Albert-Ludwigs-Universität Freiburg Wintersemester 2013/2014

Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

Übung I

29.10.2013, 30.10.2013

Anwesenheitsaufgaben

Aufgabe 1 *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 2 *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 3 *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 4 *Die Mensa* **3 Punkte**

Ein Freund von Ihnen isst oft in der Mensa, und nimmt entweder ein Schnitzel (15% der Zeit) oder die Currywurst (5% der Zeit) oder das wechselnde Tagesessen (80% der Zeit). Manchmal nach dem Mittagessen fühlt er sich nicht wohl, und Sie wissen schon, dass die Wahrscheinlichkeiten dass er sich nicht wohl fühlt 0.1, 0.2, und 0.05 sind, nachdem er das Schnitzel, die Currywurst, bzw. das Tagesessen verspeist hat. Heute Nachmittag hat er Ihnen gesagt, dass er krank ist wegen der Mensa. Was sind die Wahrscheinlichkeiten, dass er das Schnitzel, die Currywurst, oder bzw. das Tagesessen genommen hat?

Aufgabe 5 *Varianz und Kovarianz* **6 Punkte**

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

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

wobei

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

der Mittelwert ist. Zeigen Sie, dass dies zu

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

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 $(x_1,y_1),(x_2,y_2),\ldots(x_N,y_N)$ ist gegeben durch

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

a) Zeigen Sie, dass dies zu

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

äquivalent ist.

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

Aufgabe 6 *WM-Qualifikation* **2 Punkte**

Nehmen Sie an, während der WM-Qualifikation erzielt die deutsche Mannschaft die folgende Anzahl von Toren pro Spiel: 3, 2, 6, 4, 3, 4, 3, 3, 3, 5.

(i) Berechnen Sie den Mittelwert und die Standardabweichung dieser Stichprobe.

(ii) Berücksichtigen Sie nun auch das Freundschaftsspiel gegen Kanada wobei die deutsche Mannschaft zehn mal getroffen hat. Was beobachten Sie?

Aufgabe 7 *Wechseln oder nicht?* **5 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 8 *Autounfälle* **4 Punkte**

Man hört oft, dass starker Regen Autounfälle verursachen kann. Unten steht die Anzahl der Autounfälle in Freiburg an einem spezifischen Tag, zusammen mit der Niederschlagsumme für diesen Tag.

Autounfälle	Niederschlagsumme [mm]
10	0
16	0
4	0
12	0.4
12	2.8
9	3.0
18	7.3
15	7.6
20	15.8
24	25.5
22	25.0
35	30.1

(i) Berechnen Sie den Mittelwert und Standardabweichung für jeweils die Anzahl der Autounfälle und die Niederschlagsumme pro Tag.

(ii) Berechnen Sie die Schiefe für die beide Variablen.

(iii) Was ist die Korrelation zwischen den beiden Größen laut dieses Datensatzes.