FEDERAL ADMINISTRATION AVIATION (FAA) DATA ANALYSIS

EXECUTIVE SUMMARY:

The project aims to determine what factors affect the landing distance of a flight so that they can make better predictions to further enhance flight safety. We analyzed two datasets from the FAA using SAS. The data was cleaned up to make it suitable for analysis. The clean-up process including identifying and removing blank rows, duplicate rows, and abnormal values. Using data exploration and visualization we inferred key characteristics of the data. Using multiple linear regression, the data was modeled to find the influence of predictor variables on landing distance. The analysis found that the airspeed and height were the key variables that impact landing distance. As well it was found that the landing distance was dependent on the type of aircraft – Boeing and Airbus.

SUBMITTED BY:
SAGAR SAHOO
MS BUSINESS ANALYTICS
FALL 2019-20
M ID: M13433382

CONTENTS

INTRODUCTION

Background: Flight landing.

Motivation: To reduce the risk of landing overrun.

Goal: To study what factors and how they would impact the landing distance of a commercial flight. **Data**: Landing data (landing distance and other parameters) from 950 commercial flights (not real data set but simulated from statistical models). See two Excel files 'FAA-1.xls' (800 flights) and 'FAA-2.xls' (150 flights).

VARIABLE DICTIONARY:

Aircraft: The make of an aircraft (Boeing or Airbus).

Duration (in minutes): Flight duration between taking off and landing.

No_pasg: The number of passengers in a flight.

Speed_ground (in miles per hour): The ground speed of an aircraft when passing over the threshold of the runway.

Speed_air (in miles per hour): The air speed of an aircraft when passing over the threshold of the runway.

Height (in meters): The height of an aircraft when it is passing over the threshold of the runway.

Pitch (in degrees): Pitch angle of an aircraft when it is passing over the threshold of the runway.

CHAPTER 1 : Data Understanding and Exploration

1.1 IMPORTING/ COMBINING THE DATASETS:

The Datasets were imported into SAS using "PROC IMPORT". Once imported, both were joined using SET Command. The combined Dataset was named "FAA" in SAS. It had 950 Rows(with data) apart from blank rows and 8 Variables (aircraft, duration, no_pasg, speed_ground, speed_air, height, pitch and distance).

Obs	aircraft	duration	no_pasg	speed_ground	speed_air	height	pitch	distance
1	boeing	98.4790912	53	107.91568005	109.32837648	27.418924252	4.0435145715	3369.8363638
2	boeing	125.73329732	69	101.65558863	102.8514051	27.804716181	4.1174316991	2987.8039235
3	boeing	112.0170008	61	71.051960883	.	18.589385734	4.4340431286	1144.922426
4	boeing	196.82569105	56	85.813327679	.	30.744597235	3.8842361245	1664.2181584
5	boeing	90.095381357	70	59.888528183	.	32.397688062	4.0260964152	1050.2644976
6	boeing	137.59581722	55	75.014343744	.	41.21496259	4.203853398	1627.0681991
7	boeing	73.023794916	54	54.4298029	.	24.03532163	3.8376457299	805.30399317
8	boeing	52.903187872	57	57.101661737	.	19.388837508	4.6436717769	573.62178606
9	boeing	155.51861605	61	85.443624251	.	35.375389749	4.2287278648	1698.9927548
10	boeing	176.86203205	56	61.796710514	.	36.748816124	4.1843990127	1137.7457579

Fig 1.1 Snippet of Imported Data from SAS

1.2 COMPLETENESS CHECK OF VARIABLES:

The Missing Values of each variable was found out using Means in SAS. Below is the output of the Means PROC

The MEANS Procedure

Variable	Label	N Miss	N
duration	duration	51	800
no_pasg	no_pasg	1	850
speed_ground	speed_ground	1	850
speed_air	speed_air	643	208
height	height	1	850
pitch	pitch	1	850
distance	distance	1	850

Fig 1.2 Missing Values using MEANS Proc

1.3 VALIDITY CHECK OF VARIABLES:

The abnormal values were identified as those values which violated the conditions of each value (mentioned in problem statement).
The below validity checks were applied on the combined FAA Dataset to identify those abnormal values:

(DURATION >= 0 AND DURATION <=40) OR (speed_ground >= 0 AND speed_ground <=30) OR speed_ground > 140 OR (speed_air >= 0 AND speed_air <=30) OR speed_air > 140 OR (height>0 AND height < 6) OR distance > 6000

1.4 DATA CLEANING:

1. These abnormal rows were dropped from the combined FAA Dataset using DELETE Statement.
2. Apart from abnormal values, the rows with **all** missing columns were removed as well.
3. The Duplicates were identified using PROC SORT and NODUPKEY on the basis of speed_ground, speed_air, height and pitch distance
4. Although we have the option to drop the rows with missing values but we should refrain from doing so. We might lose important information if we drop such rows/columns.

Below is the snippet of the SAS Code used to DELETE the abnormal values/rows:

```
/*CLEANING THE DATA - HANDLING DURATION, SPEED_GROUND, SPEED_AIR, HEIGHT, DISTANCE*/

DATA FAA_CLEANED;
options missing = ' ';
set FAA;
IF  (DURATION >= 0 AND DURATION <= 40) OR
(speed_ground>=0 AND speed_ground <=30) OR
speed_ground > 140 OR
(speed_air>=0 AND speed_air <=30) OR
speed_air > 140 OR
(height>0 AND height < 6) OR distance > 6000 THEN DELETE;
if missing(cats(of _all_)) then delete;
PROC PRINT data=FAA_CLEANED ;
RUN;
```

Fig 1.3 Handling abnormal values in SAS

1.5 DATA EXPLORATION:

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
duration	duration	786	154.7238132	48.2377369	41.9493694	305.6217107
no_pasg	no_pasg	836	60.0933014	7.4952059	29.0000000	87.0000000
speed_ground	speed_ground	836	79.4738971	18.7330205	33.5741041	132.7846766
speed_air	speed_air	203	103.4850352	9.7362774	90.0028586	132.9114649
height	height	836	30.2620810	10.0776036	-3.5462524	59.9459639
pitch	pitch	836	4.0063964	0.5272812	2.2844801	5.9267842
distance	distance	836	1517.33	896.9943065	34.0807833	5381.96

Fig 1.4 MEANS Proc Output

In order to have a understanding of the distribution of the variables of the Problem Statement, we use PROC CHART method of SAS. Below is a snippet of the implementation:

```
proc chart data=FAA_CLEANED;
vbar distance/ type=FREQ;
vbar DURATION/ type=FREQ;
vbar speed_ground/ type=FREQ;
vbar speed_air/ type=FREQ;
vbar HEIGHT/ type=FREQ;
```

Fig 1.5 Plotting Distribution of Variables

To have an approximation of the distribution of the variables, we plot the Frequency Distributions of all the variables.

HEIGHT & SPEED AIR:

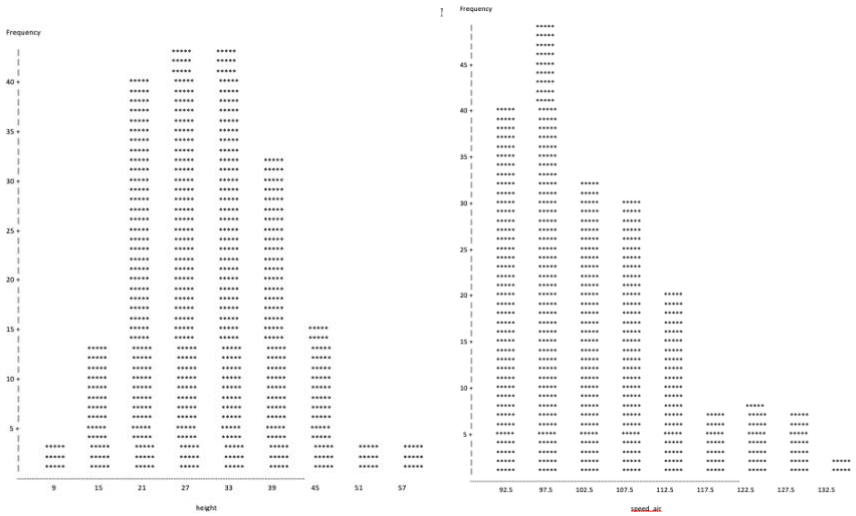

Fig 1.6 Distribution of Height and Air Speed

The distribution of Heights(above) seems to be normally distributed. Although, the speed_air column has most of the values as 0, the distribution (above) of the available values shows slightly right skewed. The Skewness is 0.88.

SPEED GROUND & DURATION:

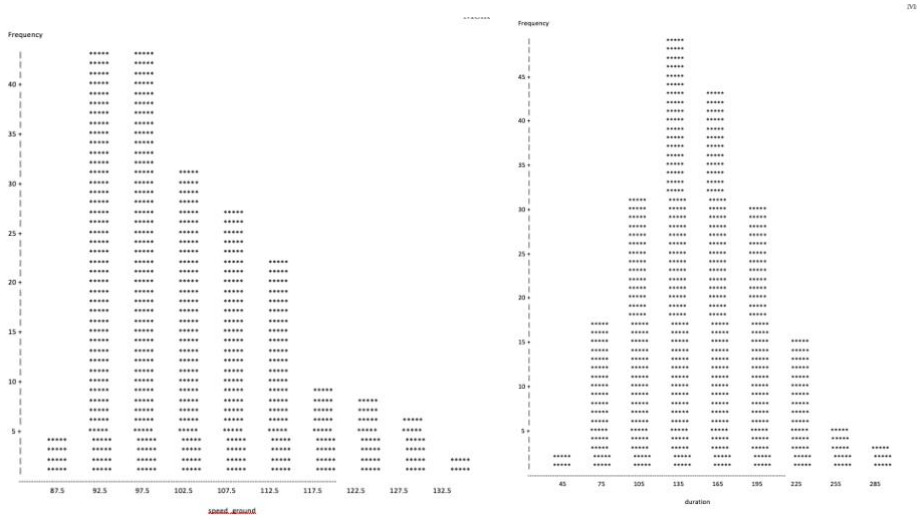

Fig 1.7 Distribution of Ground Speed and Duration

The distribution (above) of the available values of speed_ground approximately shows normally distributed. And the distribution (above) of the available values of duration shows normal distribution.

DISTANCE:

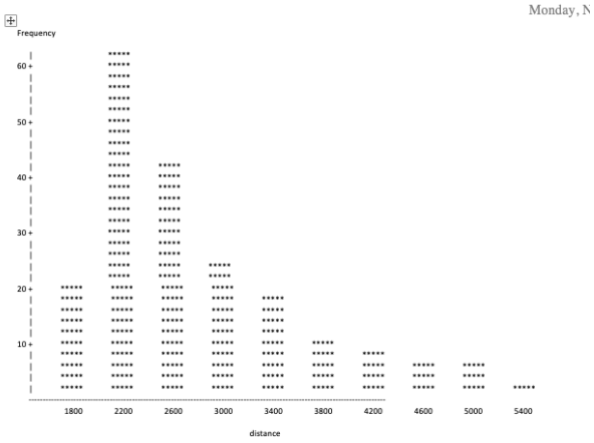

Fig 1.8 Distribution of Distance

The distribution (above) of the available values of distance shows positive skewness.

1.6 SUMMARY:

1. There were 50 rows with all column values missing which was removed after importing the data.
2. The duplicate values on the basis of speed_ground, speed_air, height and pitch distance were removed. And the abnormalities were as well handled as per problem statement.
3. The rows with specific columns missing were not dropped as we might lose information. Post data cleaning (Missing Rows, Duplicates and Abnormalities) we have 836 Observations.
4. The frequency distribution of height, ground speed, duration looks normal. The air speed and distance slows slight positive skewness.

CHAPTER 2 : Descriptive Study

2.1 XY PLOTS

In order to have a understanding of the relationship of the predictors with response, we use PROC PLOT method of SAS. Below is a snippet of the implementation:

```
proc plot data=FAA_CLEANED;
    plot distance*no_pasg;
    plot distance*speed_ground;
    plot distance*speed_air;
    plot distance*height;
    plot distance*pitch;
    plot duration*distance;
    plot height*distance;
    plot speed_ground*speed_air;
```

Fig 2.1 SAS Code Snippet for the XY Plots

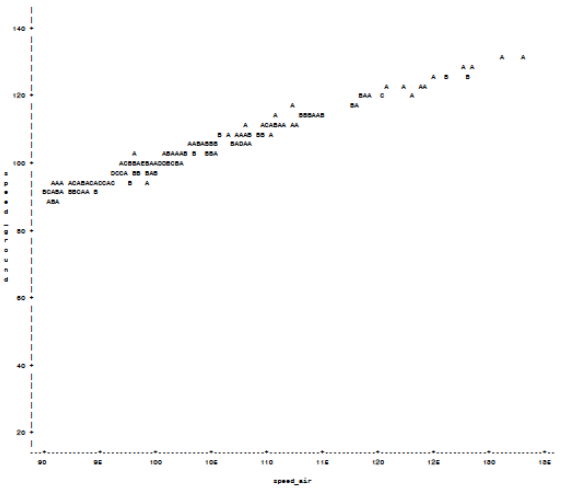

Fig 2.2: Speed_air vs speed_ground

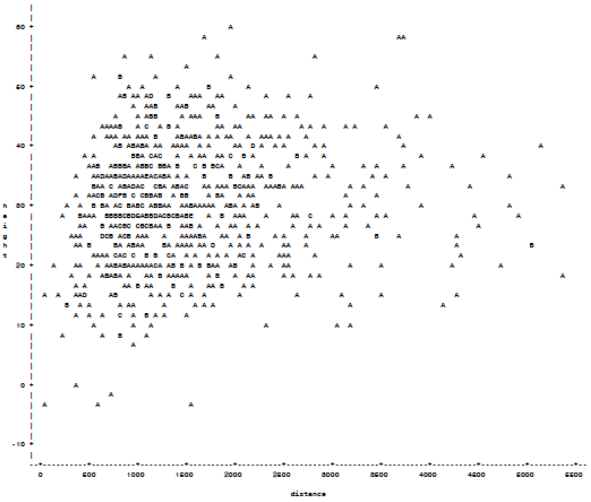

Fig 2.3: Distance vs height

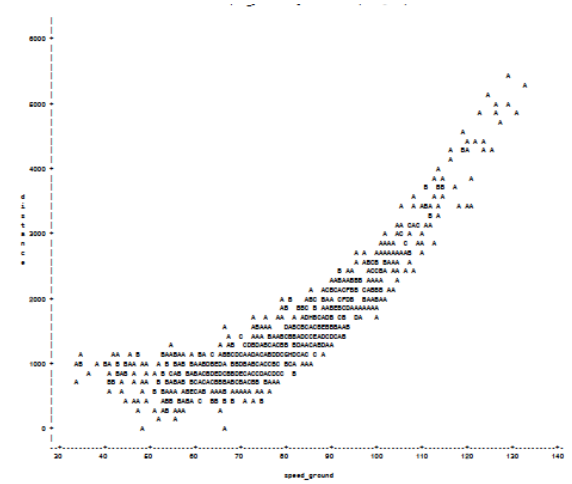

Fig 2.4: Distance vs speed_ground

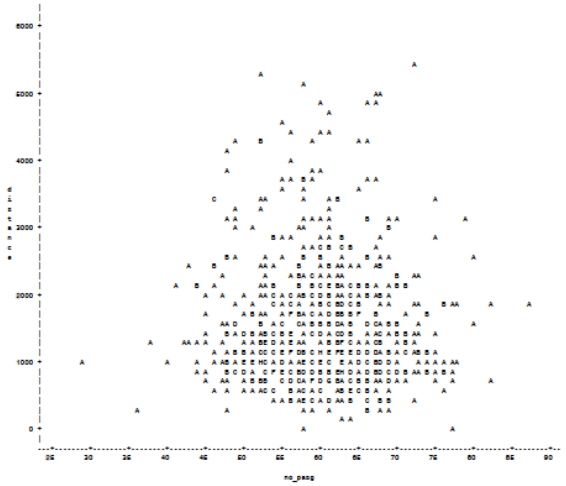

Fig 2.5: Distance vs no_pasg

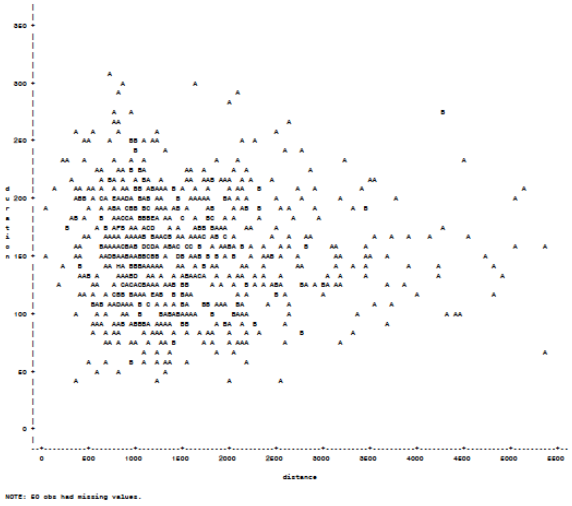

Fig 2.6: Distance vs duration

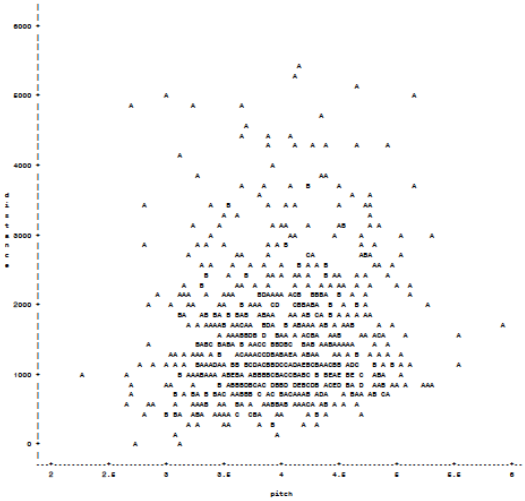

Fig 2.7: Distance vs pitch

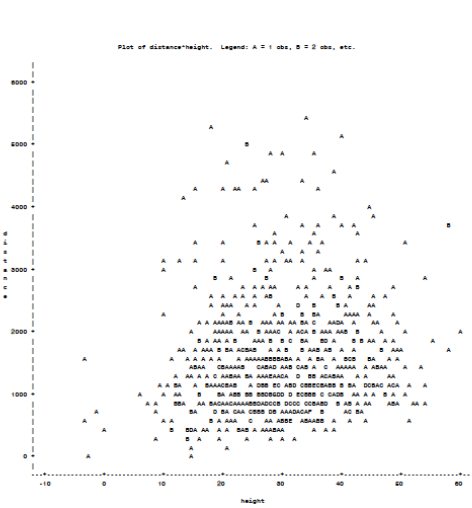

Fig 2.8: Distance vs height

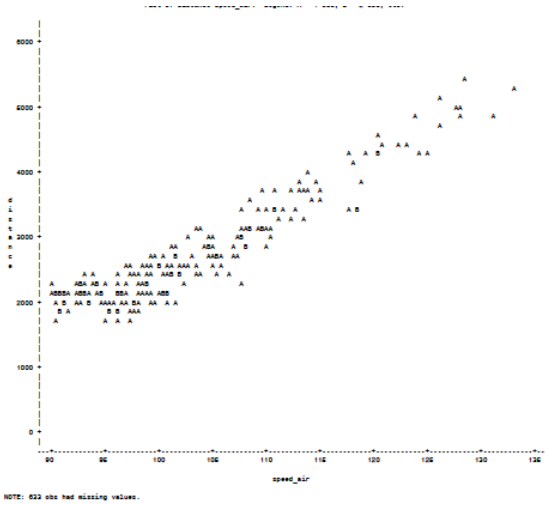

Fig 2.9: Distance vs speed_air

2.2 CORRELATION ANALYSIS

Using PROC CORR, we find the correlation coefficient among all variables in the FAA_CLEANED dataset:

```
proc corr data=FAA_CLEANED;
    var duration no_pasg speed_ground speed_air height pitch distance;
```

Fig 2.10: SAS Code Snippet for the Correlation Coefficents

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	duration	no_pasg	speed_ground	speed_air	height	pitch	distance
duration	1.00000	-0.03844	-0.05045	0.04454	0.01430	-0.04822	-0.05148
		0.2818	0.1576	0.5364	0.6890	0.1768	0.1493
	786	786	786	195	786	786	786
no_pasg	-0.03844	1.00000	-0.00146	-0.00616	0.02907	-0.01159	-0.02048
	0.2818		0.9663	0.9305	0.4013	0.7380	0.5544
	786	836	836	203	836	836	836
speed_ground	-0.05045	-0.00146	1.00000	0.98794	-0.04391	-0.03887	0.86609
	0.1576	0.9663		<.0001	0.2047	0.2616	<.0001
	786	836	836	203	836	836	836
speed_air	0.04454	-0.00616	0.98794	1.00000	-0.07933	-0.03927	0.94210
	0.5364	0.9305	<.0001		0.2606	0.5780	<.0001
	195	203	203	203	203	203	203
height	0.01430	0.02907	-0.04391	-0.07933	1.00000	0.01482	0.11432
	0.6890	0.4013	0.2047	0.2606		0.6687	0.0009
	786	836	836	203	836	836	836
pitch	-0.04822	-0.01159	-0.03887	-0.03927	0.01482	1.00000	0.08689
	0.1768	0.7380	0.2616	0.5780	0.6687		0.0120
	786	836	836	203	836	836	836
distance	-0.05148	-0.02048	0.86609	0.94210	0.11432	0.08689	1.00000
	0.1493	0.5544	<.0001	<.0001	0.0009	0.0120	
	786	836	836	203	836	836	836

Fig 2.11 : Pearson Correlation Coefficients

2.3 TTEST and ANOVA

We use T Test and ANOVA to check if the mean value of duration varies across Aircrafts i.e Boeing and Airbus.

```
proc ttest data=FAA_CLEANED;
    class aircraft;
    var distance;
title "T-Test for Aircraft / Distance";
proc anova data=FAA_CLEANED;
    class aircraft;
    model distance = aircraft;
    means aircraft;
title "ANOVA Test for Aircraft / Distance";
```

Fig 2.12: Snippet for TTEST and ANOVA

T-Test for Aircraft / Distance

The TTEST Procedure

Variable: distance (distance)

aircraft	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
airbus		447	1318.9	792.9	37.5051	34.0808	4896.3
boeing		389	1745.4	954.6	48.4005	573.6	5382.0
Diff (1-2)	Pooled		-426.5	871.9	60.4558		
Diff (1-2)	Satterthwaite		-426.5		61.2311		

aircraft	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
airbus		1318.9	1245.2	1392.6	792.9	744.2	848.6
boeing		1745.4	1650.2	1840.5	954.6	891.9	1026.9
Diff (1-2)	Pooled	-426.5	-545.2	-307.8	871.9	832.0	915.9
Diff (1-2)	Satterthwaite	-426.5	-546.7	-306.3			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	834	-7.05	<.0001
Satterthwaite	Unequal	756.55	-6.97	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	388	446	1.45	0.0002

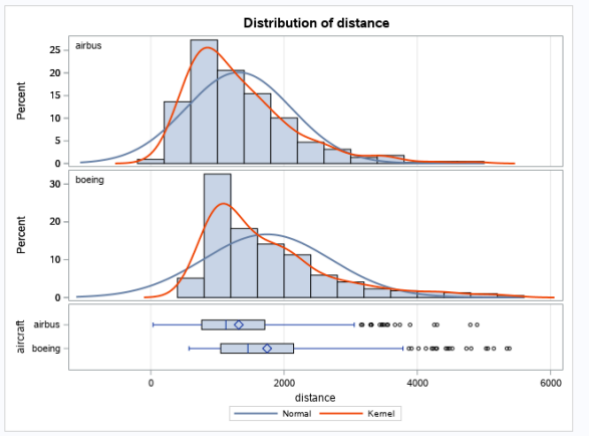

Fig 2.13 TTEST PROC Outputs

ANOVA Test for Aircraft / Distance

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
aircraft	2	airbus boeing

Number of Observations Read	836
Number of Observations Used	836

ANOVA Test for Aircraft / Distance

The ANOVA Procedure

Dependent Variable: distance distance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	37835348.3	37835348.3	49.77	<.0001
Error	834	634004637.9	760197.4		
Corrected Total	835	67183986.2			

R-Square	Coeff Var	Root MSE	distance Mean
0.056316	57.46243	871.8930	1517.327

Source	DF	Anova SS	Mean Square	F Value	Pr > F
aircraft	1	37835348.32	37835348.32	49.77	<.0001

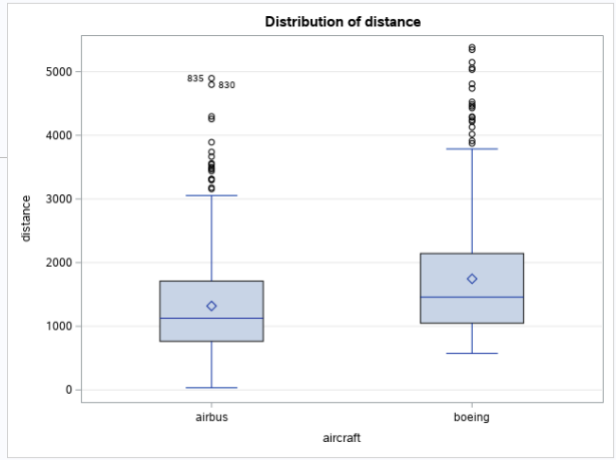

Fig 2.14 ANOVA Test Outputs

2.4 SUMMARY

1. From the XY plots and the correlations analysis, we found that Air Speed, Height and Ground Speed are **positively correlated** with distance variable. Air Speed has close to perfect linear relationship with distance as compared to Ground Speed.
2. There exists **strong positive correlation** between air speed and ground speed.
3. The results of TTEST and ANOVA indicates that the aircraft type has **high effect** on the landing distance.

Chapter 3 Statistical Modelling and Model Checking

3.1 DUMMY VARIABLES

In order to apply Linear/Multiple Regression, we create dummy variables for the categorical variables.

```
data FAA_CLEANED;
    set FAA_CLEANED;

    if aircraft="airbus" then
        aircraft_airbus=1;
    else
        aircraft_airbus=0;
run;
```

Fig 3.1 Using Dummy Variables for categorical columns

3.2 SIMPLE LINEAR REGRESSION MODELING

We apply regression of distance against each predictor variable separately to understand the relationship.

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	836
Number of Observations Used	203
Number of Observations with Missing Values	633

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	121121738	121121738	1586.41	<.0001
Error	201	15346230	76349		
Corrected Total	202	136467968			

Root MSE	276.31396	R-Square	0.8875
Dependent Mean	2774.67289	Adj R-Sq	0.8870
Coeff Var	9.95843		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5455.70878	207.54681	-26.29	<.0001
speed_air	speed_air	1	79.53210	1.99680	39.83	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	836
Number of Observations Used	836

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8780600	8780600	11.04	0.0009
Error	834	663059386	795035		
Corrected Total	835	671839986			

Root MSE	891.64748	R-Square	0.0131
Dependent Mean	1517.32705	Adj R-Sq	0.0119
Coeff Var	58.76436		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1209.39110	97.65676	12.38	<.0001
height	height	1	10.17564	3.06191	3.32	0.0009

Fig 3.2 Simple Regression of speed_air and height against distance

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	836
Number of Observations Used	836

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	37835348	37835348	49.77	<.0001
Error	834	634004638	760197		
Corrected Total	835	671839986			

Root MSE	871.89300	R-Square	0.0563
Dependent Mean	1517.32705	Adj R-Sq	0.0552
Coeff Var	57.46243		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1745.37422	44.20673	39.48	<.0001
aircraft_airbus		1	-426.50433	60.45577	-7.05	<.0001

Dependent Variable: distance distance

Number of Observations Read	836
Number of Observations Used	836

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5072437	5072437	6.34	0.0120
Error	834	666767549	799481		
Corrected Total	835	671839986			

Root MSE	894.13728	R-Square	0.0076
Dependent Mean	1517.32705	Adj R-Sq	0.0064
Coeff Var	58.92845		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	925.11566	237.13586	3.90	0.0001
pitch	pitch	1	147.81647	58.68386	2.52	0.0120

Fig 3.3 Simple Regression of aircraft_airbus ,pitch against distance

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	836
Number of Observations Used	836

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	281693	281693	0.35	0.5544
Error	834	671558293	805226		
Corrected Total	835	671839986			

Root MSE	897.34373	R-Square	0.0004
Dependent Mean	1517.32705	Adj R-Sq	-0.0008
Coeff Var	59.13977		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1664.58788	250.90319	6.63	<.0001
no_pasg	no_pasg	1	-2.45054	4.14316	-0.59	0.5544

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	836
Number of Observations Used	836

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	503950942	503950942	2503.41	<.0001
Error	834	167889044	201306		
Corrected Total	835	671839986			

Root MSE	448.67116	R-Square	0.7501
Dependent Mean	1517.32705	Adj R-Sq	0.7498
Coeff Var	29.56984		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1778.52414	67.67520	-26.28	<.0001
speed_ground	speed_ground	1	41.47086	0.82885	50.03	<.0001

Fig 3.4 Simple Regression of no_pasg, speed_ground against distance

3.3 MULTIPLE REGRESSION MODELING

3.3.1 Regressing Distance against all Predictors:

Now we use PROC REG to create multivariate linear regression (Model 1) on all the predictor and response variables:

Root MSE	134.83444	R-Square	0.9744
Dependent Mean	2784.49158	Adj R-Sq	0.9736
Coeff Var	4.84234		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5915.50780	139.04730	-42.54	<.0001
speed_ground	speed_ground	1	-3.63956	6.43387	-0.57	0.5723
speed_air	speed_air	1	85.64372	6.54021	13.09	<.0001
aircraft_airbus		1	-439.40658	21.29791	-20.63	<.0001
duration	duration	1	0.14822	0.20402	0.73	0.4684
height	height	1	13.68209	1.04149	13.14	<.0001
pitch	pitch	1	-12.94161	18.65656	-0.69	0.4887

Fig 3.5 PROC REG Output for Model 1

3.3.2 Regressing Distance against Air Speed, Height and Aircraft Type:

Now we use predictors – speed_air, aircraft_airbus and height (Model 2) to predict distance using regression.

Root MSE	134.24368	R-Square	0.9737
Dependent Mean	2774.67289	Adj R-Sq	0.9733
Coeff Var	4.83818		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5962.93395	107.93619	-55.24	<.0001
speed_air	speed_air	1	82.14852	0.97601	84.17	<.0001
aircraft_airbus		1	-427.44156	19.17339	-22.29	<.0001
height	height	1	13.70161	1.00718	13.60	<.0001

Fig 3.6 PROC REG Output for Model 2

Once the regression model is fit, we check for the residuals plot using residuals and diagnostics in PROC REG

```
proc reg data=FAA_CLEANED;
 model Distance= Speed_air Height aircraft_airbus / r;
 output out=diagnostics r=residual;
 run;

/*Checking Residual Plots*/
proc plot data=diagnostics;
 plot Residual*Speed_air;
 plot Residual*aircraft_airbus;

proc means data=diagnostics t prt;
var Residual;
run;
```

Fig 3.7 Residuals Snippet in SAS

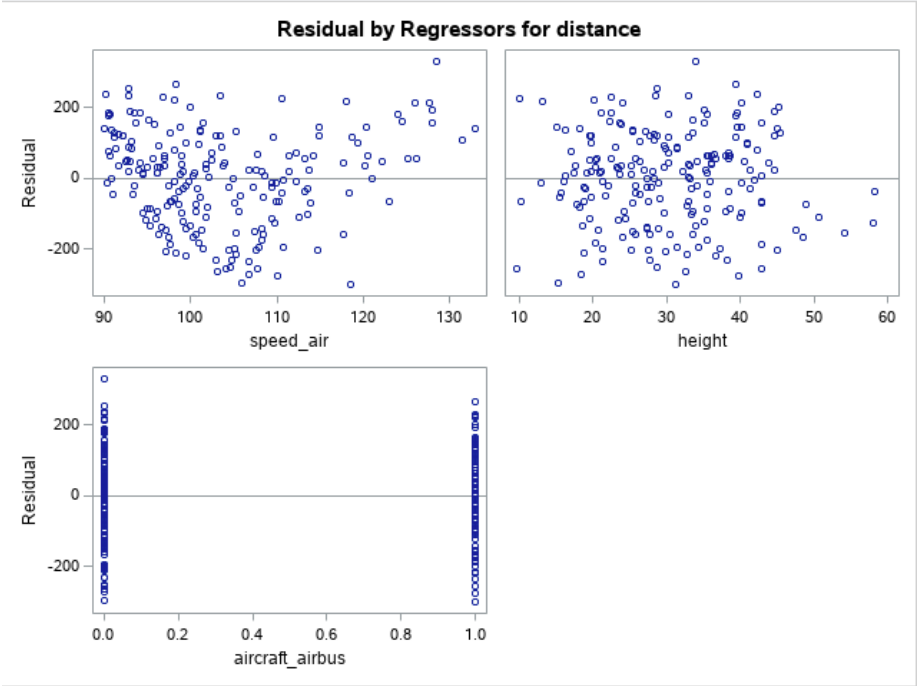

Fig 3.8 Residuals Plots for Model 2

3.3.3 Regressing Distance against Ground Speed, Height and Aircraft Type:

And lastly we use predictors – speed_ground, aircraft_airbus and height (Model 3) to predict distance using regression.

Root MSE	348.99913	**R-Square**	0.8492
Dependent Mean	1517.32705	**Adj R-Sq**	0.8486
Coeff Var	23.00092		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-2004.12627	66.06174	-30.34	<.0001
speed_ground	speed_ground	1	42.34385	0.64591	65.56	<.0001
aircraft_airbus		1	-495.28403	24.22184	-20.45	<.0001
height	height	1	13.91330	1.19972	11.60	<.0001

Fig 3.9 PROC REG Output for Model 3

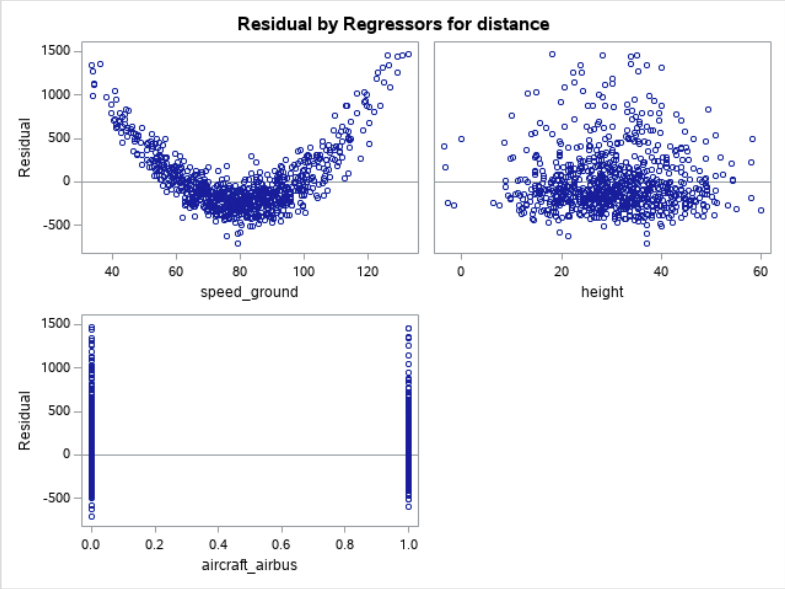

Fig 3.10 Residuals Plots for Model 3

3.4 SUMMARY

1. Dummy variable “aircraft_airbus” was introduced instead of the categorical variable “aircraft”. This would be equal to 1 for airbus and 0 for boeing.

2. On applying simple linear regression for each predictor against distance, we see that R-square value for height, pitch and no_pasg is close to 0.

3. On applying multiple linear regression against distance, we see that the p-value is more than 0.05 for duration, pitch and speed_ground. Here we cannot reject the NULL hypothesis that the coefficients of these variables is 0. Therefore we went ahead and dropped these predictors from our model.

4. Now, we regressed distance against speed_air, aircraft_airbus and height. The R-square was 0.97 and the residual analysis of mean and variance looked fine (Fig 3.8).

5. Since speed_air and speed_ground were highly correlated, we tried regressing distance against speed_ground, aircraft_airbus and height. The R-square was 0.84 but the residual analysis of speed_ground demonstrated a pattern (Fig 3.10). From Chapter 2, we found that air_speed has better linear relationship with distance.

6. Hence we use our final model as Model 2 i.e use predictors – speed_air, aircraft_airbus and height to regress distance. The final model is:

   Distance = -5962.93 + 82.14 * speed_air – 427.44 * aircraft_airbus + 13.70 *height

   Special Case:
   For scenarios with missing speed_air values, we can use speed_ground in our model instead of speed_air. In that case, the alternative model would be :

   Distance = -2004.12 + 42.34 * speed_ground – 495.28 *aircraft_airbus + 13.91 *height

Project Q&A

1. How many observations (flights) do you use to fit your **final** model? If not all 950 flights, why?

   We used 836 Observations in total for modelling the data. The FAA1.xls and FAA2.xls had 950 observations in total (excluding the blank 50 rows). During data exploration we found that there were abnormal and duplicate values. Data cleaning involved removing these rows from our dataset. In this process, the rows with specific column missing values has not been dropped and neither the missing values has been imputed with mean/median(to avoid losing variability from original data).

2. What factors and how they impact the landing distance of a flight?

   Upon applying regression and on analysis we found that Air Speed, Height and Aircraft Type were most significant in determining Landing Distance. The Model is

   Distance = -5962.93 + 82.14 * speed_air – 427.44 * aircraft_airbus + 13.70 *height

   1 unit of increase in Air Speed will result in 82.14 units of increase in Landing Distance keeping all other variables constant.

   1 unit of increase in height will result in 13.70 units of increase in Landing Distance keeping all other variables constant.

   The regression coefficient for aircraft_airbus provides a measure of the difference between airbus group identified by the dummy variable and the group that serves as a reference (boeing). Here, the regression coefficient for airbus is -427.44. This suggests that, after effects of all other variables into account, airbus will have 427.44 units of distance lower than the reference group (boeing).

   And in case of missing speed_air values we can use alternative model using ground_speed as mentioned in section 3.4

   Distance = -2004.12 + 42.34 * speed_ground – 495.28 *aircraft_airbus + 13.91 *height

3. Is there any difference between the two makes Boeing and Airbus?

   Yes there is a difference in factors and how they affect landing distance for “boeing” and “airbus” aircrafts. The TTEST shows the different mean landing distance and landing distributions for both the aircrafts- “airbus” and “boeing”.