lab9\_stat123

Koki Itagaki

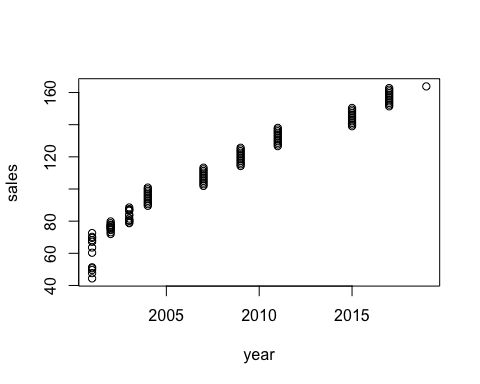
2023-03-21

#1. Load the sales.csv dataset into R and save it to df.

df<-read.csv("/Users/itagakikouki/stat123/lab9/sales.csv")  
head(df)

## id year month sales  
## 1 1 2001 1 44.39524  
## 2 2 2001 2 47.69823  
## 3 3 2001 3 49.58708  
## 4 4 2001 4 50.70508  
## 5 5 2001 5 51.29288  
## 6 6 2001 6 67.15065

year<-df$year  
sales<-df$sales  
month<-df$month  
#(a) Define the relationship between month and sales in df. Please identify direction, form, and strength of the relationship.  
plot(year, sales)



cor(year,sales)

## [1] 0.9739005

#direction is increasing,the form is linear   
#and this is strong trelationship.  
  
  
  
  
#(b) Perform a linear regression sales (y) and year (x1) and month(x2).  
model<-lm(sales~year+month)  
  
  
#(c) Predict the sales in March 2010.  
#we know exp\_y = -1141 + 5.599\*year + 1.155\*month  
y = -11141.030 + 5.599\*2010 + 1.155\*3  
coef = as.matrix(round(model$coefficients,3))  
coef

## [,1]  
## (Intercept) -11141.030  
## year 5.599  
## month 1.155

y = coef[1] + coef[2]\*2010 + coef[3]\*3  
cat(y,"is an expected value")

## 116.425 is an expected value

#2. Use the built-in data set HairEyeColor to answer this question.

#(a) Create a single table called hair\_eye\_totals which summarizes the total number of statistics students with each combination of hair and eye colour. Note: The built-in data set consists of two tables with this information (one for women and one for men). The answer to part (a) is a single table combining the information from these two tables.  
data("HairEyeColor")  
HairEyeColor

## , , Sex = Male  
##   
## Eye  
## Hair Brown Blue Hazel Green  
## Black 32 11 10 3  
## Brown 53 50 25 15  
## Red 10 10 7 7  
## Blond 3 30 5 8  
##   
## , , Sex = Female  
##   
## Eye  
## Hair Brown Blue Hazel Green  
## Black 36 9 5 2  
## Brown 66 34 29 14  
## Red 16 7 7 7  
## Blond 4 64 5 8

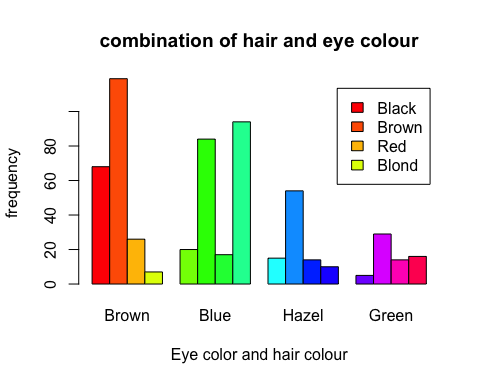
dim(HairEyeColor)

## [1] 4 4 2

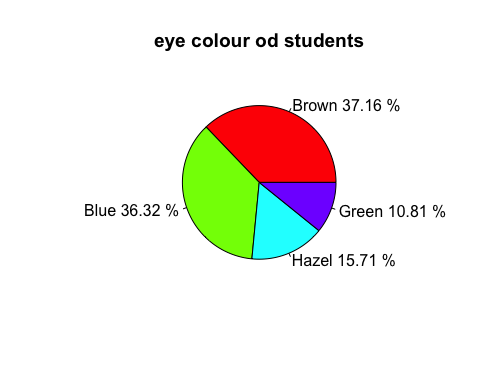
#male female tgt  
hair\_eye\_totals = HairEyeColor[,,1] + HairEyeColor[,,2]  
  
#(b) Print out the hair\_eye\_totals table.  
print(hair\_eye\_totals)

## Eye  
## Hair Brown Blue Hazel Green  
## Black 68 20 15 5  
## Brown 119 84 54 29  
## Red 26 17 14 14  
## Blond 7 94 10 16

#(c) Create a grouped bar plot which displays the information from the hair\_eye\_totals table. Your plot should include the following: - a main title - titles for the x-axis and y-axis - colours to help differentiate the bars - a legend to identify what each colour represents  
#xaxis is eye color  
barplot(hair\_eye\_totals,main = "combination of hair and eye colour",xlab = "Eye color and hair colour",ylab ="frequency" ,col = rainbow(length(hair\_eye\_totals)),legend = rownames(hair\_eye\_totals),beside = T)



#(d) Create and print out a vector called percent\_eye which contains the percent of statistics students with each eye colour (rounded to 2 decimal places). Show any additional code needed to create this vector.  
eyetotal = colSums(hair\_eye\_totals)  
  
percent\_eye =round(eyetotal/sum(eyetotal)\*100,2)  
eye\_label<- paste(names(percent\_eye),percent\_eye,"%")  
  
  
#(e) Create a pie chart displaying the information in the percent\_eye vector. Your graph should include:- a main title- labels for each wedge displaying the eye colour- a different colour for each eye colour - the percentages displaying next to each wedge.  
 pie(percent\_eye, main = "eye colour od students", labels = eye\_label,col = rainbow(length(percent\_eye)))



1. Use the mtcars data set in R to answer this question.

data(mtcars)  
dim(mtcars)

## [1] 32 11

head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

#(a) Create a multi regression model for predicting miles per gallon (mpg) using weight (wt) and horsepower (hp).  
model<-lm(mpg~wt + hp, data = mtcars)  
summary(model)

##   
## Call:  
## lm(formula = mpg ~ wt + hp, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.941 -1.600 -0.182 1.050 5.854   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.22727 1.59879 23.285 < 2e-16 \*\*\*  
## wt -3.87783 0.63273 -6.129 1.12e-06 \*\*\*  
## hp -0.03177 0.00903 -3.519 0.00145 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.593 on 29 degrees of freedom  
## Multiple R-squared: 0.8268, Adjusted R-squared: 0.8148   
## F-statistic: 69.21 on 2 and 29 DF, p-value: 9.109e-12

#(b) What is the interpretation of the coefficient estimates for the linear regression model part(a), which predicted miles per gallon (mpg) using weight (wt) and horsepower (hp) as predictor variables  
  
"The intercept coefficient of 37.22727 represents the predicted mpg when both wt and hp are zero. Which is not meaningful scenario but is included in the model for completeness."

## [1] "The intercept coefficient of 37.22727 represents the predicted mpg when both wt and hp are zero. Which is not meaningful scenario but is included in the model for completeness."

#The wt coefficient indicates that for every one-unit increase in wt.  
#The predicted mpg decreases by approximately 3.92. holding hp constant.  
  
#The hp coefficient -0.03177 of indicates that for every one-unit increases in hp, the predicted mpg decreases by approximately 0.02, holding wt constant.