**Final Exam STT 802 (Fall 2018)**

**Name**:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Rules**: The exam is strictly individual, during the examination you cannot consult (chat, e-mails, texting etc.) any person other than the instructor.

The exam includes **3 questions** (printed in pages 1 and 2) plus a bonus question (printed at the end of the exam, page 3). **Submit by 5pm to D2L (folder FINAL) an HTML or pdf file with your code and the required outcomes** (tables, plots). **Complete BY 5pm the written portion of the exam** (last 2 pages of this exam).

**Q1. Sampling of random variables (30%)**

In a population there are 52% of women and 48% of men, and BMI varies according to

[1a]

[1b]

* 1. Use composition sampling to generate 1,000 samples of BMI and SEX.

**Report a histogram of BMI**.

* 1. **Answer written question 1 (on page 2)**.

**Q2. Power Analysis-I (20%)**

A study is being designed to estimate the effect of BMI on the risk of developing a disease.

You are asked to carry out a Monte Carlo study to assess power and type-I error rate.

**Simulation setting**:

Predictors: simulate SEX and BMI using the procedure you used in question 1.

Outcome: simulate the disease random variable using a Bernulli distribution with success probability given by the following model:

[2]

where is the probability that the ith individual develop the disease. Note that solving [2] for yields: where

**Task**: develop a Monte Carlo study (at least 3,000 replicates) to estimate the power to detect the effect of BMI assuming and n=200. For testing use a logistic regression and a significance level of 0.05.

**Report your result** in the markdown.

**Answer written Question 2 (in page 3)**

**Q3. Type-I error rate and power (50)**

Estimate type-I error rate and power for N=100,500,1K and 5K and an effect size of BMI of . For testing use a logistic regression and a significance level of 0.05.

**Report** a table with type-I error rate by sample size and a table with power by sample size and effect size.

**Answer written question 3.**

**Written Answers**

**Question 1**

1.a What is the average BMI in the population? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1.b What is the average BMI in your sample? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Question 2**

2.a.Report the estimated power \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2.b. Report the estimated Type-I error rate for each sample size

N=100 \_\_\_\_\_\_\_\_\_\_\_\_\_ N=500 \_\_\_\_\_\_\_\_\_\_\_ N=1,000 \_\_\_\_\_\_\_\_\_ N=5,000 \_\_\_\_\_\_\_\_\_

**Question 2**

Among the sample sizes considered, what would be the minimum sample size required to achieve a power of at least 80% for each of the effect sizes considered:

N=100 \_\_\_\_\_\_\_\_\_\_\_\_\_ N=500 \_\_\_\_\_\_\_\_\_\_\_ N=1,000 \_\_\_\_\_\_\_\_\_ N=5,000 \_\_\_\_\_\_\_\_\_

**Bonus Question**: Although the model of questions 2 and 3 is rather simple (just two predictors) a large sample size is needed to achieve high power.

**Why do you think that happens**?

**What strategy would you suggest to increase power**?