

Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

Übung I

26.10.2012, 30.10.2012

Anwesenheitsaufgaben

Aufgabe 1 *Autounfälle*

Der Anteil guter Autofahrer sei $P(G) = 0,6$ und der Anteil schlechter Autofahrer sei $P(S) = 0,4$. Die Wahrscheinlichkeit für einen guten Fahrer, pro Jahr einen Unfall zu haben, sei $P(U|G) = 0,01$, und für schlechte Fahrer sei $P(U|S) = 0,1$.

Jemand hat einen Unfall. Wie gross ist die Wahrscheinlichkeit, dass es sich hierbei um einen schlechten Fahrer handelt?

Aufgabe 2 *Programmierübung 1*

Versuchen Sie die soeben gezeigte ROOT-Einführung anhand folgender Beispiele nachzuvollziehen:

- (i) Erstellen Sie ein Skript mit dem Sie sich die ganzen Zahlen von 0 bis 15 am Bildschirm ausgeben lassen (Tipp: Benutzen Sie eine for-Schleife).
- (ii) Berechnen Sie mittels dieses Skriptes die Summe der ersten 15 ganzen Zahlen.
- (iii) Modifizieren Sie nun dieses Skript und lassen Sie sich nur die ungeraden Zahlen ausgeben.
- (iv) Modifizieren Sie nun das Skript so, dass es, wenn die Zahl 3 erreicht wird, einen zusätzlichen Kommentar auf dem Bildschirm ausgibt (z.B. `Ich bin bei der drei!`). Tipp: Der logische Ausdruck um die Identität zweier Variablen zu testen lautet `var1==var2`.
- (v) Erstellen Sie ein Histogramm mit 16 Bins zwischen $-0,5$ und $15,5$, füllen Sie alle ungerade Zahlen ein und zeichnen es auf dem Bildschirm.

Aufgabe 3 *Programmierübung 2*

In dieser Programmierübung soll die Wahrscheinlichkeitsverteilung für die Summe der Augenzahlen zweier normaler Würfel ermittelt werden. Erstellen Sie dazu ein ROOT-Skript, das folgende Dinge durchführt:

- Ein Histogramm erzeugen (wieviele Bins muss dieses haben und welcher Wertebereich der x-Achse ist dazu sinnvoll?).
- Zufällige, ganze Zahlen generieren mithilfe eines `TRandom3` Objektes und der Funktion `Integer(6)` (es werden dann die Zahlen 0,1,2,3,4,5 zufällig erzeugt).
- Erstellen Sie eine for-Schleife von 0 bis N . Erzeugen Sie innerhalb der for-Schleife jeweils zwei Zufallszahlen und addieren Sie jeweils 1 hinzu (warum?). Tragen Sie die Summe dieser beiden erhaltenen Zahlen in das Histogramm ein mittels der `Fill`-Funktion. Probieren Sie verschiedene Werte für N : 100, 1000, 10000, ...
- Zeichnen Sie das letztendliche Histogramm mit der `Draw`-Methode.
- Normieren Sie das Histogramm auf 1 mit Hilfe der Funktionen `Scale()` und `Integral()`, so dass die relative Verteilung eine Wahrscheinlichkeitsverteilung wird und zeichnen Sie das Histogramm nochmal.
- Wie wahrscheinlich ist es, eine Summe von z.B. 10 zu erhalten? (Verwenden Sie `GetBinContent(Binnummer)`.)

Aufgabe 4 Programmierübung 3

Erstellen Sie nun ein zweidimensionales Histogramm mit 15 Bins auf der x -Achse zwischen 0 und 15 und 15 Bins auf der y -Achse zwischen 0 und 15.

- Das Erstellen eines zweidimensionalen Histogramms geht wie folgt:

```
TH2F hist = TH2F("hist","hist",15,0,15,15,0,15);
```

- Füllen Sie mittels der Funktion `hist.Fill(x,y)` die Zahlenpaare (3,5), (7,6), (1,3), (8,8), (6,4) und (7,4) in das Histogramm.
- Zeichnen Sie das Histogramm mit dem Befehl `hist.Draw("col")`.
- Zeichnen Sie das Histogramm mit dem Befehl `hist.Draw("lego")`.
- Lassen Sie sich den Korrelationskoeffizienten berechnen und ausgeben (benutzen Sie die Funktion `float x=hist.GetCorrelationFactor()`).

Hausaufgaben

Aufgabe 5 Varianz und Kovarianz

5 Punkte

- (i) Die Varianz einer Stichprobe x vom Umfang N , mit den Elementen $\{x_1, x_2, \dots, x_N\}$, ist gegeben durch

$$V(x) = \frac{1}{N} \sum_{i=0}^N (x_i - \bar{x})^2,$$

wobei

$$\bar{x} = \frac{1}{N} \sum_{i=0}^N x_i$$

der Mittelwert ist. Zeigen Sie, dass dies zu

$$V(x) = \overline{x^2} - \bar{x}^2$$

äquivalent ist. Mit anderen Worten: Die Varianz ergibt sich aus dem Mittelwert der Quadrate minus dem Quadrat des Mittelwertes.

- (ii) Die Kovarianz der N Paare von Messobservablen $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ ist gegeben durch

$$\text{cov}(x, y) = \frac{1}{N} \sum_{i=0}^N (x_i - \bar{x})(y_i - \bar{y})$$

- a) Zeigen Sie, dass dies zu

$$\text{cov}(x, y) = \overline{xy} - \bar{x}\bar{y}$$

äquivalent ist.

- b) Zeigen Sie, dass unter der Variablentransformation $x \rightarrow x + C$, wobei C eine beliebige Konstante ist, die Kovarianz $\text{cov}(x, y)$ konstant ist.

Aufgabe 6 Altersverteilung

3 Punkte

Nehmen Sie an, dass die 25 Studenten in einer Vorlesung folgende Lebensalter (in Jahren) haben:
19,0 18,7 19,3 19,2 18,9 19,0 20,2 19,9 18,6 19,4 19,3 18,8 19,3 19,2 18,7 18,5 18,6 19,7 19,9 20,0 19,5
19,4 19,6 20,0 18,9.

- (i) Berechnen Sie den Mittelwert und die Standardabweichung dieser Stichprobe.
(ii) Berücksichtigen Sie nun auch den Vorlesenden (Alter: 37,0). Was beobachten Sie?

Aufgabe 7 Wein

4 Punkte

Marc trinkt gerne Wein. Falls er nur 1 oder 2 Gläser trinkt, hat er am nächsten Tag nie Kopfschmerzen, jedoch immer, wenn er 5 oder mehr Gläser trinkt. Bei 3 oder 4 Gläsern betragen die Wahrscheinlichkeiten für Kopfschmerzen 0,4 bzw. 0,6. Angenommen, mit Wahrscheinlichkeiten von 10%, 20%, 5%, 30% und 35% trinkt er 1, 2, 3, 4 bzw. mehr als 4 Gläser. Mit welcher Wahrscheinlichkeit hat Marc am nächsten Tag Kopfschmerzen?

Aufgabe 8 Wechseln oder nicht?

3 Punkte

In einer Spielshow gibt es drei Türen. Hinter einer Tür wartet ein Gewinn, hinter den beiden anderen Türen befinden sich Nieten. Nachdem der Kandidat eine Tür ausgewählt hat, öffnet der Moderator eine Tür mit einer Niete (aber nicht die Tür, die der Kandidat gewählt hat). Der Kandidat darf erneut eine der beiden übrigen Türen wählen.

Sollte er bei der ursprünglichen Tür bleiben oder besser die andere Tür wählen, um seine Gewinnchance zu erhöhen? Benutzen Sie explizit das Bayes-Theorem, um Ihre Antwort zu begründen.

Aufgabe 9 Dopingtest

5 Punkte

Gegeben sei ein Dopingtest für Sportler. Der Test hat eine Sensitivität (Wahrscheinlichkeit für ein positives Resultat, gegeben dass der Sportler gedopt hat) von 99% und eine Spezifität (Wahrscheinlichkeit für ein positives Resultat, gegeben, dass der Sportler nicht gedopt hat) von 1%.

- (i) Nehmen Sie an, dass 1% der Sportler in einem Wettkampf leistungssteigernde Medikamente benutzen. Berechnen Sie mit Hilfe des Bayes-Theorems, die Wahrscheinlichkeit, dass ein Sportler gedopt ist, gegeben, dass der Test positiv ist.
- (ii) Berechnen Sie nun die Wahrscheinlichkeit, dass ein Sportler gedopt hat, gegeben ein positives Resultat, falls die Sensitivität 90% beträgt.
- (iii) Nehmen Sie nun für die Sensitivität des Tests 99% an und dass 10% aller Sportler in einem Wettkampf leistungssteigernde Medikamente benutzen. Berechnen Sie die neue Wahrscheinlichkeit, dass ein Sportler gedopt hat, gegeben ein positives Resultat.
- (iv) Nehmen Sie an, dass neben der A-Probe ein weiterer Test, die B-Probe, gemacht wird. Wie gross ist die Wahrscheinlichkeit, dass ein Sportler gedopt hat, wenn beide Tests positive Ergebnisse liefern? Verwenden sie die Werte für Sensitivität und Spezifität aus der Einleitung und (i).