**Assignment on Statistics and Modeling using R**

**BY**

**Radha Swathi Kadavakolanu**

Introduction

Data is imported in R and summarised and visualised as given below.

#Read data

rm(list = ls())

setwd('/Users/rsklanu/Randstat/SRM')

Cash<-read.csv("cashtransfer.csv",header=TRUE)

#Summarize

summary(Cash)

#Data visualization

par(mfrow=c(2,2))

#Histogram

hist(Cash$bplcaloriebeforect)

hist(Cash$aplcaloriebeforect)

hist(Cash$bplcalorieafterct)

hist(Cash$aplcalorieafterct)

#boxplot

boxplot(Cash$bplcaloriebeforect)

boxplot(Cash$aplcaloriebeforect)

boxplot(Cash$bplcalorieafterct)

boxplot(Cash$aplcalorieafterct)

Result:

ummary(Cash)

bplhhid bplcaloriebeforect bplcalorieafterct aplhhid aplcaloriebeforect aplcalorieafterct

Min. : 1.0 Min. :2011 Min. :1938 Min. : 1.0 Min. :2009 Min. :2013

1st Qu.:14.5 1st Qu.:2221 1st Qu.:2293 1st Qu.:14.5 1st Qu.:2212 1st Qu.:2273

Median :28.0 Median :2296 Median :2454 Median :28.0 Median :2328 Median :2435

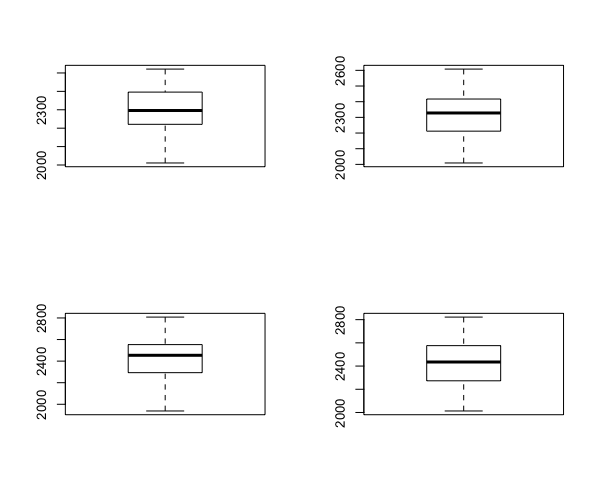
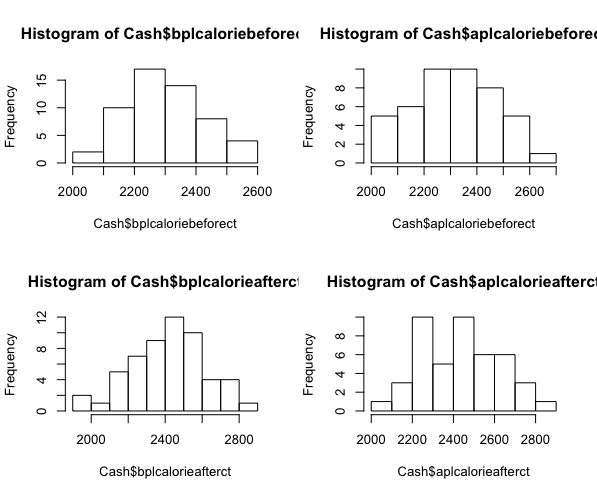
Mean :28.0 Mean :2305 Mean :2425 Mean :28.0 Mean :2315 Mean :2434

3rd Qu.:41.5 3rd Qu.:2396 3rd Qu.:2553 3rd Qu.:41.5 3rd Qu.:2417 3rd Qu.:2576

Max. :55.0 Max. :2521 Max. :2808 Max. :55.0 Max. :2608 Max. :2822

NA's :10 NA's :10

>



Question 1

Test whether BPL and APL households are consuming more than 2300 calories per day

We need to find if BPL and APL consume more than 2300 calories per day, this can be determined by one side t-test for each category.

Case 1:

BPL before cash transfer

R code:

t.test(Cash$bplcaloriebeforect,mu=2300,alternative = " greater")

Result:

One Sample t-test

data: Cash$bplcaloriebeforect

t = 0.29939, df = 54, p-value = 0.3829

alternative hypothesis: true mean is greater than 2300

95 percent confidence interval:

2277.134 Inf

sample estimates:

mean of x 2304.982

Conclusion:

Above result indicates the calories consumed per day on an average is 2304.982 and 38.29% consumes more than 2300 calories per day. As the P-Value is not lesser than 0.05 we retain the Null hypothesis.

Case 2:

BPL after cash transfer

R code:

t.test(Cash$bplcalorieafterct,mu=2300,alternative = "greater")

Result:

One Sample t-test

data: Cash$bplcalorieafterct

t = 4.7491, df = 54, p-value = 7.738e-06

alternative hypothesis: true mean is greater than 2300

95 percent confidence interval:

2380.774 Inf

sample estimates:

mean of x

2424.727

Conclusion:

Above result indicates the calories consumed per day on an average is 2424.727 , however p-value indicates the probability of picking this value is less than 5% hence we reject the null hypothesis.

Case 3:

APL before cash transfer

R code:

t.test(Cash$aplcaloriebeforect,mu=2300,alternative = "greater")

Result:

One Sample t-test

data: Cash$aplcaloriebeforect

t = 0.62675, df = 44, p-value = 0.267

alternative hypothesis: true mean is greater than 2300

95 percent confidence interval:

2275.609 Inf

sample estimates:

mean of x

2314.511

Conclusion:

Above result indicates the calories consumed per day on an average is 2314.511, p-value indicates the probability of picking this value is 26.7% hence we retain the null hypothesis.

Case 4:

APL after cash transfer

R code:

t.test(Cash$aplcalorieafterct,mu=2300,alternative = "greater")

Result:

One Sample t-test

data: Cash$aplcalorieafterct

t = 4.5776, df = 44, p-value = 1.918e-05

alternative hypothesis: true mean is greater than 2300

95 percent confidence interval:

2385.11 Inf

sample estimates:

mean of x

2434.467

Conclusion:

Above result indicates the calories consumed per day on an average is 2434.467, however p-value indicates the probability of picking this value is less than 5% hence we reject the null hypothesis.

Solution: After carefully examining above test we can say calories consumed per day is 2300 on an average for both BPL and APL.

Question-2

Is there any difference in per capita nutrition value (measured in calories per day) between pre and post implementation of

cash transfer mechanism within each group?

Case 1:

Comparison of per capita nutrition value before and after act for BPL

R code:

t.test(Cash$bplcaloriebeforect,Cash$bplcalorieafterct)

Result:

Welch Two Sample t-test

data: Cash$bplcaloriebeforect and Cash$bplcalorieafterct

t = -3.8514, df = 91.338, p-value = 0.0002178

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-181.50103 -57.98988

sample estimates:

mean of x mean of y

2304.982 2424.727

Conclusion: P value is less than 0.05 hence we reject the null hypothesis which indicates there is difference between per captia nutrition consumed before and after the act was enforced.

Case 2:

Comparison of per capita nutrition value before and after act for APL

R code:

t.test(Cash$aplcaloriebeforect,Cash$aplcalorieafterct)

Result:

Welch Two Sample t-test

data: Cash$aplcaloriebeforect and Cash$aplcalorieafterct

t = -3.2072, df = 83.446, p-value = 0.001901

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-194.34144 -45.56967

sample estimates:

mean of x mean of y

2314.511 2434.467

Conclusion: P value is less than 0.05 hence we reject the null hypothesis which indicates there is difference between per captia nutrition consumed before and after the act was enforced.

Solution: Not only we can arrive at the conclusion that there is a difference in per captia nutrition consumption we can compare the mean values and state it’s a positive difference for both BPL and APL.

Question-3

Is there any difference in the per-capita value of calorific consumption across the two segments, BPL and APL?

Case 1:

Comparison of per capita nutrition value before and after act for BPL

R code:

t.test(Cash$bplcaloriebeforect,Cash$aplcaloriebeforect)

Result:

Welch Two Sample t-test

data: Cash$bplcaloriebeforect and Cash$aplcaloriebeforect

t = -0.33422, df = 83.124, p-value = 0.7391

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-66.23788 47.17930

sample estimates:

mean of x mean of y

2304.982 2314.511

Conclusion: P value is greater than 0.05 hence we retain the null hypothesis which indicates there is difference between per captia nutrition consumed for BPL and APL

Case 2:

Comparison of per capita nutrition value before and after act for BPL

R code:

t.test(Cash$bplcalorieafterct,Cash$aplcalorieafterct)

Result:

Welch Two Sample t-test

data: Cash$bplcalorieafterct and Cash$aplcalorieafterct

t = -0.24717, df = 93.683, p-value = 0.8053

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-87.97953 68.50074

sample estimates:

mean of x mean of y

2424.727 2434.467

Conclusion: P value is greater than 0.05 hence we retain the null hypothesis which indicates there is difference between per captia nutrition consumed for BPL and APL

Solution: There is no clear evidence to prove the relation between nutrition consumption and class , that is APL and BPL.

R File:

#Read data

rm(list = ls())

setwd('/Users/rsklanu/Randstat/SRM')

Cash<-read.csv("cashtransfer.csv",header=TRUE)

#Summarize

summary(Cash)

#Data visualization

par(mfrow=c(2,2))

#Histogram

hist(Cash$bplcaloriebeforect)

hist(Cash$aplcaloriebeforect)

hist(Cash$bplcalorieafterct)

hist(Cash$aplcalorieafterct)

#boxplot

boxplot(Cash$bplcaloriebeforect)

boxplot(Cash$aplcaloriebeforect)

boxplot(Cash$bplcalorieafterct)

boxplot(Cash$aplcalorieafterct)

#Test

#question 1

t.test(Cash$bplcaloriebeforect,mu=2300,alternative = "greater")

t.test(Cash$bplcalorieafterct,mu=2300,alternative = "greater")

t.test(Cash$aplcaloriebeforect,mu=2300,alternative = "greater")

t.test(Cash$aplcalorieafterct,mu=2300,alternative = "greater")

#question 2

t.test(Cash$bplcaloriebeforect,Cash$bplcalorieafterct)

t.test(Cash$aplcaloriebeforect,Cash$aplcalorieafterct)

#question 3

t.test(Cash$bplcaloriebeforect,Cash$aplcaloriebeforect)

t.test(Cash$bplcalorieafterct,Cash$aplcalorieafterct)

#Variance

var.test(Cash$bplcaloriebeforect,Cash$aplcaloriebeforect,ratio=1)

var.test(Cash$bplcalorieafterct,Cash$aplcalorieafterct,ratio=1)