

Angewandte Statistik für Biometrie

(Abgabe: Di., 05.07.2011, vor den Übungen)

1. In einer Studie soll untersucht werden wieviel Zeit die Studenten morgens benötigen um vom Wohnort zur Uni zu kommen und welche Faktoren die Dauer der Anreise beeinflussen.

- (a) Welche Größen könnten die Zeit beeinflussen und sollten im Modell berücksichtigt (und untersucht) werden? Finde mindestens zehn und begründe kurz warum die Größe einen Einfluss haben könnte.
- (b) Es soll lediglich der Zusammenhang zwischen der Zeit und dem Wohnort der Studenten (Eselsberg, Weststadt, Innenstadt, Oststadt, Lehr) einerseits, sowie der Wahl des Verkehrsmittels (Auto, Bus, Fahrrad) andererseits, erforscht werden. Dazu geben 42 Freiwillige an 10 Tagen die Zeit an die sie für die Fahrt aus den Ortsteilen zur Uni benötigt haben, dabei werden die Startpunkte und Verkehrsmittel den Freiwilligen zufällig zugeteilt.

Beschreibe die Studie durch ein geeignetes Modell. Stelle eine Vermutung an ob Wechselwirkungen zwischen den beiden Einflussfaktoren auftreten. Kann das Modell auch verwendet werden wenn statt den Ortsteilen die Entfernung in km betrachtet wird? Kann es verwendet werden wenn die Studenten Startpunkt und Verkehrsmittel selbst wählen?

- (c) Für die Studie fanden sich lediglich fünf Studenten s_1, \dots, s_5 , die alle im selben Wohnheim wohnen. Allerdings sind sie lediglich eine Woche lang bereit an der Studie teilzunehmen (Montag – Freitag). Sie starten alle gleichzeitig um 9 Uhr, daher kann lediglich eine Anreise pro Tag durchgeführt werden. Es stehen die Verkehrsmittel Auto (A), Bus (B), Fahrrad (F) und Roller (R) zur Verfügung, ausserdem besteht die Möglichkeit zu Fuß zu gehen (P). Stelle ein Lateinisches Quadrat auf um die fünf Anreisemöglichkeiten an den verschiedenen Tagen auf die Studenten aufzuteilen. Dabei sollte s_1 am Montag nicht mit dem Auto fahren und die Diagonale des Quadrats sollte mindestens zwei verschiedene Verkehrsmittel enthalten.

(4 Punkte)

2. Der Datensatz `verkauf.dat` von der Homepage enthält die Verkaufszahlen von Sonnencreme eines Supermarktes. Dabei wurden in drei Regalfächern (oben, mitte, unten) die Creme jeweils in zwei verschiedenfarbigen Tuben angeboten (gelb oder blau). Die Zahlen wurden an fünf aufeinanderfolgenden Freitagen erhoben.

Führe eine Varianzanalyse durch um den Einfluss der Faktoren „Position“ und „Farbe“ auf die Verkaufszahlen zu analysieren.

- (a) Existieren signifikante Wechselwirkungen zwischen den Faktoren?
- (b) Haben die Faktoren einen signifikanten Einfluss auf die Verkaufszahlen?

Teste jeweils zum Niveau $\alpha = 5\%$.

(4 Punkte)

3. Der Datensatz `jobdata.dat` von der Homepage enthält die monatlichen Einkommen von Frauen und Männern in verschiedenen Berufen (Köche, Taxifahrer, Biometriker). Dabei wurden bei jeder Kombination aus Geschlecht und Beruf die Daten von 10 zufällig gewählten Personen erhoben. Man vermutet dass Wechselwirkungen zwischen Geschlecht und Beruf auftreten, sowie dass die Faktoren einen Einfluss auf das monatliche Einkommen haben.
- (a) Untersuche die Daten zunächst grafisch auf Einfluss und Wechselwirkungen. Erstelle dazu zwei Plots mit `interaction.plot()` (einmal mit den Berufen, einmal mit dem Geschlecht auf der x -Achse und jeweils mit dem Einkommen auf der y -Achse). Interpretiere die Plots.
 - (b) Führe eine Varianzanalyse durch um die Einschätzung aus (a) zu überprüfen. Passen die Ergebnisse zusammen?
 - (c) Berechne die geschätzten Mittelwerte $\hat{\mu}$ (über alle Daten), $\hat{\alpha}_i$ (Einfluss des Geschlechts), $\hat{\beta}_j$ (Einfluss des Berufs) und $\hat{\gamma}_{i,j}$ (Wechselwirkungen zwischen den Faktoren). Wie passt das zu den Ergebnissen von `lm()$coef`?

Hinweis: Die Funktionen `model.tables()`, `interaction.plot()`, `aov()` und `lm()` (sowie ihre Dokumentationen) könnten hilfreich sein.

(4 Punkte)