Aufgabe 1:

Betrachten Sie wieder den Datensatz `teengamb` (vgl. Aufgabe 1, Blatt 3) und das folgende Modell:

```
library(faraway)
mod <- lm(gamble ~ income + status + verbal, data = teengamb)
```

(a) Berechnen Sie alle sequentiellen Residuenquadratsummen dieses Modells in **R** und stellen Sie diese in einer Tabelle dar, welche in der ersten Spalte die $R(\cdot|\cdot)$-Notation der jeweiligen Quadratsumme angibt, in der zweiten den berechneten Wert und in der dritten die kumulative Summe enthält.

(b) Überprüfen Sie, dass die Summe der sequentiellen Quadratsummen der Modellstreuung (SSM) des vollen Modells (`mod`) entspricht und berechnen Sie die Gesamtstreuung (SST).

(c) Geben Sie die zur Hypothese $\beta_{status} = \beta_{verbal} = 0$ zugehörige partielle Quadratsumme zunächst allgemein in $R(\cdot|\cdot)$-Notation an. Wie hängt diese mit den Residuen des vollen Modells und des Modells unter der Null-Hypothese zusammen. Was ist der Zusammenhang zum F-Test?

Aufgabe 2:

Die Datei `buli1314.Rdata` enthält Informationen zu den Bundesliga-Platzierungen der Saison 2013/2014. Wir interessieren uns primär für den Zusammenhang zwischen erreichten Punkten (am Ende der Saisson) in Abhängigkeit des Budgets einer Mannschaft. Weitere Einflussgrößen von Interesse sind die Punkte der Vorsaison sowie der Anteil der Punkte aus Auswärtsspielen der Vorsaison.

(a) Schätzen Sie zwei multiple Regressionsmodelle mit den oben genannten Einflussgrößen. Verwenden Sie für das erste Modell die beobachteten Werte und standardisieren Sie diese für das zweite Modell.

(b) Stellen Sie die geschätzten Koeffizienten graphisch dar und interpetieren Sie diese. Welche Einflussgröße hat den stärksten Effekt?

(c) Geben Sie $X'X$ und $X'Y$ für obiges Beispiel in allgemeiner Form an.

(d) Können die Koeffizientenschätzer für die Einflussgrößen in diesem Fall als Korrelationen zwischen der jeweiligen Einflussgröße und der Zielgröße interpretiert werden?

Aufgabe 3:

Der aus der Vorlesung bekannte Datensatz `lesen` untersucht den Zusammenhang zwischen dem Verhalten von Grundschülern und der Anzahl Fehler bei einem Lesetest:

`Fehlerzahl`	Anzahl Fehler beim Lesetest
`sex`	Indikator für männlich (= 1 für männlich, 0 sonst)
`Jahrgang`	Indikator für Klassenstufe (=1 für 3. Klasse, 0 für 4.Klasse)
`Lesezeitmin`	Lesezeit in der Schule
`WieoftLesen`	Wie oft wird sonst gelesen

Sie möchten die `Fehlerzahl` nun in Abhängigkeit der Variable `WieOftLesen` modellieren, wobei `WieOftLesen` als ordinale/kategoriale Variable behandelt werden soll.

(a) Betrachten Sie folgende Kodierungsmöglichkeiten für ordinale Variablen:

- Mittelwertsmodell
- Effektkodierung
- Referenzkodierung

Stellen Sie jeweils die sich aus der Kodierung ergebenden Designmatrizen schematisch dar.

(b) Passen Sie für jede der angegebenen Kodierungsmöglichkeiten ein lineares Modell an. Interpretieren Sie jeweils die Parameterschätzer.

(c) Testen Sie die Hypothese, dass kein linearer Zusammenhang zwischen `Fehlerzahl` und `WieOftLesen` vorliegt mit einer geeigneten linearen Hypothese.