Problem Set 6

Prateek Kumar

Contents

Problem 1: Categorical relationships 2
Determine whether responses to indoor preference influenced individual response to outdoor preference .. 3
Determine whether indoor and outdoor preferences were the same 5
Problem 2: Which test to do: You decide. 7
Is there an impact of gender on the type of car purchased? 8
Is there a difference in amount paid for a car for men versus women? 9
Do people tend to buy vehicles from of the same origin as their last vehicle (US, Europe, Japan)? 12
Is there a relationship between driver age and the value of his car? 13
What is your best estimate for the value of a car driven by a 32, 52, and 62-year-old? 16
Is there a relationship between how much someone paid for their previous car and how much they paid for their current car? 16
Did people tend to pay more for their current car than their previous car? 19
Did trucks cost more than SUVs? 20

Problem 1: Categorical relationships

```r
set.seed(1000)
indoor<- sample(c("A","B","C","D"),prob=c(5,8,25,6),size=200,replace=T)
outdoor <-sample(c("A","B","C","D"),prob=c(5,20,3,6),size=200,replace=T)
tab1 <- table(indoor,outdoor)   #cross-tabulations of each pair of responses


#what responses were given for indoor and outdoor across the four options
tab2 <- table(c(indoor,outdoor),rep(c("I","O"),each=200))


#Changing the row and column names

row.names(tab1) <- c('clowns','horses','chocolate fountain','magician')
colnames(tab1) <- c('clowns','horses','chocolate fountain','magician')
row.names(tab2) <- c('clowns','horses','chocolate fountain','magician')


addmargins(tab1)

##                     outdoor
## indoor                clowns horses chocolate fountain magician Sum
##   clowns                   4     11                  3        5  23
##   horses                   9     23                  1       10  43
##   chocolate fountain      19     59                 10       19 107
##   magician                 5     12                  2        8  27
##   Sum                     37    105                 16       42 200

tab2

##
##                        I   O
##   clowns              23  37
##   horses              43 105
##   chocolate fountain 107  16
##   magician            27  42
```

Here we have two tables: tab1 and tab2, tab1 compares the children indoor response across various outdoor responses so it's a cross tabulation of each pair of responses and tab2 gives us the details as what responses were given for both indoor and outdoor across the four options.

Determine whether responses to indoor preference influenced individual response to outdoor preference

```
#chisq.test(tab2[,1],p = tab2[,2],rescale.p = T)
#install.packages("corrplot")
library(corrplot)

library(BayesFactor)

ct_t1 <- chisq.test(tab1)

ct_t1 #Chi-squared test on tab1

##
##  Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 5.0836, df = 9, p-value = 0.827



bf_t1 <- contingencyTableBF(tab1, sampleType = 'indepMulti', fixedMargin = 'cols')
bf_t1 #bayes factor contingency table tests for tab1

## Bayes factor analysis
## --------------
## [1] Non-indep. (a=1) : 0.000346898 ±0%
##
## Against denominator:
##   Null, independence, a = 1
## ---
## Bayes factor type: BFcontingencyTable, independent multinomial

bf_t1@bayesFactor$bf  #bayes factor value

## [1] -7.96648



corrplot(ct_t1$residuals, is.cor = F)
```

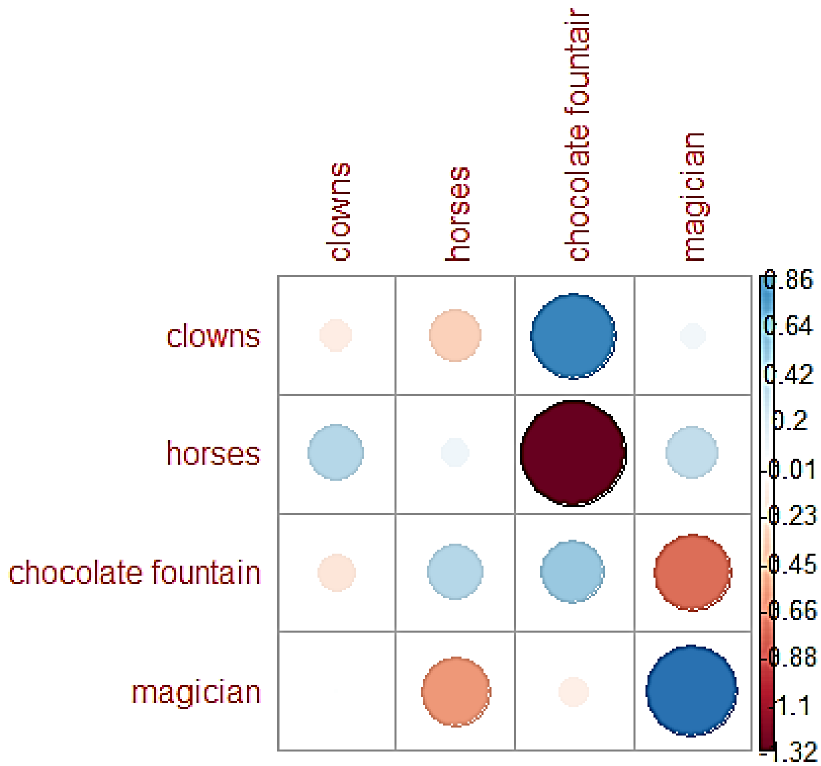

Figure 1: This is the visualization of Pearson residuals using the package corrplot. Positive residuals are in blue and negative residuals are in red. It is evident that there is an association between indoor choice of clowns and outdoor choice of chocolate fountain and indoor and outdoor choices of magicians and there is repulsion between the outdoor choice of chocolate fountain and indoor choice of horses.

Inorder to determine if responses to indoor preference influenced individual response to outdoor preference we use tab1 because its the cross tabulation of each pair of responses.

Here I performed both the chi-squared tests and bayes factor contingency table tests for tab1.

Null hypothesis(H0): The responses to indoor preference do not influence the individual response to outdoor preference.

Alternative hypothesis(H1): The responses to indoor preference influence the individual response to outdoor preference.

Chi-squared test:

- We see that the p-value = 0.827 which is much greater than 0.05
- Also we here have 9 degrees of freedom[(row-1)(column-1)] for which Chi-square (X-squared) value should be 16.92 (from the table) but here we get X-squared = 5.0836 which is less than the expected value

So this indicates weak evidence against the Null hypothesis(H0), so we fail to reject the Null hypothesis

Bayes factor contingency table test:

- We get the bayes factor value as -7.966 which is smaller than 0.00667

So this is also a very strong evidence for Null hypothesis.

And hence we see from the above results that both the Chi-squared test and bayes factor fail to reject the Null hypothesis so we can conclude that "***the responses to indoor preference do not influence the individual response to outdoor preference***".

Determine whether indoor and outdoor preferences were the same

```
ct_t2 <- chisq.test(tab2)
ct_t2 #Chi-squared test on tab2

##
##  Pearson's Chi-squared test
##
## data:  tab2
## X-squared = 99.826, df = 3, p-value < 2.2e-16



bf_t2 <- contingencyTableBF(tab2, sampleType = 'indepMulti', fixedMargin = 'cols')
bf_t2 #bayes factor contingency table tests for tab2

## Bayes factor analysis
## --------------
## [1] Non-indep. (a=1) : 1.21942e+21 ±0%
##
## Against denominator:
##   Null, independence, a = 1
## ---
## Bayes factor type: BFcontingencyTable, independent multinomial

bf_t2@bayesFactor$bf  #bayes factor value

## [1] 48.55266



corrplot(ct_t2$residuals, is.cor = F)
```

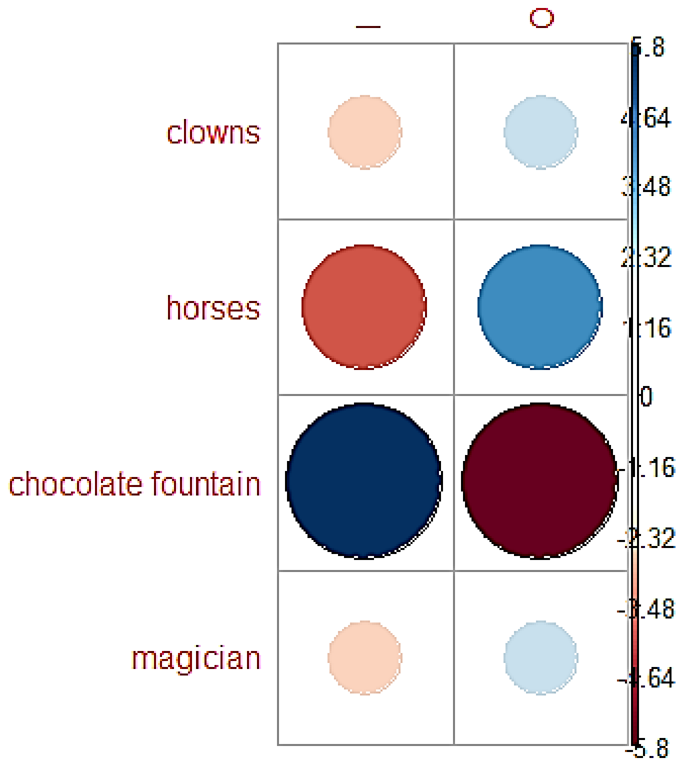

Figure 2: This is a similar figure as fig.1 but on tab2. It is evident that there is an association between indoor choice and choice of chocolate fountain and outdoor choice and choice of horses and there is repulsion between the outdoor choice and the choice of chocolate fountain.

In-order to determine if the indoor and outdoor preferences were same we use tab2 because it gives us the details as what responses were given for both indoor and outdoor across the four options.

Here I performed both the chi-squared tests and bayes factor contingency table tests for tab2.

Null hypothesis(H0): Indoor and outdoor preferences are not same.

Alternative hypothesis(H1): Indoor and outdoor preferences are same.

Chi-squared test:

- We see that the p-value = 0 which is less than 0.05
- Also we here have 3 degrees of freedom[(row-1)(column-1)] for which Chi-square (X-squared) value should be 7.81 (from the table) but here we get X-squared = 99.826 which is more than the expected value

So this indicates strong evidence against the Null hypothesis(H0), so we reject the Null hypothesis

Bayes factor contingency table test:

- We get the bayes factor value as 48.55266 which is between 20 to 150

So this gives a strong support for alternate hypothesis.

And hence we see from the above results that both the Chi-squared test and bayes factor indicate strong evidence to reject the Null hypothesis so we can conclude that “***the indoor and outdoor preferences are same***”.

Problem 2: Which test to do: You decide.

```
cardat <- read.table(text="age gender type origin origin.last carval carval.last
                     34 F SUV US US 16400 15800
                     31 M Truck US Europe 16900 16000
                     47 M Sedan US US 18800 17100
                     21 F Sedan Japan Japan 16000 15500
                     42 M SUV US Japan 16800 16100
                     43 F SUV US US 17200 16300
                     60 F Truck Europe Europe 19900 17800
                     37 M Truck Europe Europe 17100 16200
                     46 F SUV Japan Japan 16900 16300
                     27 M Sedan US US 16200 15700
                     50 M SUV US US 18800 17100
                     64 F SUV Japan US 50700 31700
                     33 M SUV Japan Japan 16500 15900
                     39 M Truck US Europe 17000 16200
                     58 F Sedan Japan US 19400 17500
                     53 F SUV US Europe 19200 17400
                     29 F Sedan US Japan 16300 15700
                     37 F Sedan US US 17300 16300
                     37 M SUV US Japan 18200 16700
                     54 F Sedan Japan Japan 24500 19800
                     46 F SUV Japan Europe 18000 16700
                     55 F SUV US Japan 28900 21700
                     46 F Truck US Europe 16600 16100
                     57 M SUV Europe Europe 24300 19700
                     40 M SUV US US 16800 16100
                     27 M Sedan Japan US 16900 16000
                     58 M SUV Europe Europe 20300 17900
                     64 M Truck US US 40600 27100
                     47 M Truck US Europe 18400 16900
                     32 M Truck US US 15900 15600
                     43 F Sedan Japan US 17200 16300
                     66 M Truck Europe Europe 19100 17500
                     36 F SUV US Japan 16900 16100
                     68 M Truck US US 69300 40100
                     54 F Sedan Japan US 17000 16400
                     64 M Truck Japan Europe 34900 24600
                     27 M SUV Japan Europe 15800 15500
                     51 F Sedan Japan Japan 29000 21700
                     69 M Sedan US Japan 54400 33400
                     25 F Sedan Japan Japan 15800 15500",header=T)
```

This is the cardat data. The column names are:

- age: age of car owner
- gender: gender of car owner
- type: type of vehicle

- origin: location car was manufactured
- origin.last: location of previous car's manufacture
- carval: purchase price of vehicle
- carval.last: purchase price of previous vehicle

Is there an impact of gender on the type of car purchased?

```
library(corrplot) #loading the corrplot and bayesfactor library

library(BayesFactor)

t1 <- table(cardat$gender, cardat$type)
t1 #extracting the table of gender w.r.t. car type

##
##     Sedan SUV Truck
##   F     9   8     2
##   M     4   8     9

ct_1 <- chisq.test(t1)
ct_1 #Chi-squared test on t1

##
##  Pearson's Chi-squared test
##
## data:  t1
## X-squared = 6.2934, df = 2, p-value = 0.04299

bf_1 <- contingencyTableBF(t1, sampleType = 'indepMulti', fixedMargin = 'cols')
bf_1 #bayes factor contingency table tests for t1

## Bayes factor analysis
## --------------
## [1] Non-indep. (a=1) : 3.723788 ±0%
##
## Against denominator:
##   Null, independence, a = 1
## ---
## Bayes factor type: BFcontingencyTable, independent multinomial

bf_1@bayesFactor$bf  #bayes factor value

## [1] 1.314741



corrplot(ct_1$residuals, is.cor = F)
```

Here I performed the non-parametric test viz. chi-squared test in-order to test the difference between the gender and type of car sample and bayes factor contingency table test for t1.

Null hypothesis(H0): There is no impact of gender on the type of car purchased

Alternative hypothesis(H1): Gender has an impact on the type of car purchased

Chi-squared test:

- We see that the p-value = 0.043 which is less than 0.05
- Also we here have 2 degrees of freedom[(row-1)(column-1)] for which Chi-square (X-squared) value should be 5.99 (from the table) but here we get X-squared = 6.29 which is more than the expected value

So this indicates strong evidence against the Null hypothesis(H0), so we reject the Null hypothesis

Bayes factor contingency table test:

- We get the bayes factor value as 1.315 which is between 0.333 and 3

So this gives an ambivalent support for alternate hypothesis.

But the Chi-sq test shows a slight bias towards the alternate hypothesis so we reject the Null hypothesis and can conclude that "***gender has an impact on the type of car purchased***".

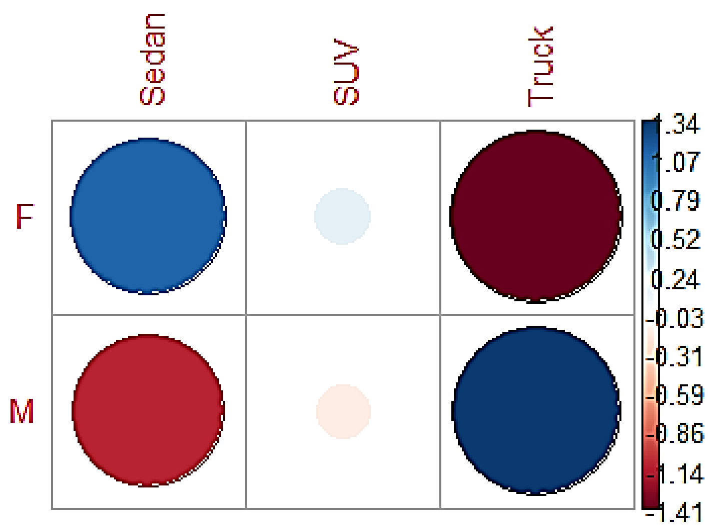

Figure 3: We can see from the plot that there is an association between females and sedans and males and trucks and vice versa.

Is there a difference in amount paid for a car for men versus women?

```
library(vioplot) #loading the vioplot and bayesfactor library

library(BayesFactor)

women <- subset(cardat, cardat$gender == "F") #extracting the female data
men <- subset(cardat, cardat$gender == "M") #extracting the male data
```

We have,

Null hypothesis(H0): There is a difference in the amount paid for a car for men vs women

Alternative hypothesis(H1): There is no difference in the amount paid for a car for men vs women

```
t.test(women$carval, men$carval, alternative = 'less') #comparing two independent samples

##
##  Welch Two Sample t-test
##
## data:  women$carval and men$carval
## t = -0.89186, df = 32.677, p-value = 0.1895
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 2925.704
## sample estimates:
## mean of x mean of y
##  20694.74  23952.38
```

T-test:

- We can see the p-value = 0.1895 which is more than 0.05

This indicates a weak evidence against the Null hypothesis.

```
t2_bf <- ttestBF(women$carval, men$carval) #bayesfactor test
t2_bf

## Bayes factor analysis
## --------------
## [1] Alt., r=0.707 : 0.4172006 ±0.01%
##
## Against denominator:
##   Null, mu1-mu2 = 0
## ---
## Bayes factor type: BFindepSample, JZS


t2_bf@bayesFactor$bf

## [1] -0.8741881
```

Bayesfactor test:

- We get the bayes factor value as -0.874 which is smaller than 0.00667

So this is also a very strong evidence for Null hypothesis.

And hence we see from the above results that both the t-test and bayes factor test fail to reject the Null hypothesis so we can conclude that "***there is a difference in the amount paid for a car for men vs women***".

```r
vioplot(women$carval, men$carval, names = c('female','male'),col = c('gold'))

title('Violin plot for car values of men and women')

grid()
```

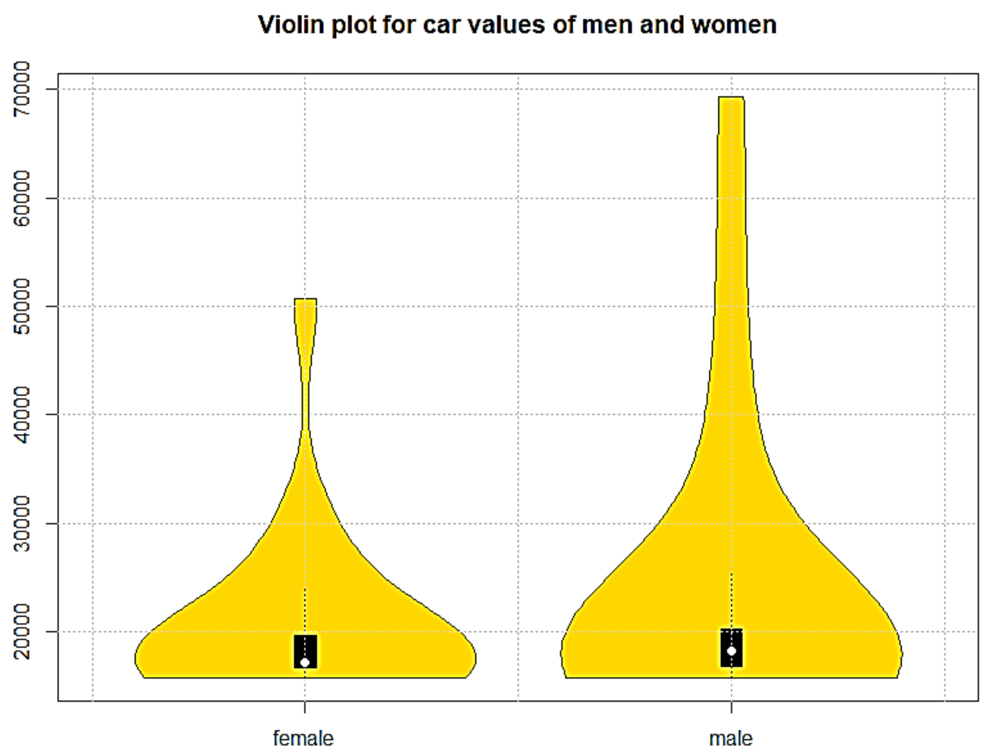

Figure 4: Car value of men vs women

```r
mean(men$carval) #mean amout of car of men and women

## [1] 23952.38

mean(women$carval)

## [1] 20694.74

length(men$carval) #number of men and women

## [1] 21

length(women$carval)

## [1] 19
```

From the above results also we can see that there are more men than women so obviously the total amount spent by men will be more but when we see then mean amount spent by both of them, we see that there is a difference of 3257.64 between them so this concludes our assumption and we can say that ***there is a difference in the amount paid for a car for men vs women***.

Do people tend to buy vehicles from of the same origin as their last vehicle (US, Europe, Japan)?

```
new <- cardat$origin #New vehicles origin
old <- cardat$origin.last #Previous vehicles origin
t3 <- table(old,new) #table for both the values
addmargins(t3)

##         new
## old      Europe Japan US Sum
##   Europe      5     3  5  13
##   Japan       0     6  6  12
##   US          0     5 10  15
##   Sum         5    14 21  40

ct_3 <- chisq.test(t3)

ct_3 #Chi-squared test on t3

##
##  Pearson's Chi-squared test
##
## data:  t3
## X-squared = 12.772, df = 4, p-value = 0.01245

bf_3 <- contingencyTableBF(t3, sampleType = 'indepMulti', fixedMargin = 'cols')
bf_3 #bayes factor contingency table tests for t3

## Bayes factor analysis
## --------------
## [1] Non-indep. (a=1) : 10.3086 ±0%
##
## Against denominator:
##   Null, independence, a = 1
## ---
## Bayes factor type: BFcontingencyTable, independent multinomial

bf_3@bayesFactor$bf  #bayes factor value

## [1] 2.332979

corrplot(ct_3$residuals, is.cor = F)
```

Here I performed the non-parametric test viz. chi-squared test in-order to test the difference between the previous car make and new car make and bayes factor contingency table test for t3.

Null hypothesis(H0): People do not buy vehicles from of the same origin as their last vehicle

Alternative hypothesis(H1): People buy vehicles from of the same origin as their last vehicle

Chi-squared test:

- We see that the p-value = 0.012 which is less than 0.05

- Also we here have 4 degrees of freedom[(row-1)(column-1)] for which Chi-square (X-squared) value should be 9.488 (from the table) but here we get X-squared = 12.772 which is more than the expected value

So this indicates strong evidence against the Null hypothesis(H0), so we reject the Null hypothesis

Bayes factor contingency table test:

- We get the bayes factor value as 2.333 which is between 0.333 and 3

So this gives an ambivalent support for alternate hypothesis.

But the Chi-sq test shows a slight bias towards the alternate hypothesis so we reject the Null hypothesis and can conclude that "***people buy vehicles from the same origin as their last vehicle***".

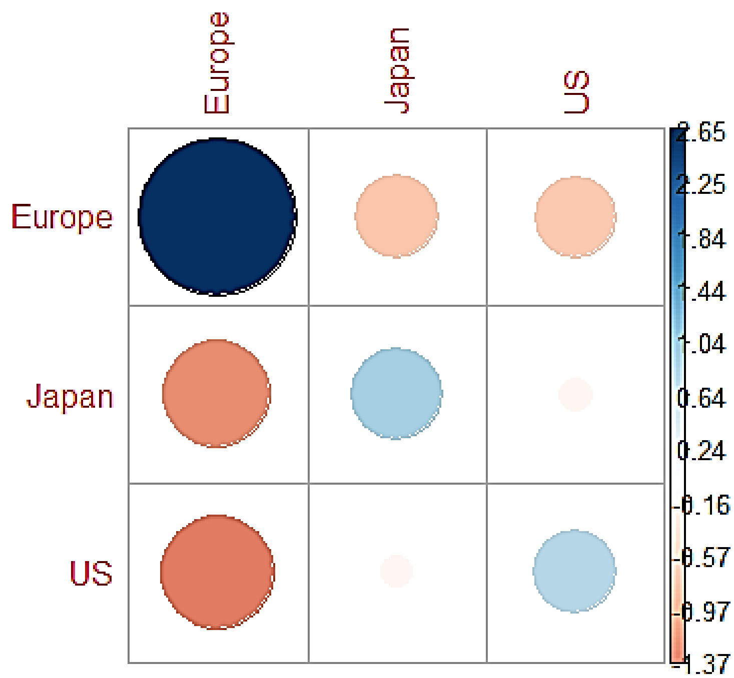

Figure 5: We can see that there is an association with the same origin of vehicles and with different origins we have red circles showing repulsion, so we can clearly accept the alternate hypothesis that most people buy vehicles from the same origin as their last vehicles

Is there a relationship between driver age and the value of his car?

```
t4 <- cardat[,c('age','carval')] #extracting age and car value

model1 <- lm(t4$carval~t4$age) #automatic fit
model1

##
## Call:
## lm(formula = t4$carval ~ t4$age)
##
## Coefficients:
```

```
## (Intercept)       t4$age
##     -4494.1        592.2
```

```
par(mfrow=c(1,1)) #plotting carval and age
plot(t4$age,t4$carval,pch=16,cex=1.5,col="gold",
     main=paste("Best-fitting line\n", "y = ",round(model1$coef[1] ,2) ," + ",
                round(model1$coef[2],3) , " * x \nPlot for age and carval",sep=""),
     type = 'p', xlab = 'Age', ylab = 'Car Value')

points(t4$age,t4$carval,pch=1,cex=1.5,col="grey20")
abline(model1$coef,lwd=2) #best fit line
```

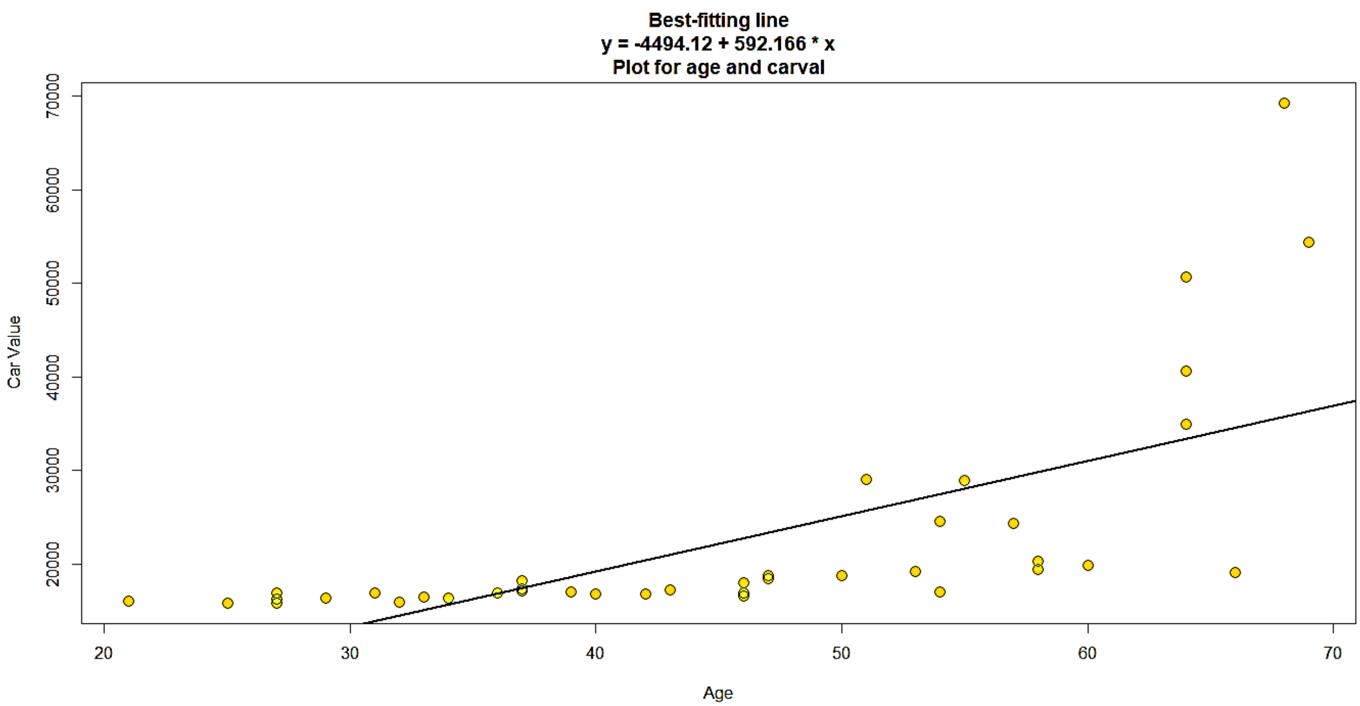

Figure 6: Plot of age and carval

```
summary(model1)

##
## Call:
## lm(formula = t4$carval ~ t4$age)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -15489  -4943   -216   3376  33527
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4494.1     5056.1  -0.889     0.38
## t4$age         592.2      106.9   5.538 2.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 8889 on 38 degrees of freedom
## Multiple R-squared:  0.4467, Adjusted R-squared:  0.4321
## F-statistic: 30.67 on 1 and 38 DF,  p-value: 2.446e-06

predictedy <- model1$coef[1] + model1$coef[2]* t4$age #predictions of value we did not see

plot(predictedy, t4$carval,cex=.5,col="grey20",pch=16,

     main = 'Plot for predictedy and carval',xlab = 'predictedy', ylab = 'carval')

points(predictedy,t4$carval,pch=1,cex=1,col="black")) #plotting them
abline(0,1)
```

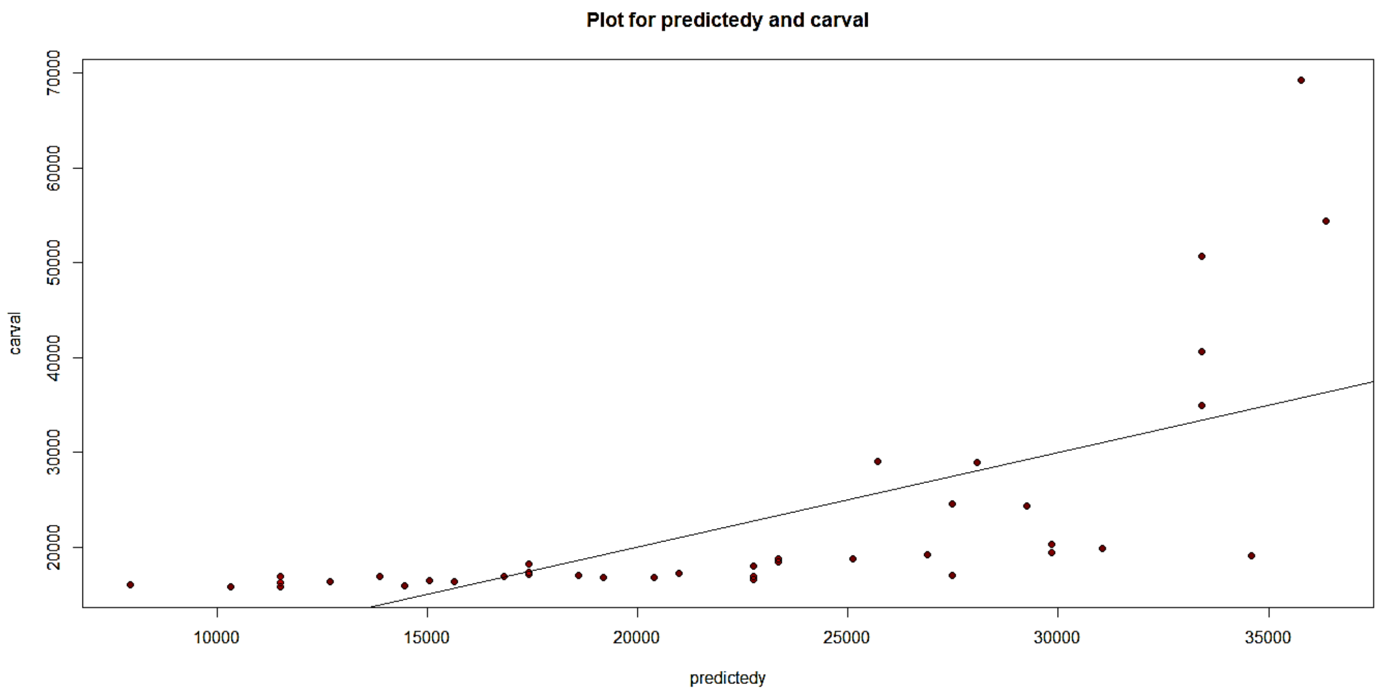

Figure 7: Plot of predicty and carval

```
cor(t4$carval,predictedy)^2

## [1] 0.4466648
```

From the above plots we can see that there is a relation between driver's age the value of his car, as the age increases the car value also increases. Moreover, we have a positive correlation of 0.45 which supports this point.

```
#Bayes factor test
x4 <- ttestBF(t4$age,t4$carval,paired = T)
x4@bayesFactor$bf

## [1] 27.11851
```

Applying bayesfactor test on the age and carval to confirm our result.

Null hypothesis(H0): There is no relationship between driver age and the value of his car

Alternative hypothesis(H1): There is a relationship between driver age and the value of his car

Bayes factor test:

- We get the bayes factor value as 27.12 which is between 20 to 150

So this gives a strong support for alternate hypothesis.

So we can conclude that "***there is a relationship between driver age and the value of his car***".

What is your best estimate for the value of a car driven by a 32, 52, and 62-year-old?

```
length(cardat$carval[cardat$age==32]) #driver with age 32

## [1] 1

length(cardat$carval[cardat$age==52]) #driver with age 52

## [1] 0

length(cardat$carval[cardat$age==62]) #driver with age 62

## [1] 0
```

We can see that we just have one value for 32-year-old and there are no values for 52 and 62 years old. So we cannot estimate the value of car because of the lack of data.

Is there a relationship between how much someone paid for their previous car and how much they paid for their current car?

```
t6 <- cardat[,c("carval","carval.last")] #extracting age and car value
cor.test(t6$carval, t6$carval.last)

##
##  Pearson's product-moment correlation
##
## data:  t6$carval and t6$carval.last
## t = 249.62, df = 38, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9994195 0.9998400
## sample estimates:
##       cor
## 0.9996952
```

The correlation is almost close to 1.

```
model2 <- lm(t6$carval~t6$carval.last) #automatic fit
model2
```

```
##
## Call:
## lm(formula = t6$carval ~ t6$carval.last)
##
## Coefficients:
##    (Intercept)  t6$carval.last
##     -18237.252           2.173
```

```
par(mfrow=c(1,1)) #plotting carval and age
plot(t6$carval.last,t6$carval,pch=16,cex=1.5,col="gold",
     main=paste("Best-fitting line\n", "y = ",round(model1$coef[1] ,2) ," + ",
                round(model1$coef[2],3) , " * x \nPlot for previous car value and current car
value ",sep=""),
xlab = 'Previous car value', ylab = 'Current car value')

points(t6$carval.last,t6$carval,pch=1,cex=1.5,col="black")
abline(model2$coef,lwd=2) #best fit line
```

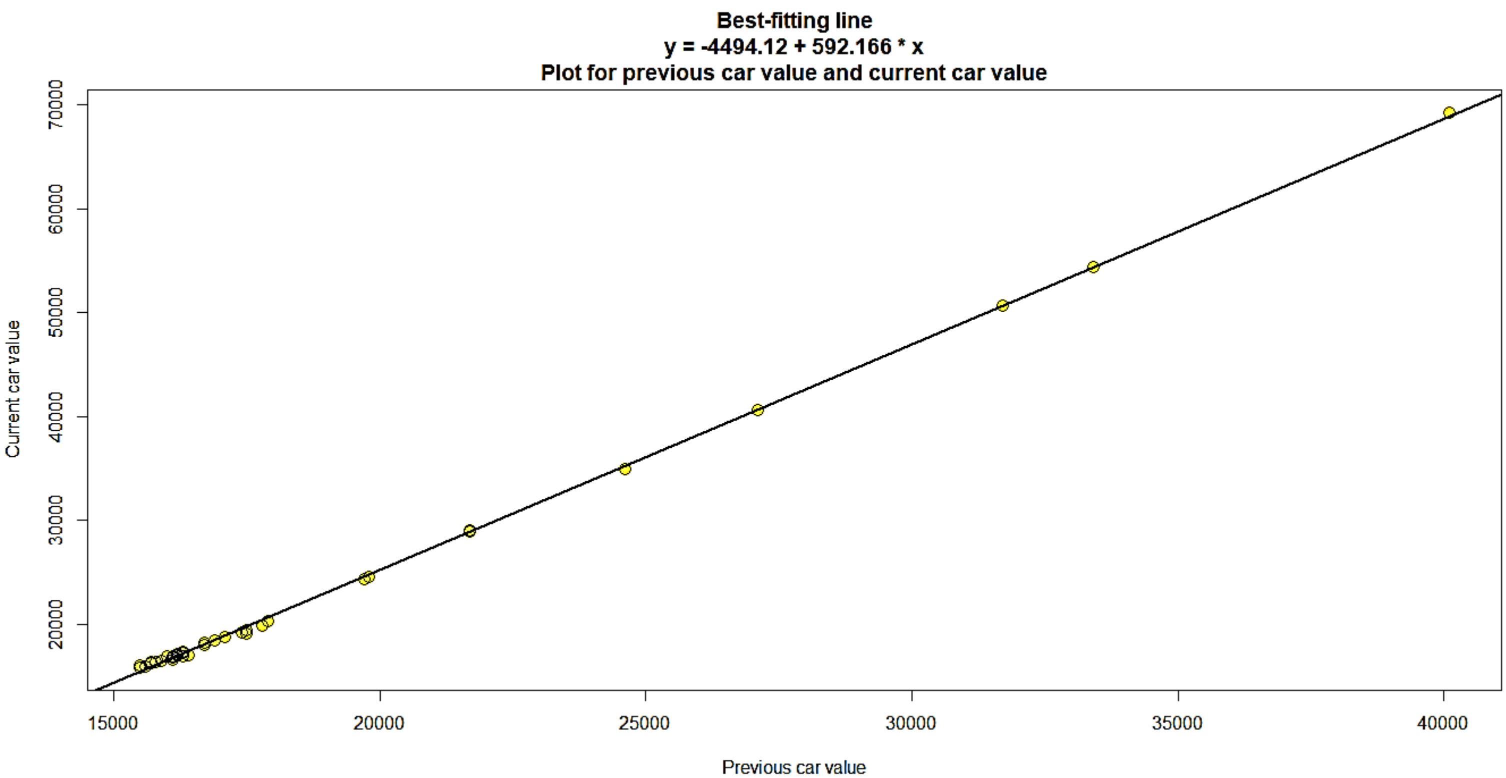

Figure 8: plot for previous car value and new car value

```
summary(model2)

##
## Call:
## lm(formula = t6$carval ~ t6$carval.last)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -696.94 -185.25    34.74   193.47   549.82
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.824e+04  1.694e+02  -107.7   <2e-16 ***
## t6$carval.last  2.173e+00  8.707e-03   249.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 295 on 38 degrees of freedom
## Multiple R-squared:  0.9994, Adjusted R-squared:  0.9994
## F-statistic: 6.231e+04 on 1 and 38 DF,  p-value: < 2.2e-16
```

```
#predictions of value we did not see
predictedy <- model2$coef[1] + model2$coef[2]* t6$carval.last


plot(predictedy, t6$carval,cex=2,col="black",pch=16, main = 'Plot for predictedy and carval',
xlab = 'predictedy', ylab = 'carval') #plotting them
points(predictedy,t6$carval,pch=1,cex=2,col="black")

abline(0,1)
```

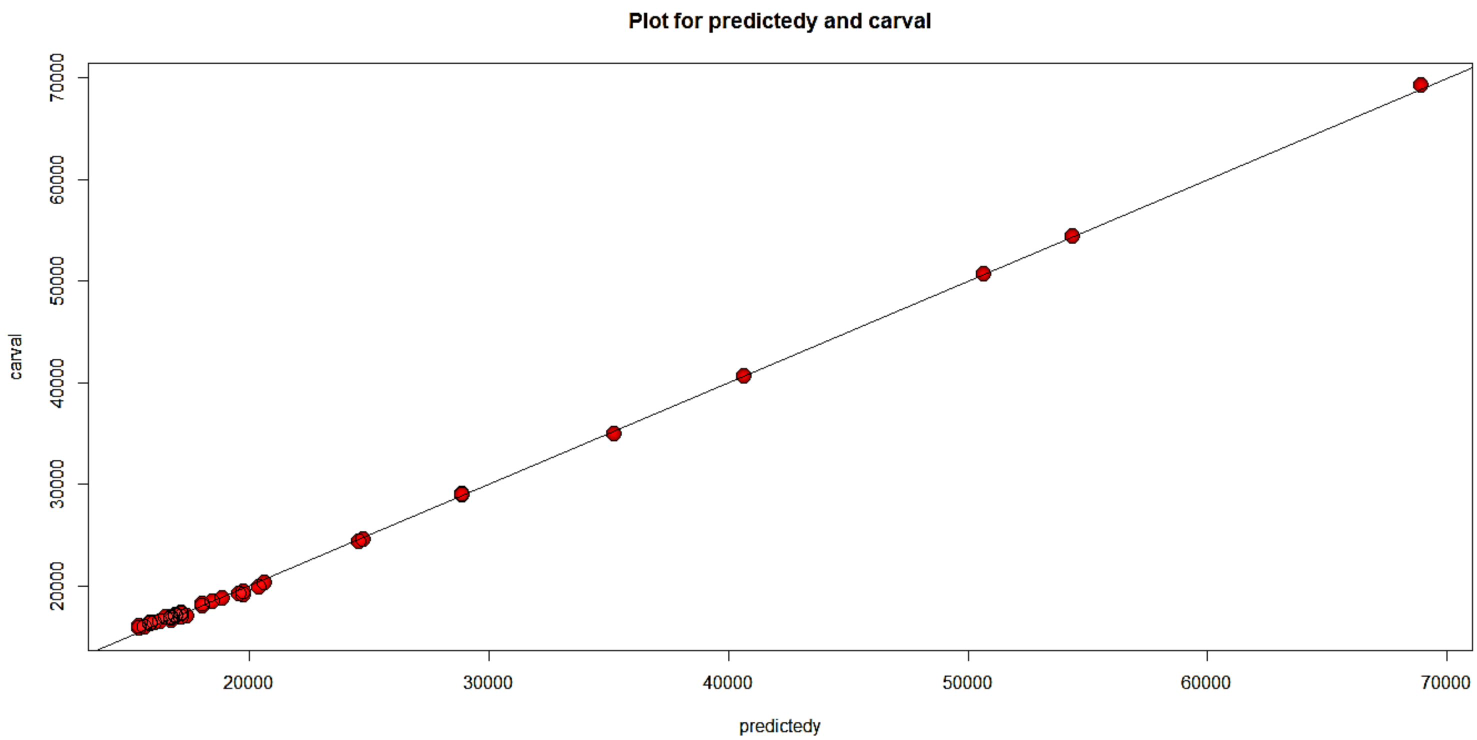

Figure 9: Plot for predicty and carval

```
cor(t6$carval,predictedy)^2

## [1] 0.9993905
```

From the above plots we can see that there is a strong almost linear relation between previous car value and the new car value, if the previous car value is more the new car value is also more and vice versa. Moreover, we have a positive correlation of 0.999 which supports this point.

```
#Bayes factor test
x6 <- ttestBF(t6$carval.last,t6$carval,paired = T)
x6@bayesFactor$bf

## [1] 3.72441
```

Applying bayesfactor test on the age and carval to confirm our result.

Null hypothesis(H0): There is no relationship between previous car value and the new car value

Alternative hypothesis(H1): There is a relationship between previous car value and the new car value

Bayes factor test:

- We get the bayes factor value as 3.72 which is between 3 to 20

So this gives a positive support for alternate hypothesis.

So we can conclude that "***there is a relationship between previous car value and the current car value***".

Did people tend to pay more for their current car than their previous car?

```
tmp1 <-
cbind(mean(cardat[which(cardat$type=='Sedan'),"carval"]),mean(cardat[which(cardat$type=='Seda
n'),"carval.last"]))
tmp2 <-
cbind(mean(cardat[which(cardat$type=='SUV'),"carval"]),mean(cardat[which(cardat$type=='SUV'),
"carval.last"]))
tmp3 <-
cbind(mean(cardat[which(cardat$type=='Truck'),"carval"]),mean(cardat[which(cardat$type=='Truc
k'),"carval.last"]))
t7 <- rbind(tmp1,tmp2,tmp3)
row.names(t7) <- c('Sedan','SUV','Truck')
colnames(t7) <- c('carval','carval.last')
```

```
t7

##         carval carval.last
## Sedan 21446.15    18223.08
## SUV   20731.25    17937.50
## Truck 25972.73    20372.73
```

Here I extracted the mean car values both previous and current car values with respect to the three car types.

```
ct_7 <- chisq.test(t7)
ct_7 #Chi-squared test on t7
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  t7
## X-squared = 58.534, df = 2, p-value = 1.948e-13
```

```
bf_7 <- contingencyTableBF(t7, sampleType = 'indepMulti', fixedMargin = 'cols')
bf_7 #bayes factor contingency table tests for t7
```

```
## Bayes factor analysis
## --------------
## [1] Non-indep. (a=1) : 407551219 ±0%
##
## Against denominator:
##   Null, independence, a = 1
## ---
## Bayes factor type: BFcontingencyTable, independent multinomial
```

```
bf_7@bayesFactor$bf  #bayes factor value
```

```
## [1] 19.82568
```

Here I performed the non-parametric test viz. chi-squared test in-order to test the difference in price between the previous car value and the current car value and bayes factor contingency table test for t7.

Null hypothesis(H0): People did not pay more for their current car as compared to their previous car

Alternative hypothesis(H1): People did pay more for their current car as compared to their previous car

Chi-squared test:

- We see that the p-value = 0 which is less than 0.05
- Also we here have 2 degrees of freedom[(row-1)(column-1)] for which Chi-square (X-squared) value should be 5.991 (from the table) but here we get X-squared = 58.534 which is more than the expected value

So this indicates strong evidence against the Null hypothesis(H0), so we reject the Null hypothesis

Bayes factor contingency table test:

- We get the bayes factor value as 19.825 which is between 3 and 20

So this gives a positive support for alternate hypothesis.

So from the results of Chi-sq test and bayes factor test we reject the Null hypothesis and we can conclude that "***People did pay more for their current car as compared to their previous car***".

Did trucks cost more than SUVs?

```
q1<-subset( cardat, cardat$type == 'Truck')
t8 <- cbind(sum(q1$carval),sum(q1$carval.last))

q2 <- subset( cardat, cardat$type == 'SUV')
temp <- cbind(sum(q2$carval),sum(q2$carval.last)) #combining the sum of prices of trucks and
suv
```

```r
t8 <- rbind(t8,temp)
row.names(t8) <- c('Truck','SUV')
colnames(t8) <- c('carval','carval.last')
t8

##       carval carval.last
## Truck 285700      224100
## SUV   331700      287000
```

We can see that trucks cost less than SUV in both previous price and current price.

```r
truck <- q1 #truck data
suv <- q2 #suv data
```

Here I just extracted truck data and suv data into truck and suv variables respectively.

We have,

Null hypothesis(H0): Trucks did not cost more than SUV

Alternative hypothesis(H1): Trucks cost more than SUV

```r
t.test(truck$carval, suv$carval, alternative = 'less') #comparing current prices

## 
##  Welch Two Sample t-test
## 
## data:  truck$carval and suv$carval
## t = 0.96255, df = 13.809, p-value = 0.8238
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 14841.94
## sample estimates:
## mean of x mean of y 
##  25972.73  20731.25
```

T-test:

- We can see the p-value = 0.8238 which is more than 0.05

This indicates a weak evidence against the Null hypothesis.

```r
t8_bf <- ttestBF(truck$carval, suv$carval) #bayesfactor test
t8_bf

## Bayes factor analysis
## --------------
## [1] Alt., r=0.707 : 0.5538505 ±0%
## 
## Against denominator:
##   Null, mu1-mu2 = 0 
## ---
## Bayes factor type: BFindepSample, JZS

t8_bf@bayesFactor$bf

## [1] -0.5908605
```

Bayesfactor test:

- We get the bayes factor value as -0.59 which is smaller than 0.00667

So this is also a very strong evidence for Null hypothesis.

And hence we see from the above results that both the t-test and bayes factor test fail to reject the Null hypothesis so we can conclude that "current price of trucks did not cost more than current price of SUV".

```
t.test(truck$carval.last, suv$carval.last, alternative = 'less') #comparing previous prices

##
##  Welch Two Sample t-test
##
## data:  truck$carval.last and suv$carval.last
## t = 0.97615, df = 13.873, p-value = 0.8271
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 6832.059
## sample estimates:
## mean of x mean of y
##  20372.73  17937.50
```

T-test:

- We can see the p-value = 0.8271 which is more than 0.05

This also indicates a weak evidence against the Null hypothesis.

```
t8_bf <- ttestBF(truck$carval.last, suv$carval.last) #bayesfactor test
t8_bf

## Bayes factor analysis
## --------------
## [1] Alt., r=0.707 : 0.5599386 ±0%
##
## Against denominator:
##   Null, mu1-mu2 = 0
## ---
## Bayes factor type: BFindepSample, JZS

t2_bf@bayesFactor$bf

## [1] -0.8741881
```

Bayesfactor test:

- We get the bayes factor value as -0.874 which is smaller than 0.00667

So this is also a very strong evidence for Null hypothesis.

And hence here as well we see from the above results that both the t-test and bayes factor test fail to reject the Null hypothesis so we can conclude that "previous price of trucks did not cost more than previous price of SUV".

Henceforth, we conclude that ***the total cost of SUV is more than total cost of trucks***.

```r
vioplot(truck$carval, suv$carval, names = c('truck','suv'),col = c('gold'))
title('Violinplot for current price of trucks and suv')
grid()
```

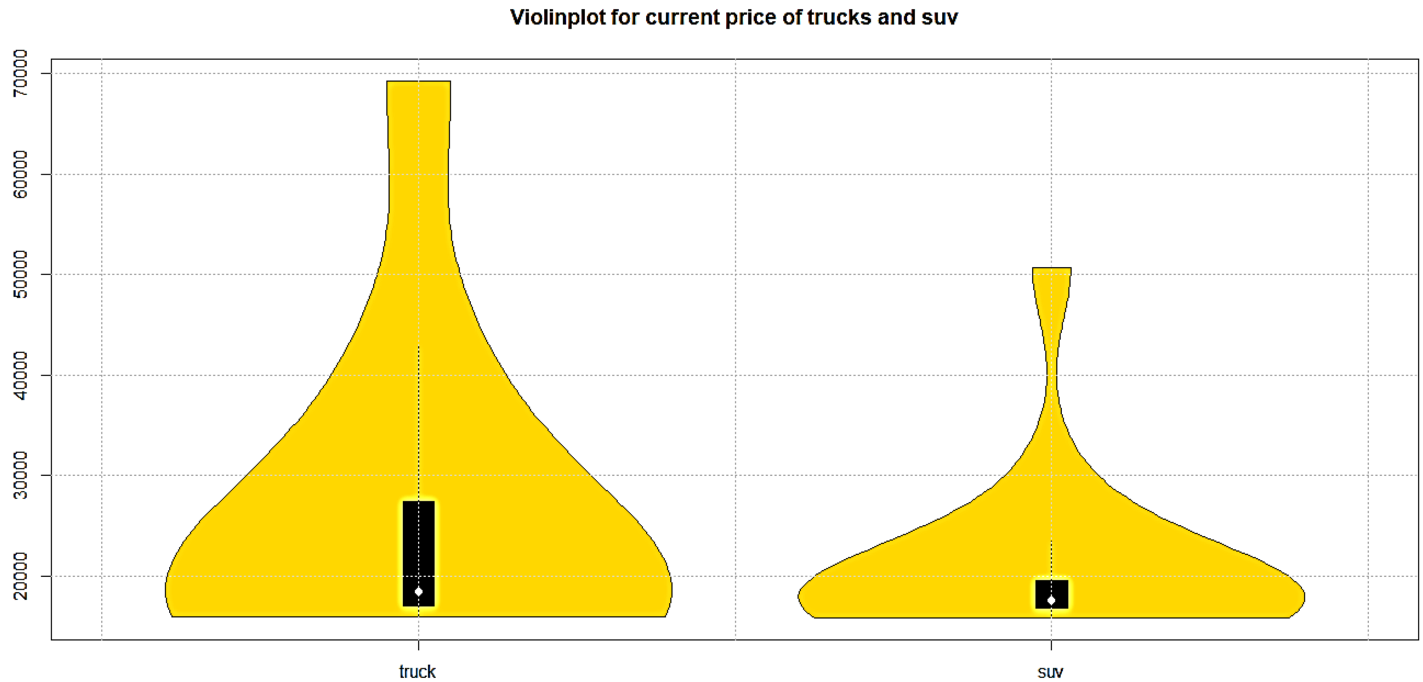

Figure 10: Violin plot of current price of Truck and SUV

```r
vioplot(truck$carval.last, suv$carval.last, names = c('truck','suv'),col = c('gold'))
title('Violinplot for previous price of trucks and suv')
grid()
```

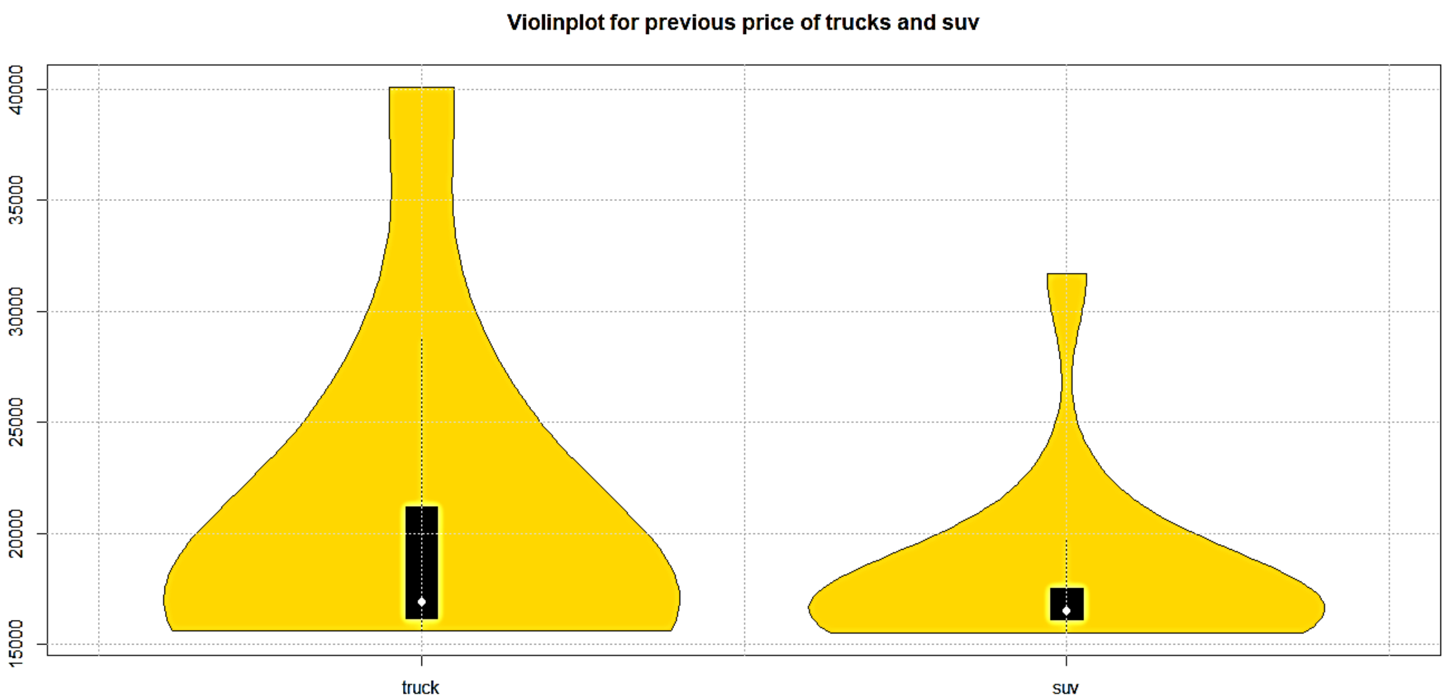

Figure 11: : Violin plot of previous price of Truck and SUV