A1: Brain Size and Intelligence

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

Oct. 13, 2016

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

Null hypothesis: True difference in means is equal to 0. We are using an two sample t-test. We are making the assumption that two groups have the normal distribution and the variance of MRI in both groups are unequal. The test statistic is 1.53. The degree of freedom is 37.324, the significance is 0.05 and the confidence interval is 95%. P-value is 0.1344. We fail to reject the null hypothesis. Therefore we don't have sufficient evidence to show a relationship.

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

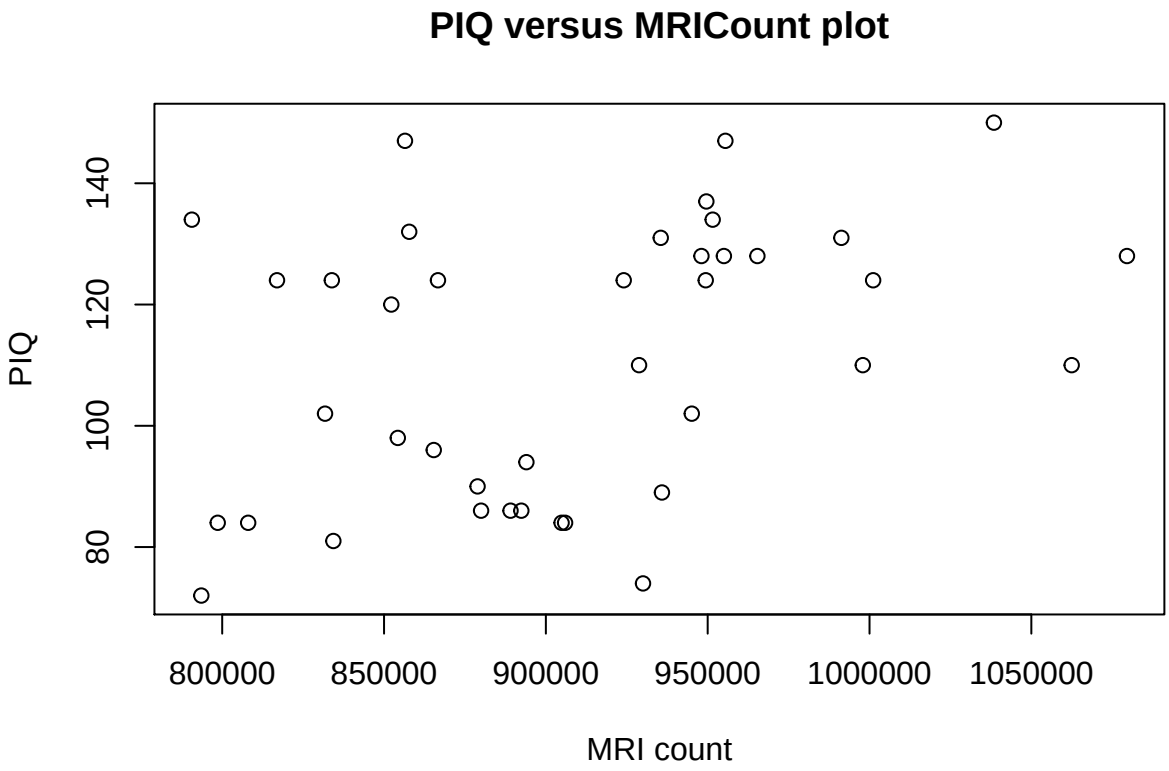
- There is a weak correlation between MRI count and FSIQ for full data whereas there's moderate correlation for high and low IQ group. Since the p-values are less than significance, the correlations are significance and there is evidence to show the relationship.
- There is a weak correlation between MRI count and VIQ for the full data and since p-value is less than significance level the correlations are significant and there is evidence to show the relationship. There is a moderate correlation between MRI count and VIQ for the high IQ group but since the p-value is greater than significance level the correlation is not significant. There is negligible correlation between MRI count and VIQ for the low IQ group but it's not significant.
- There is a weak correlation between MRI count and PIQ for full data and it is significant. There is a weak correlation between MRI count and PIQ for highIQ group but it's not significant. There is a moderate correlation between MRIcount and PIQfor lowIQ group and it's significant.
- There is a conflict between the correlation between MRI count and PIQ for high IQ groups and for the low IQ groups. When we look at the p-value for the correlation between MRI count and PIQ for the high IQ group we conclude that they're not correlated whereas when we do the same for low IQ group we conclude that they're not correlated.

Q3

Our result from t-test shows that there is no sufficient evidence for the relationship between brain size and overall IQ but when we look at the correlations we have sufficient evidence to show brain size and IQ are correlated. So t-test disagrees with the correlation. Correlation would be the preferred analysis because it uses the full information available whereas t-test doesn't take in account of the FSIQ, it just looks at the true difference in means between different groups.

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	0.0405	1.100e+02	2.265e-05	73.51348	0.3948
Low-IQ grups	0.3436	1.6363	1.003e-04	88.7552	0.0066

i.) Slope being small means that a change in the predictor variables causes a small change in the response variable. A change in MRI count will have a very small change in PIQ.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 lower MSE is more 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(MRIhighIQ, MRIlowIQ)

## 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)
linregression$coefficients #b0 and b1
summary(linregression)

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