Executive Summary

The two datasets FAA1 and FAA2 have similar columns. For this project, I have performed following actions - Data Cleaning, Descriptive Analysis, Statistical Modeling and Model Checking.

The challenge is to identify the redundant, missing, unacceptable/abnormal values in the data. After these values were removed, descriptive analysis operations such as finding means, summary, standard deviation t-test were performed. This analysis will be useful to find the relationship between landing distance and other variables and help in model creation.

Using statistical modeling, I created linear regression model between landing distance and other related variables. The Model Checking helped in validation of the models and was helpful in residual analysis and gave clearer picture about which model to use.

Chapter 1- Data Cleaning

Importing the Excel file - 1

```
PROC IMPORT
out = WORK.FAA1
datafile= "/home/u44024564/sasuser.v94/FAA1.xls"
dbms=xls replace;
sheet="FAA1";
getnames=yes;
RUN;
PROC Print Data=FAA1;
Run;
```

Importing the Excel file - 2

```
PROC IMPORT
out = WORK.FAA2
datafile= "/home/u44024564/sasuser.v94/FAA2.xls"
dbms=xls replace;
sheet="FAA2";
getnames=yes;
data=FAA2;
RUN;
PROC Print Data=FAA2;
Run;
```

Combining the data files – consists of 1000 rows

```
DATA combined;
Set FAA2 FAA1;
by;
PROC Print Data=combined;
Run;
```

Obs	aircraft	no_pasg	speed_ground	speed_air	height	pitch	distance	duration
1	boeing	53	107.91568005	109.32837648	27.418924252	4.0435145715	3369.8363638	.
2	boeing	69	101.65558863	102.8514051	27.804716181	4.1174316991	2987.8039235	.
3	boeing	61	71.051960883	.	18.589385734	4.4340431286	1144.922426	.
4	boeing	56	85.813327679	.	30.744597235	3.8842361245	1664.2181584	.
5	boeing	70	59.888528183	.	32.397688062	4.0260964152	1050.2644976	.
6	boeing	55	75.014343744	.	41.21496259	4.203853398	1627.0681991	.
7	boeing	54	54.4298029	.	24.03532163	3.8376457299	805.30399317	.
8	boeing	57	57.101661737	.	19.388837508	4.6436717769	573.62178606	.
9	boeing	61	85.443624251	.	35.375389749	4.2287278648	1698.9927548	.
10	boeing	56	61.796710514	.	36.748816124	4.1843990127	1137.7457579	.
11	boeing	61	53.778126741	.	46.355832902	5.5563991716	1075.3717411	.
12	boeing	54	141.21863535	141.72493569	23.575935009	5.2168022511	6533.0476506	.
13	boeing	54	93.391762435	92.869561214	32.223489271	3.8182761471	2128.708285	.

986	airbus	58	83.198056989	.	16.030000255	3.3080804277	1022.0840734	127.07799051
987	airbus	36	47.486765029	.	13.984809941	4.2990197162	250.68976141	172.04931209
988	airbus	66	123.31053074	124.39076754	22.327175815	4.2767104875	4295.9006131	98.500307809
989	airbus	58	89.953914147	.	30.857402547	4.1721894508	1711.6142255	165.19313237
990	airbus	63	69.083940765	.	28.397461881	4.1725671613	812.31993404	104.6341753
991	airbus	55	102.74650485	102.41247143	38.534020829	4.1697182481	2623.6512568	264.59337482
992	airbus	71	84.333339769	.	28.932963149	2.9914135514	1190.0228225	182.74887203
993	airbus	67	60.789195406	.	33.797453043	3.7235699727	563.10266864	80.477362793
994	airbus	68	91.15047186	.	15.378493757	3.5127744454	1445.1634341	194.99198538
995	airbus	67	98.02641239	99.421688766	40.993052838	4.7268210237	2440.381218	149.36036239
996	airbus	50	73.939616162	.	42.353383637	3.9571042618	1027.2134659	98.461455246
997	airbus	57	79.982703338	.	42.244751261	3.785954219	1162.404395	114.23485681
998	airbus	63	75.368171615	.	31.340776135	3.5580199527	960.25559642	118.57607182
999	airbus	63	77.148459304	.	23.602422529	3.020177825	899.43055864	200.62136624
1000	airbus	59	66.464640399	.	48.067790297	4.1656597705	853.86453785	124.14010259

Summary of each variable before performing Data Cleaning:

The UNIVARIATE Procedure
Variable: no_pasg (no_pasg)

Moments			
N	950	Sum Weights	950
Mean	60.1652632	Sum Observations	57157
Std Deviation	7.49000414	Variance	56.1001619
Skewness	-0.0187692	Kurtosis	0.2140783
Uncorrected SS	3492105	Corrected SS	53239.0537
Coeff Variation	12.4490507	Std Error Mean	0.24300782

The UNIVARIATE Procedure
Variable: speed_ground (speed_ground)

Moments			
N	950	Sum Weights	950
Mean	79.284994	Sum Observations	75320.7443
Std Deviation	19.3364178	Variance	373.897052
Skewness	0.15066504	Kurtosis	-0.0936015
Uncorrected SS	6326633.06	Corrected SS	354828.303
Coeff Variation	24.3884962	Std Error Mean	0.62735623

The UNIVARIATE Procedure
Variable: speed_air (speed_air)

Moments			
N	239	Sum Weights	239
Mean	103.730417	Sum Observations	24791.5698
Std Deviation	10.6051134	Variance	112.46843
Skewness	1.10536209	Kurtosis	0.99689462
Uncorrected SS	2598407.37	Corrected SS	26767.4864
Coeff Variation	10.2237258	Std Error Mean	0.68598776

The UNIVARIATE Procedure
Variable: height (height)

Moments			
N	950	Sum Weights	950
Mean	30.1392714	Sum Observations	28632.3078
Std Deviation	10.3593491	Variance	107.316113
Skewness	-0.1009795	Kurtosis	0.10468453
Uncorrected SS	964799.887	Corrected SS	101842.991
Coeff Variation	34.3715976	Std Error Mean	0.33610167

The UNIVARIATE Procedure
Variable: pitch (pitch)

Moments			
N	950	Sum Weights	950
Mean	4.01924722	Sum Observations	3818.28486
Std Deviation	0.52603224	Variance	0.27670992
Skewness	0.03086593	Kurtosis	-0.0642412
Uncorrected SS	15609.2285	Corrected SS	262.597716
Coeff Variation	13.08783	Std Error Mean	0.01706674

The UNIVARIATE Procedure
Variable: distance (distance)

Moments			
N	950	Sum Weights	950
Mean	1548.82326	Sum Observations	1471382.1
Std Deviation	948.681256	Variance	899996.126
Skewness	1.68786721	Kurtosis	3.83994617
Uncorrected SS	3133007144	Corrected SS	854096323
Coeff Variation	61.2517438	Std Error Mean	30.7792843

The UNIVARIATE Procedure
Variable: duration (duration)

Moments			
N	800	Sum Weights	800
Mean	154.006538	Sum Observations	123205.231
Std Deviation	49.2592338	Variance	2426.47211
Skewness	0.12147943	Kurtosis	-0.0551851
Uncorrected SS	20913162.3	Corrected SS	1938751.22
Coeff Variation	31.9851574	Std Error Mean	1.74157691

Performing Data Cleaning:

Step 1: **Checking missing values:**

proc means data=combined

NMISS N;

Run;

Speed_air has the maximum number of missing values.

The MEANS Procedure

Variable	Label	N Miss	N
no_pasg	no_pasg	50	950
speed_ground	speed_ground	50	950
speed_air	speed_air	761	239
height	height	50	950
pitch	pitch	50	950
distance	distance	50	950
duration	duration	200	800

Step 2: **Removing duplicate rows to remove redundancy:**

PROC SORT data=combined

out=datawoithoutduplicates

nodupkey;

by aircraft no_pasg speed_ground speed_air height pitch distance;

RUN;

PROC PRINT DATA = datawoithoutduplicates;

RUN;

```
: There were 851 observations read from the data set WORK.DATAWOITHOUTDUPLICATES.
: PROCEDURE PRINT used (Total process time):
  real time               1.59 seconds
  user cpu time           1.59 seconds
  system cpu time         0.00 seconds
  memory                  2952.81k
  OS Memory               30888.00k
  Timestamp               11/13/2019 01:15:30 AM
  Step Count                        68  Switch Count  0
  Page Faults                       0
  Page Reclaims                     378
  Page Swaps                        0
  Voluntary Context Switches        0
  Involuntary Context Switches      1
  Block Input Operations            0
```

Step 3: **Checking variable dictionary and removing abnormal values:**

```
DATA new_Data;
SET datawoithoutduplicates;
if aircraft ="" then DELETE;
if duration<40 then DELETE;
if no_pasg<0 then DELETE;
```

```
if speed_ground<0 then DELETE;
if speed_ground<30 or speed_ground>140 then DELETE;
if height<6 then DELETE;
if distance<0 or distance>6000 then DELETE;
Run;
```

```
NOTE: There were 686 observations read from the data set WORK.NEW_DATA.
NOTE: PROCEDURE PRINT used (Total process time):
      real time               1.24 seconds
      user cpu time           1.23 seconds
      system cpu time         0.01 seconds
      memory                  2978.28k
      OS Memory               30888.00k
      Timestamp               11/13/2019 01:37:15 AM
      Step Count                        95  Switch Count  0
```

Obs	aircraft	no_pasg	speed_ground	speed_air	height	pitch	distance	duration
1	airbus	36	47.486765029	.	13.984809941	4.2990197162	250.68976141	172.04931209
2	airbus	38	85.180842251	.	37.028793691	4.1216901717	1257.0092519	188.01797726
3	airbus	40	80.627416679	.	28.60255713	3.6234201886	1021.0888117	93.540807771
4	airbus	41	97.568203986	96.978436701	38.409192953	3.5322719834	2167.7576915	123.30242152
5	airbus	43	82.483044979	.	30.140024889	4.0896284195	1321.0000654	109.19713407
6	airbus	44	99.596841547	99.160266345	35.187030092	3.8402667146	2116.080919	139.31381028
7	airbus	45	72.490616757	.	33.228125197	4.3693164876	748.7667918	214.22048507
8	airbus	45	77.805502137	.	20.189958388	4.178015403	905.49788375	182.7116757
9	airbus	45	81.375317855	.	46.285569727	4.2052754575	1459.5022976	197.43183449
10	airbus	45	86.875879978	.	34.838071106	3.7997683715	1262.1538907	115.86922387
11	airbus	45	91.618595738	.	38.324199382	4.7436314527	1967.6109937	216.87640251
12	airbus	46	69.037688441	.	48.559845058	4.1973990963	1127.8005331	149.65108927
13	airbus	47	92.907591785	95.762981617	23.784897183	3.9086740504	1955.3034911	168.79553003
672	boeing	72	74.833473963	.	26.186971031	3.9863265825	1449.0794019	158.16277689
673	boeing	72	76.999831013	.	13.801193211	4.2019377631	1273.6644967	78.24709997
674	boeing	73	52.360449116	.	44.121081799	4.4970887384	1078.0988933	236.19293349
675	boeing	73	57.426894092	.	48.868703509	4.1846317788	1154.4436143	213.77917717
676	boeing	74	79.257984189	.	37.19716518	4.3370030837	1158.8376732	112.31707003
677	boeing	74	86.852747395	.	16.894461754	3.8308960194	1725.3804918	168.23013919
678	boeing	75	69.880248247	.	31.31135869	4.6879165411	1045.0302857	124.54353753
679	boeing	75	106.7461226	106.73317595	18.346201583	4.8074017332	2785.855295	79.705863144
680	boeing	76	63.597942325	.	36.489042355	4.4917734289	1051.9369604	147.03191592
681	boeing	76	88.103462433	.	42.085495821	4.6540097977	1927.0536775	219.72115595
682	boeing	77	55.086685785	.	38.032817792	4.0971206341	998.09700633	130.16891519
683	boeing	77	82.29713755	.	44.758716354	4.2293090445	1809.27205	172.56012205
684	boeing	79	106.93389135	108.42651323	30.457709156	4.8421492	3203.3188407	128.93810992
685	boeing	80	82.509055403	.	36.680194026	4.685310032	1590.3719225	161.82569155
686	boeing	82	40.815188666	.	22.618444074	4.8765952309	761.4850777	194.4671661

Step 4: Checking missing values after Data cleaning:

```
proc means data=new_DATA
        NMISS N;
        Run;
```

The MEANS Procedure

Variable	Label	N Miss	N
no_pasg	no_pasg	0	686
speed_ground	speed_ground	0	686
speed_air	speed_air	521	165
height	height	0	686
pitch	pitch	0	686
distance	distance	0	686
duration	duration	0	686

Step 4: Summary of each variable after performing Data Cleaning

```
PROC UNIVARIATE data=new_DATA;
```

Run;

The UNIVARIATE Procedure
Variable: no_pasg (no_pasg)

Moments			
N	686	Sum Weights	686
Mean	59.9985423	Sum Observations	41159
Std Deviation	7.51319992	Variance	56.4481731
Skewness	-0.0244209	Kurtosis	0.45274178
Uncorrected SS	2508147	Corrected SS	38666.9985
Coeff Variation	12.5223041	Std Error Mean	0.28685531

The UNIVARIATE Procedure
Variable: speed_ground (speed_ground)

Moments			
N	686	Sum Weights	686
Mean	79.821828	Sum Observations	54757.774
Std Deviation	18.7090654	Variance	350.029127
Skewness	0.05935343	Kurtosis	-0.1889996
Uncorrected SS	4610635.57	Corrected SS	239769.952
Coeff Variation	23.4385329	Std Error Mean	0.71431544

The UNIVARIATE Procedure
Variable: speed_air (speed_air)

Moments			
N	165	Sum Weights	165
Mean	103.778858	Sum Observations	17123.5116
Std Deviation	9.69395004	Variance	93.9726674
Skewness	0.86484898	Kurtosis	0.28846318
Uncorrected SS	1792469.99	Corrected SS	15411.5174
Coeff Variation	9.34096812	Std Error Mean	0.75467299

The UNIVARIATE Procedure
Variable: height (height)

Moments			
N	686	Sum Weights	686
Mean	30.395991	Sum Observations	20851.6498
Std Deviation	9.66090875	Variance	93.3331578
Skewness	0.11366768	Kurtosis	-0.2727078
Uncorrected SS	697739.773	Corrected SS	63933.2131
Coeff Variation	31.7834965	Std Error Mean	0.36885521

The UNIVARIATE Procedure
Variable: pitch (pitch)

Moments			
N	686	Sum Weights	686
Mean	4.00472333	Sum Observations	2747.2402
Std Deviation	0.52641015	Variance	0.27710764
Skewness	-0.0431676	Kurtosis	-0.1505883
Uncorrected SS	11191.7557	Corrected SS	189.818735
Coeff Variation	13.1447319	Std Error Mean	0.02009843

The UNIVARIATE Procedure
Variable: distance (distance)

Moments			
N	686	Sum Weights	686
Mean	1514.67553	Sum Observations	1039067.42
Std Deviation	890.672771	Variance	793297.985
Skewness	1.44816725	Kurtosis	2.45684501
Uncorrected SS	2117259115	Corrected SS	543409120
Coeff Variation	58.8028756	Std Error Mean	34.0060444

The UNIVARIATE Procedure
Variable: duration (duration)

Moments			
N	686	Sum Weights	686
Mean	155.974421	Sum Observations	106998.453
Std Deviation	48.772079	Variance	2378.71569
Skewness	0.1776281	Kurtosis	-0.2058045
Uncorrected SS	18318442	Corrected SS	1629420.25
Coeff Variation	31.2692804	Std Error Mean	1.86212663

Observation from Data Cleaning:

All the duplicate and abonormal values are deleted.
Missing columns are not replaced with any other value because the case doesn't specify explicitly what value can be used in place of missing values.

Chapter 2 - Descriptive Study

Goal: After data cleaning, we have the desired data, we will start doing analysis using the descriptive approach.

First goal is to find the correlation between the variables, analyze the plots.

Step 1: Check the mean, Median, Std. Dev., Variance, Quartile, Min, Max Range.

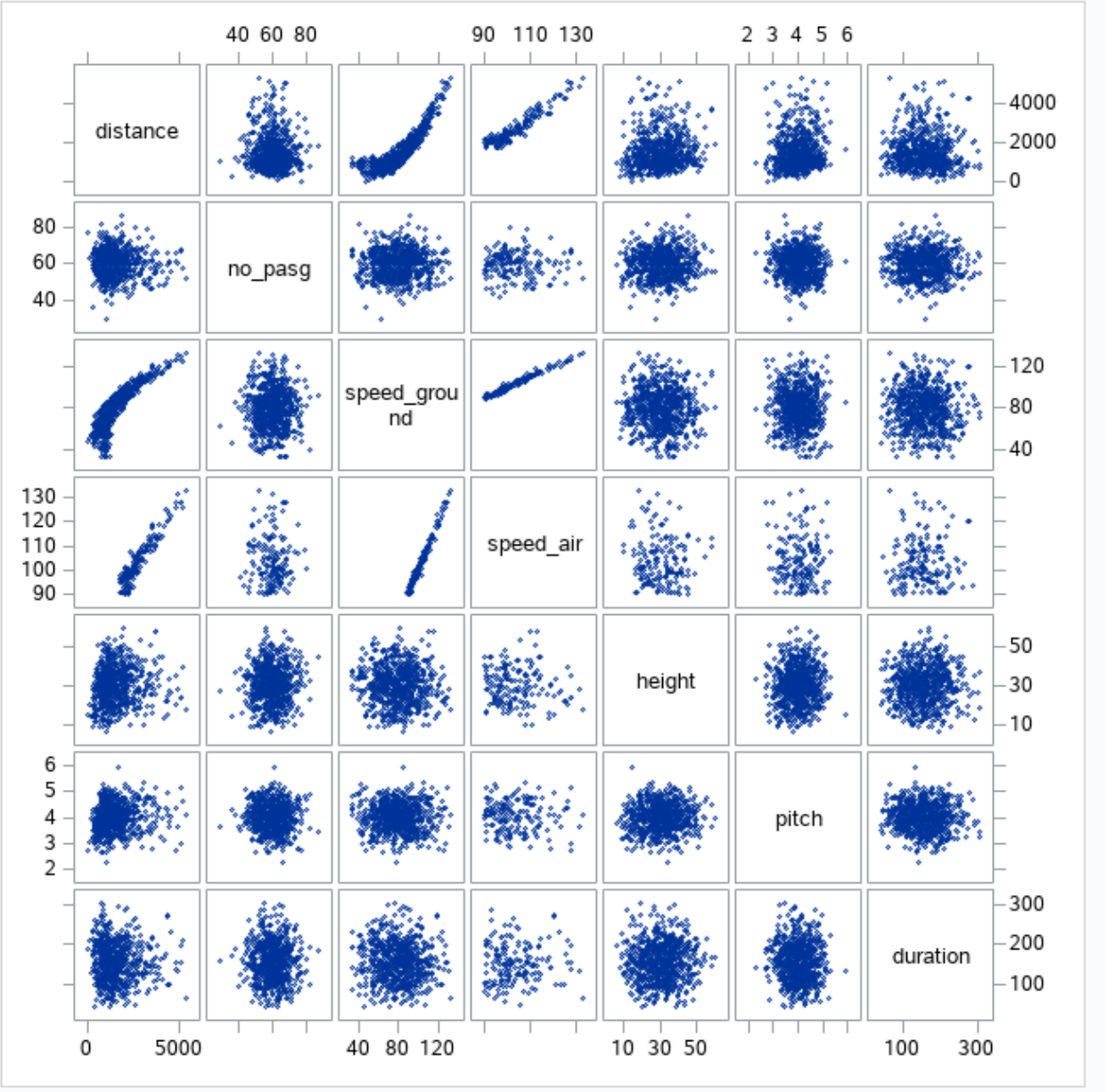
PROC Means DATA = new_Data nmiss mean median std var q1 q3 min max range;
TITLE Descriptive Study;
Run;

Statistical Analysis

The MEANS Procedure

Variable	Label	N Miss	Mean	Median	Std Dev	Variance	Lower Quartile	Upper Quartile	Minimum	Maximum	Range
no_pasg	no_pasg	0	59.9985423	60.0000000	7.5131999	56.4481731	55.0000000	65.0000000	29.0000000	87.0000000	58.0000000
speed_ground	speed_ground	0	79.8218280	80.1667143	18.7090654	350.0291274	66.4308787	92.0982905	33.5741041	132.7846766	99.2105726
speed_air	speed_air	521	103.7788580	101.3356142	9.6939500	93.9726674	96.7295938	109.4581269	90.0028586	132.9114649	42.9086063
height	height	0	30.3959910	30.2330307	9.6609087	93.3331578	23.5944766	36.8211847	6.2275178	59.9459639	53.7184462
pitch	pitch	0	4.0047233	4.0053227	0.5264101	0.2771076	3.6365363	4.3710717	2.2844801	5.9267842	3.6423041
distance	distance	0	1514.68	1259.89	890.6727708	793297.98	898.3072330	1935.07	41.7223127	5343.20	5301.48
duration	duration	0	155.9744209	155.6978936	48.7720790	2378.72	120.3527974	191.0184614	41.9493694	305.6217107	263.6723414

Step 2: We will create scatter plot to find about the potential associations between the variables.
PROC SGSCATTER data=new_data;
MATRIX distance no_pasg speed_ground speed_air height pitch duration;
Run;



Step 3: We will perform T test.

Proc ttest data= new_data;
class= distance;
var speed_ground;
Title Ttest between distance and duration variables;

test for aitcraft and distance

The TTEST Procedure

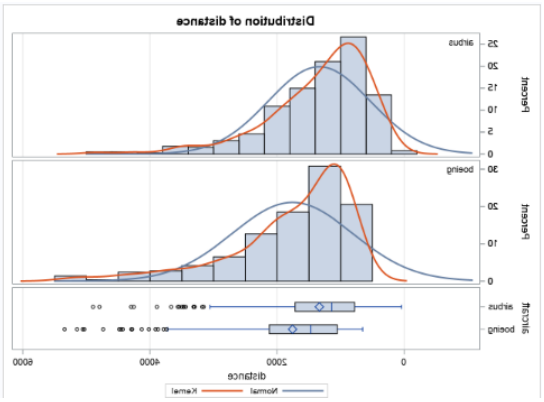
Variable: distance (distance)

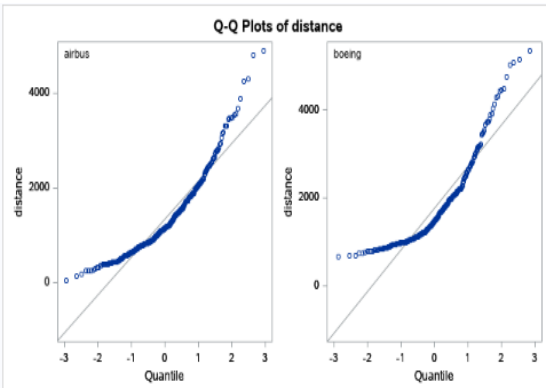
aircraft	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
airbus		394	1335.2	802.8	40.4464	41.7223	4896.3
boeing		292	1756.9	945.7	55.3444	653.6	5343.2
Diff (1-2)	Pooled		-421.8	866.5	66.9110		
Diff (1-2)	Satterthwaite		-421.8		68.5486		

aircraft	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
airbus		1335.2	1255.6	1414.7	802.8	750.4	863.2
boeing		1756.9	1648.0	1865.8	945.7	874.7	1029.4
Diff (1-2)	Pooled	-421.8	-553.1	-290.4	866.5	822.9	915.0
Diff (1-2)	Satterthwaite	-421.8	-556.4	-287.1			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	684	-6.30	<.0001
Satterthwaite	Unequal	565.42	-6.15	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	291	393	1.39	0.0026





Step 4:

To incorporate aircraft type to the data model, we will have to convert it to numeric value.

```
DATA model_data;
Set new_Data;
if aircraft= "airbus" then aircraftValue = 1;
```

else aircraftValue = 0;

Step 5: Check correlation between the variables:

```
PROC Corr data= model_Data;
var duration aircraftValue no_pasg speed_ground speed_air height
pitch distance;
Title Correlation in Data;
Run;
```

Correlation in Data

The CORR Procedure

8 Variables:	duration aircraftValue no_pasg speed_ground speed_air height pitch distance

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
duration	686	155.97442	48.77208	106998	41.94937	305.62171	duration
aircraftValue	686	0.57434	0.49480	394.00000	0	1.00000	
no_pasg	686	59.99854	7.51320	41159	29.00000	87.00000	no_pasg
speed_ground	686	79.82183	18.70907	54758	33.57410	132.78468	speed_ground
speed_air	165	103.77886	9.69395	17124	90.00286	132.91146	speed_air
height	686	30.39599	9.66091	20852	6.22752	59.94596	height
pitch	686	4.00472	0.52641	2747	2.28448	5.92678	pitch
distance	686	1515	890.67277	1039067	41.72231	5343	distance

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

	duration	aircraftValue	no_pasg	speed_ground	speed_air	height	pitch	distance
duration duration	1.00000 686	0.02214 0.5627 686	-0.04681 0.2208 686	-0.05351 0.1615 686	0.02572 0.7429 165	0.02267 0.5533 686	-0.04557 0.2333 686	-0.04556 0.2334 686
aircraftValue	0.02214 0.5627 686	1.00000 686	0.04460 0.2434 686	0.04412 0.2484 686	0.06452 0.4104 165	0.02456 0.5207 686	-0.39283 <.0001 686	-0.23430 <.0001 686
no_pasg no_pasg	-0.04681 0.2208 686	0.04460 0.2434 686	1.00000 686	-0.00305 0.9365 686	-0.00668 0.9322 165	0.03852 0.3137 686	-0.04648 0.2241 686	-0.02799 0.4642 686
speed_ground speed_ground	-0.05351 0.1615 686	0.04412 0.2484 686	-0.00305 0.9365 686	1.00000 686	0.98815 <.0001 165	-0.05682 0.1371 686	-0.03115 0.4153 686	0.86563 <.0001 686
speed_air speed_air	0.02572 0.7429 165	0.06452 0.4104 165	-0.00668 0.9322 165	0.98815 <.0001 165	1.00000 165	-0.09328 0.2334 165	-0.04170 0.5949 165	0.94249 <.0001 165
height height	0.02267 0.5533 686	0.02456 0.5207 686	0.03852 0.3137 686	-0.05682 0.1371 686	-0.09328 0.2334 165	1.00000 686	0.03949 0.3017 686	0.09700 0.0110 686
pitch pitch	-0.04557 0.2333 686	-0.39283 <.0001 686	-0.04648 0.2241 686	-0.03115 0.4153 686	-0.04170 0.5949 165	0.03949 0.3017 686	1.00000 686	0.10024 0.0086 686
distance distance	-0.04556 0.2334 686	-0.23430 <.0001 686	-0.02799 0.4642 686	0.86563 <.0001 686	0.94249 <.0001 165	0.09700 0.0110 686	0.10024 0.0086 686	1.00000 686

Observation:
A linear relationship was observed between distance and speed_ground, distance and speed_air and speed_air and speed_ground.

T-test suggests that there is significance of aircraft on the distance.

A strong correlation was observed between distance and speed_ground, distance and speed_air and speed_air and speed_ground. I will incorporate them in the model creation.

Chapter – 3 Statistical Modelling

The next goal is to find mathematical relation between the landing distance and other variables that are related. This can be achieved through the linear model. I performed regression between distance and other variables.

```
PROC REG DATA= model_data;
Model distance= speed_air;
Run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	686
Number of Observations Used	165
Number of Observations with Missing Values	521

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	96517152	96517152	1296.03	<.0001
Error	163	12138799	74471		
Corrected Total	164	108655951			

Root MSE	272.89404	R-Square	0.8883
Dependent Mean	2768.33404	Adj R-Sq	0.8876
Coeff Var	9.85770		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5444.41279	229.11595	-23.76	<.0001
speed_air	speed_air	1	79.13699	2.19822	36.00	<.0001

```
PROC REG DATA= model_data;
Model distance=speed_ground;
Run;
```

Root MSE	446.27723	R-Square	0.7493
Dependent Mean	1514.67553	Adj R-Sq	0.7489
Coeff Var	29.46355		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1774.73625	74.71808	-23.75	<.0001
speed_ground	speed_ground	1	41.20943	0.91140	45.22	<.0001

```
PROC REG DATA= model_data;
Model distance=aircraftValue;
Run;
```

Root MSE	866.51242	R-Square	0.0549
Dependent Mean	1514.67553	Adj R-Sq	0.0535
Coeff Var	57.20779		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1756.90985	50.70880	34.65	<.0001
aircraftValue		1	-421.75822	66.91097	-6.30	<.0001

```
PROC REG DATA= model_data;
Model distance=height;
Run;
```

Root MSE	887.12039	**R-Square**	0.0094
Dependent Mean	1514.67553	**Adj R-Sq**	0.0080
Coeff Var	58.56834		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1242.84872	111.89327	11.11	<.0001
height	height	1	8.94285	3.50848	2.55	0.0110

```
PROC REG DATA= model_data;
Model distance=aircraftValue height speed_air;
Run;
```

Root MSE	133.95895	**R-Square**	0.9734
Dependent Mean	2768.33404	**Adj R-Sq**	0.9729
Coeff Var	4.83897		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5937.94568	121.03310	-49.06	<.0001
aircraftValue		1	-421.07152	21.00737	-20.04	<.0001
height	height	1	13.75809	1.13294	12.14	<.0001
speed_air	speed_air	1	81.77047	1.08657	75.26	<.0001

```
PROC REG DATA= model_data;
Model distance=aircraftValue height speed_ground;
Run;
```

Root MSE	348.71033	R-Square	0.8474
Dependent Mean	1514.67553	Adj R-Sq	0.8467
Coeff Var	23.02211		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1999.99249	74.77724	-26.75	<.0001
aircraftValue		1	-498.99627	26.96314	-18.51	<.0001
height	height	1	14.21499	1.38186	10.29	<.0001
speed_ground	speed_ground	1	42.20881	0.71404	59.11	<.0001

Observation:

Model 1:

Landing Distance = -5937.946 + 12.75 Height + 81.77 Speed_air – 421.072 Aircraft_value

Model 2:

Landing Distance = -1999.99 + 14.22 Height + 42.208 Speed_ground – 498.996 Aircraft_value

Chapter 4- Model Checking

```
PROC REG DATA = model_Data;
Model distance = aircraftValue height speed_ground;
output out= diagnostic r=residual;
Run;
```

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1999.99249	74.77724	-26.75	<.0001
aircraftValue		1	-496.99627	26.96314	-18.51	<.0001
height	height	1	14.21499	1.38186	10.29	<.0001
speed_ground	speed_ground	1	42.20881	0.71404	59.11	<.0001

```
PROC REG DATA = model_Data;
Model distance = aircraftValue height speed_ground;
output out= diagnostic r=residual;
Run;
```

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-5937.94568	121.03310	-49.06	<.0001
aircraftValue		1	-421.07152	21.00737	-20.04	<.0001
height	height	1	13.75809	1.13294	12.14	<.0001
speed_air	speed_air	1	81.77047	1.08657	75.26	<.0001

Final model -

Landing Distance = -5937.946 + 12.75 Height + 81.77 Speed_air – 421.072 Aircraft_value

Observation:
In section 3, I created two models, one with speed_air, another with speed_ground. But, after checking the correlation values, I decided to go ahead with speed_ground. The reason being that speed_air has more stronger correlation with landing distance as compare to speed_ground