STA302/1001HF: Methods of Data Analysis I **Fall 2016**

Assignment 1 : Brain Size and Intelligence

Out: Sept. 22,2016 *Due: Oct. 13, 2016*

Reminder : You MUST write your solution independently and turn in your own write-up.

Your assignment solution is due 12 :00pm, Oct. 13, 2016. Submit your solution as instructed by Crowdmark, namely, one pdf file for each question.

Late assignments will be subject to a deduction of 4% of the total marks for the assignment for each day late. Any late assignment after the day I post the solution will get zero mark.

Presentation of solutions is very important. A Rmarkdown template for the solution is provided. Please following the instruction in that template to complete your solution. You should produce a PDF file, save it and submit it on the portal together with the corresponding Rmd file. Also, make sure the source R code at the end is complete. Marks will be deducted if the instructions herein are not followed.

Data

Are the size and weight of your brain indicators of your mental capacity ? In this study by Willerman et al. (1991) the researchers use Magnetic Resonance Imaging (MRI) to determine the brain size of the subjects. The researchers take into account gender and body size to draw conclusions about the connection between brain size and intelligence.

The data were collected on 40 right-handed introductory psychology students at a large southwestern university. 20 males among the 40 subjects. For both genders, half of the students had full-scale IQs of at least 130, and the rest had full-scales IQs of at most 103. In addition to taking IQ tests, the MRI Scans were performed at the same facility for all 40 subjects. The scans consisted of 18 horizontal MR images. The computer counted all pixels with non-zero gray scale in each of the 18 images and the total count served as an index for brain size.

The variables in the dataset are :

- Gender : male or female.
- FSIQ : Full-scale IQ scores (an overall score of intelligence).
- VIQ : verbal IQ score.
- PIQ : performance IQ score.
- Weight : body weight in pounds.
- Height : height in pounds.
- MRIcount : total pixel Count from the 18 MRI scans.

The weights of two subjects and the height of one subject were withheld for confidentiality. The missing data symbol in R is recoded as NA.

More detail about this published student can be found in the article : Willerman, L. Schultz, R., Rutledge, J.N., and Bigler, E. (1991), "In vivo brain size and intelligence", *Intelligence*,**15**, 223-8.

Questions

You will need to create a variable that indicates whether the subject is in the high- or low-intelligence group(as classified by FSIQ).

Using R to do all the analysis for the following. Uncompleted R code is given at the end of the assignment.

1. A straightforward method of considering the research question is to conduct a t-test to compare the mean MRI count between the high intelligence and low intelligence groups. Assume the variance of the MRI accounts in the two groups are different. What is the null hypothesis of this test ? What is the p-value for this test ? What can you conclude from this test ?
2. Alternatively, we could consider whether the MRI count and IQ score are correlated. Find the correlations between MRI count and all the three IQ scores (FSIQ, VIQ, PIQ) for both the entire dataset, and for the high and low intelligence groups separately. Summarize your results in a table and give your comment in a short paragraph.
3. Does the result of the t-test in question 1 agree with the relevant correlation in question 2 ? Why or why not ? Under the assumption that the data have a bivariate normal distribution, why would the correlation be the preferred analysis ? (The answer to the last part requires you to use your common sense. People in medical science often like to dichotomized variables, such as high IQ group and low IQ groups, while statisticians tend to utilize all the information given).
4. Now we focus on investigating the relationship between performance IQ (PIQ) and MRI count. For a regression analysis, we choose the PIQ as the response variable and the MRI count as the predictor.
   (a) Give the scatterplot of PIQ versus MRI count on all 40 observations.
   (b) Carry out the regression analysis seperately for the high IQ and low IQ groups. For each groups, find and summarize the following in a table.
      - R^2.
      - the intercept and slope.
      - the estimate of the variance of the error term.
      - the p-value for the test with null hypothesis that the slope is 0.

      i. The slopes are very small. Is this an indication of no relationship between PIQ and MRI count ? Why or why not ?
      ii. Compare the value of R^2 for the two regressions. Which regression gives a better fit if we use R^2 as the criteria ?
      iii. Compare the value of MSE for both regressions, which regression gives a better fit if we use this as the criteria ?
      iv. Given only the values of R^2 and MSE, I would almost always choose MSE as my criteria. Why is this reasonable ?

Incomplete R code

```
#
# R code for STA302 or STA1001H1F assignment 1
# copyright by YourName
# date: Sept. #, 2016
#

## Load in the data set
brain = read.table("???/A1/BrainData.csv",sep=" ",header=T)

## create an indicator for high-IQ (value =1) and low-IQ (value=0)
highIQ = ifelse(brain$FSIQ>=130,1, 0)

## Q1: t-test on MRI count between high- and low IQ groups


## Q2: correlation analysis
# cor.test() : missing value is suppressed, default setting:
#      mu = 0, alternative = c("two.sided"),  paired = FALSE,var.equal = FALSE
# - find correlation between MRI count and 3 IQ variables


# - find correlation between MRI count and 3 IQ variables in high-IQ group


# - find correlation between MRI count and 3 IQ variables in low-IQ group


## Q4:
# - Scatterplot of PIG vs MRI count


# - find R-square, b0, b1, MSE and p-value for b1 in high-IQ group


# - find R-square, b0, b1, MSE and p-value for b1 in low-IQ group
```

Listing 1 – Incomplete code for the data analysis