PS6Q2

Prateek Kumar

21 October 2018

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)

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

library(corrplot)

## corrplot 0.84 loaded

library(BayesFactor)

## Loading required package: coda

## Loading required package: Matrix

## \*\*\*\*\*\*\*\*\*\*\*\*  
## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).  
##   
## Type BFManual() to open the manual.  
## \*\*\*\*\*\*\*\*\*\*\*\*

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

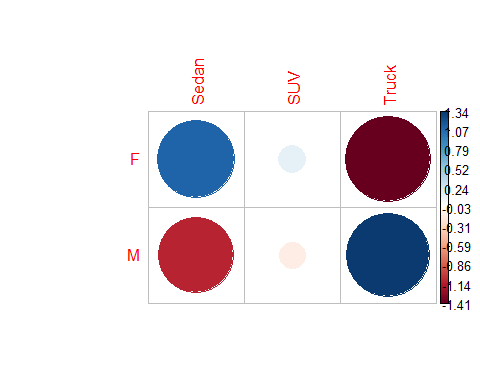
bf\_1 <- contingencyTableBF(t1, sampleType = 'indepMulti', fixedMargin = 'cols')   
bf\_1 #bayes factor contingency table tests for t1

## 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)



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

library(vioplot)

## Loading required package: sm

## Package 'sm', version 2.2-5.5: type help(sm) for summary information

library(BayesFactor)  
  
t2 <- table(cardat$carval,cardat$gender)   
  
women <- subset(cardat, cardat$gender == "F")  
men <- subset(cardat, cardat$gender == "M")  
#overall differences  
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

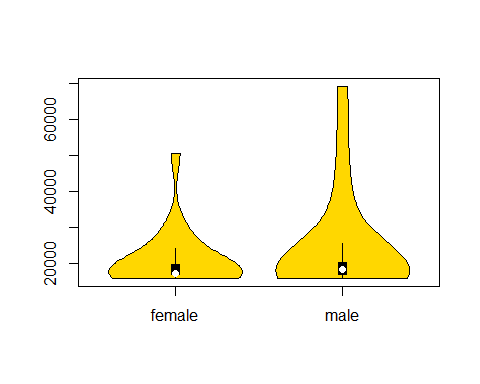
t2\_bf <- ttestBF(women$carval, men$carval) #bayesfactor test  
t2\_bf

## 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

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



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

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

new <- cardat$origin  
old <- cardat$origin.last  
t3 <- table(old,new)  
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)

## Warning in chisq.test(t3): Chi-squared approximation may be incorrect

ct\_3 #Chi-squared test on t3

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

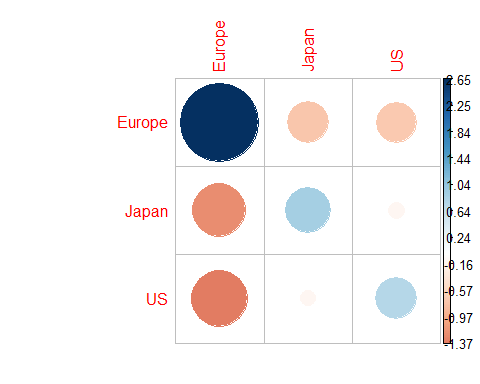
bf\_3 <- contingencyTableBF(t3, sampleType = 'indepMulti', fixedMargin = 'cols')   
bf\_3 #bayes factor contingency table tests for t3

## 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)

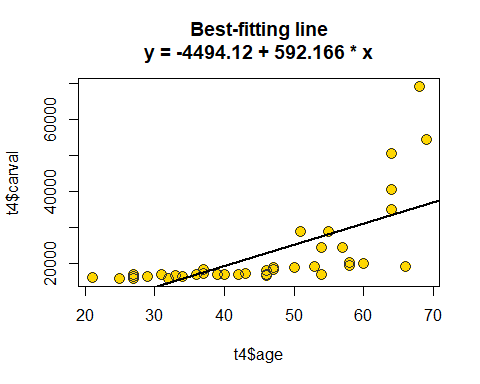


# Is there a relationship between driver age and the value of his car?

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

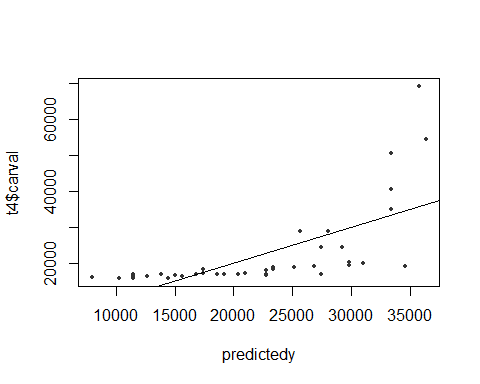
par(mfrow=c(1,1)) #plotting carval and age   
plot(t4$age,t4$carval,pch=16,cex=1.5,col="gold",  
 main=paste("Best-fitting line\n", "y = ",round(model1$coef[1] ,2) ," + ",   
 round(model1$coef[2],3) , " \* x",sep=""))  
  
points(t4$age,t4$carval,pch=1,cex=1.5,col="grey20")  
abline(model1$coef,lwd=2) #best fit line



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) #plotting them  
abline(0,1)



cor(t4$carval,predictedy)^2

## [1] 0.4466648

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

## [1] 27.11851

# 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])

## [1] 1

length(cardat$carval[cardat$age==52])

## [1] 0

length(cardat$carval[cardat$age==62])

## [1] 0

# 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)

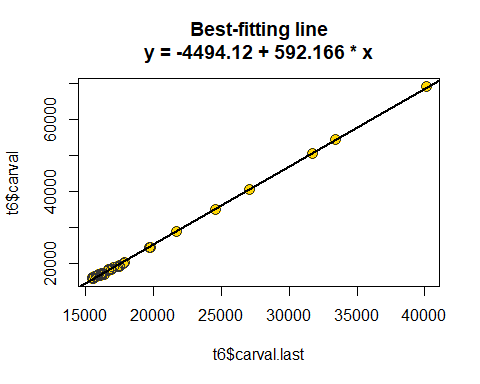
##   
## Pearson's product-moment correlation  
##   
## data: t6$carval and t6$carval.last  
## t = 249.62, df = 38, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9994195 0.9998400  
## sample estimates:  
## cor   
## 0.9996952

The correlation is almost perfect

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

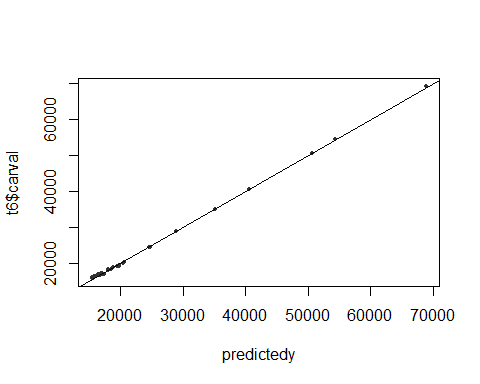
par(mfrow=c(1,1)) #plotting carval and age   
plot(t6$carval.last,t6$carval,pch=16,cex=1.5,col="gold",  
 main=paste("Best-fitting line\n", "y = ",round(model1$coef[1] ,2) ," + ",   
 round(model1$coef[2],3) , " \* x",sep=""))  
  
points(t6$carval.last,t6$carval,pch=1,cex=1.5,col="grey20")  
abline(model2$coef,lwd=2) #best fit line



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

predictedy <- model2$coef[1] + model2$coef[2]\* t6$carval.last #predictions of value we did not see  
  
plot(predictedy, t6$carval,cex=.5,col="grey20",pch=16) #plotting them  
abline(0,1)



cor(t6$carval,predictedy)^2

## [1] 0.9993905

ascnlan

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

## [1] 3.72441

# 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=='Sedan'),"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=='Truck'),"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

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

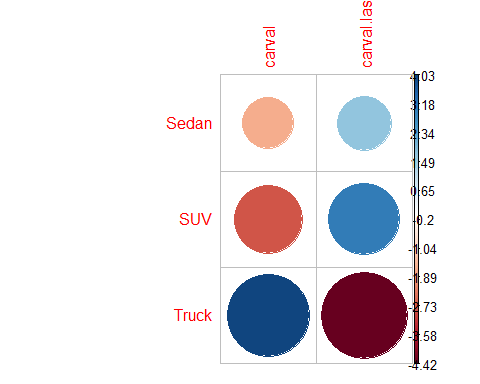
bf\_7 <- contingencyTableBF(t7, sampleType = 'indepMulti', fixedMargin = 'cols')   
bf\_7 #bayes factor contingency table tests for t7

## 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

corrplot(ct\_7$residuals, is.cor = F)



# 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  
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.

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

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

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

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

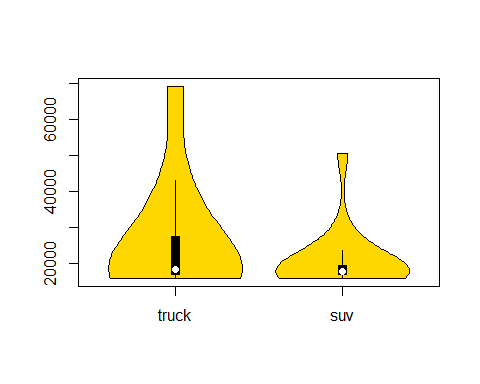
t8\_bf <- ttestBF(truck$carval.last, suv$carval.last) #bayesfactor test  
t8\_bf

## 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

vioplot(truck$carval, suv$carval, names = c('truck','suv'),col = c('gold'))



vioplot(truck$carval.last, suv$carval.last, names = c('truck','suv'),col = c('gold'))

