Final Exam Report

Block 1)

Description:

Dataset Boeing has 300 records, having variables aircraft, no of passenger, ground speed, air speed, height, error.

Air speed calculated from the true airspeed whose value is more than 90.

Distance is non linear equation depends on true air speed and height and error.

R Program:

aircraft<-rep("boeing",300)

no\_psag<-sample.int(60, 300, replace = TRUE)

speed\_ground<-rnorm(300,mean=80,sd=20)

speed\_air\_true<-speed\_ground+rnorm(300,mean=0,sd=1.5)

speed\_air<-rep('.',300)

speed\_air[speed\_air\_true>90]<-speed\_air\_true[speed\_air\_true>90]

height<-rnorm(300,mean=30,sd=10)

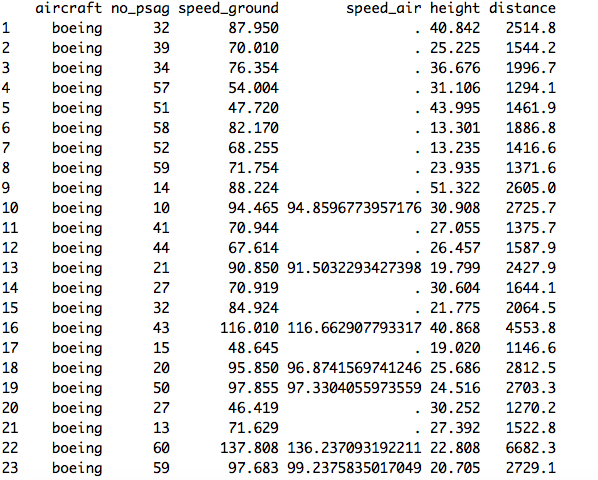
error<-rnorm(300,mean=0,sd=100)

distance=900+0.7\*(speed\_air\_true-50)^2+13\*height+error;

boeing<-data.frame(aircraft,no\_psag,speed\_ground,speed\_air,height,distance)

boeing

R-Output: (total 300 observations)



Questions:

A. There are 300 observations and 6 variables in the dataset Boeing.

B.Statistical model for response variable distance

distance=900+0.7\*(speed\_air\_true-50)^2+13\*height+error;

Distance=a+b1(speed\_air\_true)^2+b2\*height+e

Where e is error

Model in R is

Lm(distance~speed\_air\_true speed\_air\_true +height)

Block 2)

Description:

Dataset Airbus has 200 records, having variables aircraft, no of passenger, ground speed, air speed, height, error.

Air speed calculated from the true airspeed whose value is more than 95.

Distance is non linear equation depends on true air speed and height and error.

R Program:

aircraft<-rep("airbus",200)

no\_psag<-sample.int(70, 200, replace = TRUE)

speed\_ground<-rnorm(200,mean=80,sd=15)

speed\_air\_true<-speed\_ground+rnorm(200,mean=0,sd=1.5)

speed\_air<-rep('.',200)

speed\_air[speed\_air\_true>95]<-speed\_air\_true[speed\_air\_true>95]

height<-rnorm(200,mean=30,sd=10)

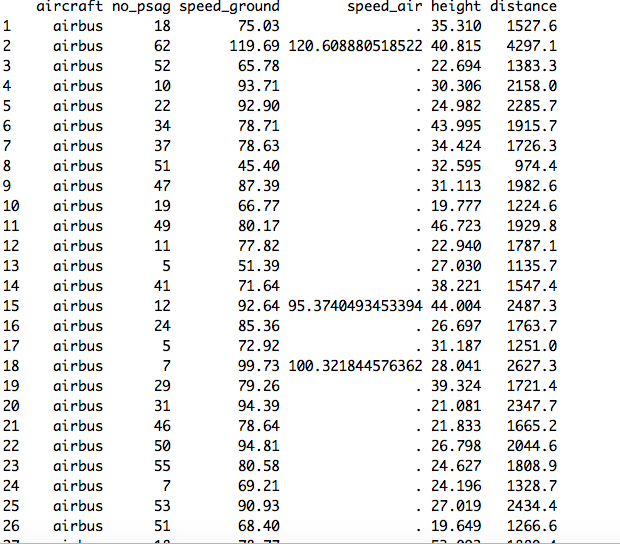
error<-rnorm(200,mean=0,sd=100)

distance=800+0.6\*(speed\_air\_true-50)^2+10\*height+100+error;

airbus<-data.frame(aircraft,no\_psag,speed\_ground,speed\_air,height,distance)

airbus

R-Output: (total 200 observations)



Questions: written in paper

Block 3

Description: load the datasets boeing and airbus. Faa dataset contains boeing dataset followed by airbus

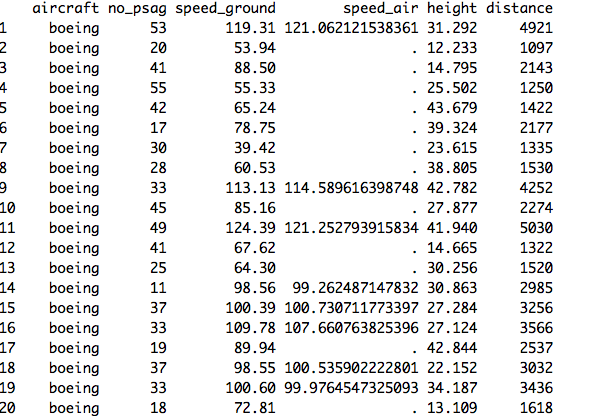
R-code:

faa<-rbind(boeing, airbus)

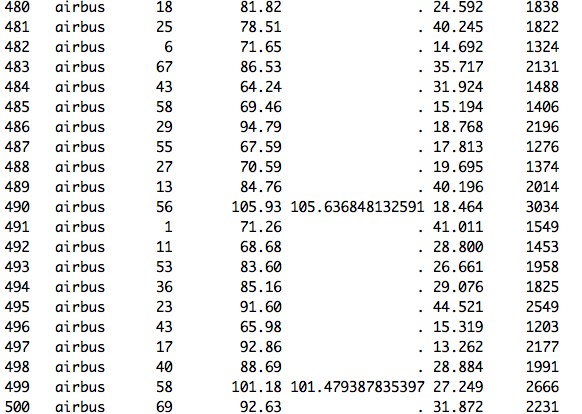
faa

R-Output: totally there are 500 observations in the dataset faa

First 20 observations



last 20 observations



dim(faa)

[1] 500 6

**Block4**

Description:

It results number of missing values in the variable speed\_air in faa dataset.

R-Code:

sum(faa[,4]==".")

R-Output:

[1] 381

**Block 5**

Description:

A new dataset faa\_new is created from faa. In that new variable type is added so that the value of type is 1 if the aircraft is boeing other wise it is set to 0. In final dataset speed\_air is removed.

aircraft, no of passenger, ground speed, height, distance and type are variables in faa\_new dataset

R-Code:

faa\_new<-data.frame(faa$aircraft,faa$no\_psag,faa$speed\_ground,faa$height,faa$distance)

type<-rep(0,times=dim(faa\_new)[1])

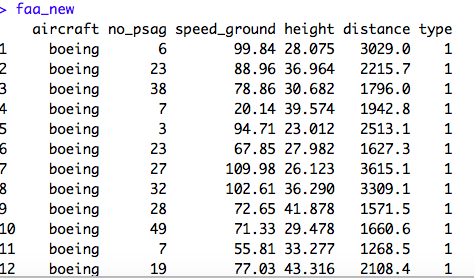
type[faa\_new$faa.aircraft=="boeing"]<-1

faa\_new<-data.frame(faa\_new,type)

colnames(faa\_new)<-c("aircraft","no\_psag","speed\_ground","height","distance","type")

faa\_new

R-Output:



dim(faa\_new)

[1] 500 6

**Block 6**

Description:

it finds the mean of no\_pasg, speed\_ground height and distance.That is from column no. 2 to 5 it finds mean value on each variable

R-Code:

colMeans(faa\_new[,2:5])

R-Output

no\_psag speed\_ground height distance

32.50 79.46 29.42 2028.31

**Block 7**

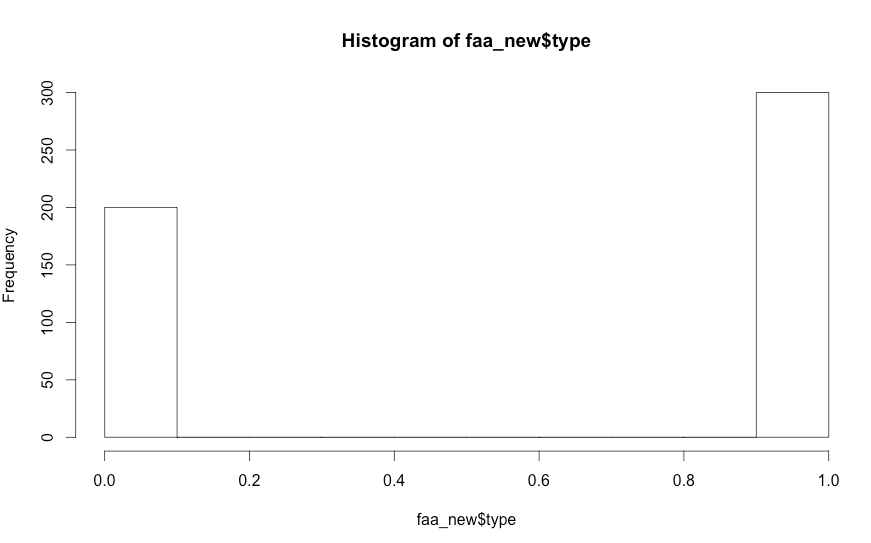
Description:

Plots vertical bars of type variable in faa\_new dataset

R-Code:

hist(faa\_new$type)

R-Output:



**Block 8**

Description:

Set of plots are drawn by taking faa\_new

Distance vs no\_psag

Distance vs height

Distance vs speed\_ground are drawn separately

R-Code:

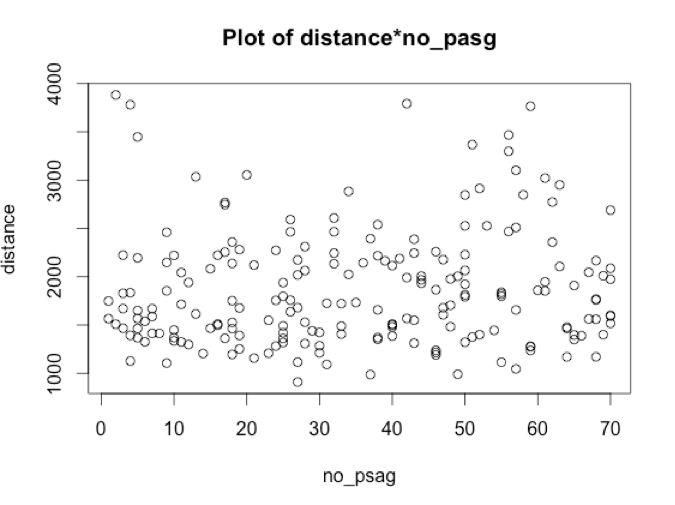
> plot(distance~no\_psag,main="Plot of distance\*no\_pasg")

> plot(distance~height,main="Plot of distance\*height")

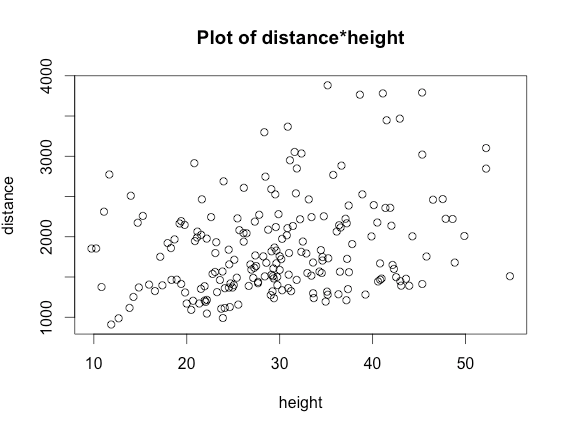
> plot(distance~speed\_ground,main="Plot of distance\*speed\_ground")

R-Output:

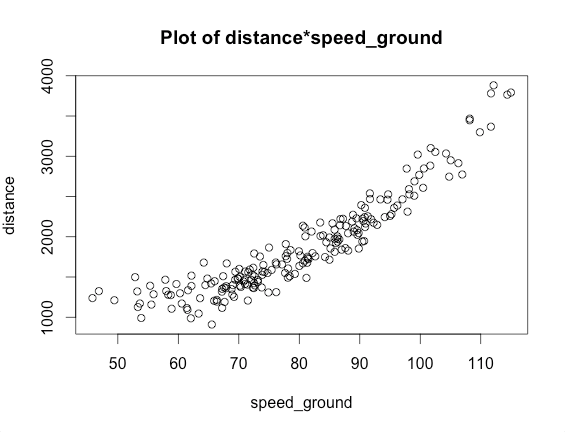
Distance vs no\_psg

****

distance vs height



distance vs ground speed



**Block 9:**

Description:

Correlation among variables is checked with this proc procedure. The result is a correlation matrix.

R-Code:

cor(faa\_new[,2:6], method = "pearson")

R-Output:

no\_psag speed\_ground height distance type

no\_psag 1.000000 0.007334 -0.05521 -0.03331 -0.11583

speed\_ground 0.007334 1.000000 -0.02290 0.87321 -0.01759

height -0.055206 -0.022899 1.00000 0.13053 -0.04096

distance -0.033313 0.873212 0.13053 1.00000 0.17378

type -0.115828 -0.017590 -0.04096 0.17378 1.00000

**Block 10:**

Description:

shows the fit diagnostic for distance with speed\_ground and height

R-Code:

model<-lm(faa\_new$distance~faa\_new$speed\_ground+faa\_new$height)

summary(model)

R-Output:

Call:

lm(formula = faa\_new$distance ~ faa\_new$speed\_ground + faa\_new$height)

Residuals:

Min 1Q Median 3Q Max

-578.8 -214.2 -62.1 112.6 2252.6

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1382.735 89.714 -15.41 < 2e-16 \*\*\*

faa\_new$speed\_ground 38.494 0.913 42.15 < 2e-16 \*\*\*

faa\_new$height 11.979 1.654 7.24 1.7e-12 \*\*\*

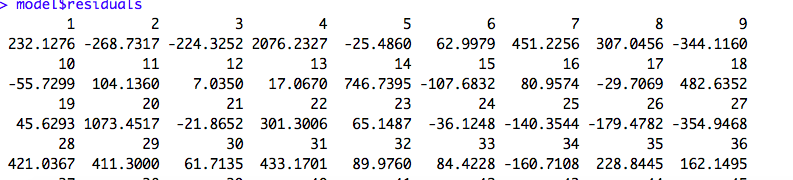
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 357 on 497 degrees of freedom

Multiple R-squared: 0.785, Adjusted R-squared: 0.784

F-statistic: 908 on 2 and 497 DF, p-value: <2e-16



Block 11:

Description:

Diagnostic Plots of residals vs speed\_ground residual vs height is plotted

R-Code:

plot(model$residuals~faa\_new$speed\_ground)

plot(model$residuals~faa\_new$height)

Output:-

