UPPSALA UNIVERSITET
Matematiska Institutionen
Statistics for biology. Spring semester 2015.
Jordi-Lluís Figueras (figueras@math.uu.se)

Computer Lab 1.

1. Given the following sample compute its range, mean, median, mode, variance and standard deviation.

   1 -1 2 4 3 3 7 10 -2

   (Hint) Use the following commands:

   `c(), range(), mean(), median(), var(), sd(), quantile().`

2. Plot the histogram of the data in the first exercise.

   (Hint) Use the following command: `hist()`.

3. Consider the data file `data1.csv2`. Read this file. As you will see, the file contains two columns: *sample* and *height*. Store the data in the second column into a variable with name `height` and compute its mean, median, variance, standard deviation, quantiles. Plot its histogram.

In `R` there are several commands related to the binomial distribution. These are:

- `dbinom(y, s, p)`: This function computes the density function:

$$f(y) = \binom{s}{y} p^y (1-p)^{s-y}.$$

- `pbinom(q, s, p)`: This function computes the (cumulative) distribution function:

$$F(q) = \sum_{i=0}^{q} f(i).$$

- `rbinom(n, s, p)`: This function creates a vector of n numbers simulating the output results of n simulations of a binomial random variable with size s and probability p.

---

4. Given a biased coin with probability of heads 0.3, compute the probability of:
   - getting 2 heads out of 10 trials.
   - getting at least 2 heads out of 10 trials.
   - getting at least 3 or at most 5 heads out of 10 trials.
5. The probability of being left handed is $p = 0.1$. What is the probability that a group of 100 has exactly 50 left handed? And at most 50 left handed? Are your results correct?
6. Simulate random numbers with binomial distribution with different sizes and probabilities. Compute the mean and standard deviation of the simulations. What do you obtain? Do you expect these results?

   (Hint) You can use values $2, 10, 100$ for the sizes and $0.2, 0.4, 0.7$ for the probabilities.
7. With the help of `R` discuss the following problem:

   Some of my students write me emails asking me all sort of questions. More concretely, I received 20 emails last week, 18 of which were sent after 12 o'clock (so just 2 during the morning). This distribution of the emails' timing attracted my attention. I wondered if it was true that my students send me emails with *equal distribution*: I asked myself if the probability that they send me an email during the morning was 0.5. Hence, I constructed a hypothesis test.

$$H_0: \qquad p = 0.5.$$

$$H_1: \qquad p > 0.5.$$

   Test if I should accept H_0. ($\alpha = 0.05$). Discuss the meaning of your solution.