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

Computer Lab 3.

1. In `R` there are several commands for working with the Student's t distribution and with the chi square distribution. These are `pt, qt, rt, pchisq, qchisq, rchisq`. With the help of `Rstudio` discover how they work.

2. Let $T(n), \chi(n)$ represent a Student's t and a chi square distributions with n degrees of freedom. Compute for $n = 1, 5, 10, 20$ the values t_n, χ_n such that

$$\begin{aligned} P(T(n) < t_n) &= 0.95 \\ P(\chi(n) < \chi_n) &= 0.95. \end{aligned}$$

3. [Exercise 7 from the third set of exercises.]

   A school board is responsible for two elementary schools. It wants to determine how the mean IQ of the students at school *Gryffindor* compares with the mean IQ of those at school *Slytherin*. It chooses a random sample of 90 students from each school. At Gryffindor the sample mean IQ is 109, and at Slytherin it is 98.

   - Compute a 95 percent confidence interval if we know that both schools have the same standard deviation, which is equal to 5.
   - Compute a 95 percent confidence interval if we know that both schools have the same unknown standard deviation. The sample standard deviations are 11 and 9 respectively.
   - Compute a 95 percent confidence interval if we don't know that the schools have the same (unknown) standard deviation. The sample standard deviations are 11 and 9 respectively.
   - In the former case, can we accept the hypothesis that the true means of both schools are the same?

4. In `R` there is a very useful command: `t.test`. This command is used for performing hypothesis test, both with one and two samples.

   Remark: Among all possible options that the `t.test` command has, one very important is `var.equal`, which can be `TRUE` or `FALSE`. This option specifies if the variances of the two samples should be thought of being equal.

5. Download the file `Grad1.csv2`. This file contains the gradings of the four exercises that appeared in the first inlämninsuppgift. Read this file using the command `read.csv2`. With the help of the command `t.test`, check:

   - the means of the first and second exercise are equal.
   - the means of the second and third exercise are equal.
   - the means of the first and fifth exercise are equal.

For the previous tests, did you suppose that the variances where equal? Why? Perform the tests also without knowing that the variances are equal (in this case the test is done by using the Welch test).

6. In `R` there is a very useful command: `chisq.test`. This command is used for performing χ^2 tests. In order to use it, we need to use as input a table.

   Remark: If we are interested in performing a **goodness of fit** test, then the command `chisq.test` has an option, `p`, that should be equal to a vector with the probabilities of the test.

7. [Exercise 1 from the fourth set of exercises.]

   We casted a 6 sided die and got the following results:

1	2	3	4	5	6
32	28	20	10	50	30

   Is the die fair? ($\alpha = 0.05$).

8. [Exercise 4 from the fourth set of exercises.]

   The following table represents the number of bike accidents in Uppsala in 2013. The data is sorted by age group and gender:

	Male	Female
$14 - 18$	3250	4321
$19 - 30$	2241	1441
$31 - 65$	3245	3753
> 65	532	212

   Are the gender and age associated (independent)? ($\alpha = 0.05$).

9. [Exercise 5 from the fourth set of exercises.]

   Students from a high school where asked about their ice-cream preferences. From this poll we got the following data:

	Chocolate	Vanilla	Berries
$12 - 14$	234	512	123
$14 - 16$	112	243	78
$16 - 18$	80	212	324

   Is the preference of the students homogeneous with respect to their age? ($\alpha = 0.05$).