LANDING DISTANCE OF A COMMERCIAL FLIGHT

IDENTIFYING FACTORS IMAPCTING LANDING DISTANCE TO REDUCE THE RISK OF OVERRUNS

Ameya Jamgade | BANA6043

EXECUTIVE SUMMARY

This report provides an analysis and evaluation of flight landing data of commercial aircrafts to identify the nature and magnitude of factors affecting landing distance.

Two datasets (FAA1 and FAA2) containing flight landing information of commercial aircrafts Airbus and Boing were provided. The goal of this project was to conduct statistical analysis using SAS to study what factors and how they would impact the landing distance thereby mitigating risk of overruns.

The data preparation stage consisted of combining the two datasets to make handling data much easier. Duplicates were identified and removed from the combined dataset. Validations were performed to make sure that all the values were in accordance with the rules specified in the project document. Statistical analysis was performed on columns containing large number of missing values to check if any column can be removed completely. All missing values in every column were replaced by median value of non-missing values in that column respectively. Finally, all the outliers were flagged. The result consisted of 851 rows.

The statistical analysis process began with plotting graphs of response variable (*Landing Distance)* against each of the predictor variables. This helped in getting a basic understanding of connection (linear, exponential, no relation etc.) between the response variable Distance with remaining explanatory variables. In order to better understand the data and measure the strength and direction of association that exists between two variables, pairwise correlation of data was performed. *SpeedGround* and *SpeedAir* had slightly strong positive correlation between them.

Next step in the analysis was to measure the strength of the relationship between dependent variable (Distance) and a series of other changing variables using regression. Airbus was coded as 1 and Boing as 0. P-values of all predictor variables were noted to single out those values which were statistically insignificant (*Duration, No. of passengers and Pitch*). These three variables were removed from the model and VIF was calculated for all variables to test multi-collinearity. Multi-collinearity was absent in thus generated model and high value of adjusted R^2 indicated goodness of model fit.

Final model thus obtained is as below,

Distance = -6954.39 – 498.24(Aircraft) + 36.06(Speed_Ground) + 52.99(Speed_Air) + 14.71(Height)

where Aircraft =1 for Airbus and Aircraft=0 for Boeing

The Landing distance was calculated for both Airbus and Boeing. To understand the distribution of the two values of Landing Distance better, a Box-Whiskers plot was created. From the box-plot, we see that both mean and median for Airbus are lower than those of Boeing are. In addition, the IQR range is higher for Boeing than that of Airbus. Since range of Landing Distance Airbus aircrafts is lower that Boeing, this indicates that the Airbus has higher risk of overruns. Boeing aircrafts give pilots more room about landing distance to prevent overruns.

PROJECT INFORMATION

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 Provided: Two datasets FAA1&FAA2 were provided consisting of information about flight statistics of generic Boeing and Airbus aircrafts.

DATA PREPARATION PROCESS

Concatenation- After briefly examining the data, the two data sets were first COMBINED to form a single dataset. This would make handling data much easier. Since the data was different in both datasets, hence merging function was avoided.

Identifying and removing duplicate values- It is important to check if the combined data set has any duplicate values. It was found that there are 100 duplicate records. These were identified and removed to avoid erroneous results.

Counting missing values- Missing values in the dataset were identified by creating a separate column (count). This helped in identifying an initial estimate of the quality of data provided. Specific pointer was placed to avoid counting those rows that would have all columns with missing values.

Validations to check and prevent bad data- Specific validations were done to check if character and numeric rules were followed for specific columns. Hence, in the final dataset, all character columns have character values and all numeric columns have numeric values.

Operations on columns with large number of missing values- Statistical distributions of columns with many missing values was plotted to understand the nature of existing data in these columns. All such columns had specific type of distributions. Hence, a decision was made not to exclude these columns for future analysis.

Replacing missing values with median- An option was to replace missing values with 0, but that would have rendered statistical analysis of the data incurred/biased. Hence, the most appropriate solution was to replacing missing values in all columns with median.

Identifying outliers- In order to see if each row contains any outliers that aren't in accordance with the conditions specified in metadata project document, a separate column was created.

Total number of rows in dataset after completing above operations= 851

DISCRIPTIVE STUDY

1. X-Y PLOTS

In order to study the relationship between Distance and other variables, simple X-Y plots were obtained. This helped in getting a basic understanding of connection between the response variable *Distance* with remaining explanatory variables. From these plots it can be seen that *Distance v. SpeedGround* has exponential distribution. Also *Distance v. SpeedAir* has a distorted linear distribution. All other plots have a lot of noise or there isn't any relationship. In order to test if there is any relationship among explanatory variables themselves, X-Y plots of some of these were also plotted. It can be seen that there is a relationship between *SpeedGround v. SpeedAir* (linear).

2. PAIRWISE CORRELATION

In order to better understand the data and measure the strength and direction of association that exists between two variables, pairwise correlation of data was performed. From the correlation table thus generated it can be seen that *Distance* has positive strong correlation with both *SpeedGround* and *SpeedAir*. Also *SpeedGround* and *SpeedAir* have slightly strong positive correlation between them. All other independent variables have correlation coefficient values close to 0 indicating that there isn't any correlation amongst themselves.

STATISTICAL MODELLING

REGRESSION ANALYSIS

The next step was to measure the strength of the relationship between dependent variable (*Distance*) and a series of other changing variables using regression. Airbus was coded as 1 and Boing as 0.

Distance = -6942.74– 484.23(Aircraft) – 0.41(Duration) – 2.09 (No.ofPasg) + 36.59(Speed_Ground) + 52.98(Speed_Air) + 14.67(Height) + 32.98(Pitch)

Where Aircraft =1 for Airbus and Aircraft=0 for Boeing

MODEL EVALUATION

a. p-values: Analyzing the output of regression it could be seen that the p-value for *Duration and No. of passengers* was greater for alpha=0.05. This indicated that they are not statistically significant. Both of these variables had negative coefficients indicating that they are negatively affecting the distance.

b. Multi-collinearity: From the above analysis and previous analysis of relationship between *SpeedGround* and *SpeedAir,* the next step was to identify if there is any quantifiable relationship

between the independent variables. Variance Inflation Factor was calculated for all the variables. It could be seen that VIF values are approx. = 1 for all independent variables. This indicates that no multi-collinearity is affecting our model.

c. Final Model: *Duration, No. of passengers and Pitch* were removed from final model for being statistically insignificant. The final model has all independent variables with VIF=1 indicating multi-collinearity is absent. The sign of independent variables did not change between the two models. Also, the adjusted R^2 = 86% indicates that this is a good model fit.

Distance = -6954.39 – 498.24(Aircraft) + 36.06(Speed_Ground) + 52.99(Speed_Air) + 14.71(Height)

d. Box-Plot: The Landing distance was calculated for both Airbus and Boeing. In order to better understand the distribution of the two values of Landing Distance, a box-plot was created. From the box-plot, we see that both mean and median for Airbus are lower than those of Boeing are. In addition, the IQR range is higher for Boeing than Airbus.

```
/*Importing file 1*/
PROC IMPORT OUT=FirstDataSet
                DATAFILE='~/Stat computing class/Stat Comp Project/FAA1' DBMS=xls REPLACE;
        SHEET="FAA1";
        GETNAMES=YES;
RUN;

/*Importing file 2*/
PROC IMPORT OUT=SecondDataSet
                DATAFILE='~/Stat computing class/Stat Comp Project/FAA2' DBMS=xls REPLACE;
        SHEET="FAA2";
        GETNAMES=YES;
RUN;

/*Appending both files into a single dataset*/
DATA FAA1FAA2COMBINED;
        SET FirstDataSet SecondDataSet;
RUN;

/*Summarizing the distribution of each variable*/
PROC MEANS DATA=FAA1FAA2COMBINED N MEAN MEDIAN STD RANGE NMISS;
        /*TITLE'SUMMARY STATS OF COMBINED UNCLEAN DATASET';*/
        VAR DURATION;
        VAR NO_PASG;
        VAR SPEED_GROUND;
        VAR SPEED_AIR;
        VAR HEIGHT;
        VAR PITCH;
        VAR DISTANCE;
        ;
RUN;

/* Finding and removing duplicates from the combined set*/
PROC SORT data=FAA1FAA2COMBINED NODUPKEY;
        BY aircraft duration no_pasg;
RUN;
```

```
DATA TEST1;
        SET FAA1FAA2COMBINED;

        /* Counting the missing values in all columns*/
        count=nmiss(OF DURATION NO_PASG SPEED_GROUND SPEED_AIR HEIGHT PITCH DISTANCE);

        if count<7;

        /*To avoid choosing rows where all columns are blank*/
        /* Check to see if dataset contains any Invalid characters removing those values in final data set*/
        IF AIRCRAFT NOT IN ('BOEING' , 'AIRBUS');

        IF VERIFY(DURATION, '0123456789');

        IF VERIFY(NO_PASG, '0123456789');

        IF VERIFY(SPEED_GROUND, '0123456789');

        IF VERIFY(SPEED_AIR, '0123456789');

        IF VERIFY(HEIGHT, '0123456789');

        IF VERIFY(PITCH, '0123456789');

        IF VERIFY(DISTANCE, '0123456789');
RUN;

/* Since Duration and Speed_air contain alot of missing values
understanding the distribution of these two columns to decide whether to exclude these columns or not     */
PROC UNIVARIATE DATA=TEST1;
        VAR DURATION;
        HISTOGRAM DURATION / NORMAL;

        /* HISTOGRAM variables / <OPTIONS> */
        VAR SPEED_AIR;
        HISTOGRAM SPEED_AIR / NORMAL;

        /* HISTOGRAM variables / <OPTIONS> */
Run;

/*Since the two columns have normal and right-skewed distribution, we cannot omit these.
Hence, replacing all missing values with MEDIAN*/
proc stdize data=TEST1 out=NEW1 MISSING=median reponly;
        VAR DURATION;
        VAR NO_PASG;
        VAR SPEED_GROUND;
        VAR SPEED_AIR;
        VAR HEIGHT;
        VAR PITCH;
        VAR DISTANCE;
run;

/* To check outliers for each columns with give specifications*/
DATA NEW2;
        SET NEW1;

        IF DURATION <=40 THEN
                Outlier='Present';
        else
                Outlier="Not present";

        IF SPEED_GROUND <=30 OR SPEED_GROUND>=140 THEN
                Outlier='Present';
        else
                Outlier="Not present";

        IF SPEED_AIR <=30 OR SPEED_AIR>=140 THEN
                Outlier='Present';
        else
```

```
                Outlier="Not present";

        IF HEIGHT < 6 THEN
                Outlier='Present';
        else
                Outlier="Not present";
        ;
RUN;

DATA FINALCOMBINEDDATA;
        SET NEW2;

        /* Referencing to cancatinated data set*/
        /* Labelling all given columns*/
        LABEL AIRCRAFT='AIRCRAFT';
        LABEL DURATION='DURATION';
        LABEL NO_PASG='NO. OF PASSENGERS';
        LABEL SPEED_GROUND='SPEED ON GROUND';
        LABEL SPEED_AIR='SPEED IN AIR';
        LABEL HEIGHT='HEIGHT';
        LABEL PITCH='PITCH';
        LABEL DISTANCE='DISTANCE';
        ;
RUN;

proc print data=FINALCOMBINEDDATA(obs=10);
        title 'FIRST 10 OBSERVATIONS OF DATASET';
RUN;

/*Summarizing the distribution of each variable*/
PROC MEANS DATA=FINALCOMBINEDDATA N MEAN MEDIAN STD RANGE NMISS;
        TITLE'SUMMARY STATS OF COMBINED DATASET AFTER COMPLETING DATA PREPARATION OPERATIONS';
        VAR DURATION;
        VAR NO_PASG;
        VAR SPEED_GROUND;
        VAR SPEED_AIR;
        VAR HEIGHT;
        VAR PITCH;
        VAR DISTANCE;
RUN;

/*CODING AIRBUS=1 AND BOEING=0*/
DATA TESTING;
        SET finalcombineddata;
        LABEL CODING='AIRCRAFT';

        IF AIRCRAFT='airbus' THEN
                CODING=1;
        ELSE if AIRCRAFT='boeing' THEN
                CODING=0;
RUN;

/*PROC PRINT DATA= TESTING;
RUN;*/
/* PLOTS OF LANDING DISTANCE WITH EACH OF OTHER VARIABLES*/
proc plot data=FINALCOMBINEDDATA;
        title 'BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES';

        /*PLOT OF DEPENDENT VARIABLE WITH ALL OTHER VARIABLES*/
        plot DISTANCE*AIRCRAFT;
        plot DISTANCE*DURATION;
        plot DISTANCE*NO_PASG;
        plot DISTANCE*SPEED_GROUND;
        plot DISTANCE*SPEED_AIR;
        plot DISTANCE*HEIGHT;
        plot DISTANCE*PITCH;

        /*PLOTTING DEPENDENT VARIABLES WITH EACH OTHER TO SEE IF THERE IS ANY CONNECTION*/
        plot SPEED_GROUND*SPEED_AIR;
```

```
        /*plot HEIGHT*PITCH;
        plot DURATION*SPEED_AIR;
        plot SPEED_GROUND*PITCH;
        plot HEIGHT*SPEED_AIR; */
        run;

        /* IDENTIFYING IF THERE IS ANY CORRELATION BETWEEN THE VARIABLES*/
proc corr data=FINALCOMBINEDDATA;
        var DISTANCE DURATION NO_PASG SPEED_GROUND SPEED_AIR HEIGHT PITCH;
        title 'PAIRWISE CORRELATION';
run;

/* Performing Regression analysis for all variables*/
PROC REG DATA=TESTING;
        MODEL DISTANCE=CODING DURATION NO_PASG SPEED_GROUND SPEED_AIR HEIGHT PITCH /
                VIF;
        title 'REGRESSION ANALYSIS INCLUDING ALL VARIABLES';
        run;

        /* Performing Regression analysis for remaining variables to test multicollinearity*/
PROC REG DATA=TESTING;
        MODEL DISTANCE=CODING SPEED_GROUND SPEED_AIR HEIGHT / VIF;
        title 'REGRESSION ANALYSIS FOR REMAINING VARIABLES TO TEST MULTICOLLINEARITY';
        run;



DATA TESTINGNEW;
        SET TESTING;
        /*DO I=1 TO 851;*/
        LANDINGDISTANCE = -6954.39 - (498.24 * CODING) + (36.06 * Speed_Ground) + (52.99*Speed_Air) +
(14.71*Height);
        /*FIELD2 = -6954.39 + (36.06 * Speed_Ground) + (52.99*Speed_Air) + (14.71*Height);*/
        /*OUTPUT;
        END;*/
RUN;



title 'Box Plot ';
proc boxplot data=TESTINGNEW ;
   plot LANDINGDISTANCE*AIRCRAFT;
   inset min mean max stddev /
      header = 'Overall Statistics'
      pos    = tm;
   insetgroup  mean ;

run;

PROC MEANS DATA=TESTINGNEW N MEAN MEDIAN STD RANGE ;
        by  descending coding;
        var LANDINGDISTANCE;
run;
```

The MEANS Procedure

Variable	Label	N	Mean	Median	Std Dev	Range	N Miss
duration	duration	800	154.0065385	153.9480975	49.2592338	290.8575036	150
no_pasg	no_pasg	950	60.1652632	60.0000000	7.4900041	58.0000000	0
speed_ground	speed_ground	950	79.2849940	79.4129094	19.3364178	113.4829200	0
speed_air	speed_air	239	103.7304174	100.8916770	10.6051134	51.7220771	711
height	height	950	30.1392714	29.9044945	10.3593491	63.4922163	0
pitch	pitch	950	4.0192472	4.0153874	0.5260322	3.6423041	0
distance	distance	950	1548.82	1267.44	948.6812561	6498.97	0

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

Basic Statistical Measures			
Location		Variability	
Mean	154.0065	Std Deviation	49.25923
Median	153.9481	Variance	2426
Mode	.	Range	290.85750
		Interquartile Range	69.44330

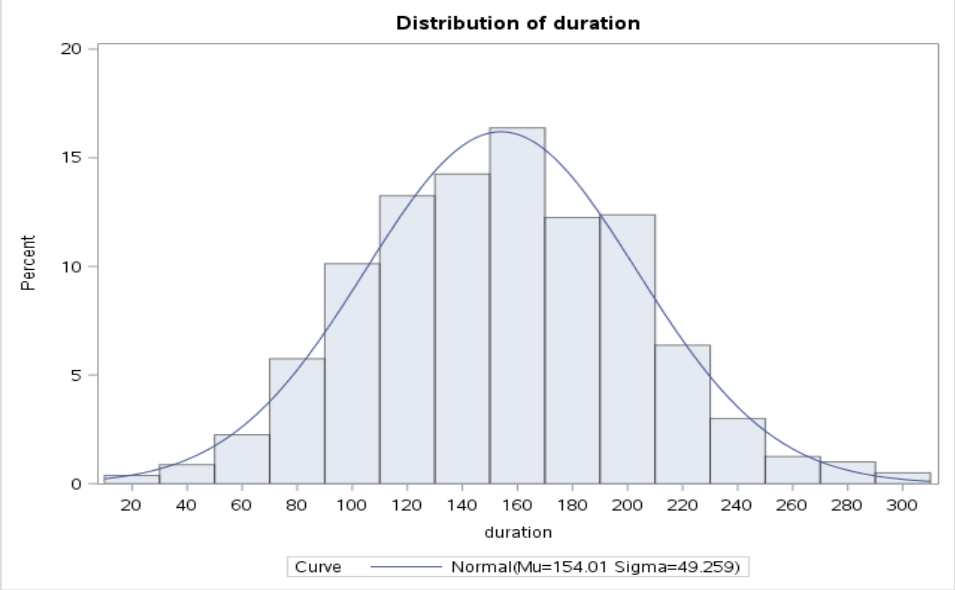
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	88.42936	Pr > |t|	<.0001
Sign	M	400	Pr >= |M|	<.0001
Signed Rank	S	160200	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	305.6217
99%	275.6969
95%	234.1229
90%	214.4738
75% Q3	188.9179
50% Median	153.9481
25% Q1	119.4746
10%	92.0313
5%	74.4080
1%	45.5691
0% Min	14.7642

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
14.7642	452	289.320	420
16.8935	23	293.230	850
17.3755	453	298.522	851
31.3910	454	302.967	421
31.7017	24	305.622	422

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	51	5.99	100.00

The UNIVARIATE Procedure



Fitted Normal Distribution for duration (duration)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	154.0065
Std Dev	Sigma	49.25923

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.02307107	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.05904902	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.41681999	Pr > A-Sq	>0.250

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	45.5691	39.4124
5.0	74.4080	72.9823
10.0	92.0313	90.8783
25.0	119.4746	120.7817
50.0	153.9481	154.0065
75.0	188.9179	187.2314
90.0	214.4738	217.1348
95.0	234.1229	235.0308
99.0	275.6969	268.6007

The UNIVARIATE Procedure
Variable: speed_air (speed_air)

Moments			
N	212	Sum Weights	212
Mean	103.899733	Sum Observations	22026.7433
Std Deviation	10.4434468	Variance	109.06558
Skewness	1.02447209	Kurtosis	0.71666251
Uncorrected SS	2311585.58	Corrected SS	23012.8374
Coeff Variation	10.0514664	Std Error Mean	0.71725887

Basic Statistical Measures			
Location		Variability	
Mean	103.8997	Std Deviation	10.44345
Median	101.1070	Variance	109.06558
Mode	92.1535	Range	51.72208
		Interquartile Range	13.32165

Note: The mode displayed is the smallest of 9 modes with a count of 2.

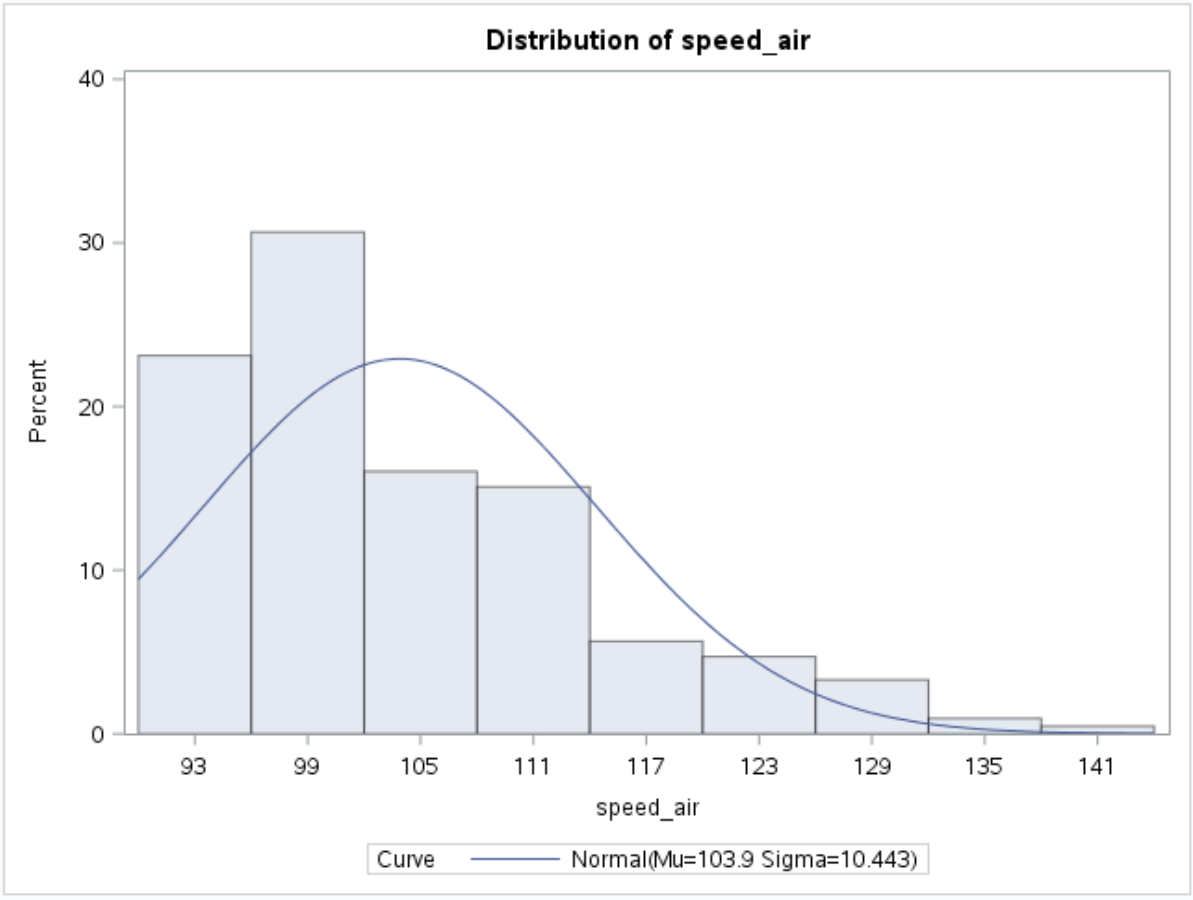
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	144.8567	Pr > |t|	<.0001
Sign	M	106	Pr >= |M|	<.0001
Signed Rank	S	11289	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	141.7249
99%	132.9115
95%	125.9869
90%	119.4021
75% Q3	109.5074
50% Median	101.1070
25% Q1	96.1857
10%	92.7526
5%	91.0725
1%	90.3674
0% Min	90.0029

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
90.0029	515	128.418	694
90.1110	497	131.338	154
90.3674	478	132.911	459
90.4767	729	136.423	553
90.5033	652	141.725	743

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	639	75.09	100.00

The UNIVARIATE Procedure



The UNIVARIATE Procedure
Fitted Normal Distribution for speed_air (speed_air)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	103.8997
Std Dev	Sigma	10.44345

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.11554382	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.73739925	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	4.61829345	Pr > A-Sq	<0.005

The UNIVARIATE Procedure
Fitted Normal Distribution for speed_air (speed_air)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	103.8997
Std Dev	Sigma	10.44345

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.11554382	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.73739925	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	4.61829345	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	90.3674	79.6046
5.0	91.0725	86.7218
10.0	92.7526	90.5159
25.0	96.1857	96.8557
50.0	101.1070	103.8997
75.0	109.5074	110.9437
90.0	119.4021	117.2835
95.0	125.9869	121.0777
99.0	132.9115	128.1948

FIRST 10 OBSERVATIONS OF DATASET

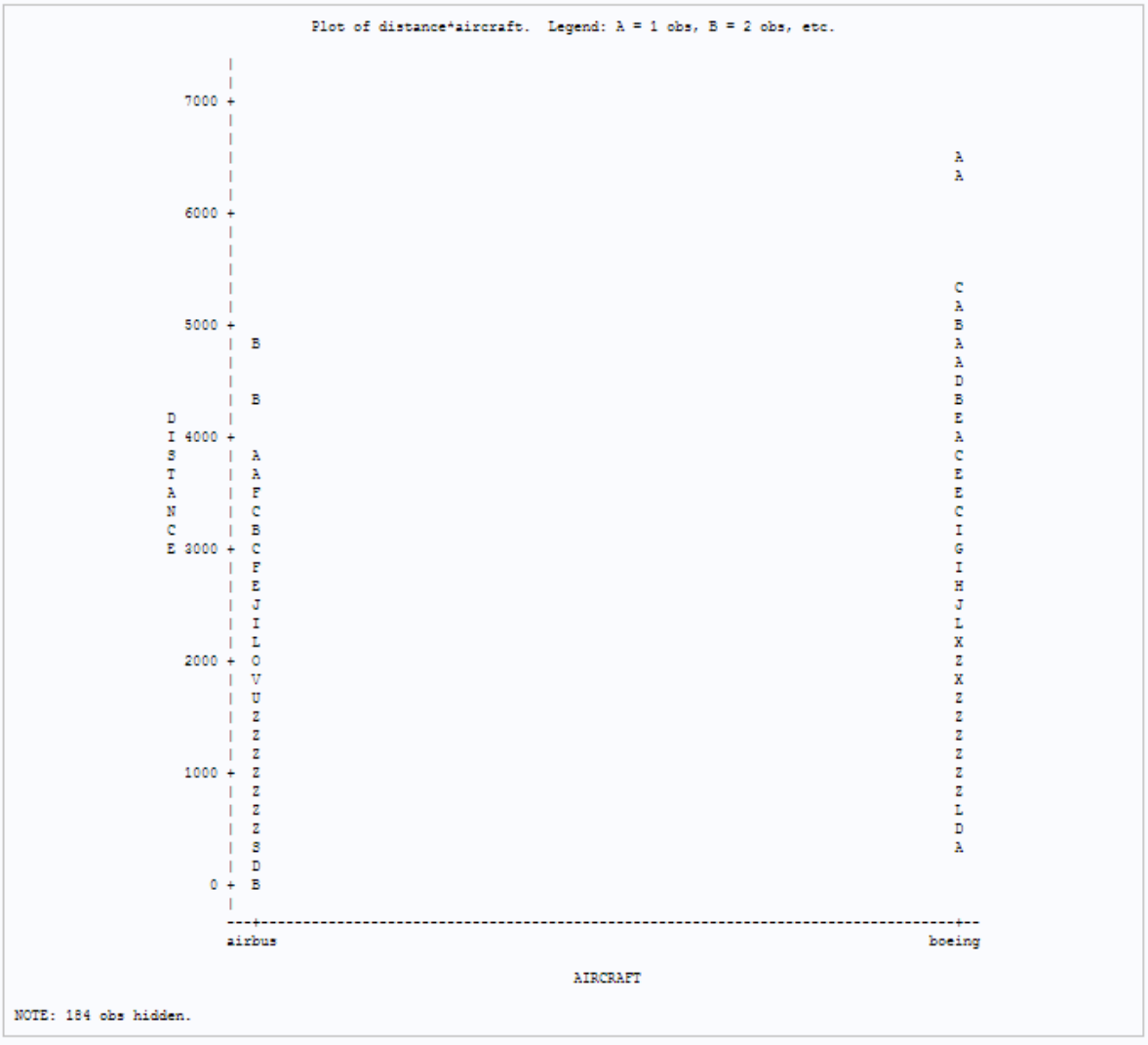
Obs	aircraft	duration	no_pasg	speed_ground	speed_air	height	pitch	distance	count	Outlier
1	airbus	153.94809754	46	40.801786477	101.10702128	24.400127629	3.9682093233	620.09051196	2	Not pre
2	airbus	153.94809754	48	61.570704648	101.10702128	21.785707448	4.3511947442	560.53392302	2	Not pre
3	airbus	153.94809754	50	84.219908138	101.10702128	32.542946798	3.318828622	1485.4400456	2	Not pre
4	airbus	153.94809754	51	62.484050366	101.10702128	26.53804471	3.8228939729	749.48028928	2	Not pre
5	airbus	153.94809754	52	73.761115944	101.10702128	9.688307724	3.3585464091	554.16098701	2	Not pre
6	airbus	153.94809754	54	83.071912777	101.10702128	37.317578277	3.4734612582	1338.6101651	2	Not pre
7	airbus	153.94809754	55	68.751529748	101.10702128	48.277120042	4.2626359629	1079.1170993	2	Not pre
8	airbus	153.94809754	56	86.528840828	101.10702128	40.94901507	3.7270256473	1437.6338566	2	Not pre
9	airbus	153.94809754	57	88.631154153	101.10702128	17.087228104	3.2746103159	1117.6839116	2	Not pre
10	airbus	153.94809754	58	101.09983213	101.46762268	42.952395762	4.2257015342	2688.9856974	1	Not pre

SUMMARY STATS OF COMBINED DATASET AFTER COMPLETING DATA PREPARATION OPERATIONS

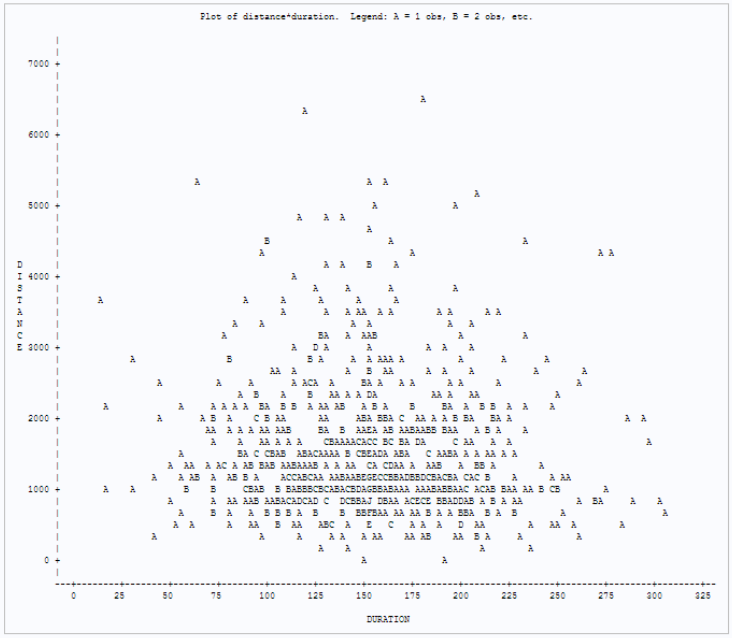
The MEANS Procedure

Variable	Label	N	Mean	Median	Std Dev	Range	N Miss
duration	DURATION	851	154.0030362	153.9480975	47.7586011	290.8575036	0
no_pasg	NO. OF PASSENGERS	851	60.1551116	60.0000000	7.6013563	58.0000000	0
speed_ground	SPEED ON GROUND	851	79.4323592	79.4155233	19.2620326	113.4829200	0
speed_air	SPEED IN AIR	851	101.8027379	101.1070213	5.3417747	51.7220771	0
height	HEIGHT	851	30.0928884	30.0479902	10.3355585	63.4922163	0
pitch	PITCH	851	4.0160081	4.0209904	0.5243580	3.6423041	0
distance	DISTANCE	851	1543.26	1262.15	943.8219715	6498.97	0

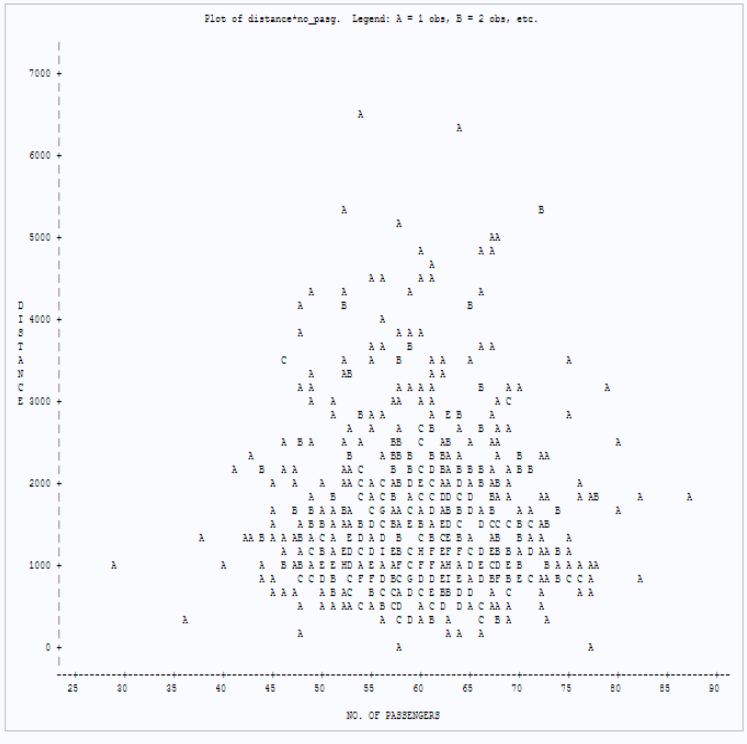
BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



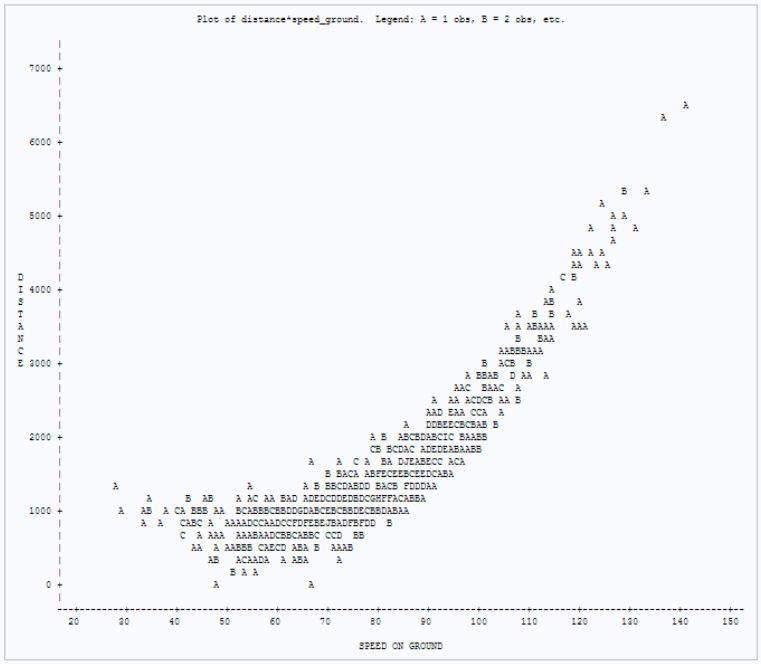
BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



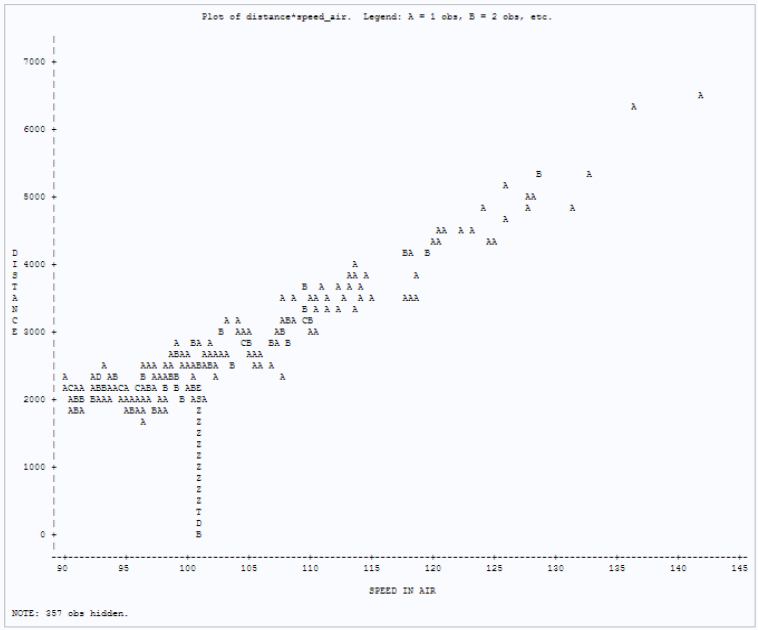
BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



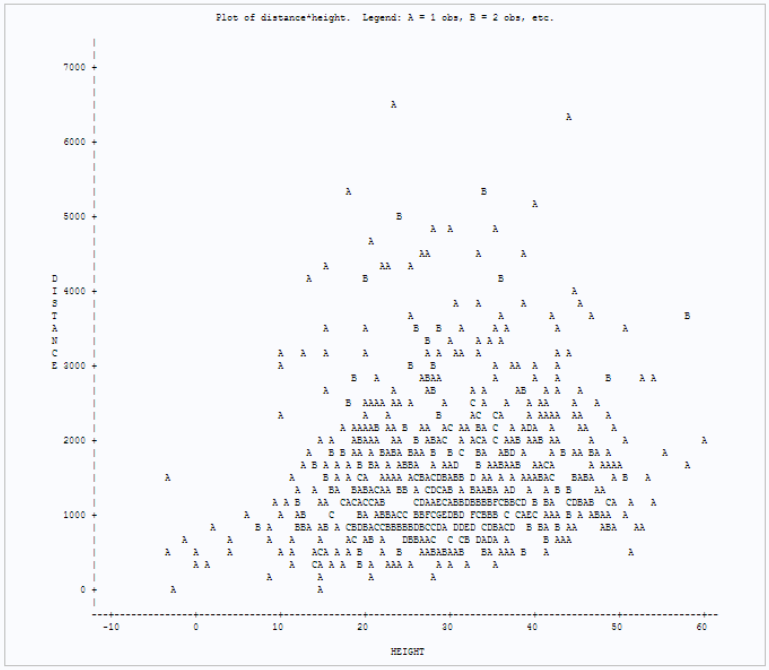
BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



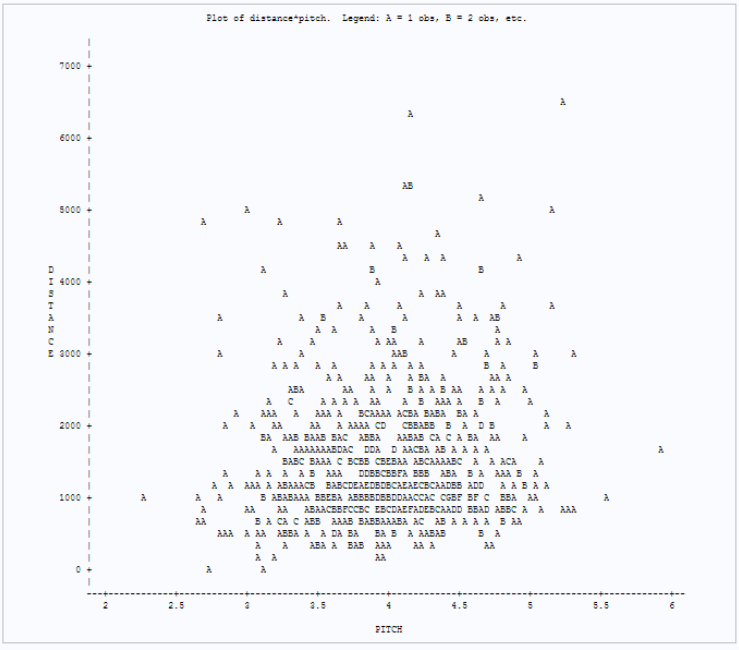
BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



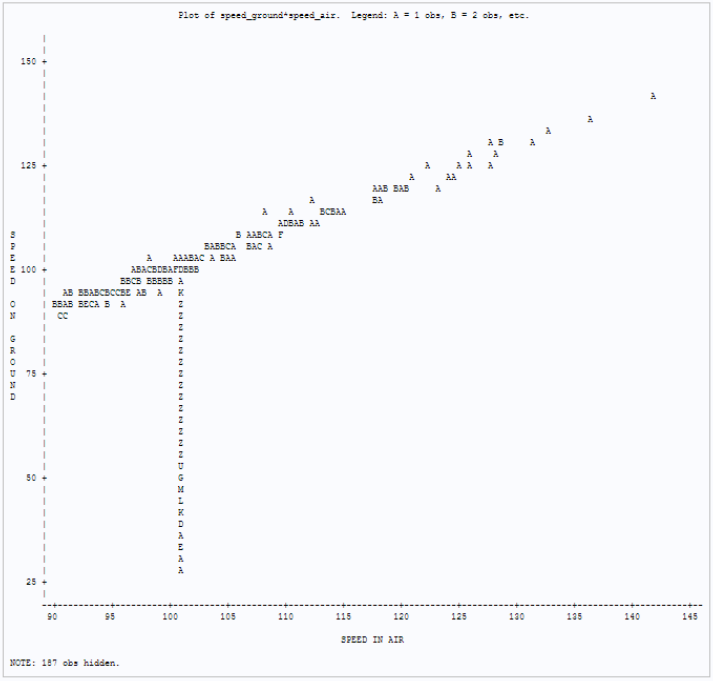
BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



BASIC PLOTS TO IDENTIFY ANY RELATIONSHIP AMONG VARIABLES



PAIRWISE CORRELATION

The CORR Procedure

7 Variables:	distance duration no_pasg speed_ground speed_air height pitch

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
distance	851	1543	943.82197	1313317	34.08078	6533	DISTANCE
duration	851	154.00304	47.75860	131057	14.76421	305.62171	DURATION
no_pasg	851	60.15511	7.60136	51192	29.00000	87.00000	NO. OF PASSENGERS
speed_ground	851	79.43236	19.26203	67597	27.73572	141.21864	SPEED ON GROUND
speed_air	851	101.80274	5.34177	86634	90.00286	141.72494	SPEED IN AIR
height	851	30.09289	10.33556	25609	-3.54625	59.94596	HEIGHT
pitch	851	4.01601	0.52436	3418	2.28448	5.92678	PITCH

Pearson Correlation Coefficients, N = 851
Prob > |r| under H0: Rho=0

	distance	duration	no_pasg	speed_ground	speed_air	height	pitch
distance DISTANCE	1.00000	-0.05984 0.0811	-0.01820 0.5960	0.86322 <.0001	0.62185 <.0001	0.14211 <.0001	0.09628 0.0049
duration DURATION	-0.05984 0.0811	1.00000	-0.03256 0.3428	-0.05870 0.0870	0.01081 0.7529	-0.00654 0.8490	-0.03780 0.2707
no_pasg NO. OF PASSENGERS	-0.01820 0.5960	-0.03256 0.3428	1.00000	0.00413 0.9041	-0.00178 0.9587	0.00560 0.8705	-0.00671 0.8449
speed_ground SPEED ON GROUND	0.86322 <.0001	-0.05870 0.0870	0.00413 0.9041	1.00000	0.42909 <.0001	-0.00990 0.7731	-0.03301 0.3361
speed_air SPEED IN AIR	0.62185 <.0001	0.01081 0.7529	-0.00178 0.9587	0.42909 <.0001	1.00000	-0.02244 0.5133	0.01410 0.6814
height HEIGHT	0.14211 <.0001	-0.00654 0.8490	0.00560 0.8705	-0.00990 0.7731	-0.02244 0.5133	1.00000	0.01825 0.5950
pitch PITCH	0.09628 0.0049	-0.03780 0.2707	-0.00671 0.8449	-0.03301 0.3361	0.01410 0.6814	0.01825 0.5950	1.00000

REGRESSION ANALYSIS INCLUDING ALL VARIABLES

The REG Procedure
Model: MODEL1
Dependent Variable: distance DISTANCE

Number of Observations Read	851
Number of Observations Used	851

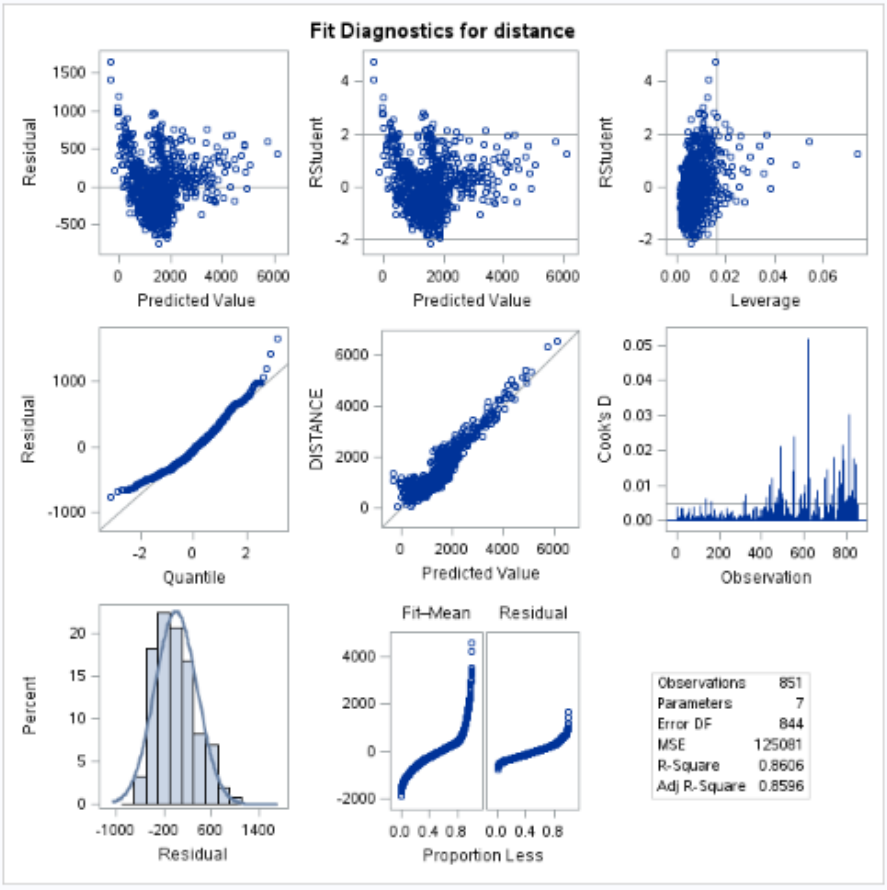
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	694474705	99210672	1333.77	<.0001
Error	843	62705222	74383		
Corrected Total	850	757179927			

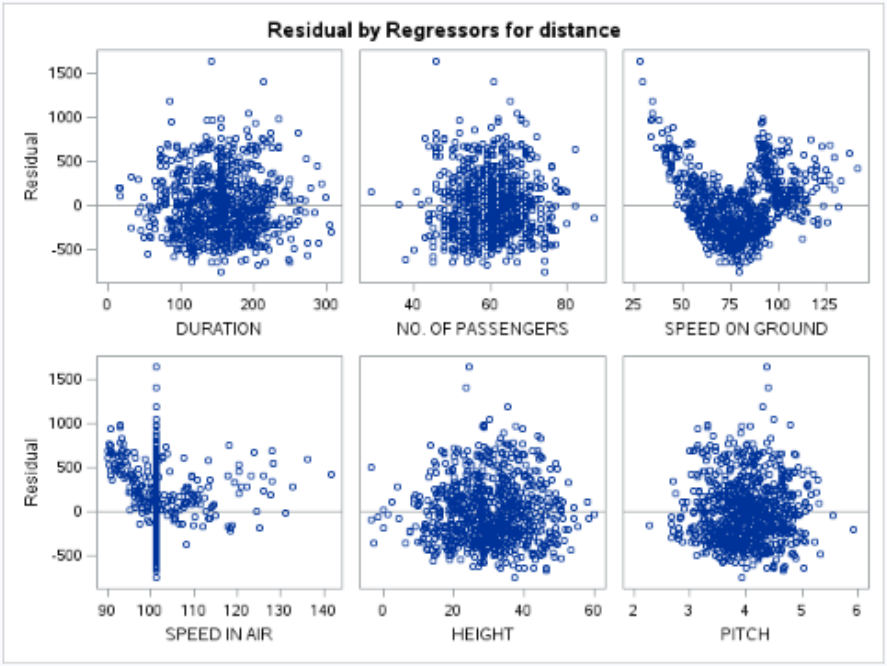
Root MSE	272.73324	R-Square	0.9172
Dependent Mean	1543.26272	Adj R-Sq	0.9165
Coeff Var	17.67251		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	-6942.74772	217.26215	-31.96	<.0001	0
CODING	AIRCRAFT	1	-484.23379	20.17205	-24.01	<.0001	1.16376
duration	DURATION	1	-0.14642	0.19675	-0.74	0.4570	1.00900
no_pasg	NO. OF PASSENGERS	1	-2.09658	1.23159	-1.70	0.0891	1.00151
speed_ground	SPEED ON GROUND	1	36.59834	0.54008	67.76	<.0001	1.23672
speed_air	SPEED IN AIR	1	52.98139	1.94279	27.27	<.0001	1.23074
height	HEIGHT	1	14.67474	0.90584	16.20	<.0001	1.00164
pitch	PITCH	1	32.98221	19.22629	1.72	0.0866	1.16142

REGRESSION ANALYSIS INCLUDING ALL VARIABLES

The REG Procedure
Model: MODEL1
Dependent Variable: distance DISTANCE





REGRESSION ANALYSIS FOR REMAINING VARIABLES TO TEST MULTICOLLINEARITY

The REG Procedure
Model: MODEL1
Dependent Variable: distance DISTANCE

Number of Observations Read	851
Number of Observations Used	851

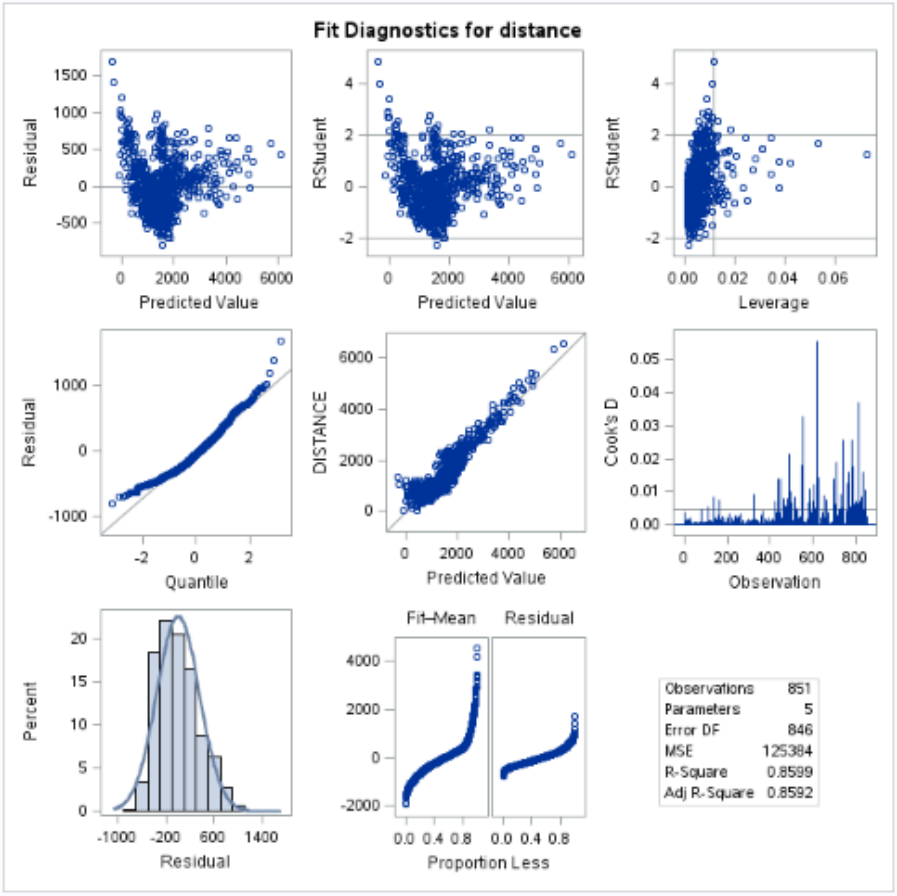
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	694000049	173500012	2323.22	<.0001
Error	846	63179878	74681		
Corrected Total	850	757179927			

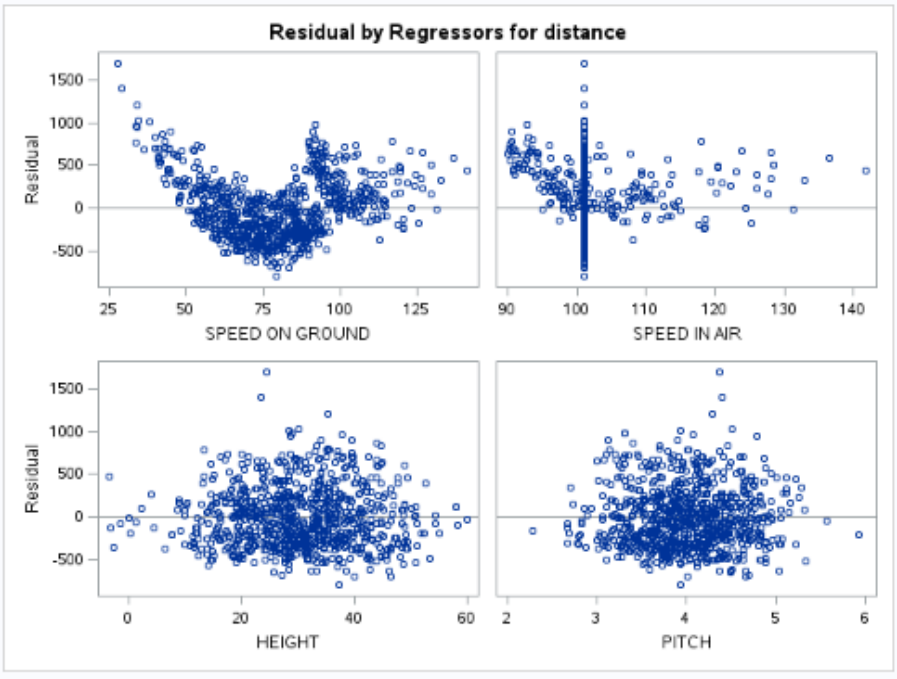
Root MSE	273.27771	R-Square	0.9166
Dependent Mean	1543.26272	Adj R-Sq	0.9162
Coeff Var	17.70779		

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	-6954.39003	186.96604	-37.20	<.0001	0
CODING	AIRCRAFT	1	-498.24000	18.77074	-26.54	<.0001	1.00368
speed_ground	SPEED ON GROUND	1	36.60053	0.53954	67.84	<.0001	1.22932
speed_air	SPEED IN AIR	1	52.99147	1.94463	27.25	<.0001	1.22816
height	HEIGHT	1	14.71379	0.90728	16.22	<.0001	1.00083

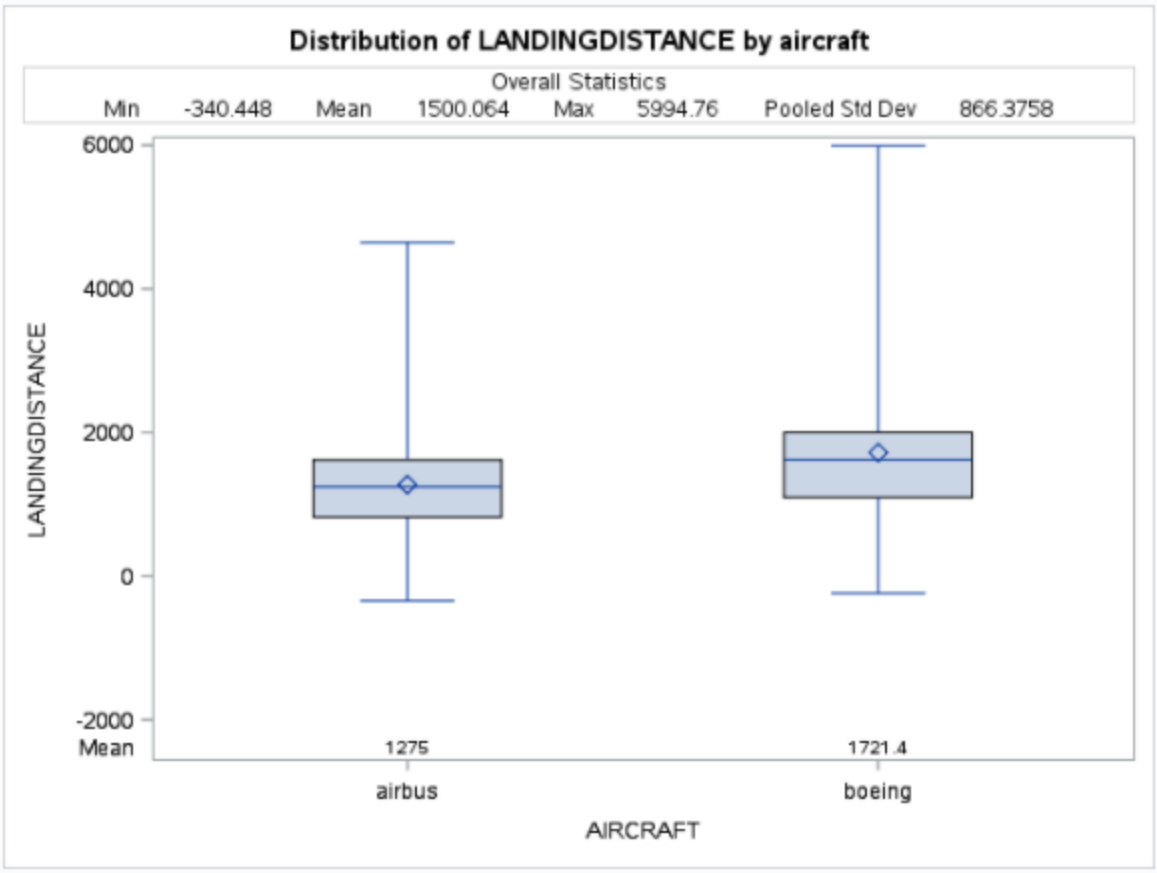
REGRESSION ANALYSIS FOR REMAINING VARIABLES TO TEST MULTICOLLINEARITY

The REG Procedure
Model: MODEL1
Dependent Variable: distance DISTANCE





Box Plot



Box Plot

The MEANS Procedure

AIRCRAFT=1

Analysis Variable : LANDINGDISTANCE				
N	Mean	Median	Std Dev	Range
422	1275.04	1243.14	735.4092110	4988.51

AIRCRAFT=0

Analysis Variable : LANDINGDISTANCE				
N	Mean	Median	Std Dev	Range
429	1721.42	1625.03	978.2415235	6232.41