Problem 1	**Score**
Main Point	
Improperly Formatted Tables, No references to tables, No informative titles for tables	-1 (each)
Rounding	-1

Introduction

Topic or population introduced:

- "Infants with Thrombocytopaneia"
    - Thrombocytopaenia is a condition of abnormally low platelets that often occurs during necrotizing enterocolids (NEC)- a serious illness in infants that can cause tissue damage to the intestines.

Research question aptly described:

- Patients were grouped into four categories of NEC status. Group 0 referred to infants with no gangrene, group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, group 2 referred to patients with the majority of small and large bowel involved
- Which (if any) of the four groups has a different mean platelet count (on the log-scale given in the Table)?
    - Testing to see whether at least of one of the mean in one of the research groups is different than the means of any of the other research groups

Methods

Description of Outcome:

- Outcome Variable: logarithmic platelet level count
- Predicator Variable: infant "group status"
    - Patients were grouped into four categories of NEC status. Group 0 referred to infants with no gangrene, group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, group 2 referred to patients with the majority of small and large bowel involved

Description of Group:

- Group A (or group 0) refers to infants with no gangrene
- Group B (or group 1) refers to subjects with in whom gangrene was limited to a single intestinal segment
- Group C (or group 2) refers to subjects with two or more intestinal segments of gangrene
- Group D (or group 3) refers to patients with the majority of small/large bowel involved

Description of Data Summary

- *for each group:* sample size, mean, standard deviation, 95% Confidence Interval
- *for each difference*: difference of means, standard error, adjusted 95% Confidence Interval

Verification of Normality:

- Sample sizes were assessed (check to see if each group's sample size > 30)
- Distribution of data was assessed with histograms and QQPlots.

Statement of H_0: "The four group means are equal."

- There is no difference between the group means.

Statement of H_A: "At least two of the means differ."

- There is at least one mean difference between the groups.

Statistical Method for testing difference in means:

- A one- factor, equal-variance analysis of variance (ANOVA, F-test)

Statistical Method for testing equality of variances:

- Bartlett's Test

Method for multiple comparisons:

- Tukey test

Decision Rule:

- If the resulting p-value from the ANOVA model is less than 0.05, we will reject the null hypothesis in favor of the alternative hypothesis, and multiple comparisons between group-level means will be made using the Tukey method, and observed difference, standard errors, adjusted confidence intervals and adjusted p-values will be reported; otherwise we will fail to reject the null hypothesis.

Method of computation: R statistical software

- Version 2.11.1

Significance level: $\alpha=0.05$

Results

Data Summary

Table of Group Summaries

Table 1.1: Summaries for infant logarithmic platelet level count and group status.

NEC Status Group	n	mean	SD	95% CI
0 (no gangrene)	67	2.2	0.40	(2.1, 2.3)
1 (single seg)	30	2.1	0.44	(1.9, 2.2)
2 (≥ 2 seg)	54	2.0	0.46	(1.9, 2.1)
3 (major involved)	27	1.8	0.37	(1.7, 2.0)

			Adjusted using the Tukey HSD	
Comparison	Difference	SE	CI	*p*-value
2 - 3	0.18	0.07	(-0.1, 0.4)	0.2895
1 - 3	0.25	0.07	(-0.04, 0.5)	0.1116
0 - 3	0.35	0.07	(0.1, 0.6)	0.0017
1 - 2	0.08	0.06	(-0.2, 0.3)	0.8558
0 - 2	0.18	0.06	(-0.02, 0.4)	0.0996
0 - 1	0.10	0.08	(-0.1, 0.3)	0.6917

Assumptions:

- Samples is representative, independent, sample size is large enough (check data distribution and sample size for each group), assume that variances are equal based on

Decision:

- We will reject the null hypothesis in favor of the alternative hypothesis that there is at least one mean difference between the groups.

Results from Multiple Comparisons: “”

- Using the Tukey method, of all the tested groups, group 0 and group 3 group mean platelet counts were significantly different. The p-value was 0.0017.

Discussion

Substantive statement of conclusion:

- The largest difference in means between the groups is between group 0 and group 0 with a difference of 0.35. The smallest difference between groups is group 2 and group 1 with a difference of 0.08. There was a significant difference in mean blood platelet levels between patients with no gangrene and patients with the majority of their intestinal tract affected with gangrene.

Implication of results

- Going forward pediatric health care professional should take into account the patients condition when considering treatment for platelet count.

Problem 1	**Score**
Main Point	
Improperly Formatted Tables, No references to tables, No informative titles for tables	-1 (each)
Rounding	-1

Introduction

Topic or population introduced:

- "Infants with Thrombocytopaneia"
    - Thrombocytopaenia is a condition of abnormally low platelets that often occurs during necrotizing enterocolids (NEC)- a serious illness in infants that can cause tissue damage to the intestines.

Research question aptly described:

- Patients were grouped into four categories of NEC status. Group 0 referred to infants with no gangrene, group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, group 2 referred to patients with the majority of small and large bowel involved
- Which (if any) of the four groups has a different mean platelet count (on the log-scale given in the Table)?
    - Testing to see whether at least of one of the mean in one of the research groups is different than the means of any of the other research groups

Methods

Description of Outcome:

- Outcome Variable: logarithmic platelet level count
- Predicator Variable: infant "group status"
    - Patients were grouped into four categories of NEC status. Group 0 referred to infants with no gangrene, group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, group 2 referred to patients with the majority of small and large bowel involved

Description of Group:

- Group A (or group 0) refers to infants with no gangrene
- Group B (or group 1) refers to subjects with in whom gangrene was limited to a single intestinal segment
- Group C (or group 2) refers to subjects with two or more intestinal segments of gangrene
- Group D (or group 3) refers to patients with the majority of small/large bowel involved

Description of Data Summary

Verification of Linear Association:

- Checked with three scatter plots of age against each variable.

Statement of Null Hypotheses H_0:

(i) There is no association between age and attitude.

(ii) There is no association between age and social behavior.

(iii) There is no association between age and hyperactivity.

Statement of Alternative Hypotheses H_A:

(i) There is an association between age and attitude.

(ii) There is an association between age and social behavior.

(iii) There is an association between age and hyperactivity.

Statistical Method for testing for association:

- Pearson's linear correlation coefficient

Decision Rule:

- If the resulting p-value from the Pearson's correlation test is less than 0.05, we will reject the null hypothesis in favor of the alternative hypothesis; otherwise, we will fail to reject the null hypothesis.

Method of computation: R statistical software

- Version 2.11.1

Significance level: $\alpha = 0.05$

Results

Data Summary

Variable	n	Mean	SD	95% CI
Age	31	9.55	1.88	(8.86, 10.24)
Attitude	31	0.33	0.69	(0.80, 0.58)
Hyperactivity	31	-0.45	0.74	(-0.72, -0.18)
Social Behavior	31	0.36	0.51	(0.17, 0.54)

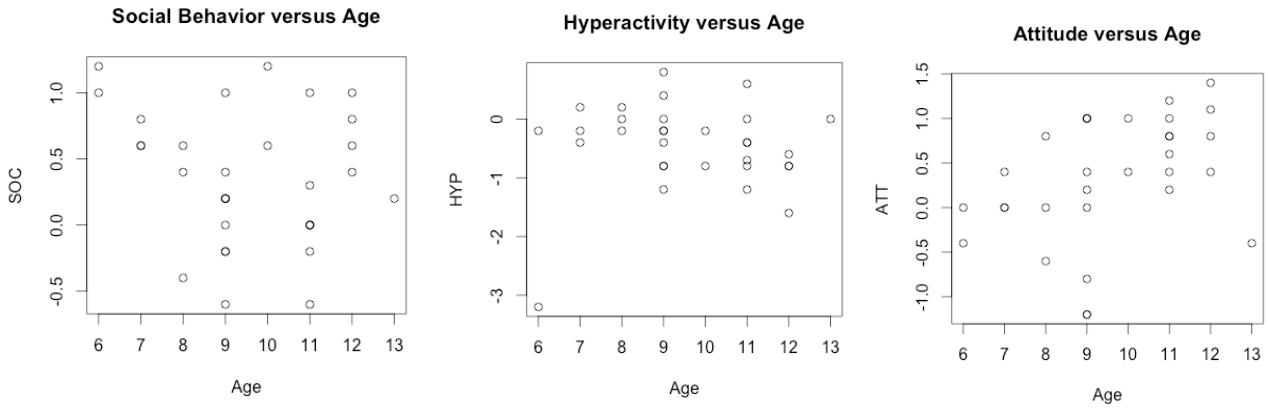
Assumptions

- Samples are representative, independent, sample size is adequate (n >30), linearity of Relationship, and heteroskedasticity.

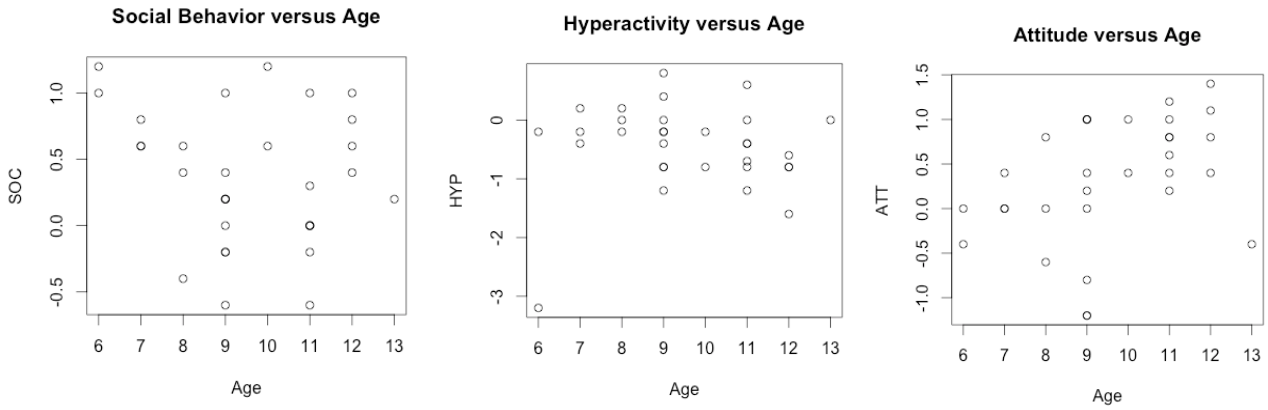
Test Results:

Variable	r	*p*-value
Age and ATT	0.399	0.0262
Age and HYP	0.010	0.9566
Age and SOC	-0.185	0.3187

Include 3 figures here for Scatterplots between age and ATT, age and HYP, and age and SOC



Include 3 figures here for Scatterplots between age and ATT, age and HYP, and age and SOC



Decision:

- Attitude is the only variable of the three that resulted in a p-value than was less than the alpha level of 0.05. That being said, there seems to be an association between Age and Attitude with regards to treatment therapy in children with hyperactivity disorder. There were no associations found between Age & Social Behavior and Age & Hyperactivity.

Description of Association:

- The correlation coefficient for Age and Attitude is 0.399, which is significantly different than 0

R-Code:

```
################################################################################################
#Skyler Kuhn
#STAT 543: Assignment 7
#HW5: Analysis of Variance (ANOVA) and Correlation
################################################################################################
#########
#Problem 1.) ANOVA question


#Reading in the dataset
thrombocytopaenia.data <- read.csv(file.choose(), header = TRUE)
##Output:
#    NECGroup Platelet
# 1         A     1.97
# 2         A     0.85
# 3         A     1.79
# ...
# 68        B     1.38
# 69        B     1.86
# 70        B     2.26
# ...
# 98        C     1.87
# 99        C     1.90
# 100       C     2.43
# ...
# 152       D     1.77
# 153       D     1.68
# 154       D     1.46
# ...
# 178       D     2.36
```

```
##About the groups:

# Group 0 referred to infants with no gangrene and is labeled as Group A

# Group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, Group 1 is labeled as Group B

# Group 2 referred to patients with two or more intestinal segments of gangrene, Group2 is labeled as Group C and

# Group 3 referred to patients with the majority of small and large bowel involved, Group 3 is labeled as GroupD.


table(thrombocytopaenia.data)


#Creating table for each group (A, B, C, D):

groupA.data <- thrombocytopaenia.data[thrombocytopaenia.data$NECGroup == "A", ]

groupB.data <- thrombocytopaenia.data[thrombocytopaenia.data$NECGroup == "B", ]

groupC.data <- thrombocytopaenia.data[thrombocytopaenia.data$NECGroup == "C", ]

groupD.data <- thrombocytopaenia.data[thrombocytopaenia.data$NECGroup == "D", ]

##############################################

#####GroupA data:

#About: Group 0 referred to infants with no gangrene and is labeled as Group A

library(car)


#Finding the sample size:

nrow(groupA.data)

##Output:

# 67  (greater than 30)


#Checking the distribution of Data:

hist(groupA.data$Platelet,

    main = "Histogram of groupA data", # changes the Legend Header/title #comma is neeeded

    xlab = "log10 platelet counts",   #changes the X-axis label

    col = "lightblue")

#data is normally distributed
```

```
qqPlot(groupA.data$Platelet)

##data is normally distributed


#Determining the Variance (looking at sd similarty and bartlett test results):

sd(groupA.data$Platelet) #Output: 0.3971375

sd(groupB.data$Platelet) #Output: 0.4383884

sd(groupC.data$Platelet) #Output: 0.4635171

sd(groupD.data$Platelet) #Output: 0.3706927

bartlett.test(thrombocytopaenia.data$Platelet ~ thrombocytopaenia.data$NECGroup)

# p-value = 0.5077, Result is highly indicative (much greater than 0.05), it suggests equal variances

################################################

#####GroupB data:

#About: Group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, Group 1 is labeled as Group B


library(car)


#Finding the sample size:

nrow(groupB.data)

##Output:

#30  (sample size is equal to 30?)


#Checking the distribution of Data:

hist(groupB.data$Platelet,

    main = "Histogram of groupB data", # changes the Legend Header/title #comma is neeeded

    xlab = "log10 platelet counts",   #changes the X-axis label

    col = "lightblue")

#data is normally distributed

qqPlot(groupB.data$Platelet)

##data is normally distributed
```

```
#Determining the Variance (looking at sd similarty and bartlett test results):

sd(groupA.data$Platelet) #Output: 0.3971375

sd(groupB.data$Platelet) #Output: 0.4383884

sd(groupC.data$Platelet) #Output: 0.4635171

sd(groupD.data$Platelet) #Output: 0.3706927

#all standard deviations are silimar, equal variance

bartlett.test(thrombocytopaenia.data$Platelet ~ thrombocytopaenia.data$NECGroup)

# p-value = 0.5077, Result is highly indicative (much greater than 0.05), it suggests equal variances

################################################

#####GroupC data:

#About: Group 2 referred to patients with two or more intestinal segments of gangrene, Group2 is labeled as Group C


library(car)


#Finding the sample size:

nrow(groupC.data)

##Output:

#54 (sample size is equal to 30)


#Checking the distribution of Data:

hist(groupC.data$Platelet,

    main = "Histogram of groupC data", # changes the Legend Header/title #comma is neeeded

    xlab = "log10 platelet counts",   #changes the X-axis label

    col = "lightblue")

#data is normally distributed

qqPlot(groupC.data$Platelet)

##data is normally distributed


#Determining the Variance (looking at sd similarty and bartlett test results):

sd(groupA.data$Platelet) #Output: 0.3971375
```

```
sd(groupB.data$Platelet) #Output: 0.4383884

sd(groupC.data$Platelet) #Output: 0.4635171

sd(groupD.data$Platelet) #Output: 0.3706927

#all standard deviations are silimar, equal variance

bartlett.test(thrombocytopaenia.data$Platelet ~ thrombocytopaenia.data$NECGroup)

# p-value = 0.5077, Result is highly indicative (much greater than 0.05), it suggests equal variances

################################################

#####GroupD data:

#About: Group 3 referred to patients with the majority of small and large bowel involved, Group 3 is
labeled as GroupD.


library(car)


#Finding the sample size:

nrow(groupD.data)

##Output:

#27 (sample size is less than to 30!!!!)


#Checking the distribution of Data:

hist(groupD.data$Platelet,

    main = "Histogram of groupD data", # changes the Legend Header/title #comma is neeeded

    xlab = "log10 platelet counts",   #changes the X-axis label

    col = "lightblue")

#data is normally distributed

qqPlot(groupD.data$Platelet)

##data is normally distributed


#Determining the Variance (looking at sd similarty):

sd(groupA.data$Platelet) #Output: 0.3971375

sd(groupB.data$Platelet) #Output: 0.4383884

sd(groupC.data$Platelet) #Output: 0.4635171
```

```
sd(groupD.data$Platelet) #Output: 0.3706927
#all standard deviations are silimar, equal variance
bartlett.test(thrombocytopaenia.data$Platelet ~ thrombocytopaenia.data$NECGroup)
# p-value = 0.5077, Result is highly indicative (much greater than 0.05), it suggests equal variances
################################################

########DATA SUMMARY:
#Since the sample size/normality is equal for each group, we summarize with:
##sample size, mean standard deviation and 95%CI

#Group A Data Summary:
nrow(groupA.data)          #sample size = 67
mean(groupA.data$Platelet)   #mean = 2.182985
sd(groupA.data$Platelet)     #standard deviation = 0.3971375
t.test(groupA.data$Platelet, alternative = "two.sided", var.equal = TRUE)
#95% CI: (2.086116, 2.279855)

#Group B Data Summary:
nrow(groupB.data)          #sample size = 30
mean(groupB.data$Platelet)   #mean = 2.081333
sd(groupB.data$Platelet)     #standard deviation = 0.4383884
t.test(groupB.data$Platelet, alternative = "two.sided", var.equal = TRUE)
#95% CI: (1.917636, 2.245030)

#Group C Data Summary:
nrow(groupC.data)          #sample size = 54
mean(groupC.data$Platelet)   #mean = 2.004815
sd(groupC.data$Platelet)     #standard deviation = 0.4635171
t.test(groupC.data$Platelet, alternative = "two.sided", var.equal = TRUE)
#95% CI: (1.878299, 2.131331)
```

```
#Group D Data Summary:

nrow(groupD.data)           #sample size = 27

mean(groupD.data$Platelet)   #mean = 1.828519

sd(groupD.data$Platelet)     #standard deviation = 0.3706927

t.test(groupD.data$Platelet, alternative = "two.sided", var.equal = TRUE)

#95% CI: (1.681877, 1.975160)


#########Equal Variance ANOVA TEST:

thrombocytopaenia.data.ANOVA <- aov(thrombocytopaenia.data$Platelet ~
thrombocytopaenia.data$NECGroup)

#Ouput:

# Call:

#aov(formula = thrombocytopaenia.data$Platelet ~ thrombocytopaenia.data$NECGroup)

# Terms:

#   thrombocytopaenia.data$NECGroup Residuals

# Sum of Squares                  2.637573 30.942439

# Deg. of Freedom                        3       174

# Residual standard error: 0.4216991

# Estimated effects may be unbalanced

summary(thrombocytopaenia.data.ANOVA)

##Output:

#                                Df Sum Sq  Mean Sq  F value  Pr(>F)

# thrombocytopaenia.data$NECGroup   3  2.638   0.8792   4.944     0.00255 **   #####p-value less than
0.05 (reject Ho)

# Residuals                    174 30.942  0.1778

##Becasue the F-test gave us a p-value less than 0.05, we will reject the Null Hypothesis in favor of the
alternative

#Now than we know one of the samples means are different, we will use a multiple comparision test to find
out which one it is.


#Tukey Test (multiple comparsion test, to find the different sample):
```

```
TukeyHSD(thrombocytopaenia.data.ANOVA, ordered = TRUE)   #diff show the difference of the means, report the standard error (not the sd here)

##Output:

# Tukey multiple comparisons of means

# 95% family-wise confidence level

# Fit: aov(formula = thrombocytopaenia.data$Platelet ~ thrombocytopaenia.data$NECGroup)

# $`thrombocytopaenia.data$NECGroup`

#     diff        lwr       upr     p adj

# C-D 0.17629630 -0.08153854 0.4341311 0.2895170

# B-D 0.25281481 -0.03736872 0.5429984 0.1115784

# A-D 0.35446656  0.10510902 0.6038241 0.0017074    #Significant (< 0.05)!!

# B-C 0.07651852 -0.17257360 0.3256106 0.8557834

# A-C 0.17817026 -0.02187868 0.3782192 0.0996042

# A-B 0.10165174 -0.13865486 0.3419583 0.6916619


#################################################################################################

#Problem 2.) Correlation question: perform analysis to determine if there is evidience to indiciate that the age (years)

##is correlated with any of the threee outcome variables. alpha = 0.05


#Ho: ρ = 0, there is no association.

#Ha: ρ ≠ 0, there is an association.


#Reading in the dataset

hyperchildren.data <- read.csv(file.choose(), header = TRUE)

##Output:

#    Subject AGE  ATT  HYP  SOC

# 1        1   9 -1.2 -1.2  0.0     #negative HYP (hyperactivity) score indicates an improvement in hyperactivity

# 2        2   9  0.0  0.0  1.0     #postive ATT (attitude) or SOC (social behavior) score indicates an improvement in hyperactivity
```

```
# 3       3  13 -0.4  0.0  0.2
# 4       4   6 -0.4 -0.2  1.2
# 5       5   9  1.0 -0.8  0.2
# 6       6   8  0.8  0.2  0.4
# 7       7   8 -0.6 -0.2  0.6
# 8       8   9 -1.2 -0.8 -0.6
# 9       9   7  0.0  0.2  0.8
# 10     10  12  0.4 -0.8  0.4
#Size:
nrow(hyperchildren.data) #Sample Size: 31
mean(hyperchildren.data$AGE)
mean(hyperchildren.data$HYP)
mean(hyperchildren.data$SOC)


#Seeing if there is an Assoication (3 outcomes) with a scatterplots:
plot(hyperchildren.data$AGE, hyperchildren.data$ATT,
    xlab = "Age",
    ylab = "ATT",
    main = "Attitude versus Age")
#slight positive correlation
plot(hyperchildren.data$AGE, hyperchildren.data$HYP,
    xlab = "Age",
    ylab = "HYP",
    main = "Hyperactivity versus Age")
#graph indicates a very slight correlation
plot(hyperchildren.data$AGE, hyperchildren.data$SOC,
    xlab = "Age",
    ylab = "SOC",
    main = "Social Behavior versus Age")
#graph indcates no correlation
```

############## Hypothesis Testing:

#Test of Correlation between ages the three outcome variables (attention, hyperactivity, and social behavior):

#Ho: $\rho = 0$, there is no association.

#Ha: $\rho \neq 0$, there is an association.

#Testing Attitude versus Age:

```
cor.test(hyperchildren.data$AGE,hyperchildren.data$ATT,
        alternative = "two.sided",
        method = "pearson")
```

##Output:

```
#Pearson's product-moment correlation
# data:  hyperchildren.data$AGE and hyperchildren.data$ATT
# t = 2.3436, df = 29, p-value = 0.02616      #####p-value is less than 0.05, reject the null#####
# alternative hypothesis: true correlation is not equal to 0
# 95 percent confidence interval:
#  0.05207162 0.66005750
# sample estimates:
#       cor
# 0.3990489
```

##Reject the null in favor of the alternative hypothesis#####

#Testing Hyperactivity versus Age:

```
cor.test(hyperchildren.data$AGE,hyperchildren.data$HYP,
        alternative = "two.sided",
        method = "pearson")
```

##Output:

```
# Pearson's product-moment correlation
# data:  hyperchildren.data$AGE and hyperchildren.data$HYP
# t = 0.054905, df = 29, p-value = 0.9566      #####p-value is greater than 0.05, fail to reject the null#####
```

```
# alternative hypothesis: true correlation is not equal to 0
# 95 percent confidence interval:
#  -0.3453927  0.3632231
# sample estimates:
#        cor
# 0.01019513
```

##Fail to reject the Null hypothesis##

#Testing Social Behavior versus Age:

```
cor.test(hyperchildren.data$AGE,hyperchildren.data$SOC,
        alternative = "two.sided",
        method = "pearson")
```

##Output:

```
# Pearson's product-moment correlation
# data:  hyperchildren.data$AGE and hyperchildren.data$SOC
# t = -1.0147, df = 29, p-value = 0.3187      #####p-value is greater than 0.05, fail to reject the null#####
# alternative hypothesis: true correlation is not equal to 0
# 95 percent confidence interval:
#  -0.5062822  0.1810605
# sample estimates:
#        cor
# -0.1851589
```

##Fail to reject the Null hypothesis##