Lab 04

Download the `lab04.Rmd` file and open it using RStudio. Then, use the R programming language to help you answer the questions below. **Don't forget to fill out the worksheet form before the next class!**

In this lab we are going to collect a dataset in a spreadsheet program, read the dataset into R, and run an hypothesis test. To simplify things, I will describe the experiment that I want you to run. If you go to the following link:

`https://en.wikipedia.org/wiki/Special:Random`

It will open a random Wikipedia page from the English language version of Wikipedia. If you change the subdomain "en" at the beginning to another language (such as "fr" for french or "zh" for Chinese), you will get a random page on the respective version of Wikipedia. I want you to collect data about 100 random pages from English Wikipedia and 100 pages from another language of your choosing. Record whether a page has at least one image or not and test the null-hypothesis that the proportion of English Wikipedia pages with images is the same as your chosen language.

1. Pick a language that you want to compare to English and record your data for 100 pages. The easiest way to do this is to open up pages using the link on the course notes an just count on a piece of paper. You can do this step in pairs, but I want you to make and read the data in the next question yourself.

2. Now, convert the count data into a tabular data format in your preferred spreadsheet program. Make sure to use column names follow the rules covered in class.

3. Save the dataset as an `xlsx` or `csv` file somewhere easy to find on your computer. Read the dataset into R. If you need help describing absolute paths on your computer, just ask for it!

4. Describe the null-hypothesis and alternative hypothesis in terms of the data you have collected here.

5. Run the function `tmod_z_test_prop` on your dataset. What is

the p-value of the hypothesis test? Is the result significant at a 0.05 percent level? Which proportion is larger? Does this match or conflict with your own expectations?

6. Finally, run the function `tmod_prop_by_group` using the same data and formula argument as in question 4. It gives more summary information about the proportions. Which proportion