A1: Brain Size and Intelligence

Last name:
First name:
Student ID: 000000
Course section: STA302/1001H1F-All sections

October 13th, 2016

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

The null hypothesis assume we have equal means of MRI count in the high IQ and low IQ groups. The p-value for the two-side t-test is 0.1344 (assume variance for two groups are different). The test result gives no evidence of a difference in the mean brain size between the high and low intelligence groups.

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 evidence that FSIQ is moderately correlated with MRI count for both IQ groups as well as for the full data set.
- There is moderately correlation between MRI count and VIQ in the full data set. However, ther is only weak evidence for the non-zero correlation in the high-IQ group and no evidence of correlation for the low-IQ group.
- When considering the PIQ and MRI count. We have evidence of non-zero correlation for all observations, but no evidence of a linear relationship for high-IQ group and strong evidence of a a linear relationship for the low-IQ group.

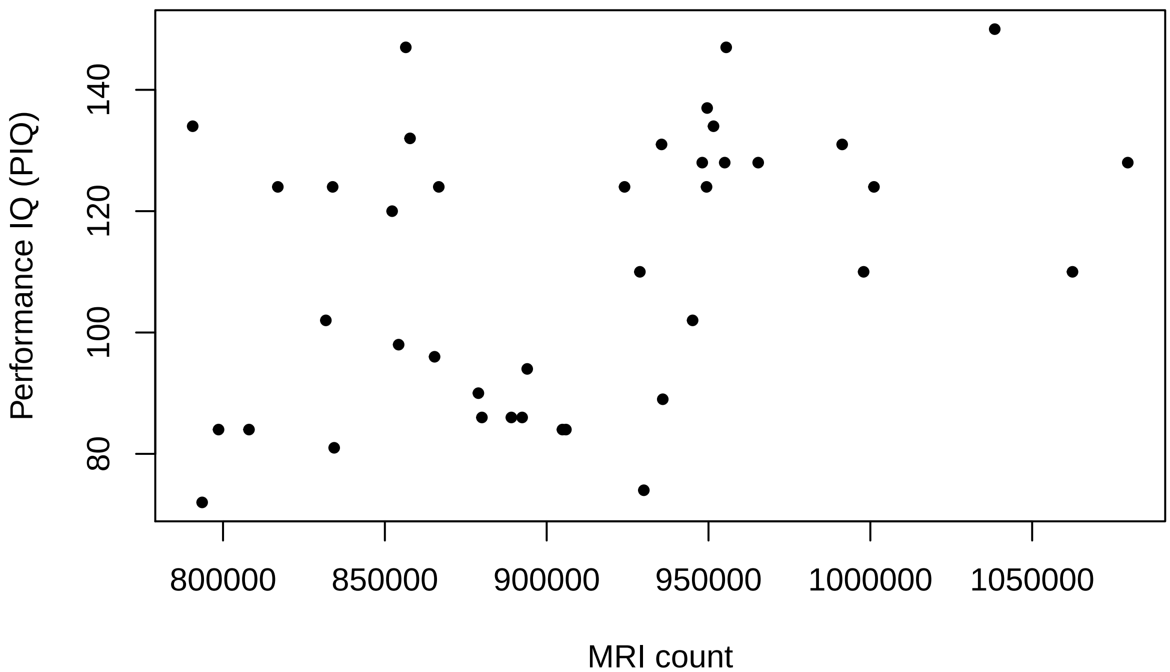
Q3

The t-test shows no evidence to reject the equal means of MRI count between the two groups, but there is evidence (p=0.0235) that the correlation between FSIQ and MRI count is not zero. So the t-test indicates no relationship between IQ and brain size, while the correlation analysis suggests that there is a linear relationship.

The correlation analysis is preferred because we use all of the information given while the t-test in Q1 treats all FSIQs over 130 (or under 103) the same, regardless of their value.

Q4(a) Scatter plot of PIQ versus MRI count

```r
# scatter plot of PIQ versus MRI count
brain = read.table("/Users/Wei/TA/Teaching/0-STA302-2016F/HW/A1/BrainData.csv",sep=" ",header=T)
plot(brain$MRIcount, brain$PIQ,type="p",pch=20,xlab="MRI count", ylab="Performance IQ (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	110	0.0000227	73.96	0.3948
Low-IQ grups	0.3436	1.6363	0.0001003	88.75	0.0066

i.) No. The slopes are very small because the MRI counts are large. A change of one unit in MRI count should result in a tiny change in IQ on average. Whether the change is significant or not, we assess it by the p-value for the test that $\beta_1 = 0$.

ii.) The fitted regression line for the low-IQ group has a much higher R^2 than the estimated regression line in the high-IQ group (0.3436 versus 0.0405).

iii.) The estimated variance MSE is lower in the high-IQ group than the low-IQ group (73.96 versus 88.75).

iv.) R^2 indicates the percentage of variation in the data that can be explained by the regression line. It is not used except as an inidicator of how closely the estimated line fits the given data. However, MSE is used in the hypothesis testing and confidence interval construction for the intercept and the slope. The lower the value of MSE, the more precise the estimates are(as a result, the more precise any predictions will be). So having lower MSE is useful for the practical purpose of carrying out a regression analysis.

Apendix: Source R code

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

## Load in the data set
brain = read.table(file.choose(),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(brain$MRIcount[highIQ==1], y = brain$MRIcount[highIQ==0], mu = 0,
        alternative = c("two.sided"),  paired = FALSE,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)  # or use: with(brain, cor.test(MRIcount, 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
cor.test(brain$MRIcount[highIQ==1], brain$FSIQ[highIQ==1])
# above is equivalent to: with(subset(brain,highIQ==1),cor.test(MRIcount,FSIQ))
cor.test(brain$MRIcount[highIQ==1], brain$VIQ[highIQ==1])
cor.test(brain$MRIcount[highIQ==1], brain$PIQ[highIQ==1])

# - find correlation between MRI count and 3 IQ variables in low-IQ group
cor.test(brain$MRIcount[highIQ==0], brain$FSIQ[highIQ==0])
cor.test(brain$MRIcount[highIQ==0], brain$VIQ[highIQ==0])
cor.test(brain$MRIcount[highIQ==0], brain$PIQ[highIQ==0])

## Q4:
# - Scatterplot of PIG vs MRI count
plot(brain$MRIcount, brain$PIQ,type="p",pch=20,xlab="MRI count", ylab="Performance IQ (PIQ)")
abline(h=130,lty=2,col="gray")

# - find R-square, b0, b1, MSE and p-value for b1 in high-IQ group
fit1 = lm ( PIQ~ MRIcount, data=subset(brain, highIQ==1) )
summary(fit1)

# - find R-square, b0, b1, MSE and p-value for b1 in low-IQ group
fit0 = lm ( PIQ~ MRIcount, data=subset(brain, highIQ==0) )
summary(fit0)
```