A1: Brain Size and Intelligence

Last name: Zheng
First name: Siyuan
Student ID: 1000726814
Course section: STA302H1F-L0101

Oct. 11, 2016

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

Type your concise and clear answer here.

Let a denotes the MRI data of high-intelligence group and b denotes the MRI data of the low-intelligence group

Null Hypothesis: $H_o : \mu_a = \mu_b$

The P-value is 0.1344475

Since the p-value is greater than 0.05, which is the usual common significance level, we cannot reject the null hypothesis. Thus, we accept the fact that the two means equal.

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)	0.5482853(0.0123)	0.5273002(0.01689)
VIQ	0.3374777(0.0332)	0.4066862(0.07516)	0.1463655(0.5381)
PIQ	0.3868173(0.01367)	0.2012682(0.3948)	0.5861888(0.0066)

From the correlation analysis, we conclude that

-Generally speaking, people's full-scale IQ, verbal IQ and performance IQ have direct relationship(positive and weak correlation) with their brain size, respectively.

-People with higher IQ have moderate positive correlation between full-scale IQ and their brain size, as well as the one between verbal IQ and their brain size.

-People with lower IQ have moderate positive correlation between full-scale IQ and their brain size, as well as the one between performance IQ and their brain size.

-People with higher IQ have weak positive correlation between performance IQ and their brain size while people with lower IQ have weak positive correlation between verbal IQ and their brain size.

-People with higher IQ has stronger positive correlation between verbal IQ and their brain size than lower IQ people.

-People with lower IQ has stronger positive correlation between performance IQ and their brain size than higher IQ people.

Q3

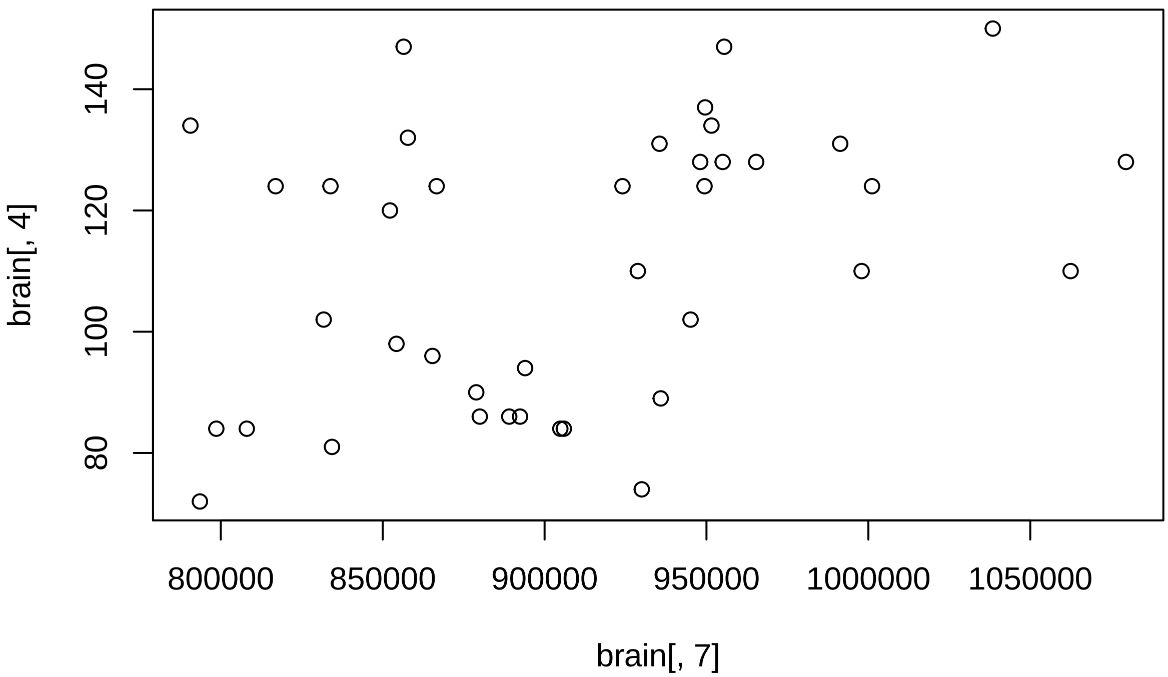
Type your answer here. Please make it as concise as possible.

No, it doesn't.q2 gives us the information that both low-IQ and high-IQ people have direct relationship between IQ and brain size. So q2 let us realize that high IQ people should have larger brain size, which is a contradiction to q1's information.

Regression assumes the independent variable is fixed or can be fixed, but neither a person's brain size nor IQ is fixed or controlable. Correlation analysis assumes two variables are both random.

Q4(a) Scatter plot of PIQ versus MRI count

```r
# scatter plot of PIQ versus MRI count
brain = read.table("/Users/raymondz/Desktop/UofT/STA302H1/302A1/BrainData.csv",sep=" ",header=T
#complete the following plot() command to get the scatter plot
plot(brain[,7],brain[,4])
```



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	109.97725257	0.00002265	73.5	0.39483
Low-IQ grups	0.3436	1.6363092	0.0001003	89	0.0066

i.) Type your answer here.

No. Because slope just gives the information of "the extent of proportion" between the assumed dependent and independent variables instead of the existence of correlation between the two variables.

ii.) Type your answer here.

The regression of the low-IQ people gives a better fit, since R^2 is bigger than the high-IQ people's, although a high R^2 does not necessarily imply a good line fit.

iii.) Type your answer here.

The high-IQ's regression gives a better fit since it has a smaller MSE. MSE is the unbiased estimator of variance.

iv.) Type your answer here.

In the case that the true regression association between X and Y is curvilinear might generate a large number of R's square, thus R^2 will gives us the 'false positive' information of fitness of the regression. Also, while R^2 is large, MSE may still be too large,thus we don't consider the regression is a good fit.

Apendix: Source R code

```r
# ---------> complete and run the following code for this assignment  <-------
#
# R code for STA302 or STA1001H1F assignment 1
# copyright by Siyuan Zheng
# date: Oct. 11th, 2016
#
rm(list = ls())
help("t.test")
## Load in the data set
brain = read.csv("/Users/raymondz/Desktop/STA302H1/302A1/BrainData.csv",sep=" ",header=T)
## create an indicator for high-IQ (value =1) and low-IQ (value=0)
highIQ = ifelse(brain$FSIQ>=130,1, 0)
## Q1: t-test on MRI count between high- and low IQ groups
high = subset(brain,brain[,2] >= 130)
low = subset(brain,brain[,2] <= 103)
highMRI <- high[,7,drop=FALSE]
lowMRI <- low[,7,drop=FALSE]
question1 <- t.test(highMRI,lowMRI,var.equal=FALSE)
names(question1)
question1$statistic
question1$p.value



## 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
allMRI <- brain[,7]
allFSIQ <- brain[,2]
allVIQ <- brain[,3]
allPIQ <- brain[,4]
cor.test(allMRI,allFSIQ,method='pearson')$estimate
cor.test(allMRI,allFSIQ,method='pearson')$p.value
cor.test(allMRI,allVIQ,method='pearson')$estimate
cor.test(allMRI,allVIQ,method='pearson')$p.value
cor.test(allMRI,allPIQ,method='pearson')$estimate
cor.test(allMRI,allPIQ,method='pearson')$p.value

# - find correlation between MRI count and 3 IQ variables in high-IQ group
highMRI <- high[,7]
highFSIQ <- high[,2]
highVIQ <- high[,3]
highPIQ <- high[,4]
cor.test(highMRI,highFSIQ,method='pearson')$estimate
cor.test(highMRI,highFSIQ,method='pearson')$p.value
```

```r
cor.test(highMRI,highVIQ,method='pearson')$estimate
cor.test(highMRI,highVIQ,method='pearson')$p.value
cor.test(highMRI,highPIQ,method='pearson')$estimate
cor.test(highMRI,highPIQ,method='pearson')$p.value
# - find correlation between MRI count and 3 IQ variables in low-IQ group
lowMRI <- low[,7]
lowFSIQ <- low[,2]
lowVIQ <- low[,3]
lowPIQ <- low[,4]
cor.test(lowMRI,lowFSIQ,method='pearson')$estimate
cor.test(lowMRI,lowFSIQ,method='pearson')$p.value
cor.test(lowMRI,lowVIQ,method='pearson')$estimate
cor.test(lowMRI,lowVIQ,method='pearson')$p.value
cor.test(lowMRI,lowPIQ,method='pearson')$estimate
cor.test(lowMRI,lowPIQ,method='pearson')$p.value

## Q4:
# - Scatterplot of PIQ vs MRI count
brain = read.table("/Users/raymondz/Desktop/STA302H1/302A1/BrainData.csv",sep=" ",header=T)
plot(brain[,7],brain[,4])

# - find R-square, b0, b1, MSE and p-value for b1 in high-IQ group
#remember  lm(Y ~ X)  while  plot(X,Y)....
options("scipen"=100, "digits"=4)
highfit <- lm(high[,4] ~ high[,7])
names(highfit)
highfit$coefficients
summary(highfit)$r.squared
anova(highfit)
summary(highfit)
# - find R-square, b0, b1, MSE and p-value for b1 in low-IQ group
lowfit <- lm(low[,4] ~ low[,7])
names(lowfit)
lowfit$coefficients
summary(lowfit)$r.squared
anova(lowfit)
summary(lowfit)
```