

1. 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	15424	21375	16352	9238	54623	11852	11740	22489	8981	21868
weights	537	521	692	485	494	642	494	459	538	508

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. Ist er ausserdem empfindlich gegenüber Ausreisser?

(Lösungsblatt: \widehat{QSR})

(2.5 + 2)

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 haben sich 62 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 30 Paare getrennt haben.

- (a) Schätzen Sie die Anteile Personen die von einer Trennung 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. Im Rahmen der PISA-Studie wurde in den einzelnen Ländern die durchschnittliche Klassengröße ermittelt. Wir bilden drei Gruppen, wobei die Länder der ersten Gruppe eine kleine, die Länder der zweiten Gruppe eine mittlere und die Länder der Gruppe Klasse eine hohe Klassengröße besitzen. Aus jeder Gruppe wurden jeweils 4 Länder ausgewählt. Die folgende Tabelle zeigt die Ergebnisse im Bereich naturwissenschaftliche Grundbildung dieser Länder:

- a) Testen Sie (unter Normalverteilungsannahme) für die Gruppe 1, ob die mittlere Punktzahl signifikant kleiner als 13⁵⁰⁰ ist (Signifikanzniveau $\alpha = 0.05$). (2)

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.

	1	2	3	4
klein	500	481	479	496
mittel	376	550	553	424
hoch	482	517	527	495

- b) Nehmen Sie an, dass die Daten in den einzelnen Gruppen normalverteilt sind mit der gleichen Varianz. Stimmen die mittleren Punktezahlen in den drei Gruppen überein (Signifikanzniveau $\alpha = 0.05$)? (3)
- c) Zeichnen Sie einen Boxplot von allen Ergebnissen als einzige Messreihe aufgefasst (der Tabelle oben). Geben Sie eine kurze Interpretation ihres Boxplots. (3)

(Lösungsblatt: Wert der Teststatistik aus b)

4. 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
-21.939  -4.995   0.002   4.898  15.204
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.921879   3.249760  -0.28   0.78
income       0.001077   0.000213   5.05  2e-06 ***
education    3.824044   0.366509  10.43 <2e-16 ***
women       -0.016460   0.028049  -0.59   0.56
```

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

Residual standard error: 7.88 on 98 degrees of freedom

Multiple R-squared: 0.821, Adjusted R-squared: 0.815

F-statistic: 150 on 3 and 98 DF, p-value: <2e-16

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

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

(2.5)

(Lösungsblatt eintragen! Wenn zutreffend, dann ein Plus im Lösungsblatt zur Frage eintragen, wenn nicht zutreffend ein Minus eintragen. Ein Punkt Abzug für das Beispiel pro falsche Antwort.)