

LEHRSTUHL FÜR STATISTIK UND ÖKONOMETRIE  
ÜBUNG ZUR DATENANALYSE

**Aufgabenserie 4: Varianzanalyse, Kovarianzanalyse**

**Aufgabe 8**

Es soll die Auswirkung von  $I = 3$  verschiedenen Unterrichtsmethoden A, B und C auf den Lernerfolg ( $Y$ ) untersucht werden, wobei letzterer über die bei einem Abschlusstest erzielten Punkte gemessen wird. 12 Teilnehmer werden per Zufall so auf drei Gruppen aufgeteilt, dass jede Gruppe genau  $J = 4$  Personen aufweist. Es ergaben sich nach den Unterrichtsmethoden getrennt folgende Resultate:<sup>1</sup>

A	B	C
16	16	2
18	12	10
20	10	9
15	12	10

1. Welche Annahmen treffen Sie, wenn Sie eine einfaktorielle Varianzanalyse zur Untersuchung der Fragestellung durchführen?
2. Stellen Sie das Modell in Effektkodierung auf, und schätzen Sie die Parameter nach der OLS-Methode. Welcher Verteilung folgt der Vektor der Parameterschätzer?
3. Sie sind der Überzeugung, dass die drei Unterrichtsmethoden nicht alle gleich effektiv sind. Überprüfen Sie Ihre Vermutung bei einer Irrtumswahrscheinlichkeit  $\alpha$  von 5%.
4. Testen Sie bei einer Irrtumswahrscheinlichkeit  $\alpha$  von 5% die Hypothese, dass die Testergebnisse von Schülern, die nach der Methode A unterrichtet wurden, im Mittel doppelt so hoch sind wie diejenigen von Schülern, bei denen Methode C zur Anwendung kam.

**Aufgabe 9**

Online-Versandhändler wie amazon.com bieten ihren Kunden die Möglichkeit, nach ihrem Einkauf eine Produktbewertung abzugeben. Es soll der Einfluss bestimmter Produkteigenschaften auf die Abgabe von Produktbewertungen untersucht werden.

Zunächst interessiert der Einfluss des Preissegments (Faktor  $A$ ) des Produkts auf dessen Bewertungsquote (d.h. den Anteil der Produktkäufe, bei denen anschließend eine Bewertung abgegeben wurde). Die Datei `Produkte.csv` enthält für je 100 zufällig gewählte Produkte aus drei unterschiedlichen Preissegmenten die Bewertungsquote im Jahr 2014.

1. Überprüfen Sie in **R** mithilfe einer einfaktoriellen Varianzanalyse, ob das Preissegment einen Einfluss auf die Bewertungsquote hat.

---

<sup>1</sup>Vgl. Stier: Empirische Forschungsmethoden, Springer, Berlin, 1996, S. 272.

2. Prüfen Sie, ob tatsächlich mit den Annahmen der Varianzhomogenität und der Normalverteilung der Residuen gearbeitet werden kann.
3. Sie möchten bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.01$  testen, ob sich die Bewertungsquote von Segment zu Segment um jeweils fünf Prozentpunkte erhöht. Sie wollen also folgende Hypothesen testen:

- $H_0 : \mu_1 - \mu_2 = -0.05$  gegen  $H_1 : \mu_1 - \mu_2 \neq -0.05$ ;
- $H_0 : \mu_2 - \mu_3 = -0.05$  gegen  $H_1 : \mu_2 - \mu_3 \neq -0.05$ ;
- $H_0 : \mu_1 - \mu_3 = -0.10$  gegen  $H_1 : \mu_1 - \mu_3 \neq -0.10$ .

Erläutern Sie das Problem, welches sich hierbei ergibt. Testen Sie die Hypothesen simultan mithilfe einer geeigneten Methode.

Als zusätzlicher Einflussfaktor interessiert die Art der Verwendung (Faktor  $B$ ) des Produkts, d.h., ob es sich bei dem Produkt um ein Gebrauchs- ( $b_1$ ) oder ein Unterhaltungsgut ( $b_2$ ) handelt. Die Effektdarstellung des entsprechenden Modells lautet wie folgt:

$$Y_{ijk} = \mu_0 + a_i + b_j + (ab)_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2).$$

4. Überprüfen Sie, ob ein sog. *balanciertes Design* vorliegt, d.h. ob für jede Zelle gleich viele Beobachtungen vorhanden sind.
5. Interpretieren Sie die Parameter des Modells.
6. Erstellen Sie in R die Designmatrix  $\mathbf{X}$  für das Modell  $\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\epsilon}$  mit  $\boldsymbol{\mu} = (\mu_0, a_1, a_2, b_1, (ab)_{11}, (ab)_{21})'$ .
7. Berechnen Sie in R den Wert des OLS-Schätzers  $\hat{\boldsymbol{\mu}}$  für vorliegende Stichprobe. Wie können bspw.  $\hat{\mu}_0$  und  $\hat{a}_1$  alternativ in R berechnet werden?
8. Stellen Sie die mithilfe von Boxplots die Verteilung der Bewertungsquote getrennt nach Preissegment und Verwendungsart grafisch dar, und interpretieren Sie die Grafiken in Bezug auf die Haupteffekte des Modells.
9. Testen Sie jeweils bei  $\alpha = 0.1$  die Hypothesen,
  - (a) dass keine Interaktionseffekte zwischen Preissegment und Verwendungsart vorliegen;
  - (b) dass die einzelnen Faktoren keinen Einfluss haben.
10. Veranschaulichen Sie das Ergebnis aus Teilaufgabe 9a mithilfe einer geeigneten Grafik.
11. Welche metrische Kovariable könnte sinnvollerweise in das vorliegende Modell aufgenommen werden?

## Aufgabe 10

Sie wollen mithilfe der Anzahl der Produktverkäufe im Jahr 2013 ( $Z$ ) die Untersuchung des Einflusses von Preissegment und Verwendungsart auf die Bewertungsquote präzisieren. Die vollständigen Beobachtungen der 300 Produkte aus der vorherigen Aufgabe liegen in der Datei `Produkte2.csv` vor.

1. Untersuchen Sie den Einfluss der beiden Faktoren auf die Bewertungsquote unter Berücksichtigung der Kovariablen  $Z$ .
2. Testen Sie in `R`, ob die Interaktion zwischen Preissegment und Anzahl der Verkäufe signifikant ist ( $\alpha = 0.05$ ).