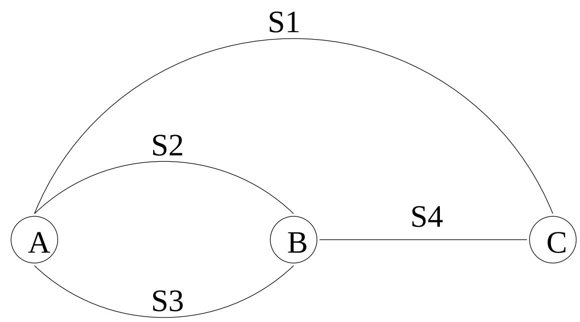
Übung 8

1. Die Strassen zwischen der Stadt A und dem Skigebiet C sind in folgenden Plan eingezeichnet:



Jede der 4 Strassen $S_1, \ldots, S_4$ ist unabhängig von den anderen Strassen mit der Wahrscheinlichkeit $p = \frac{1}{3}$ verschneit und deshalb nicht befahrbar.

a) Wie gross ist die Wahrscheinlichkeit, dass man von A nach C kommt, wenn bekannt ist, dass S_1 verschneit ist?

b) Wie gross ist die Wahrscheinlichkeit, dass man von A nach C kommt, wenn nichts über den Zustand der Strassen bekannt ist?

c) Wie gross ist die Wahrscheinlichkeit, dass S_1 offen ist, falls bekannt ist, dass es eine Möglichkeit gibt, von A nach C zu gelangen?

2. Der Standardtest, um HIV (human immunodeficiency virus) nachzuweisen, ist der Wellcome-Elisa-Test. Bei diesem Test werden 0.7% der HIV-positiven Personen fälschlich für gesund und 0.01% der Gesunden fälschlich für krank erklärt.

a) Der Anteil der HIV-positiven Personen in einer Population ohne bekannte Risikofaktoren (z.B. Blutspender/-innen) ist 0.0025%. Wie gross ist dann die Wahrscheinlichkeit, dass jemand, der für HIV-positiv erklärt wird, wirklich HIV-positiv ist?

b) Der Anteil der HIV-positiven Personen in der Bevölkerung ist bei einer zufälligen Auswahl höher als in Teilaufgabe a). Die WHO (World Health Organisation) schätzte den Anteil in Westeuropa auf 0.07% bei Frauen und 0.5% bei Männern. Wie gross ist jetzt die Wahrscheinlichkeit, dass eine Frau aus der zufälligen Auswahl, die für HIV-positiv erklärt wird, wirklich HIV-positiv ist? Wie gross ist diese Wahrscheinlichkeit für einen Mann?

Die Zahlen stammen aus S. Chatterjee, M. Handcock und J. Simonoff (1995), *"A Casebook for a First Course in Statistics and Data Analysis"* (J. Wiley & Sons, Inc).

3. In einer Studie an 592 Personen wurde der Zusammenhang zwischen Haar- und Augenfarbe untersucht. In der folgenden Tabelle sind die erhobenen Daten dargestellt:

	Haarfarbe				
Augenfarbe	schwarz	brünett	rot	blond	total
braun	68	119	26	7	220
blau	20	84	17	94	215
haselnuss	15	54	14	10	93
grün	5	29	14	16	64
total	108	286	71	127	592

a) Wir betrachten zuerst nur die Zelle brünettes Haar/blaue Augen unter der Nullhypothese H_0 : "Haar- und Augenfarbe sind unabhängig". Wie viele Personen wären unter der Nullhypothese in dieser Zelle zu erwarten gewesen? Welchen Beitrag liefert diese Zelle zum χ^2-Wert?

b) Berechne die Tabelle der unter H_0 erwarteten Anzahlen und den Wert der χ^2-Statistik. Gib den kritischen Wert (Faustregel) und den Testentscheid an. Ist der Zusammenhang zwischen Haar- und Augenfarbe auf dem 5%-Niveau signifikant?

c) Vergleiche die beiden Tabellen. Falls es einen Zusammenhang gibt, welcher Art scheint er zu sein? Welche Zellen deuten auf einen signifikanten Zusammenhang hin? (Verwende dazu in Zweifelsfällen Überschlagsrechnungen mit der Faustregel für einzelne Zähldaten.)

4. Johannsen hat in seinen berühmten Selektionsexperimenten (um 1903) die Breite von 12000 Bohnen gemessen. Die Daten sind im File `bohnen.2.dat` (am üblichen Ort) abgelegt.

a) Stelle die Werte graphisch (z.B. Histogramm) dar.

b) Berechne Mittelwert, Median, Varianz und Standardabweichung.

c) Wieviele Prozent der Daten liegen ausserhalb des Intervalls $\bar{x} \pm 2s_x$?

d) (Von Hand) Wie gross sind Varianz und Streuung vom Mittelwert von 16 zufällig ausgewählten Bohnen, unter der Zusatzannahme, dass die Bohnen unabhängig sind? (Berechne zuerst die Varianz der Summe der Bohnen und berücksichtige dann, dass Mittelwert = Summe/16.)

e) Das theoretische Resultat aus d) wollen wir nun und in f) auch mittels Simulation anschauen, zuerst nur für 2, dann auch für 4 und 16 Bohnen.
Wir wählen also zufällig zwei Bohnen aus und berechnen das arithmetische Mittel der zugehörigen Messungen. Um das zufällige Verhalten des arithmetischen Mittels empirisch kennenzulernen, wiederholen wir dieses Vorgehen 150 mal. Stelle die arithmetischen Mittel in einem Histogramm dar und bestimme den Mittelwert, die empirische Varianz und die Standardabweichung dieser 150 Mittel.

f) Wiederhole Teilaufgabe e) für die Fälle, wo 4, beziehungsweise 16, Bohnen gezogen werden.

g) Stelle (für $n = 1, 2, 4, 16$) die oben in b), e) und f) berechneten Varianzen $\mathrm{Var}(\bar{x}_n)$ des Mittels von n Werten dar gegen *i)* n, *ii)* $1/n$, und kommentiere die Grafiken.

R-Anleitung:

a)
```
> bohnen <- scan.url("http://stat.ethz.ch/Teaching/Datasets/bohnen.2.dat")
                                              # Datensatz via Internet einlesen
> par(mfrow=c(3,2))                           # Sechs Graphiken ins gleiche Fenster
> hist(bohnen, breaks=seq(6.5,11,by=.125))    # Histogramm
```

b)
```
> xq <- mean(bohnen)        # arithmetisches Mittel
> median(bohnen)            # Median
> s2 <- var(bohnen)         # empirische Varianz
> s <- sqrt(s2)             # Standardabweichung
```

c)
```
> sum(bohnen > xq+2*s)      # Anzahl Daten > x̄ + 2s_x
> ...
```

e)
```
> n <- 2
> mittel2 <- numeric(150)           # Initialisieren des Vektors mittel2
> for(i in 1:150) mittel2[i] <- mean(sample(bohnen, size=n))
                  # Die Funktion sample zieht zufällig size=n Elemente aus bohnen.
> hist(mittel2, breaks=seq(6.5,11,by=.125))
> mean(mittel2)        # Mittelwert der 150 "Stichproben" (≙ Mittel von zwei Bohnen)
> s22 <- var(mittel2)  # Varianz der 150 "Stichproben"
> ...
```

f) Setze `> n <- 4` und `> n <- 16` in d).

g)
```
> plot(c(1,2,4,16), c(s2,s22,?,?), xlab="n", ylab="Varianz")
> plot(c(1,1/2,1/4,1/16), c(s2,s22,?,?), xlab="1/n", ylab="Varianz")
```

Abgabe: Montag, 29. Mai, in der Pause der Vorlesungsstunde