Homework Assignment 5
(Due Wednesday, March 16, 2022)

The homework is due at 10.30am in the dropbox on the Course Plus page (you can find the dropbox under the 'Resources' tab in the upper right). For exercises involving R code, please knit a document from your R markdown (Rmd) file. Generate a single pdf file for your entire submission and give it a name that makes it identifiable (calling it `140.615.HW.Number.Lastname.Firstname` or similar). Show your work.

1. You want to detect differences in protein abundances between affected probands and normal controls, measured by ELISA. Since you are interested in fold changes, you record your measurements as $\log_2$(abundance). Thus, a two-fold change in abundance corresponds to a difference Δ of 1 on the $\log_2$ scale, a four-fold change corresponds to Δ of 2, etc. Also assume that in each group the $\log_2$(abundances) are independent measurements from a Normal distribution.

   (a) Assuming you plan on running ELISAs for the same number of cases and controls, how many subjects do you need to have 80% power to detect a two-fold change in abundance between cases and controls, i. e. $\Delta = 1$, if the within-group standard deviation of the $\log_2$(abundances) is the same as the differences in means of the $\log_2$(abundances), i. e. $\sigma = \Delta$?

   (b) How much power would you have if you only ran ten ELISAs per group?

   (c) How much power would you have if you only ran ten ELISAs per group, but knew that in truth the protein your are interested could not be down-regulated in the affected probands?

2. A physician wants to estimate the average body mass index (BMI) for her adult patients. She decides to draw a sample of clinical records and retrieve this information from them. She wants an estimate with a margin of error of 1.5 units of BMI with 95% confidence (i.e. the confidence interval should be 3 units of BMI in length), and believes that the national population standard deviation of adult BMI of 4.5 also applies to her patients. She knows that BMI is approximately normally distributed for adults. How large a sample does she need to draw?

3. Consider data on the treatment response of 12 mice from strain A and 9 mice from strain B. Assume that the true within-group standard deviations are the same.

   Strain A `55.2 58.1 41.7 44.9 44.8 48.9 47.5 48.1 48.4 51.6 40.6 48.0`

   Strain B `48.7 52.6 65.2 70.4 44.2 54.7 44.0 66.5 56.8`

   The data are also available here.

   (a) Test the hypothesis $H_0 : \mu_A = \mu_B$ versus the alternative $H_a : \mu_A \neq \mu_B$ using the function t.test().

   (b) Use the wilcox.test() function to perform a rank-sum test on these data.

   (c) Give an interpretation of your results.

4. These are problems 5.1 – 5.3 from Peter Dalgaard's book.

   You have to install the R package `ISwR` written by Peter Dalgaard. The easiest way to do this is to open an R session and type `install.packages("ISwR")` at the prompt. To load the package into your workspace, type `library(ISwR)`. The data sets `react` and `vitcap` will be availabe in your workspace.

   Type `react` and `vitcap` at the prompt to see the data. Type `?react` and `?vitcap` at the prompt to open the help files, to see what these data represent.

   (a) Do the values of the `react` data set (notice that this is a single vector, not a data frame) look reasonably normally distributed? Does the mean differ significantly from zero according to a t-test?

   (b) In the data set `vitcap`, use a t-test to compare the vital capacity for the two groups. Can we conclude that the population means differ? Calculate a 99% confidence interval for the difference. The result of this comparison may be misleading. Why?

   (c) Perform the analyses of the `react` and `vitcap` data using nonparametric techniques.