Exercises for the course

"An introduction to R"

Sheet 02

Exercise 6: *Here is more practice in handling vectors.* Define the the vector `data` as

```
data <- 90*1:100 - (1:100)^2 + 1000
```

- What is the first, the second and the last entry of the vector `data`?
- What is the *max*imum of the vector `data`? At *which* index is the *max*imum attained?
- Plot the vector `data` with `plot(data)` and visually confirm your last result.
- At *which* indices are the entries of `data` between 2000 and 2500?
- Define a new vector `p` by `p <- data/sum(data)`. Calculate

$$\sum_{i=1}^{100} \mathrm{p}[i], \quad m <- \sum_{i=1}^{100} i\cdot \mathrm{p}[i], \quad s <- \sum_{i=1}^{100} i^2\cdot \mathrm{p}[i], \quad \text{and} \quad \sqrt{s-m^2}$$

(4 points)

Exercise 7: Create the following matrices:

$$\begin{pmatrix}1&3&5\\2&4&6\end{pmatrix} \quad \begin{pmatrix}0&0&0\\0&0&0\end{pmatrix} \quad \begin{pmatrix}1&1&1\\1&1&1\end{pmatrix} \quad \begin{pmatrix}2&3&4\\7&8&9\end{pmatrix}$$

$$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \quad \begin{pmatrix}0&1&1&1\\1&0&1&1\\1&1&0&1\\1&1&1&0\end{pmatrix} \quad \begin{pmatrix}1&4&9&16\\1&4&9&16\\1&4&9&16\\1&4&9&16\end{pmatrix}$$

Define a new matrix `m` by

```
m <- matrix( 11:35, nrow=5, byrow=TRUE )
```

What is the entry in the third row and forth column? Briefly describe in words what

```
m[2:4,3:5]
```

returns. Calculate the matrix product of `m` with itself. (4 points)

Exercise 8: *The most important distribution is the normal distribution. So let's study it.*

- Plot the density of the (standard) normal distribution between -3 and 3.
- In statistics, null hypotheses are rejected if a certain probability is below 5%. Thus it is important to know which centered interval supports 95% of the mass of the normal distribution. Find the value x such that `pnorm(x)-pnorm(-x)`$= 0.95$. Recall the symmetry of the (standard) normal density and explain why this value x is equal to `qnorm(0.975)`. Round the value x to 3 significant digits and remember this number.
- Sample 1000 random values from the normal distribution with the command `rnorm` and denote the vector of theses values as `x`. Calculate the mean, the variance, the standard deviation and the quartiles of this vector. Then visualize the quartiles of `x` with a boxplot. Finally plot the histogram and the empirical distribution function of `x`.

(5 points)

Exercise 9: Let `p` be the vector of the weights on $0, 1, \ldots, 100$ of a binomial distribution with 100 trials and success probability $\frac{1}{2}$.

- Use the command `dbinom()` to define `p`. Plot the vector `p`.
- Recall from the script the true mean and variance of this binomial distribution. Sample 100 random values from this binomial distribution and denote the resulting vector as `y`. Calculate the mean and the variance of `y` and compare these values with the respective true quantities.
- What is the correlation between `y` and `-2*y+4`? What is the correlation between `y` and `3*y`?
- Sample another 500 random values and another 10000 random variables from the binomial distribution at hand (100 trials and success probability $\frac{1}{2}$) and denote the resulting vectors as `z1` and `z2`, respectively. One method to compare `y`, `z1` and `z2` is to compare the boxplots. This is done with the command `boxplot(y,z1,z2)`. Briefly describe in words the differences and similarities of the three boxplots (you might want to find out whether your observations are generally true by repeating the sampling of `y`, `z1` and `z2`)

(5 points)