Homework 1

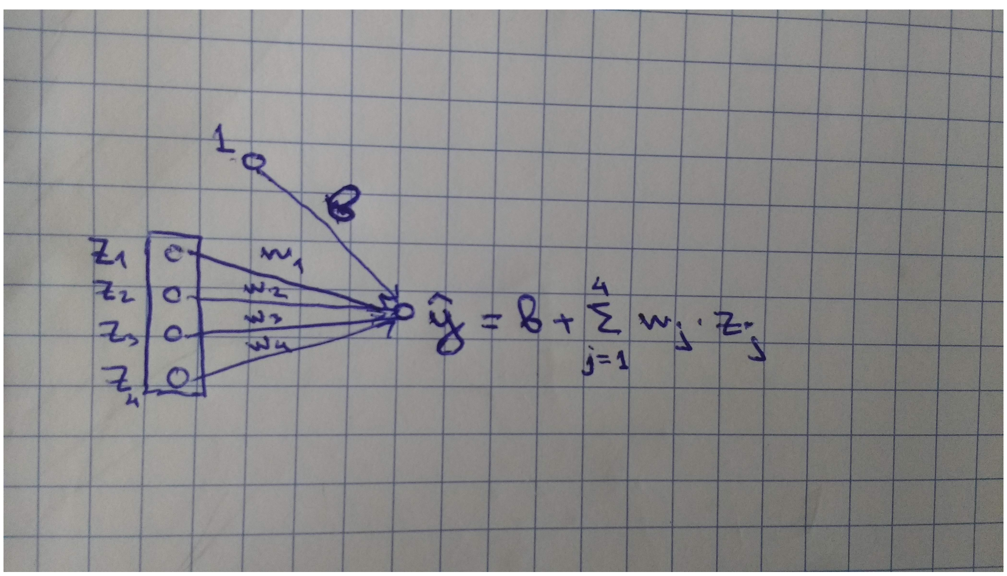
Submit the solution in the form of R Markdown report, knitted into either of the available formats (HTML, pdf or Word). Provide all relevant code and output. Goal of this homework is to have you 1) familiarized with sample size calculation procedures; 2) concepts of power & Type II error for significance tests; 3) practice your R coding.

Whenever asked to do mathematical derivations, which would include heavy usage of formulas and algebraic notation, you may either:

1. **Type it up in LaTeX mode (check the .Rmd source file for some examples), like this**

$$z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \cdots \implies \ldots$$

2. **Write the solutions by hand, take a picture, insert it into this R markdown document (DON'T supply it separately), like this**



Check the *.Rmd* source file for examples, and also, see https://stackoverflow.com/questions/25166624/insert-picture-table-in-r-markdown for reference on how to insert images and manipulate image size in R markdown.

Problem #1

1. Write a $prop.sample.size()$ function that will output the sample size needed for a one-sample proportion test to achieve

   - a desired margin of error (argument #1)
   - for a given confidence level (argument #2),

in the "worst-case scenario" (as was explained in class). What is meant by the "worst-case scenario"?

2. Use your $prop.sample.size()$ from part 1 to do **exercise** 8.50 from the Agresti book.

3. Write a $mean.sample.size()$ function that will output the sample size needed for a one-sample mean test to achieve

   - a desired margin of error (argument #1)
   - for a given confidence level (argument #2),
   - for a given standard deviation (argument #3).

Proceed to use that function in order to do **exercise** 8.53 from the Agresti book.

4. For a **two-sample proportion test**, presuming equal sample sizes ($n_1 = n_2 \equiv n$), proceed to derive the mathematical formula for the sample sizes needed to achieve

   - a desired margin of error (argument #1)
   - for a given confidence level (argument #2).

Do it in similar fashion to how it was derived in class for **one-sample proportion** test. How is the obtained formula different from the case of **one-sample proportion** test?

Problem #2

1. Use the https://istats.shinyapps.io/power/ app to do **exercise** 9.66 from Agresti book.

2. Presume you've developed a skin cancer treatment and you were granted permission to try it out on patients. You would like to test if its accuracy differs from the golden standard method which has 28% cure rate. In particular, you'd want the ability to correctly detect a difference of 10% (practically significant effect).

Hint: To answer parts $(c), (d), (e)$, you can use the in-class R code from *Slides_1.R*, playing around with the parameters.

Proceed to

a. Formulate the hypotheses for the one-sample proportion test.
b. Interpret the statement: "At a 0.05 significance level, the significance test will have 0.77 power when detecting a difference of 10%."
c. Presume we witness your treatment's results for 100 patients. Obtain the power of one-sample proportion test at $\alpha = 0.05$ significance level when detecting a difference of 10%.
d. Presume we witness your treatment's results for 100 patients. Obtain the α significance level needed for your test to have the power of 0.85 when detecting a difference of 10%.

e. What # of patients is needed for your test to have the power of 0.85 when detecting a difference of 10% (at $\alpha = 0.05$ significance level)?
f. In this case, do you think it is more important for your test to have lower significance level or higher power? Explain.
g. Provided that you don't have the resources to recruit more than $n = 100$ patients, what can you do in order to increase the power of your test for detecting 10% difference (**Hint**: See part (d)).

Problem #3 (bonus 3 pts)

Derive the "actual formula" for calculating sample size that would yield:

- a specific Type II error ($beta$)

for a proportion hypothesis test with a specific

- null value $p.0$,
- alternative value $p.a$,
- α significance level.

Note: This formula is provided in $Slides\_1.R$ source code, but in this problem you need to **derive** this formula from scratch.