Final Exam EPI 853b, Fall 2017

(Instructor: Gustavo de los Campos)

Your name:

Rules:

* You have 2 hours (10:00am-noon) to complete your midterm
* You can consult your notes, previous scripts
* You cannot consult with anyone else except the instructor.

Reporting

* Write your results and answers in the blank space
* E-mail me prior to the end of the exam the scripts you used for analyses. Please keep your scripts as neat as possible and add comments to indicate what part of the code was used to answer which question.

**1. Estimating the effect of the number of cans of beer drank on serum urate values**

**DATA.** The dataset DATA1.dat contain 11 variables, of those we will focus on serum urate (response), sex (M/F), age (years), race (B/W), bmi (kg/cm2) and beer (self-reported number of cans of beer drank in the last week).

> str(DATA)

'data.frame': 1000 obs. of  12 variables:

 $ beer       : int  0 NA NA 0 0 2 NA 0 NA NA ...

 $ serum\_urate: num  6.8 6.9 4.9 7.4 4.3 ...

 $ sex        : Factor w/ 2 levels "F","M": 2 2 1 2 1 2 1 2 1 2 ...

 $ age        : int  64 48 52 54 56 54 52 64 48 53 ...

 $ height     : num  1.73 1.83 1.65 1.87 1.62 1.77 1.59 1.73 1.64 1.81 ...

 $ weight     : num  82.1 93.9 60.8 94.3 60.8 ...

 $ drinker    : int  1 2 2 1 1 1 2 1 3 2 ...

 $ race       : Factor w/ 2 levels "B","W": 2 2 2 2 2 2 2 2 1 2 ...

 $ creatinine : num  1.2 1.2 1 1.2 1 ...

 $ glucose    : int  109 102 98 101 105 121 103 97 183 113 ...

 $ sbp        : int  139 134 117 115 128 135 87 160 171 151 ...

 $ bmi        : num  27.4 28 22.3 27 23.2 ...

**Research goal**: to estimate the effect of the number of cans of beer drank on serum urate, while accounting for differences due to sex, age, race and bmi. Keep in mind this goal when deriving your conclusions.

* 1. ***Provide in this page up to 3 items (tables, figures) that you consider are the most important descriptive statistics/graphs for the problem of interest.***
  2. ***Fit a linear model and report estimates, SE, p-values, and summarize your findings.***
  3. ***Is there evidence of sex-by-beer or race-by-beer interaction? Propose a test, implement it, report results and conclusions.***
  4. ***Estimate the p-value for serum urate using permutations, report: (i) a histogram of the sampling distribution of the t-statistic for serum urate under the null, with a vertical line supper-imposed indicating the t-value you obtained for serum urate in the data set and (ii) the permutation p-value. (iii) Do you see any evident violations of the assumptions used in lm() to produce p-values? Explain why/why not.***

1. **Power and Type-I error Rate Estimation**

A researcher wants to know determine sample size is needed to achieve at least 80% power for covariate effect of 0.3. The statistical test will be based on a linear model with Y as response, the two factors (F1 and F2) and X as predictor. The researcher plans to test for the effect of X on Y while accounting for differences due to F1 and F2. Rejection will be done at 0.05 significance.

The following function can be used to simulate data for this experiment.

getData=function(b){

x1=1.0\*(runif(1)>.5)

x2=ifelse(x1==1,runif(1)>.4,runif(1)>.5)

x3=x1+x2+rnorm(1)

eta=100+x1\*b[1]-x2\*b[2]+x3\*b[3]

y=eta+rexp(1)

x=c(x1,x2,x3,y)

names(x)=c('F1','F2','X','y')

return(x)

}

# The following code illustrates how to use getData()

# 1st two are the effects of F1 and F2, last the covariate effect

effects=c(0.5,.2,.1)

sampleSize=100

DATA=t(replicate(expr=getData(effects),n=sampleSize))

Conduct a Monte Carlo Study to estimate power and type-I error rate for effects of factors F1 and F2 equal to 0.5 and 0.2, possible values for the effect of X between 0 and 0.5, and sample size equal to N=50, 100, 200 and 300.

NOTE: to test your code I suggest you use 1000 MC replicates, then if you have time increase the number of replicates to obtain more precise estimates.

* 1. ***Report a power plot with rejection rate in the vertical axis, effects size in the horizontal axis and different curves by sample size.***
  2. **What is the minimum sample size required to achieve at least 80% power for an effect of X of 0.35?**
  3. **The errors in the simulation are not normal. Does the test use have the appropriate type-I error rate? Explain why you think the test has / does not have the right type-I error rate.**