

Aufgabe 1:

Wir betrachten ein lineares Regressionsmodell mit zwei binären und einer metrischen Einflussgröße. Mit einer singulären Designmatrix X lässt sich das Modell folgendermaßen schreiben:

$$Y = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & x_1 \\ 1 & 1 & 0 & 1 & 0 & x_2 \\ 1 & 1 & 0 & 0 & 1 & x_3 \\ 1 & 0 & 1 & 1 & 0 & x_4 \\ 1 & 0 & 1 & 0 & 1 & x_5 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_1 \\ \gamma_2 \\ \beta \end{pmatrix} + \epsilon.$$

- (a) Interpretieren Sie das Modell.
- (b) Geben Sie das Modell in Referenzkodierung (jeweils die zweite Kategorie als Referenzkategorie) und in Effektkodierung (zwei Varianten: $\sum_{i=1}^n \tau_i = \sum_{i=1}^n \gamma_j = 0$ bzw. $\sum_{i=1}^n n_i \tau_i = \sum_{i=1}^n n_j \gamma_j = 0$) an und interpretieren Sie die Parameter.
- (c) Wie lautet die Designmatrix X , wenn Sie alle Zweier-Interaktionen hinzufügen?

Aufgabe 2:

Datensatz 4¹ auf der Übungshomepage enthält Daten einer Studie zur Gewichtszunahme (weightgain) von Ratten, die anhand der Ernährung in vier Gruppen eingeteilt werden können: Proteinmenge (wenig, viel) (amount (Low, High)) und Quelle der Proteinzufuhr (Fleisch, Getreide) (source (Beef, Cereal)). Führen Sie in **R** eine (zweifaktorielle) Varianzanalyse durch und interpretieren Sie die Parameterschätzungen.

Aufgabe 3 (Nach einer früheren Klausuraufgabe):

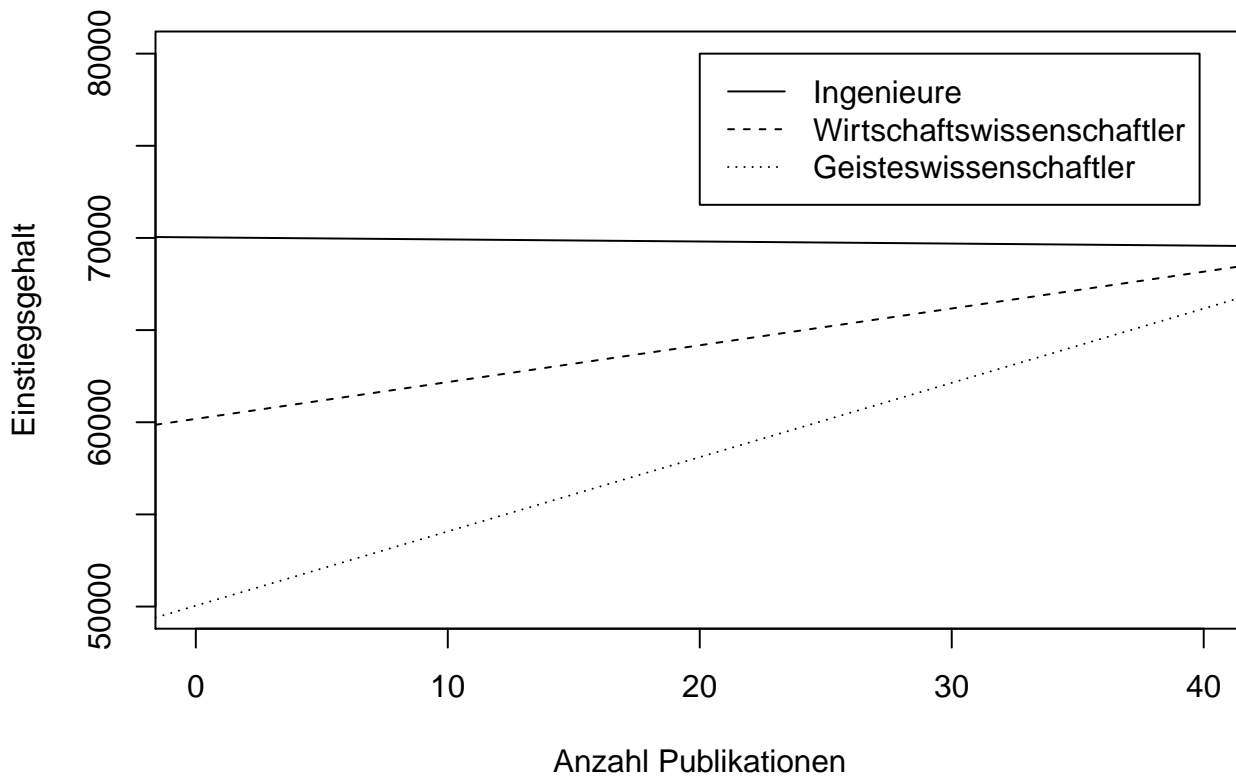
Sie möchten untersuchen, ob die Anzahl an Publikationen (X) für drei verschiedene Gruppen von Doktoranden (Gruppe 1: Ingenieure, Gruppe 2: Wirtschaftswissenschaftler, Gruppe 3: Geisteswissenschaftler) das Einstiegsjahresgehalt Y beeinflusst. Sie befragen also in jeder der drei Gruppen jeweils 20 frische Doktoren nach Ihrem Einstiegsgehalt.

- (a) Stellen Sie ein geeignetes Regressionsmodell unter Einbeziehung von Interaktionen auf. Nehmen Sie für die kategoriale Variable, die das studierte Fach beschreibt, eine Referenzkodierung vor, wobei Sie Gruppe 1 (Ingenieure) als Referenzkategorie angeben. Geben Sie hierbei auch die Designmatrix X an, die alle Prädiktorvariablen enthält.
- (b) Schreiben Sie für jede der drei Gruppen allgemein die Gleichung auf, nach der sich die prognostizierten Einstiegsgehälter berechnen lassen und fassen Sie jeweils den Teil, der die Steigung beschreibt, zusammen.
- (c) Nehmen Sie an, Sie erhalten folgende Schätzungen für die Koeffizienten:

¹Aus Everitt, Hothorn: „A Handbook of Statistical Analyses using R“, Chapman & Hall/CRC: 2006

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70031.1789	2033.7050	34.44	0.0000
as.factor(Gruppe)2	-9848.2987	2938.7922	-3.35	0.0015
as.factor(Gruppe)3	-19984.4464	2794.0068	-7.15	0.0000
X	-11.2147	82.8739	-0.14	0.8929
as.factor(Gruppe)2:X	210.8903	131.4754	1.60	0.1145
as.factor(Gruppe)3:X	414.2095	115.6276	3.58	0.0007

Erklären Sie, wie Sie anhand der Parameterschätzungen auf die Koeffizienten der Gerade für die Geisteswissenschaftler (siehe Abbildung) kommen.



- (d) Interpretieren Sie die Parameter der drei Geradengleichungen der in der Abbildung dargestellten Geraden.

Aufgabe 4:

Betrachten Sie wieder den Datensatz `teengamb` (Blatt 3/Aufgabe 1). Bisher haben wir nur Modelle ohne die Geschlechtsvariable betrachtet.

- (a) Nehmen Sie nun die Variable `sex` zusätzlich in das Modell mit auf. Interpretieren Sie den Effekt des Geschlechts auf das Glücksspielverhalten. Welche Veränderung der Glücksspielausgaben würden Sie bei einer weiblichen Jugendlichen mit einem um 10 Pfund/Woche erhöhten Einkommen im Vergleich zu einem männlichen Jugendlichen mit gleichem Status und gleicher Punktzahl beim verbalen Test-Score erwarten?
- (b) Passen Sie nun ein Modell an, in welchem alle zweifach Interaktionen der metrischen Variablen mit dem Geschlechtsindikator enthalten sind. Testen Sie anschließend $H_0 : \beta_{status} = \beta_{verbal} = \beta_{status:sex} = \beta_{verbal:sex} = 0$.
- (c) Fitten Sie nun das reduzierte Modell und stellen Sie den Effekt von `income` und `sex` graphisch dar. Diskutieren Sie den Einfluss dieser Variablen auf das Glücksspielverhalten. Wie unterscheiden sich die Glücksspielausgaben zwischen einem männlichen Jugendlichen mit 15 Pfund Einkommen und einer weiblichen Jugendlichen mit 10 Pfund Einkommen.
Hinweis: Betrachten Sie die Funktion `effect` im `effects` package.