# Exercise 5

*Riccardo De Bin and Vindi Jurinovic*

*November 18, 2015*

We have seen some statistical tools useful to compare groups in case of discrete data. In particular, the first two questions refer to the use of the chi-square and the Fisher exact tests.

1. Use a chi-square test in order to test whether the presence of chronic bronchitis and the current smoking status are independent.

2. Use a Fisher test to verify independence between sex and the presence of any liver disease.

In the previous exercises, we have transformed the variable `hdl` in order to fulfill the normality assumption necessary to perform the t-test. Now, we know some non-parametric tests which allow us to test out hypothesis without performing any transformation.

3. Perform a sign test both on `hdl` and on `logHdl` to test the hypothesis that the median of the cholesterol level is 1.30. Is the median significantly different from 1.30? Do you obtain the same results using `hdl` and `logHdl`? (remember to use the logarithm of 1.30 in the latter case).

4. Use a Mann-Whitney test to test the null hypothesis $H_0 : male\, weight = female\, weight$.

Now let us go back to the NHANES data from your R Data Project. For these new questions, it can be useful to draw a mosaic plot. In R, this can be done through the function `mosaicplot()`. As usual, for further information, it is possible to consult the related help page: `> ?mosaicplot`.

5. Income and health.
   a) It has been shown that there is a "social gradient" in health, such that the richer you are, the more likely you are to have better health. Plot general self-rated health against relative income so that you can get an impression whether this is confirmed by our data.

   b) Test the relation for statistical significance using an appropriate test.

6. Unemployment and depression
   a) Recode the variable `jobstat` into a new dichotomous variable "Unemployment yes/no".

   b) Calculate a depression score by taking the mean across all items of the depression scale ("PHQ-9"). If the values for more than 3 items are missing, assign the person a missing value. The resulting score may be treated as a metric variable.

   c) Investigate the hypothesis that unemployed people are more depressed than employed people. Check also whether the assumptions that your method of choice underlies are met.