537	616

Die Beurteilung von Ungleichheit von Einkommen (Einkommensschere) ist von zentraler Bedeutung für die Politik. Ein oft verwendeter und beliebter Indikator, die Ungleichheit von oberen und unteren Einkommen zu beurteilen, ist der sogenannte Quintile-Share-Ratio (QSR).

Sei $q_{0.2}$ und $q_{0.8}$ das gewichtete 20% und 80% Quantile von den Stichprobenwerten $x_i, i = 1, \dots, n$ und Gewichten $w_i, i = 1, \dots, n$.

Mit $I_{\leq q_{0.2}} := \{i \in \{1, \dots, n\} : x_i \leq q_{0.2}\}$ und $I_{\geq q_{0.8}} := \{i \in \{1, \dots, n\} : x_i \geq q_{0.8}\}$ lautet die Schätzung des QSR:

$$\widehat{QSR} := \frac{\sum_{i \in I_{\geq q_{0.8}}} w_i x_i}{\sum_{i \in I_{\leq q_{0.2}}} w_i x_i}$$

- ☒ (a) Schätzen Sie den QSR mit ihren Daten⁴.
- ☒ (b) Betrachten Sie die Formel und geben Sie eine Interpretation des QSR.
- ☒ (c) Ist der QSR empfindlich gegenüber Ausreißer? Wenn ja/nein: Warum?

(Lösungsblatt: \widehat{QSR})

(2.5 + 1.5 + 1.5)

2. Pärchen aufgepasst! So lautete die Schlagzeile zeitgerecht vor Weihnachten von CNN und in einer großen deutschen Zeitung basierend auf einer empirischen Studie. Die gefährliche Zeit für Beziehungen beginnt laut dieser Studie zwei Wochen vor Weihnachten - ab 10. Dezember. Grundlage dieser Studie ist die Änderung des Beziehungsstatus von 10000 Facebook-Mitgliedern in Facebook.

Die wenigsten Trennungen gibt es laut deren Studie übrigens zu Weihnachten, ein solcher Schritt wird laut CNN als *zu grausam* bewertet.

Von 10000 Facebook-Mitgliedern (wir wollen annehmen, dass sich unter den 10000 Mitgliedern 5000 Pärchen befinden) haben sich 59 Paare in der 'beziehungstechnisch schlimmsten' Woche vor Weihnachten getrennt (z. B. von *verheiratet* oder *in einer Beziehung zu Single* oder *es ist kompliziert*). Die Studie berichtet, dass im Gegensatz dazu es in der beziehungstechnisch besten Zeit - in der ersten Juliwoche - sich nur 32 Paare getrennt haben.

- (a) Schätzen Sie den Anteil der Pärchen die betroffen waren und bestimmen Sie die 95%-Konfidenzintervalle zB mit Hilfe einer Approximation durch eine Normalverteilung. Überdecken sich die Konfidenzintervalle?
- ☒ (b) Wenn Sie die Überschrift mit dem Inhalt der Studie vergleichen, welche Kritikpunkte fallen Ihnen ein?

(Lösungsblatt: *Überdeckung ja/nein*)

(3.5 + 1.5)

3. Eine Stichprobe unter den bestbezahlten Fußballspielern wurde im Jahr 2010 erhoben und unter anderem das Jahreseinkommen (in Mio. Pfund) erfragt.

☒ (a) Mit den log-normalverteilten Daten: Überprüfen Sie grafisch mit einem Wahrscheinlichkeitsnetz, ob die logarithmierten Jahreseinkommen der Spieler mit Herkunftsland Europa normalverteilt sind. Schätzen Sie das Mittel und die Standardabweichung der log-Jahreseinkommen dieser Spieler unter Verwendung des Wahrscheinlichkeitsnetzes. (3)

²Seriöse Aussagen können erst bei viel größeren Stichproben getätigt werden, aber zur einfacheren Berechnung mit dem Taschenrechner während der Prüfung wurde nur ein kleines Subsample gezogen.

³Ein Stichprobengewicht repräsentiert, vereinfacht gesprochen, die erwartete Anzahl von Personen mit gleichen Eigenschaften.

⁴Da jede Person/Haushalt eine unterschiedliche Auswahlwahrscheinlichkeit hat, müssten streng genommen auch die Quantile in gewichteter Form berechnet werden. Da dies in der Vorlesung und Übung nicht besprochen wurde, dürfen Sie die Schätzung der Quantile ungewichtet durchführen.

	Name	Einkommen	Herkunft
1	Ronaldo	11.3	Europa
2	Ibrahimovic	10.4	Europa
3	Eto-o	9.1	Afrika
4	Messi	9.1	Amerika
5	Kaka	8.7	Amerika
6	Adebayor	7.4	Afrika
7	Bezema	7.4	Europa
8	Tevez	7.0	Amerika
9	Ronaldinho	6.5	Amerika
10	Gerrard	6.5	Europa
11	Henry	6.5	Europa
12	Terry	6.5	Europa
13	Lampard	6.5	Europa
14	Xavi	6.5	Europa
15	Alves	6.1	Europa
16	Ballack	5.6	Europa
17	Toure	5.6	Europa
18	Raul	5.6	Europa
19	Ferdinand	5.6	Europa
20	Kanoute	5.2	Afrika
21	Robinho	5.2	Amerika
22	Casillas	5.2	Europa
23	Rooney	5.2	Europa
24	Valdez	5.2	Europa
25	Deco	5.2	Europa
26	Drogba	4.8	Afrika

Überprüfen Sie die logarithmierten Jahreseinkommen (unter Normalverteilungsannahme), ob für die amerikanischen und afrikanischen Fußballer die Varianzen übereinstimmen (Signifikanzniveau $\alpha = 0.1$). (2)

Testen Sie (unter Normalverteilungsannahme der logarithmierten Werte) für die Einkommen der europäischen Spieler, ob das mittlere Jahreseinkommen kleiner als 7.5 Mio Pfund ist (Signifikanzniveau $\alpha = 0.1$). (2)

(Lösungsblatt: Wert der Teststatistik aus c, Wert der Teststatistik aus d)

Mit Daten von Statistics Canada wurde das Ansehen (Prestige, abhängige Variable) von Personen bzgl. verschiedener Berufsgruppen anhand von den unabhängigen Variablen - Einkommen (in Kanadischen Dollars), Ausbildungszeit (in Jahren) und Frauenanteil - modelliert. Das Resultat (der Standardoutput) dieser Regression ist hier ersichtlich:

Call:

`lm(formula = prestige ~ income + education + women, data = Prestige)`

Residuals:

Min	1Q	Median	3Q	Max
-17.172	-4.561	0.291	4.055	14.714

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.309301	2.848334	-2.92	0.0044 **
income	0.001600	0.000288	5.55	2.4e-07 ***
education	4.100863	0.369734	11.09	< 2e-16 ***
women	0.029057	0.027694	1.05	0.2967

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.27 on 98 degrees of freedom
Multiple R-squared: 0.846, Adjusted R-squared: 0.841
F-statistic: 179 on 3 and 98 DF, p-value: <2e-16

Welche der folgenden Aussagen sind korrekt (Signifikanzniveau 5%)?

(a) Der Wert der Teststatistiken (t value) ist sehr gross (im Ablehnungsbereich) bzw der

p -Wert ist sehr klein (signifikant), daher tragen die Variablen income und education nicht zur Erklärung des Modells signifikant bei.

- + (b) Ein um 1000 Dollar höheres Durchschnittseinkommen bedeutet ungefähr 1.3 mal mehr Prestige.
- (c) Die Variable women trägt am meisten zur Erklärung von Prestige bei.
- (d) Der hohe Wert des Bestimmtheitsmasses (multiple R-square) bedeutet, dass das beste Modell gefunden wurde.
- (e) Ein zusätzliches Ausbildungsjahr bringt mehr Prestige als ein um 2000 Dollar höheres Jahreseinkommen.

(2.5)

(Lösungsblatt eintragen! Wenn zutreffend, dann ein Plus im Lösungsblatt zur Frage eintragen, wenn nicht zutreffend ein Minus eintragen.)