A1: Brain Size and Intelligence

Last name: Akad
First name: Dogan
Student ID: 1001386083
Course section: STA302H1F-L0101

Oct. 12, 2016

Q1: t-test for MRIcount between high and low intellegince groups

Null hypothesis: True difference in means is equal to 0. P-value is 0.1344. We fail to reject the null hypothesis since the p-value isn't too small.

Q2: correlation analysis among the MRI count and IQ variables

Correlations of the IQ measurements with MRI count (p-value for test of $\rho = 0$ is in brakets):

-	Full data	High-IQ group	low-IQ group
FSIQ	0.3576(0.0235)	0.5483(0.0123)	0.5273(0.0169)
VIQ	0.3375(0.0332)	0.4067(0.0752)	0.1464(0.5381)
PIQ	0.3868(0.0137)	0.2013(0.3948)	0.5862(0.0066)

From the correlation analysis, we conclude that

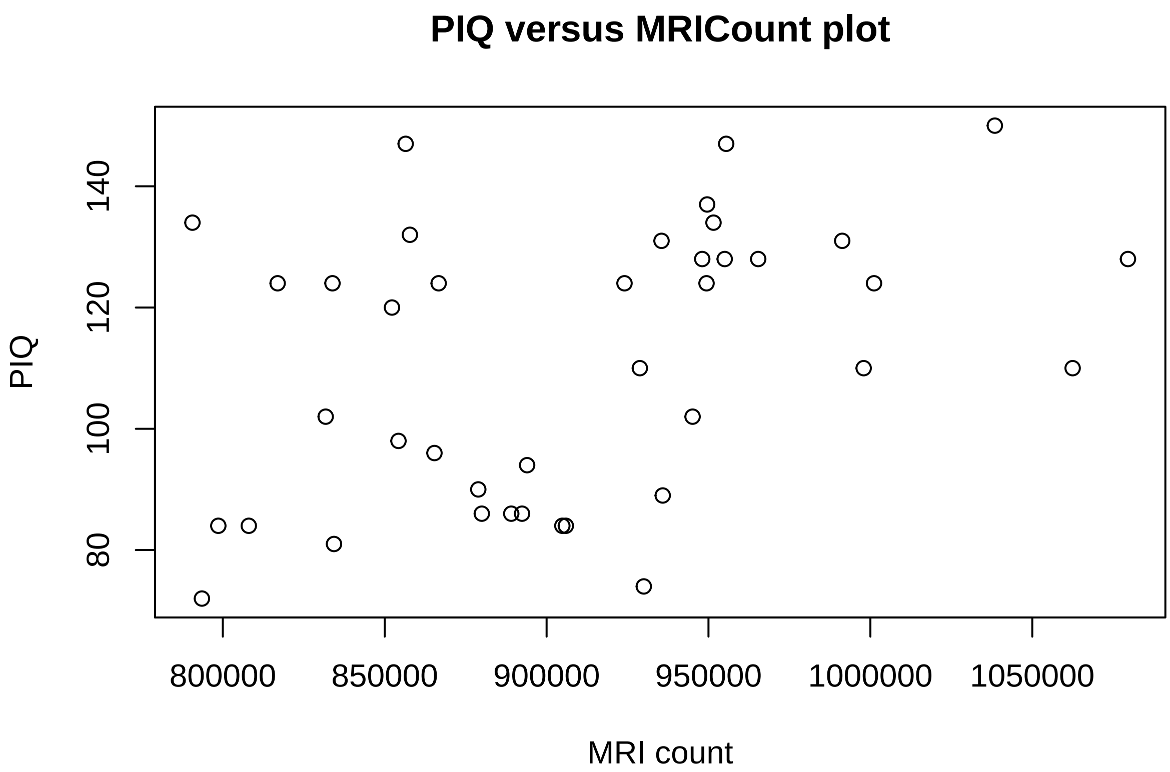
- There is a weak correlation between MRI count and all 3 IQ scores for the entire dataset.
- There is a moderate correlation between MRI count and FSIQ, VIQ whereas there's a weak correlation between MRI count and PIQ for high intelligence groups.
- There is a moderate correlation between MRI count and FSIQ, PIQ whereas there's negligible or no correlation between MRI count and VIQ for low intelligence groups.

Q3

Our result from t-test shows that brain size and overall IQ are directly proportional whereas when we look at the correlations we can see that both MRI count between FSIQ for both high and low intelligence groups are moderately correlated which means that we can't say brain size and overall IQ are directly proportional. So t-test disagrees with the correlation. Correlation would be the preferred analysis because it describes the association between two variables and we can get the exact number to determine how these two variables are correlated. Also we know that the data have a bivariate normal distribution which means that there are two linear functions of the same two independent random variables, so we can accurately calculate the correlations between these two variables.

Q4(a) Scatter plot of PIQ versus MRI count

```
# scatter plot of PIQ versus MRI count
brain = read.table("/Users/doganakad/Desktop/uoft/first semester/STA302/A1/BrainData.csv",sep=" ",header=T)
#complete the following plot() command to get the scatter plot
plot(brain$MRIcount, brain$PIQ, main="PIQ versus MRICount plot", xlab="MRI count", ylab = "PIQ")
```



Q4(b) Regression analysis for two groups

Regression	R^2	Intercept (b_0)	Slope(b_1)	MSE	p-value for $H_0 : \beta_1 = 0$
High-IQ groups	0.04051	1.100e+02	2.265e-05	73.51348	0.3948
Low-IQ grups	0.3436	1.636e+00	1.003e-04	88.75524	0.006602

i.) Slope being small means that a change in the predictor variables causes a small change in the response variable. Only looking at the slopes, we can't exactly say anything. We also need to take in account of their intercepts.

ii.) Since low-IQ groups have better R^2 value, they give a better fit if we use R^2 as a criteria.

iii.) Since high-IQ groups have lower MSE, they give a better fit if we use MSE as a criteria.

iv.) The value of R^2 indicates how close the data is to the fitted regression line whereas MSE measures the average of the squares of the difference between the estimator and what is estimated. It is reasonable because there is always a statistical error when fitting a regression line since the relationship between variables is not perfect, therefore R^2 is not very accurate.

Apendix: Source R code

```
# ---------> complete and run the following code for this assignment  <-------
#
# R code for STA302 or STA1001H1F assignment 1
# copyright by Dogan Akad
# date: Sept. #, 2016
#

## Load in the data set
brain = read.table("/Users/doganakad/Desktop/uoft/first semester/STA302/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
t.test(formula = brain$MRIcount ~ highIQ, data = brain, var.equal=FALSE)

## 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
cor.test(brain$MRIcount, brain$FSIQ)
cor.test(brain$MRIcount, brain$VIQ)
cor.test(brain$MRIcount, brain$PIQ)

# - find correlation between MRI count and 3 IQ variables in high-IQ group
MRIhighIQ <- subset(brain$MRIcount, highIQ == 1)
FSIQhigh <- subset(brain$FSIQ, highIQ == 1)
VIQhigh <- subset(brain$VIQ, highIQ == 1)
PIQhigh <- subset(brain$PIQ, highIQ == 1)

cor.test(MRIhighIQ, FSIQhigh)
cor.test(MRIhighIQ, VIQhigh)
cor.test(MRIhighIQ, PIQhigh)

# - find correlation between MRI count and 3 IQ variables in low-IQ group
MRIlowIQ <- subset(brain$MRIcount, highIQ == 0)
FSIQlow <- subset(brain$FSIQ, highIQ == 0)
VIQlow <- subset(brain$VIQ, highIQ == 0)
PIQlow <- subset(brain$PIQ, highIQ == 0)

cor.test(MRIlowIQ, FSIQlow)
cor.test(MRIlowIQ, VIQlow)
cor.test(MRIlowIQ, PIQlow)

## Q4:
# - Scatterplot of PIG vs MRI count
plot(brain$MRIcount, brain$PIQ, main="PIQ vs MRICount", xlab="MRI count", ylab = "PIQ")

# - find R-square, b0, b1, MSE and p-value for b1 in high-IQ group
linregression <- lm(PIQhigh ~ MRIhighIQ)
summary(linregression)

# - find R-square, b0, b1, MSE and p-value for b1 in low-IQ group
linregression2 <- lm(PIQlow ~ MRIlowIQ)
summary(linregression2)
```