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)

	aircraft	no_psag	speed_ground	speed_air	height	distance
1	boeing	32	87.950	.	40.842	2514.8
2	boeing	39	70.010	.	25.225	1544.2
3	boeing	34	76.354	.	36.676	1996.7
4	boeing	57	54.004	.	31.106	1294.1
5	boeing	51	47.720	.	43.995	1461.9
6	boeing	58	82.170	.	13.301	1886.8
7	boeing	52	68.255	.	13.235	1416.6
8	boeing	59	71.754	.	23.935	1371.6
9	boeing	14	88.224	.	51.322	2605.0
10	boeing	10	94.465	94.8596773957176	30.908	2725.7
11	boeing	41	70.944	.	27.055	1375.7
12	boeing	44	67.614	.	26.457	1587.9
13	boeing	21	90.850	91.5032293427398	19.799	2427.9
14	boeing	27	70.919	.	30.604	1644.1
15	boeing	32	84.924	.	21.775	2064.5
16	boeing	43	116.010	116.662907793317	40.868	4553.8
17	boeing	15	48.645	.	19.020	1146.6
18	boeing	20	95.850	96.8741569741246	25.686	2812.5
19	boeing	50	97.855	97.3304055973559	24.516	2703.3
20	boeing	27	46.419	.	30.252	1270.2
21	boeing	13	71.629	.	27.392	1522.8
22	boeing	60	137.808	136.237093192211	22.808	6682.3
23	boeing	59	97.683	99.2375835017049	20.705	2729.1

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)

	aircraft	no_psag	speed_ground	speed_air	height	distance
1	airbus	18	75.03	.	35.310	1527.6
2	airbus	62	119.69	120.608880518522	40.815	4297.1
3	airbus	52	65.78	.	22.694	1383.3
4	airbus	10	93.71	.	30.306	2158.0
5	airbus	22	92.90	.	24.982	2285.7
6	airbus	34	78.71	.	43.995	1915.7
7	airbus	37	78.63	.	34.424	1726.3
8	airbus	51	45.40	.	32.595	974.4
9	airbus	47	87.39	.	31.113	1982.6
10	airbus	19	66.77	.	19.777	1224.6
11	airbus	49	80.17	.	46.723	1929.8
12	airbus	11	77.82	.	22.940	1787.1
13	airbus	5	51.39	.	27.030	1135.7
14	airbus	41	71.64	.	38.221	1547.4
15	airbus	12	92.64	95.3740493453394	44.004	2487.3
16	airbus	24	85.36	.	26.697	1763.7
17	airbus	5	72.92	.	31.187	1251.0
18	airbus	7	99.73	100.321844576362	28.041	2627.3
19	airbus	29	79.26	.	39.324	1721.4
20	airbus	31	94.39	.	21.081	2347.7
21	airbus	46	78.64	.	21.833	1665.2
22	airbus	50	94.81	.	26.798	2044.6
23	airbus	55	80.58	.	24.627	1808.9
24	airbus	7	69.21	.	24.196	1328.7
25	airbus	53	90.93	.	27.019	2434.4
26	airbus	51	68.40	.	19.649	1266.6

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

	aircraft	no_psag	speed_ground	speed_air	height	distance
1	boeing	53	119.31	121.062121538361	31.292	4921
2	boeing	20	53.94	.	12.233	1097
3	boeing	41	88.50	.	14.795	2143
4	boeing	55	55.33	.	25.502	1250
5	boeing	42	65.24	.	43.679	1422
6	boeing	17	78.75	.	39.324	2177
7	boeing	30	39.42	.	23.615	1335
8	boeing	28	60.53	.	38.805	1530
9	boeing	33	113.13	114.589616398748	42.782	4252
10	boeing	45	85.16	.	27.877	2274
11	boeing	49	124.39	121.252793915834	41.940	5030
12	boeing	41	67.62	.	14.665	1322
13	boeing	25	64.30	.	30.256	1520
14	boeing	11	98.56	99.262487147832	30.863	2985
15	boeing	37	100.39	100.730711773397	27.284	3256
16	boeing	33	109.78	107.660763825396	27.124	3566
17	boeing	19	89.94	.	42.844	2537
18	boeing	37	98.55	100.535902222801	22.152	3032
19	boeing	33	100.60	99.9764547325093	34.187	3436
20	boeing	18	72.81	.	13.109	1618

last 20 observations

480	airbus	18	81.82	.	24.592	1838
481	airbus	25	78.51	.	40.245	1822
482	airbus	6	71.65	.	14.692	1324
483	airbus	67	86.53	.	35.717	2131
484	airbus	43	64.24	.	31.924	1488
485	airbus	58	69.46	.	15.194	1406
486	airbus	29	94.79	.	18.768	2196
487	airbus	55	67.59	.	17.813	1276
488	airbus	27	70.59	.	19.695	1374
489	airbus	13	84.76	.	40.196	2014
490	airbus	56	105.93	105.636848132591	18.464	3034
491	airbus	1	71.26	.	41.011	1549
492	airbus	11	68.68	.	28.800	1453
493	airbus	53	83.60	.	26.661	1958
494	airbus	36	85.16	.	29.076	1825
495	airbus	23	91.60	.	44.521	2549
496	airbus	43	65.98	.	15.319	1203
497	airbus	17	92.86	.	13.262	2177
498	airbus	40	88.69	.	28.884	1991
499	airbus	58	101.18	101.479387835397	27.249	2666
500	airbus	69	92.63	.	31.872	2231

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:

```
> faa_new
    aircraft no_psag speed_ground height distance type
1     boeing       6        99.84 28.075   3029.0    1
2     boeing      23        88.96 36.964   2215.7    1
3     boeing      38        78.86 30.682   1796.0    1
4     boeing       7        20.14 39.574   1942.8    1
5     boeing       3        94.71 23.012   2513.1    1
6     boeing      23        67.85 27.982   1627.3    1
7     boeing      27       109.98 26.123   3615.1    1
8     boeing      32       102.61 36.290   3309.1    1
9     boeing      28        72.65 41.878   1571.5    1
10    boeing      49        71.33 29.478   1660.6    1
11    boeing       7        55.81 33.277   1268.5    1
12    boeing      19        77.03 43.316   2108.4    1
```

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

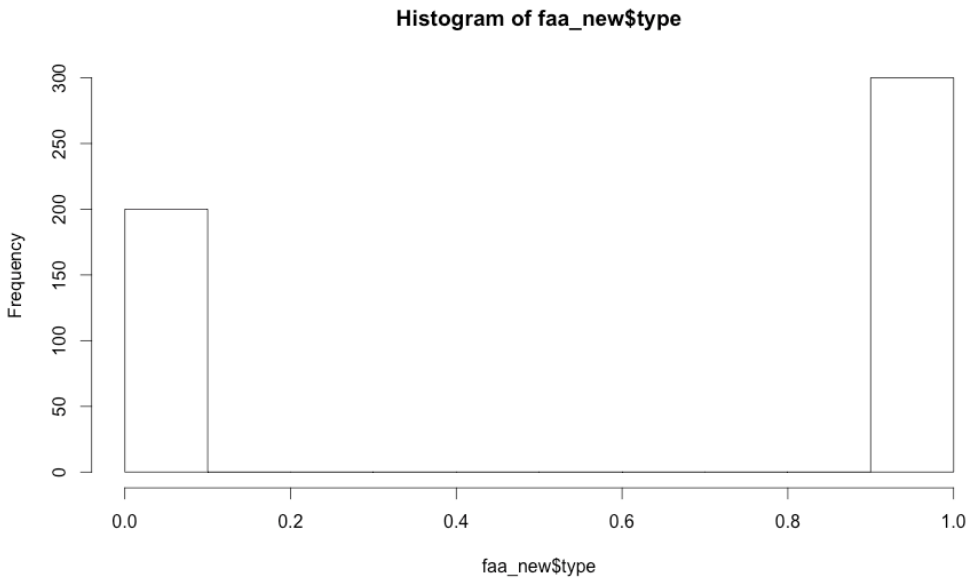
Description:

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

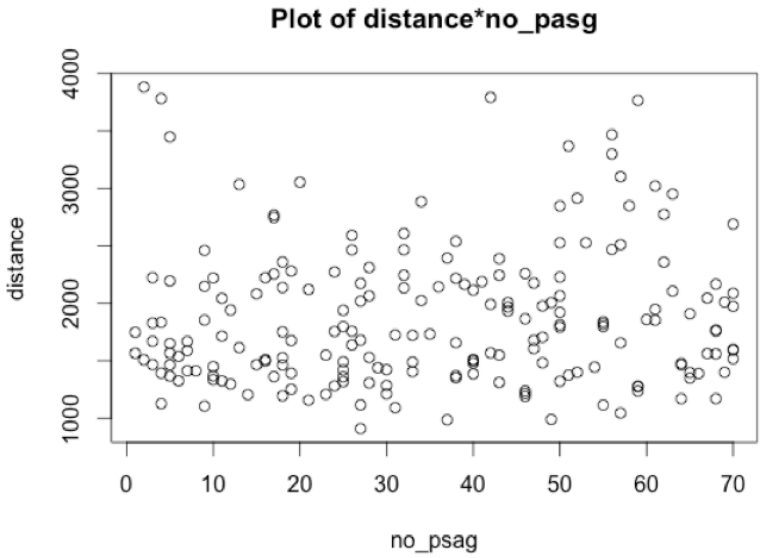
Distance vs speed_ground are drawn separately

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")
```

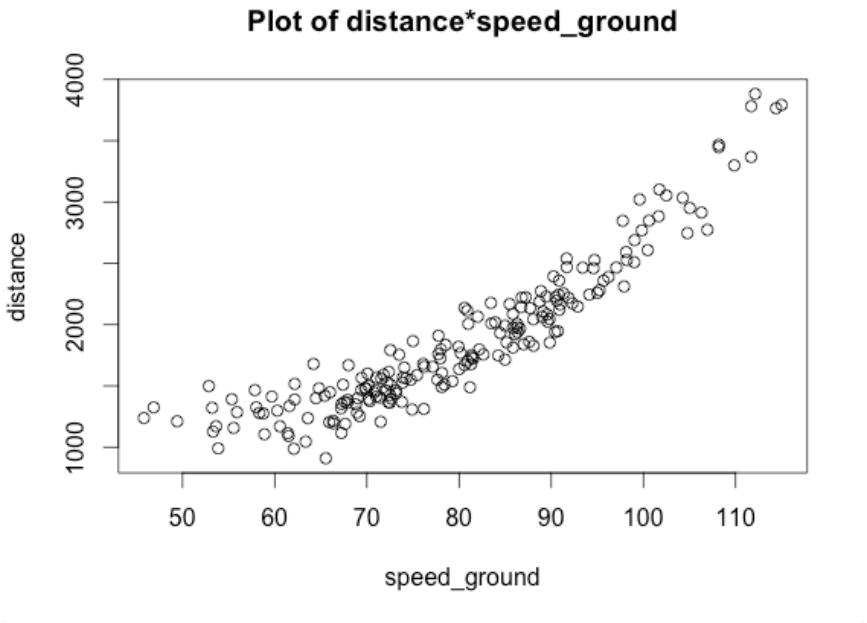
R-Output:

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
```

```
> model$residuals
       1        2        3        4        5        6        7        8        9
 232.1276 -268.7317 -224.3252 2076.2327  -25.4860   62.9979  451.2256  307.0456 -344.1160
      10       11       12       13       14       15       16       17       18
 -55.7299  104.1360    7.0350   17.0670  746.7395 -107.6832   80.9574  -29.7069  482.6352
      19       20       21       22       23       24       25       26       27
  45.6293 1073.4517  -21.8652  301.3006   65.1487  -36.1248 -140.3544 -179.4782 -354.9468
      28       29       30       31       32       33       34       35       36
 421.0367  411.3000   61.7135  433.1701   89.9760   84.4228 -160.7108  228.8445  162.1495
```

Block 11:

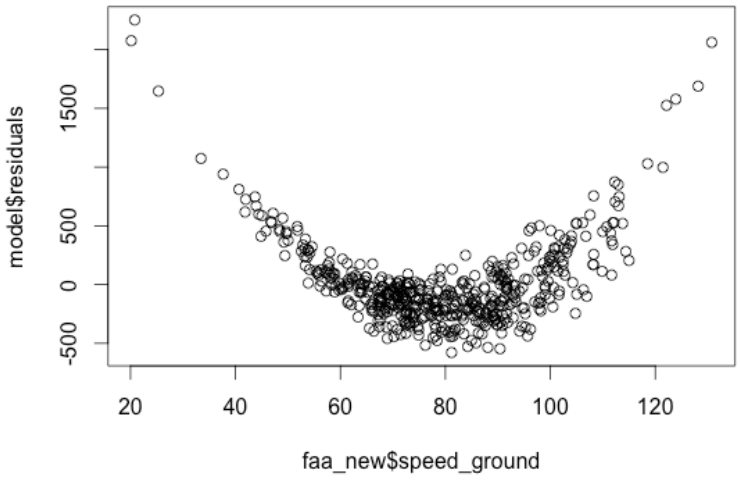
Description:

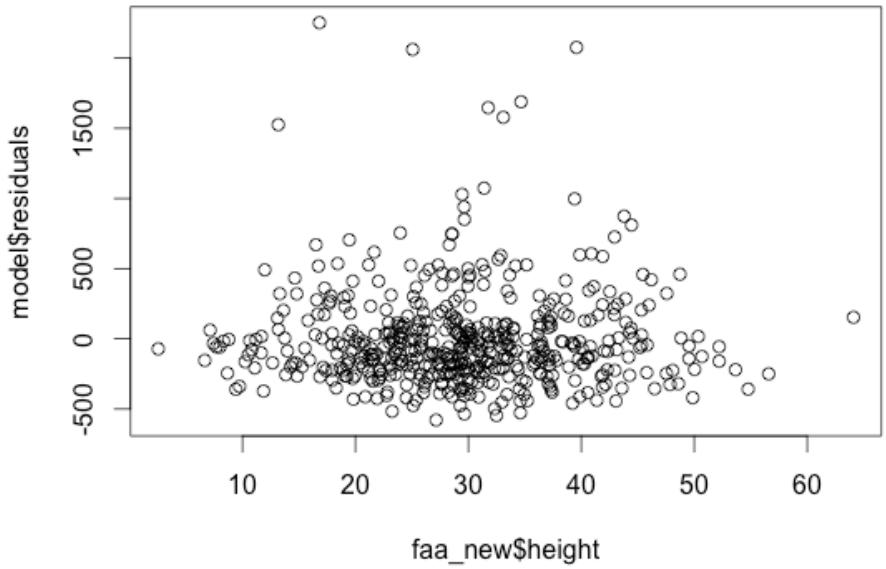
Diagnostic Plots of residals vs speed_ground residual vs height is plotted

R-Code:

```
plot(model$residuals~faa_new$speed_ground)
plot(model$residuals~faa_new$height)
```

Output:-


model$residuals
-500
0
500
1500
20
40
60
80
100
120
faa_new$speed_ground


model$residuals
-500
0
500
1500
10
20
30
40
50
60
faa_new$height