

BANA 5143/6043 PROJECT

Name : Prathiba Swamykannu

UCID – M13470183

EXECUTIVE SUMMARY

This project is about finding the factors that influence the landing distance of two aircrafts. The analysis of data is done in several steps. Initially, the data present is examined and explored. Missing values are identified and duplicate values are removed. The data is then combined and then cleaned and made into a dataset. Next, X-Y plots are drawn and correlation analysis is performed to find which variables are linearly correlated with distance. Finally, regression analysis filters out the final parameters that affect the landing distance of the aircrafts. Speed of the wind and Height of the aircraft are identified as the factors that impact Distance.

CHAPTER :1

DATA UNDERSTANDING AND EXPLORATION

Data understanding and data exploration is the initial step for analysis of any dataset. Here, the multiple data sources are combined together and made into a single dataset. Then the missing values are identified. Furthermore, the duplicate values are removed, and the data is cleaned using the constraints specified. The abnormalities present in the values are identified and then deleted. Finally, a statistical summary of the cleaned data presents the mean, median value. Without performing this step, we would be unable to proceed and provide a dependable data analysis and might end up giving incorrect conclusions. Therefore, this is the fundamental step in analyzing the data.

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.

Variable dictionary:

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

Duration(in minutes): Flight duration between taking off and landing. The duration of a normal flight should always be greater than 40min.

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.If its value is less than 30MPH or greater than 140MPH, then the landing would be considered as abnormal

Speed_air (in miles per hour):The air speed of an aircraftwhen passing over the threshold of the runway. If its value is less than 30MPH or greater than 140MPH, then the landing would be considered as abnormal.

Height(in meters): The height of an aircraft when it is passing over the threshold of the runway.The landing aircraft is required to be at least 6 meters high at the threshold of the runway.

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

Distance(in feet):The landing distance of an aircraft. More specifically, it refers to the distance between the threshold of the runway and the point where the aircraft can be fully stopped. The length of the airport runway is typically less than 6000 feet

1. Combining data sets from different sources

In this step, the two different data sources are first imported to the SAS program and then combined by using SET.

```
FILENAME REFFILE '/folders/myfolders/FAA1.xls';
PROC IMPORT DATAFILE=REFFILE
  DBMS=XLS
  OUT=WORK.IMPORT;
  GETNAMES=YES;
  SHEET="Faa1";
RUN;
PROC CONTENTS DATA=WORK.IMPORT;
RUN;
FILENAME REFFILE '/folders/myfolders/FAA2.xls';
PROC IMPORT DATAFILE=REFFILE
  DBMS=XLS
  OUT=WORK.IMPORT1;
  GETNAMES=YES;
  SHEET="Faa2";
RUN;
Data COMBINED;
SET Import Import1;
RUN;
proc print data=COMBINED;
run;
```

The screenshot shows the SAS Studio interface with multiple tabs open at the top: 'SAS Information Center', 'SAS Studio', 'Combining SAS Data Sets Men...', 'Cody's Data Cleaning Techniques...', 'How to Deal With Missing Values...', and several others. The main window is titled 'SAS® Studio' and has tabs for 'CODE', 'LOG', and 'RESULTS'. The 'RESULTS' tab is selected, displaying the output of a PROC CONTENTS run. The output includes two tables: one for 'Data Set Information' and one for 'Engine/Host Dependent Information'. The 'Data Set Information' table shows details for the 'WORK.IMPORT' data set, such as 800 observations, 8 variables, and no indexes. The 'Engine/Host Dependent Information' table provides details about the file, including its location and permissions. At the bottom of the results pane, there is a link to 'Alphabetic List of Variables and Attributes'. The Windows taskbar at the bottom shows various application icons and the system tray.

Data Set Name	WORK.IMPORT	Observations	800
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	11/04/2019 18:20:45	Observation Length	72
Last Modified	11/04/2019 18:20:45	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	908
Obs in First Data Page	800
Number of Data Set Repairs	0
Filename	\tmp\\$\\$SAS_workD29C00006293_\\$localhost.localdomain\SAS_work5F9600006293_\\$localhost.localdomain\import.sas7bdat
Release Created	9.401M6
Host Created	Linux
Inode Number	671575
Access Permission	rwxr--r--
Owner Name	sasdemo
File Size	128kB
File Size (bytes)	131072

SAS Information Center SAS Studio Combining SAS Data Sets Method Cody's Data Cleaning Techniques How to Deal With Missing Values

localhost:10080/SASStudio/38/main?locale=en_US&zone=GMT-05%253A00

SAS® Studio

CODE LOG RESULTS

Table of Contents

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
1	aircraft	Char	12	\$12.	\$12.	aircraft
8	distance	Num	8	BEST12		distance
2	duration	Num	8	BEST12		duration
6	height	Num	8	BEST12		height
3	no_pass	Num	8	BEST12		no_pass
7	pitch	Num	8	BEST12		pitch
5	speed_air	Num	8	BEST12		speed_air
4	speed_ground	Num	8	BEST12		speed_ground

Data Set Name	WORK.IMPORT1	Observations	200
Member type	DATA	Variables	7
Engine	V9	Indexes	0
Created	11/04/2019 18:20:45	Observation Length	64
Last Modified	11/04/2019 18:20:45	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64 LINUX_X86_64 ALPHA_TRU64 LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1

Messages: 38 User: sasdemo 11:28 PM 11/4/2019

SAS Information Center SAS Studio Combining SAS Data Sets Method Cody's Data Cleaning Techniques How to Deal With Missing Values

localhost:10080/SASStudio/38/main?locale=en_US&zone=GMT-05%253A00

SAS® Studio

CODE LOG RESULTS

Table of Contents

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1024
Obs in First Data Page	200
Number of Data Set Repairs	0
Filename	/tmp/SAS_workD29C00006293_localhost.localdomain/SAS_work5F9600006293_localhost.localdomain/import1.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	671576
Access Permission	rw-rw-r-
Owner Name	sasdemo
File Size	128KB
File Size (bytes)	131072

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
1	aircraft	Char	12	\$12.	\$12.	aircraft
7	distance	Num	8	BEST12		distance
5	height	Num	8	BEST12		height
2	no_pass	Num	8	BEST12		no_pass
6	pitch	Num	8	BEST12		pitch
4	speed_air	Num	8	BEST12		speed_air
3	speed_ground	Num	8	BEST12		speed_ground

Messages: 38 User: sasdemo 11:28 PM 11/4/2019

Obs	aircraft	duration	no_pasp	speed_ground	speed_air	height	pitch	distance
1	boeing	98.4790912	53	107.91568005	109.32837648	27.418924252	4.0435145715	3369.8363638
2	boeing	125.7329732	69	101.65558863	102.8514051	27.804716181	4.1174315991	2987.8059235
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	41.214962599	4.203853398	1627.0681991	
7	boeing	73.023794916	54	54.4298029	24.09532163	3.8376457298	805.30399317	
8	boeing	52.903187872	57	57.101661737	19.388837508	4.6436717769	573.62178606	
9	boeing	155.51861605	61	85.445624251	35.375389749	4.2287278648	1698.9927548	
10	boeing	176.86203205	56	61.796710514	36.748816124	4.1843990127	1137.7457579	
11	boeing	158.4616994	61	53.778126741	46.355832902	5.5565991716	1075.3717411	
12	boeing	180.61655753	54	141.21863535	141.72493569	23.57935009	5.216802251	6533.0476506
13	boeing	72.286633216	54	93.391762435	92.8669561214	32.223489271	3.8182761471	2128.708285
14	boeing	187.59954737	58	94.036412942	96.196460585	33.661226156	4.6361847249	2304.857574
15	boeing	154.36870049	63	63.540513533	26.402991875	3.8565854965	1089.0729531	
16	boeing	165.54194536	69	48.774673273	31.228664837	3.9020460338	943.06840443	
17	boeing	153.54633587	61	83.556493271	29.897473262	3.519783729	1793.5628232	
18	boeing	107.11331938	78	86.807962025	25.477015381	4.414218798	1910.8768699	
19	boeing	233.80249791	69	104.80843448	103.86845794	43.882731896	3.2450978263	3213.985265
20	boeing	163.90650312	55	119.3804635	120.44470797	38.558536007	3.7014493887	4524.2788621
21	boeing	97.477625266	63	73.53976336	29.152465311	4.0140064267	1332.0387485	
22	boeing	118.63054039	55	79.994815042	29.36686101	4.4071812572	1515.9652753	
23	boeing	126.54028789	70	94.781230282	91.142068839	39.476298784	3.5940361476	2182.2207374
24	boeing	179.91591838	66	63.671165314	19.574699606	4.2867337712	873.4408921	

2. To identify the missing values and remove duplicates :

CODE:

```
data missing;
set combined;
options missing="";
if missing(cats(of _all_))then delete;
run;
proc print data=missing;
run;
```

RESULT :

SAS® Studio

Server Files and Folders

- Folder Shortcuts
- My Folders
 - sasuser.v94
 - AAUP.txt
 - Assignment 1.sas
 - Assignment 1.sas~
 - Chapter1.sas**
 - Chapter1.sas~
 - FAA1.xls
 - FAA2.xls
 - Lecture 2.sas
 - Lecture 2.sas~
 - Lecture 3(2).sas
 - Lecture 3(2).sas~
 - Lecture 3(3).sas
 - Lecture 3(3).sas~
 - Lecture 3.sas
 - Lecture 3.sas~
- Tasks and Utilities
- Snippets
- Libraries
- File Shortcuts

RESULTS

Table of Contents

Obs	aircraft	duration	no_pasp	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	101.65585863	102.8514051	27.804716181	4.1174316991	2987.8039235
3	boeing	112.0170008	61	71.051960883	18.599085734	4.4340431266	1144.022426	
4	boeing	196.82569105	56	85.813327679	30.744597235	3.8842361245	1664.2181584	
5	boeing	90.095361357	50	59.888528183	32.397688062	4.026094152	1050.2644976	
6	boeing	137.95981722	55	75.014334744	41.21496259	4.203653398	1627.0681991	
7	boeing	73.023794916	54	54.4298029	24.03532163	3.8376457299	805.30399317	
8	boeing	52.903167872	57	57.101651737	19.388637508	4.6436717769	753.62178606	
9	boeing	155.18161605	61	85.443642451	35.757589749	4.2287278648	1698.9927548	
10	boeing	176.66203205	56	61.796715014	36.748816124	4.1843990127	1137.7457579	
11	boeing	158.4619964	61	53.778126741	46.355832902	5.5563991716	1075.3717441	
12	boeing	180.61655753	54	141.21863535	141.72493569	23.7576935009	21568022511	6533.0476506
13	boeing	72.289633216	54	93.391762435	92.869561214	32.223489271	3.8182761471	2128.708285
14	boeing	187.59954737	58	94.036412942	96.196460585	33.61262156	4.6361847249	2304.857574
15	boeing	154.36870049	63	63.540613553	26.402991875	3.8566384986	1089.9729531	
16	boeing	165.54194936	69	48.774637373	31.228664837	3.9020460339	943.06404443	
17	boeing	153.54633587	63	81.556493971	29.897473262	3.519783726	1793.5628232	
18	boeing	107.1131938	78	86.807962025	25.477015381	4.4142187966	1910.876699	
19	boeing	233.80249791	69	104.808463448	103.86845794	43.882731896	3.2450978263	3213.985265
20	boeing	163.90650312	55	119.3804635	120.44470797	38.555836007	3.7014493887	4524.2788621
21	boeing	97.477623266	63	73.533976536	29.152465111	4.014064257	1332.0987485	
22	boeing	118.53054039	55	79.994815042	29.56666101	4.4071812572	1515.9652755	
23	boeing	126.54028769	70	94.781230282	91.142068839	39.476298784	3.5949361476	2182.2207374
24	boeing	179.91591838	66	63.671165514	19.574699606	4.2867337712	873.4408921	

Code:

```
proc sort data=missing Out=missing_noduplicate Nodupkey;
```

After the missing values are removed, we end up with 950 records from the initial 1000 values.

CODE :

```
proc sort data=missing Out=missing_noduplicate Nodupkey;
```

```

by aircraft no_pasg speed_ground speed_air height pitch distance;
run;
proc print data=missing_noduplicate;
run;

```

The screenshot shows the SAS Studio interface with multiple tabs open. The 'RESULTS' tab is selected, displaying a table titled 'Table of Contents'. The table has columns for ID, Boeing, I.D., L.C., DURATION, F.L., AVG.DUR, AVG.DUR, AVG.DUR, and AVG.DUR. The data consists of 850 rows of flight information. The bottom right corner of the interface shows the message 'Messages: 80 User: sasdemo'.

ID	Boeing	I.D.	L.C.	DURATION	F.L.	AVG.DUR	AVG.DUR	AVG.DUR	AVG.DUR
828	boeing	124.37864547	72	60.367043725		3.7889195211	3.706088819	641.5995622	
829	boeing	158.16277689	72	74.8334773963		26.186671031	3.9863265825	1448.0794019	
830	boeing	78.24709907	72	76.999631013		13.001193211	4.2019377631	1273.6644967	
831	boeing	133.5569265	72	94.271392942	92.877735531	33.046424149	3.785704046	2351.0709569	
832	boeing	161.8924678	72	129.26491833	128.41773098	33.948998825	4.1399014138	5381.9588622	
833	boeing	236.19293349	73	52.360449116		44.21061799	4.497087384	1078.0688933	
834	boeing	133.45985625	73	57.045299494		1.2538552556	4.7153842391	371.27726086	
835	boeing	213.7791717	73	57.426804092		48.668703509	4.1846517788	1154.4436143	
836	boeing	112.31707003	74	79.257984189		37.19716518	4.3370030837	1158.8376732	
837	boeing	168.23013919	74	86.852747395		16.894461754	3.8308960194	1725.3804918	
838	boeing	124.54353753	75	69.880248247		31.31135669	4.6879165411	1048.0302857	
839	boeing	118.26426229	75	70.168245294		17.743340834	4.2669763429	830.71474719	
840	boeing	79.705683144	75	106.7461226	106.73317595	18.346201583	4.8074017332	2765.855295	
841	boeing	130.94961924	76	44.732763125		32.82904952	4.861881592	874.79864397	
842	boeing	147.03191592	76	63.597942325		36.489042355	4.4917734289	1051.9369604	
843	boeing	219.72115995	76	88.103462433		42.985495821	4.6540097977	1927.0536775	
844	boeing	139.68891519	77	55.086685785		38.932817792	4.0971206341	998.09706833	
845	boeing	172.56012205	77	82.29713755		44.758716354	4.2293090445	1809.27205	
846	boeing	228.17710591	78	61.220375598		21.772286622	4.5955283685	970.04651856	
847	boeing	107.11331938	78	86.807962025		25.477015381	4.4142187986	1910.8768699	
848	boeing	128.93810992	79	106.93389135	108.42651323	30.457709156	4.8421492	3203.3188407	
849	boeing	161.82569155	80	82.509055403		36.680194026	4.685510032	1590.3719225	
850	boeing	194.4671661	82	40.815188666		22.618444074	4.8765952309	761.4850777	

OBSERVATION :

The result displays 850 records instead of the initial 950. From this we can infer that there were 100 duplicate values and they have been removed

3. Next, the Validity check is performed to identify the abnormal values.

The values that doesn't meet the pre-planned conditions are considered as abnormal and other values are labeled normal.

- The duration of the flight between and takeoff and landing should always be greater than 40. Else it is considered as abnormal.
- If the ground speed is less than 30 mph or greater than 140 mph, then the landing would be considered as abnormal.
- If the air speed is less than 30mph or greater than 140 mph, then the landing would be considered as abnormal.
- The aircraft should be at least 6 meters high at the threshold of the runway; else it is considered as abnormal.

e. The landing distance should not exceed 6000 ft which is the length of the airport runway. Else it is considered as abnormal.

CODE:

```
DATA abnormal;
SET missing_noduplicate;
IF duration <40 THEN DURATION_ABNORMAL='YES';
ELSE DURATION_ABNORMAL='NO'
;
if speed_air < 30
or speed_air > 140
then AIRSPEED_abnormal = 'YES';
else AIRSPEED_abnormal = 'NO'
;
if height < 6
then HEIGHT_ABNORMAL = 'YES';
else HEIGHT_ABNORMAL = 'NO'
;
if distance >6000
then DISTANCE_ABNORMAL ='YES';
else DISTANCE_ABNORMAL = 'NO'
;
RUN;
proc print data=abnormal;
run;
```

The screenshot shows the SAS Studio interface with several tabs open in the background:

- SAS Information Center
- SAS Studio
- Combining SAS Data Sets: M...
- Cody's Data Cleaning Techniques
- How to Deal With Missing Values

The main window has the following tabs:

- CODE
- LOG
- RESULTS**
- OUTPUT DATA

The RESULTS tab displays a table of data with the following columns:

Obs	aircraft	duration	no_psg	speed_ground	speed_air	height	pitch	distance	DURATION_ABNORMAL	AIRSPEED_abnormal	HEIGHT_ABNORMAL	DISTANCE_ABNORMAL
1	boeing	98 4790912	53	107 91568005	109 32837648	27 18924252	4.0432145715	3369 8363838	NO	NO	NO	NO
2	boeing	125 73329732	69	101 65558063	102 8514051	27 894716181	4.1174316991	2987 8039235	NO	NO	NO	NO
3	boeing	112 0170008	61	71 851960883	18 589385734	4.4340431286	1144 922426	NO	YES	NO	NO	NO
4	boeing	196 82659195	56	85 81327679	30 744597235	3.8842361245	1654 2181584	NO	YES	NO	NO	NO
5	boeing	90 695381357	70	59 888628183	32 397688062	4.026094152	1050 2644876	NO	YES	NO	NO	NO
6	boeing	137 59581722	55	75 791434744	41 21496259	4.203853396	1627 0681991	NO	YES	NO	NO	NO
7	boeing	73 023794916	54	51 4298029	24 03532163	3.8376457299	865 3039931	NO	YES	NO	NO	NO
8	boeing	52 903187872	57	57 101661737	19 388837508	4.6436717768	573 62178606	NO	YES	NO	NO	NO
9	boeing	155 51861695	61	85 443624251	35 375389749	4.2287278648	1686 9927548	NO	YES	NO	NO	NO
10	boeing	176 86203205	56	61 739710514	36 748816124	4.184399127	1337 7457579	NO	YES	NO	NO	NO
11	boeing	158 4619884	61	53 778126741	46 355822902	5.5563991716	1075 3711411	NO	YES	NO	NO	NO
12	boeing	180 61665753	54	141 21863535	141 72493569	23 57593606	5.2160022511	6533 0476506	NO	YES	NO	YES
13	boeing	72 289633216	54	93 391762435	92 869561214	32 223489271	3.8182781471	2128 708285	NO	NO	NO	NO
14	boeing	107 595947737	58	94 936412942	96 19646058	4.6361847249	2304 857574	NO	NO	NO	NO	NO
15	boeing	154 36870049	63	63 540613553	26 40295785	3.8565854986	1089 9729531	NO	YES	NO	NO	NO
16	boeing	165 54194536	69	48 774673273	31 228664837	3.9020460339	943 06840443	NO	YES	NO	NO	NO
17	boeing	153 54633587	61	83 556493271	29 897473262	3.519783726	1793 5628232	NO	YES	NO	NO	NO
18	boeing	107 1131938	78	86 80762025	25 477015381	4.4142187986	1910 8768899	NO	YES	NO	NO	NO
19	boeing	233 80249791	69	104 80843448	103 86845794	43 88731996	3.245097263	3213 985265	NO	NO	NO	NO
20	boeing	163 90650312	55	119 3804635	120 44470797	38 558536007	3.7014493887	4524 2788821	NO	NO	NO	NO
21	boeing	97 477623266	63	73 533976336	29 15246511	4.014006425	1332 0387485	NO	YES	NO	NO	NO
22	boeing	118 63064639	55	79 994615042	29 366866101	4.4071812572	1516 9652753	NO	YES	NO	NO	NO
23	boeing	126 54028799	70	94 781230262	91 14206883	39 47629874	3.5945361476	2128 2207374	NO	NO	NO	NO
24	boeing	179 91591838	66	63 671165314	19 57469968	4.286737717	873 4408921	NO	YES	NO	NO	NO
25	boeing	112 90009528	53	98 180410862	99 135830727	28 152991316	3.9874712191	2586 6650864	NO	NO	NO	NO
26	boeing	56 644648966	66	72 95368239	36 154157217	4.3876559157	1265 1280251	NO	YES	NO	NO	NO
27	boeing	86 828911312	62	91 74535792	92 874851912	28 773729478	3.305888073	2313 3356963	NO	NO	NO	NO

At the bottom of the screen, there is a taskbar with various icons and a status bar indicating 'Messages: 41' and 'User: sasdemo'.

The screenshot shows the SAS Studio interface with a results table titled "Table of Contents". The table has 85 rows and several columns. The columns include:

ID	Boeing	Tail Number	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	
828	boeing	124.37864547	72	60.367043725	3.7889195211	3.7060888319	641.59956822	NO	YES	YES	NO	
829	boeing	158.16277669	72	74.833473963	26.166971031	3.9863265625	1449.0794019	NO	YES	NO	NO	
830	boeing	78.24709997	72	76.999831013	13.801193211	4.2019377631	1273.6644967	NO	YES	NO	NO	
831	boeing	133.5569265	72	94.271329242	92.877735531	33.046424149	3.785794946	2351.0709569	NO	NO	NO	NO
832	boeing	161.8924678	72	129.26401833	128.417730998	33.948998825	4.1399514138	5381.9588622	NO	NO	NO	NO
833	boeing	236.19293349	73	52.360449116	44.121081799	4.4970887384	1078.0988933	NO	YES	NO	NO	
834	boeing	133.45965625	73	57.045299494	1.25365252056	4.7153642391	371.27726096	NO	YES	YES	NO	
835	boeing	213.7917717	73	57.426894092	48.868703509	4.1846317788	154.4436143	NO	YES	NO	NO	
836	boeing	112.31707003	74	79.257984189	37.19716518	4.3370030387	1158.8376732	NO	YES	NO	NO	
837	boeing	168.23013919	74	86.852747395	16.894461754	3.8308960194	1725.3804918	NO	YES	NO	NO	
838	boeing	124.54353753	75	69.880248247	31.31135869	4.6879165411	1045.0302857	NO	YES	NO	NO	
839	boeing	118.264266229	75	70.168245934	17.743340834	4.2669763429	830.71474719	NO	YES	NO	NO	
840	boeing	79.705863144	75	106.7461226	106.73317595	18.346201583	4.8074017332	278.855295	NO	NO	NO	
841	boeing	130.94961924	76	44.732763125	32.762994552	4.861881592	874.79864397	NO	YES	NO	NO	
842	boeing	147.03191592	76	63.597942325	36.469042355	4.4917734289	1051.9369604	NO	YES	NO	NO	
843	boeing	219.72115595	76	88.103462433	42.085495821	4.6540097977	1927.0536775	NO	YES	NO	NO	
844	boeing	130.16891519	77	55.086685785	38.032817792	4.0971206341	988.09700633	NO	YES	NO	NO	
845	boeing	172.56012205	77	82.29713755	44.758716354	4.229390445	1809.27205	NO	YES	NO	NO	
846	boeing	228.77710591	78	61.220375598	21.772286622	4.5955283685	970.04651856	NO	YES	NO	NO	
847	boeing	107.11319198	78	86.807962025	25.477015381	4.4142187996	1910.8766999	NO	YES	NO	NO	
848	boeing	128.93810992	79	106.93389135	108.42651323	30.457709156	4.8421492	3203.3188407	NO	NO	NO	NO
849	boeing	161.82569155	80	82.590955403	36.680194026	4.685310032	1590.3719225	NO	YES	NO	NO	
850	boeing	194.4671661	82	40.815188666	22.61844074	4.8765952309	761.4850777	NO	YES	NO	NO	

4. The abnormal values are deleted and the data is cleaned.

CODE:

```
data cleaning;
set missing_noduplicate;
if no_pasg =. or no_pasg<0 then delete;
  if speed_ground =. or speed_ground < 30 or speed_ground >140 then delete;
  if height =. or height<6 then delete;
  if pitch =. then delete;
  if distance =. or distance> 6000 then delete;
proc print data = cleaning;
run;
proc means data = cleaning n nmiss min max;
run;
```

SAS® Studio

Server Files and Folders

- Folder Shortcuts
- My Folders
 - sasuser.v94
 - AAUP.txt
 - Assignment 1.sas
 - Assignment 1.sas~
 - Chapter1.sas
 - Chapter1.sas~
 - FAA1.xls
 - FAA2.xls
 - Lecture 2.sas
 - Lecture 2.sas~
 - Lecture 3(2).sas
 - Lecture 3(2).sas~
 - Lecture 3(3).sas
 - Lecture 3(3).sas~
 - Lecture 3.sas
 - Lecture 3.sas~
- Tasks and Utilities
- Snippets
- Libraries
- File Shortcuts

RESULTS

Table of Contents

Obs	aircraft	duration	no_pasp	speed_ground	speed_air	height	pitch	distance
1	airbus	172.04931209	36	47.486765029		13.984809941	4.2990197162	250.68976141
2	airbus	188.01797722	38	85.180842251		37.028793691	4.1216901717	1257.092519
3	airbus	93.540807771	40	80.627416679		28.60255713	3.6234201886	1021.088817
4	airbus	123.30242152	41	97.568203996	96.978436701	38.409129053	3.5322719834	2167.7576915
5	airbus	109.19713407	43	82.483044979		30.140024899	4.0896284195	1321.0000654
6	airbus	139.31381026	44	99.596841547	99.160266345	35.187030092	3.6402667146	2116.680919
7	airbus	214.22048507	45	72.490616757		33.228125197	4.3693164876	748.7667918
8	airbus	182.7116757	45	77.805502137		20.18995838	4.178015403	905.49788375
9	airbus	197.45183449	45	81.375317855		46.285569727	4.2052754575	1459.5022975
10	airbus	155.86922387	45	86.8758779978		34.838071106	3.9797603715	1262.153807
11	airbus	216.67640251	45	91.616595738		38.324193052	4.7435614527	1967.6109937
12	airbus		46	40.801786477		24.400127629	3.9682093233	620.09051196
13	airbus	149.65109927	46	69.037688441		48.559845050	4.1973990963	1127.8005331
14	airbus		46	104.07757658	103.40921036	19.7157721	4.1043931104	2494.8046454
15	airbus	168.79553003	47	92.907591765	95.762091617	23.784807183	3.9068740504	1955.3034911
16	airbus	237.40527671	48	53.774013118		28.26080221	3.1755269566	241.16096423
17	airbus		48	61.570704648		21.78570744	4.351947442	560.53392302
18	airbus	173.57275867	48	69.440266991		21.235704144	4.2349455846	860.43708392
19	airbus	221.59245844	48	75.36854228		19.395151702	4.8756423429	932.57187751
20	airbus	181.705175046	48	79.793960442		34.932683072	3.76482161	1159.5466175
21	airbus	160.84157691	48	87.551468843		28.545582924	4.4229830982	1410.1890375
22	airbus	157.16726867	48	101.83850487	103.63497127	23.01621748	4.9429730633	2524.5676617
23	airbus	199.91969175	48	109.30116236	109.55662817	33.066513954	4.0350502581	3177.1885419
24	airbus	140.67120141	48	120.45475566	118.67260401	30.351506953	4.3710717196	3891.4718916

Messages: 83 User: sasdemo

SAS® Studio

Server Files and Folders

- Folder Shortcuts
- My Folders
 - sasuser.v94
 - AAUP.txt
 - Assignment 1.sas
 - Assignment 1.sas~
 - Chapter1.sas
 - Chapter1.sas~
 - FAA1.xls
 - FAA2.xls
 - Lecture 2.sas
 - Lecture 2.sas~
 - Lecture 3(2).sas
 - Lecture 3(2).sas~
 - Lecture 3(3).sas
 - Lecture 3(3).sas~
 - Lecture 3.sas
 - Lecture 3.sas~
- Tasks and Utilities
- Snippets
- Libraries
- File Shortcuts

RESULTS

Table of Contents

Obs	boeing	duration	no_pasp	speed_ground	speed_air	height	pitch	distance
823	boeing	168.23013919	74	86.852747395		16.894461754	3.8308960194	1725.3804918
824	boeing	124.54353753	75	69.880248247		31.31155869	4.6879165411	1045.0302857
825	boeing	118.26425229	75	70.168245294		17.743340834	4.2669753429	830.11474719
826	boeing	79.705863144	75	106.7461226	106.73317595	18.346201563	4.8074017332	2785.655295
827	boeing	130.94961924	76	44.732763132		32.782994552	4.6188181592	874.79864397
828	boeing	147.03191592	76	63.507942325		36.489042355	4.4917734289	1051.9369604
829	boeing	219.72115595	76	88.103462433		42.085495821	4.6540097977	1927.056775
830	boeing	130.16891519	77	55.086685785		38.032817792	4.0971206341	998.09700633
831	boeing	172.56012205	77	82.29713755		44.756716354	4.2293090445	1809.27205
832	boeing	228.17710591	78	61.220375598		21.77228662	4.5955263685	970.04651856
833	boeing	107.11331938	78	86.807962025		25.477015381	4.4142187986	1910.8768699
834	boeing	128.93810992	79	106.93389135	108.42651323	30.457709156	4.8421492	3203.3188407
835	boeing	161.82569155	80	82.509055403		36.680194026	4.685310032	1590.3719225
836	boeing	194.4671661	82	40.815186666		22.616444074	4.8765952309	761.4850777

The MEANS Procedure

Variable	Label	N	N Miss	Minimum	Maximum
duration	duration	786	60	14.7420704	308.6217107
no_pasp	no_pasp	836	0	29.0000000	87.0000000
speed_ground	speed_ground	836	0	33.5741041	132.7846766
speed_air	speed_air	206	630	90.0025586	132.9114649
height	height	836	0	29.5000000	58.8483939
pitch	pitch	836	0	2.2844101	5.857842
distance	distance	836	0	41.7223127	5381.96

Messages: 83 User: sasdemo

After the final cleaning of data, we have 836 records from the initial uncleaned 1000 records. Following these steps will help us in providing a reasonable and trustworthy analysis of data.

5. Summary statistics

CODE :

```

PROC MEANS DATA=cleaning N NMISS MEAN STD MIN MAX RANGE;
TITLE SUMMARY STATISTICS FOR Airline Data;
RUN;
proc print data=cleaning;
run;

```

CODE LOG RESULTS

Variable	Label	N	N Miss	Mean	Std Dev	Minimum	Maximum	Range
duration	duration	786	50	153.9337944	49.3360402	14.7642071	305.6217107	290.8575036
no_psg	no_psg	836	0	60.0406699	7.4792021	29.0000000	87.0000000	58.0000000
speed_ground	speed_ground	836	0	79.5944146	18.7327127	33.5741041	132.7846766	99.2105726
speed_air	speed_air	206	630	103.4552338	0.6926499	90.0028586	132.0114649	42.0086063
height	height	836	0	30.5104883	9.8049102	6.2275178	59.9459639	53.7184462
pitch	pitch	836	0	4.0050110	0.5273975	2.2844801	5.9267842	3.6423041
distance	distance	836	0	1526.05	898.4154244	41.7223127	5381.96	5340.24

This provides the mean, Standard Deviation, minimum and maximum values of the variables present in the dataset.

CONCLUSION:

From this Chapter, we have the following conclusions:

- By doing this step, we have reduced the initial 1000 records to 836 records at the end of the section.
- We get 836 records by removing few missing and all the abnormal and duplicate values.
- Not all the missing values are removed as it will reduce the records to a very small number and data is crucial to perform the analysis.
- Abnormal values are removed by following the constraints specified for them.
- The duplicates obtained after combining the data sources are also removed for further refinement.

CHAPTER -2

DESCRIPTIVE STUDY AND ANALYSIS

In this step, X-Y plots are drawn and correlation coefficient relationships are formed to identify the variables that are correlated with the landing distance of the aircrafts. It is helpful to identify the relationships present between the variables to further filter out the values that are not needed. Correlation coefficient is a statistical measure that calculates the strength of relationship between the relative movements of two variables. Here distance is one variable that is related with all the other variables to find the variables that linearly relate with it.

1. Plot X-Y Graphs:

CODE :

```
proc plot data = cleaning;
plot distance*duration = '$';
run;

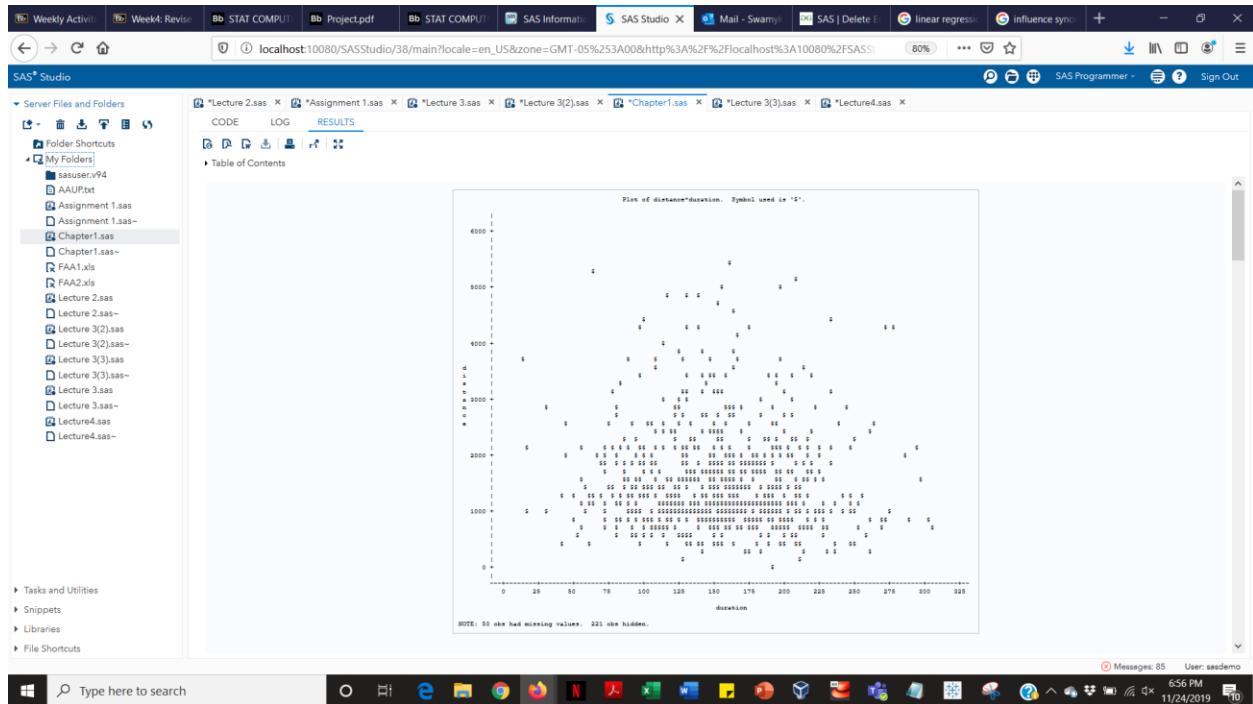
proc plot data=cleaning;
plot distance*no_pasg = '@';
run;

proc plot data=cleaning;
plot distance*speed_air = '*';
run;

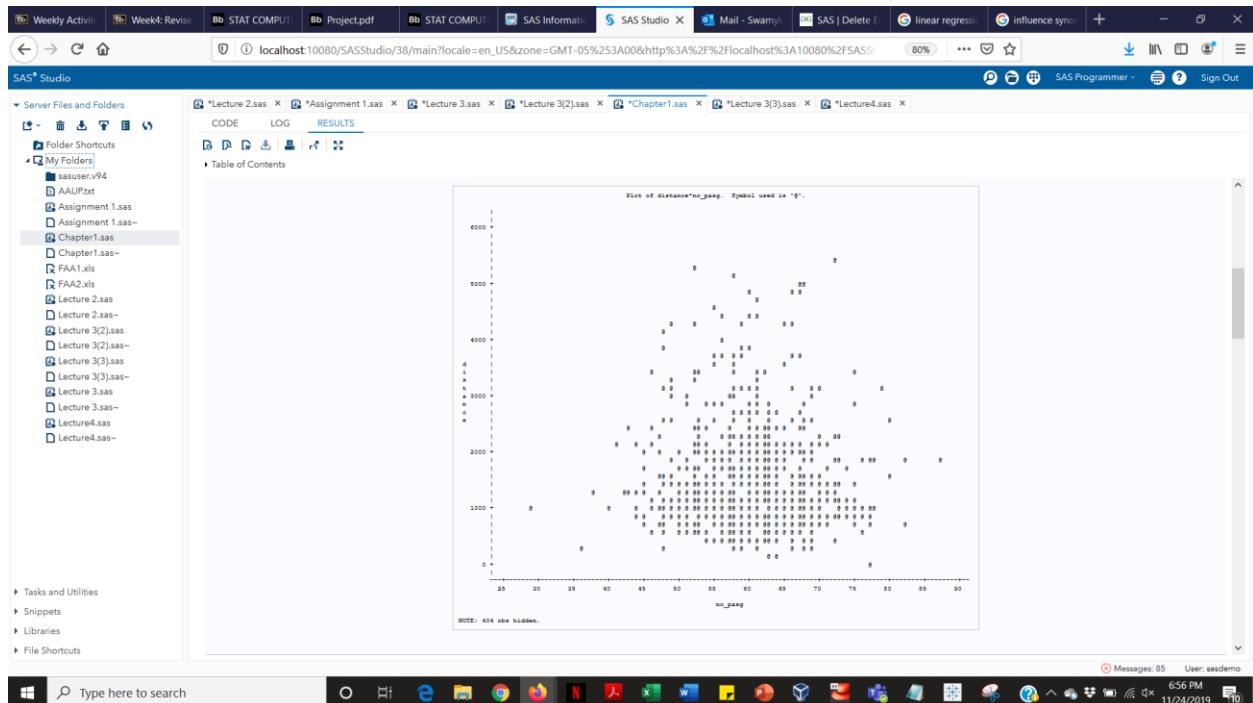
proc plot data=cleaning;
plot distance*speed_ground ;
run;

proc plot data=cleaning;
plot distance*height = '@';
run;

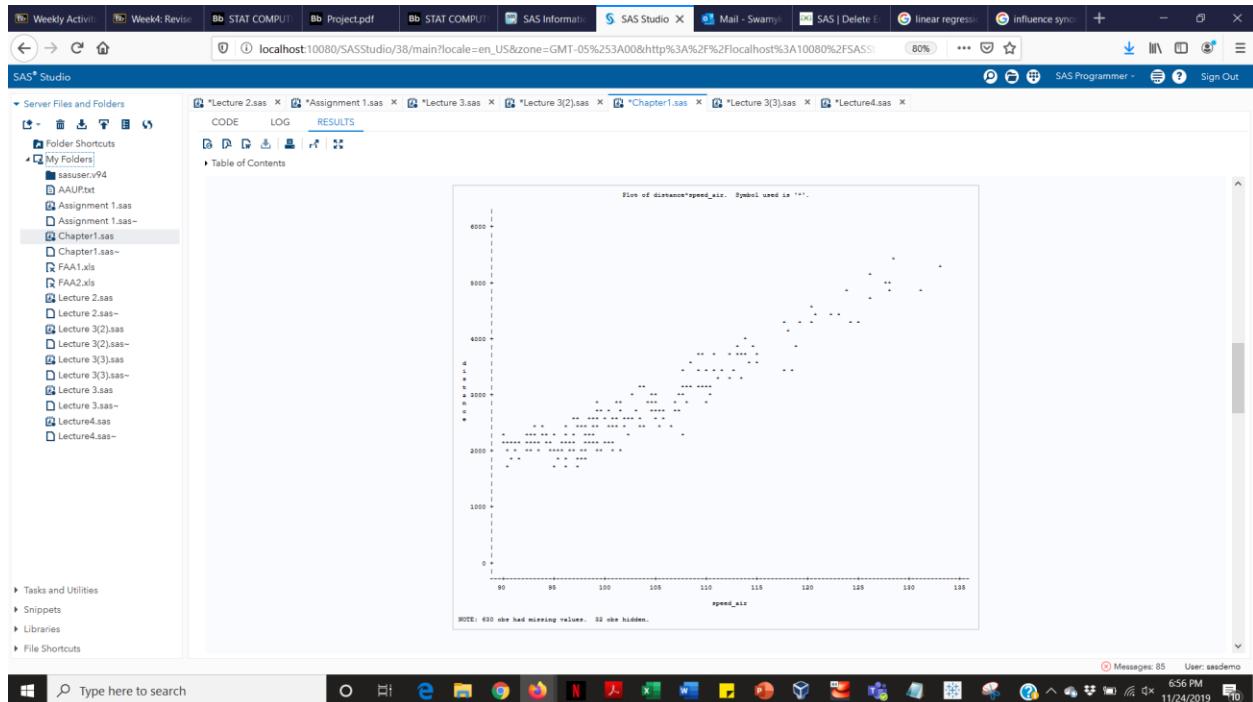
proc plot data=cleaning;
plot distance*pitch = '**';
run;
```



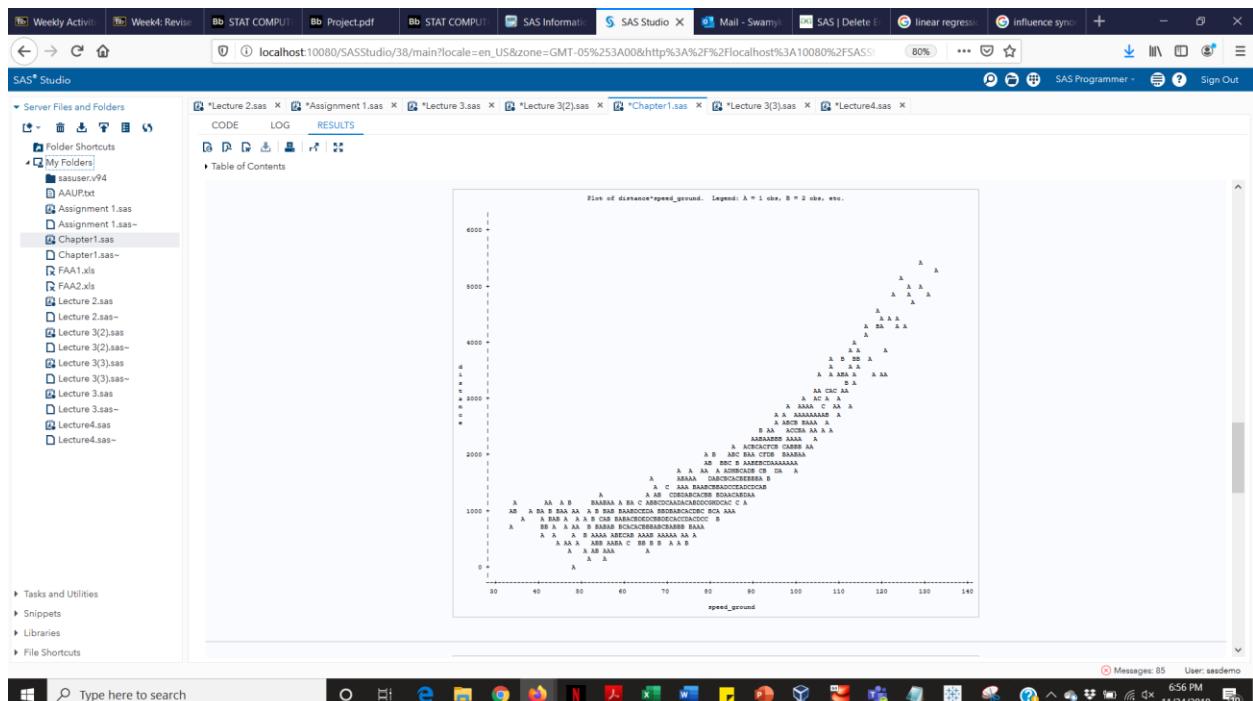
Messages: 85 User: saademo 6:56 PM 11/24/2019



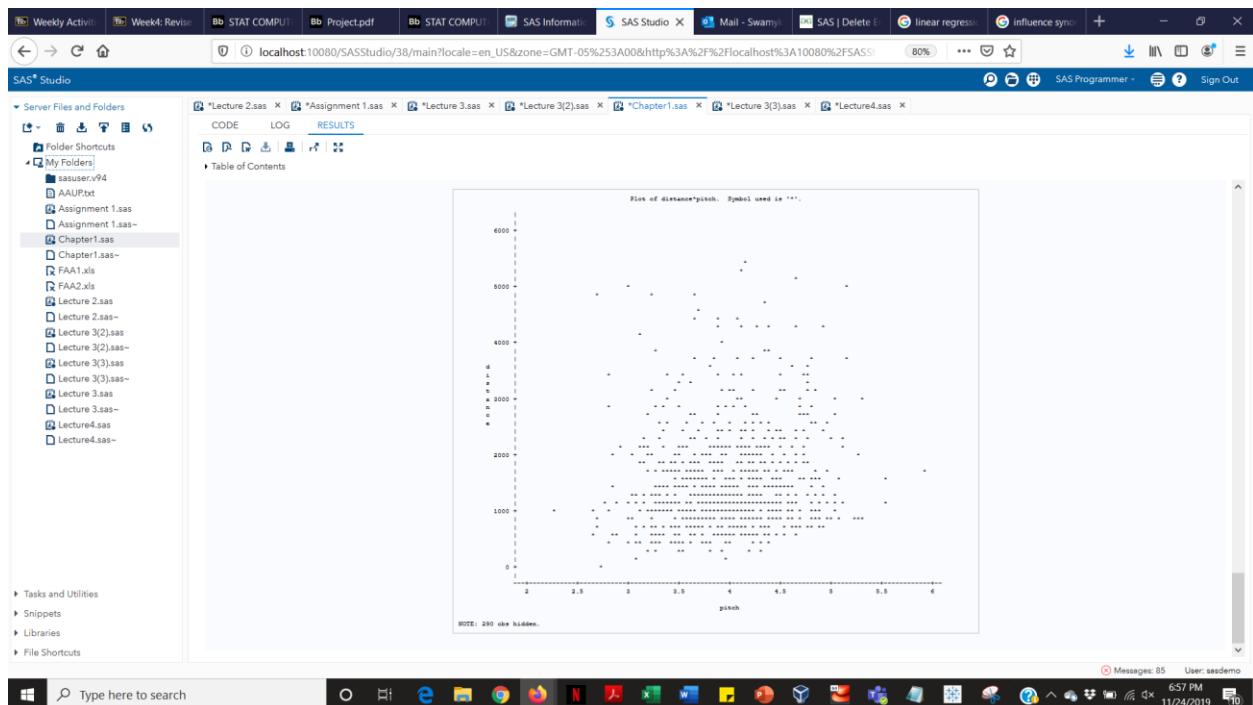
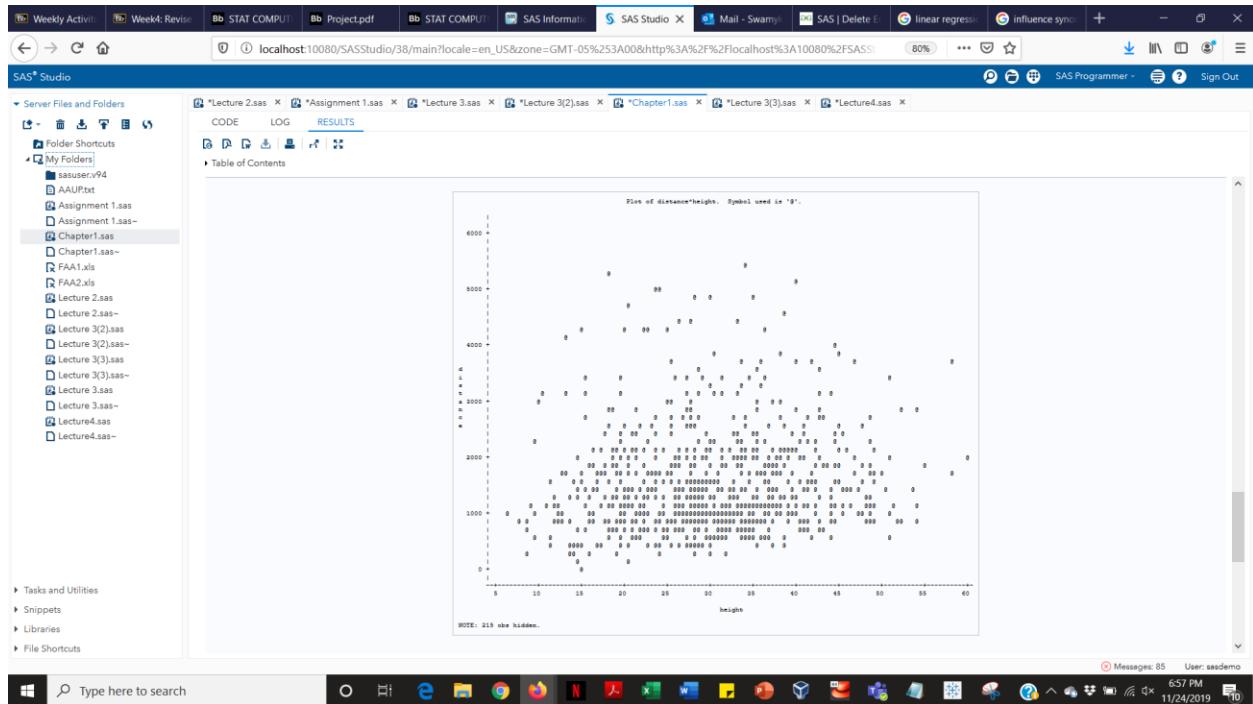
Messages: 85 User: saademo 6:56 PM 11/24/2019



Messages: 85 User: sasdemo 6:56 PM 11/24/2019



Messages: 85 User: sasdemo 6:56 PM 11/24/2019



From these X-Y graphs plotted, we can see that only speed_ground and speed_air correlate linearly with Distance compared to the others. So we can safely say that speed_ground and speed_air are two main factors the landing distance of the aircrafts.

2. Correlation Coefficient of Distance with the other 6 variables :

CODE :

```
proc corr data=cleaning;
var distance;
with duration speed_air speed_ground height pitch no_pasg;
title CORRELATION COEFFICIENTS WITH DISTANCE;
run;
```

The screenshot shows the SAS Studio interface with multiple open files in the top bar. The left sidebar displays a file tree with several SAS files. The main area shows the output of a 'CORR' procedure. The first part of the output is a 'Simple Statistics' table:

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
duration	793	155.0000	8.02065	12112	0.0000	2911	
speed_air	206	103.4623	8.82065	21112	90.0236	132.9146	speed_air
speed_ground	836	79.9941	18.73271	66541	33.07410	132.78498	speed_ground
height	836	30.91049	9.80491	25097	6.22762	59.94956	height
pitch	836	4.00501	0.52740	3348	2.28448	5.92678	pitch
no_pasg	836	60.04067	7.47920	50194	29.00000	87.00000	no_pasg
distance	836	1520	898.11542	1275781	41.72231	5382	distance

The second part of the output is a 'Pearson Correlation Coefficients' table:

		Number of Observations					
		793					
		788					
		206					
		836					
duration		-0.06107					
duration		0.00000					
speed_air		0.41339					
speed_air		-0.00000					
speed_ground		0.86661					
speed_ground		-0.00000					
height		0.10767					
height		0.00000					
pitch		0.09308					
pitch		0.00000					
no_pasg		-0.02115					
no_pasg		0.00000					
distance		0.94139					
distance		0.00000					

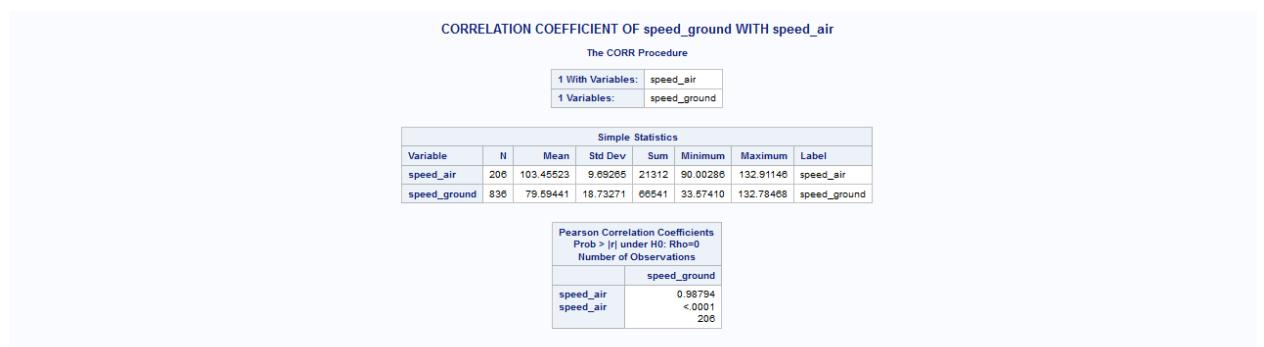
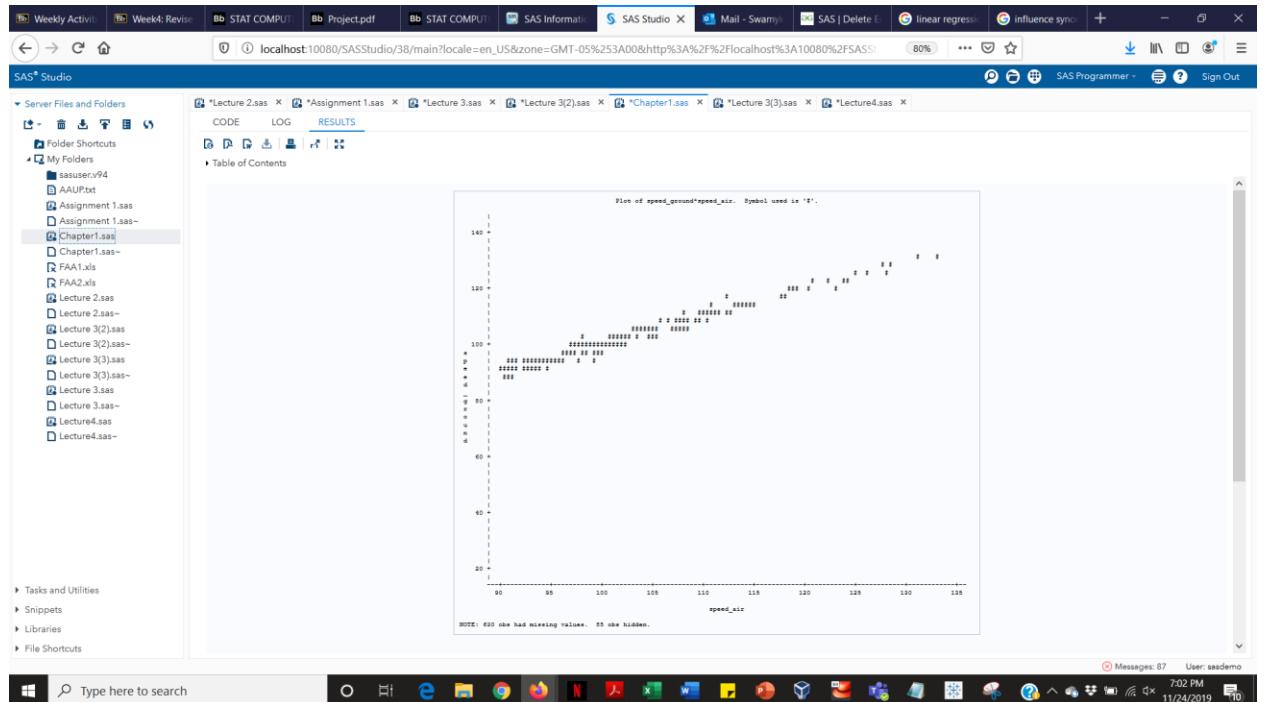
This shows how speed_air with a value of 0.86661 and speed_ground with a value of 0.94139 linearly correlate with distance. We can also infer how the other variables are not showing any dependency with distance.

3. Find the correlation coefficient between speed_ground and speed_air:

CODE :

```
proc plot data=cleaning;
plot speed_ground*speed_air = '#';
run;
```

```
proc corr data=cleaning;
var speed_ground;
with speed_air;
title CORRELATION COEFFICIENT OF speed_ground WITH speed_air;
run;
```



The X-Y plot and the correlation coefficients for these 2 variables shows how closely related they are. From this we can conclude that one out of the two variables can be taken into final consideration.

CONCLUSION:

From this step we can filter out the variables that are not dependent on distance and shows the two variables that are linearly related to distance. The further analysis between speed_ground and speed_air shows how they are interrelated and that only one of them is enough for the final observation.

CHAPTER-3

STATISTICAL MODELLING

In this step we develop a linear regression model that attempts to model a relationship between two variables. One is the explanatory variable and the other is the dependent variable. Here we predict Distance using the other variables.

1. Regression plot between distance and all the other variables :

CODE :

```
/*Regression Analysis */  
proc reg data = cleaning;  
model distance = duration no_pasg speed_ground speed_air height pitch;  
title Regression Analysis of DISTANCE using the cleaned dataset;  
run;
```

The screenshot shows the SAS Studio interface with multiple open files in the top bar. The left sidebar displays a file tree with various SAS and Excel files. The main workspace contains two tables of results from a REG procedure:

Regression Analysis of DISTANCE using the cleaned dataset

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read				
859	859	859	859	859
Number of Observations Used				
190	190	190	190	190
Number of Observations with Missing Values				
0	0	0	0	0

Analysis of Variance

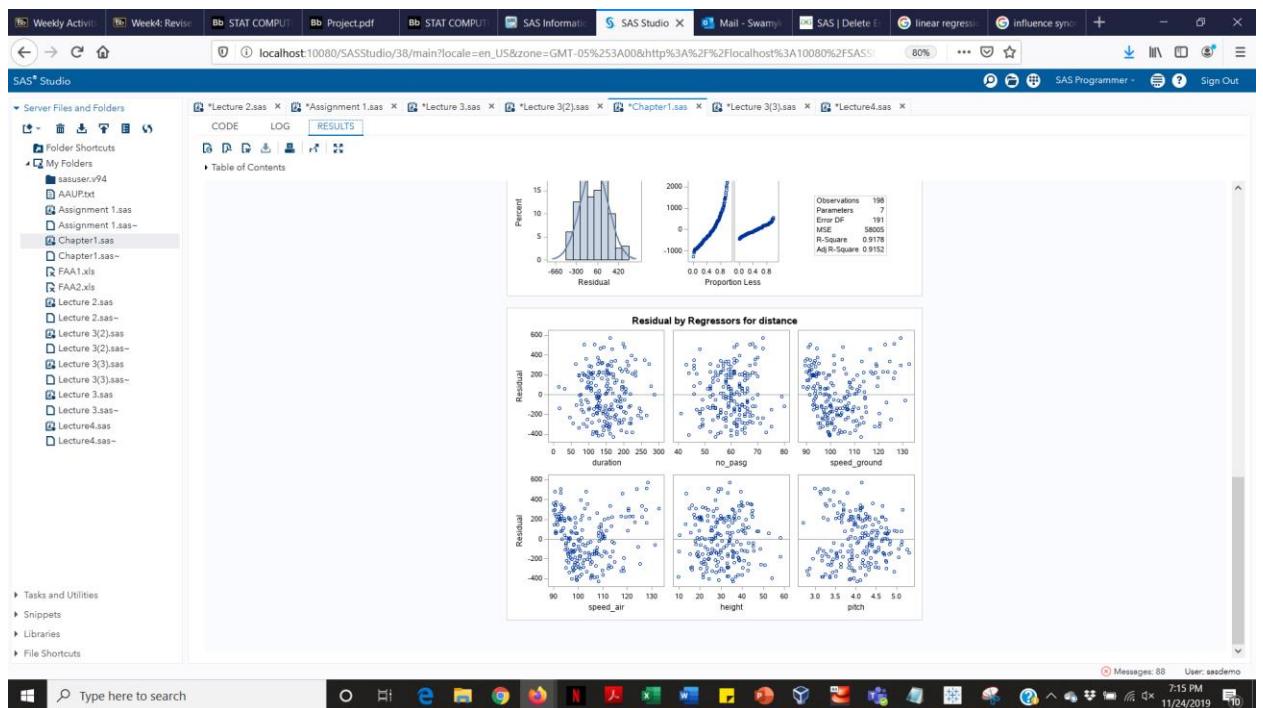
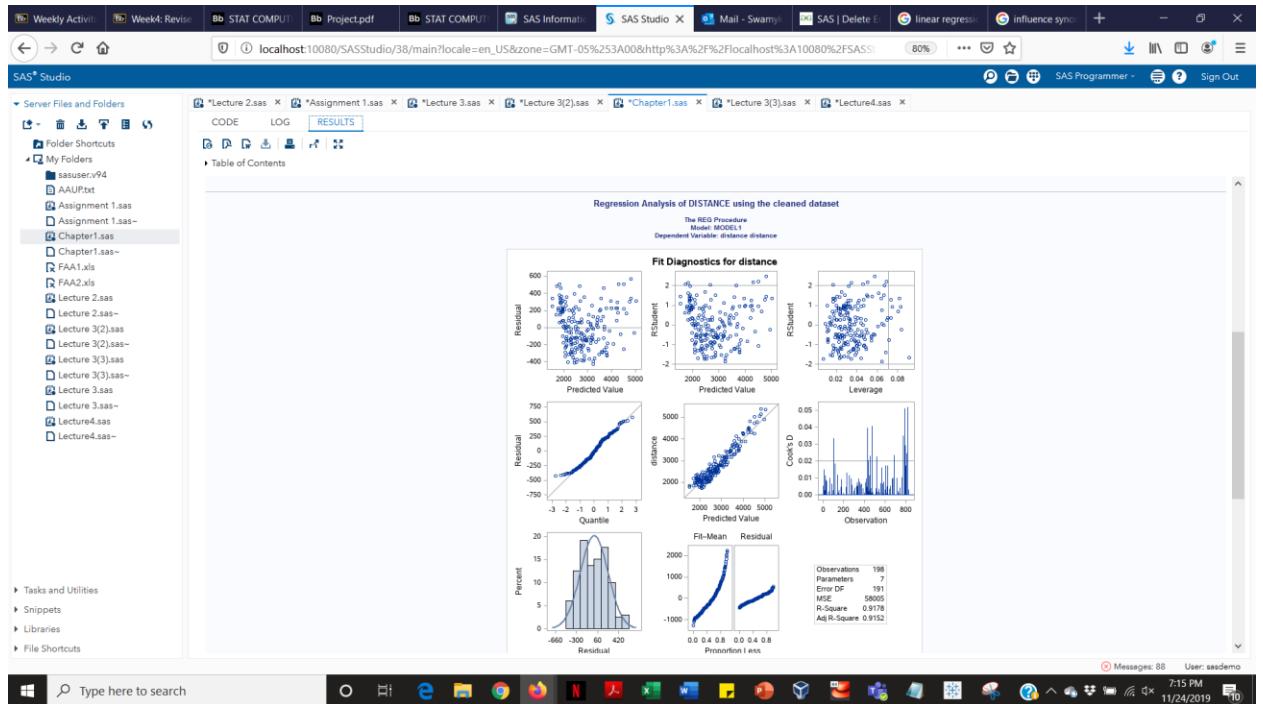
Source	DF	Sum of Squares	Mean Square	F Value	P > F
Model	6	1209.0487	201.5081	368.50	<.0001
Error	191	1107.9038	58.0008		
Corrected Total	197	1346.9525			

Root MSE: 24.04317 **R-Square**: 0.9178
Dependent Mean: 2795.82178 **Adj R-Sq**: 0.9152
Coeff Var: 8.64532

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	P > t
Intercept	Intercept	1	-6269.88431	287.39195	-21.81	<.0001
duration	duration	1	0.02260	0.34434	0.07	0.9480
no_pasg	no_pasg	1	-0.31415	2.44655	-0.35	0.7775
speed_ground	speed_ground	1	2.07002	11.43221	0.18	0.8952
speed_air	speed_air	1	62.79823	11.62617	7.12	<.0001
height	height	1	12.79770	1.87318	7.04	<.0001
pitch	pitch	1	126.76187	31.01197	4.02	<.0001

Fit Diagnostics for distance



2. Further regression analysis is performed with residuals :

CODE :

```
/*Regression Analysis with Residuals */
proc reg data = cleaning;
model distance = speed_air height/r;
```

```
output out = cleaneddata_residuals r = residuals;
title Regression Analysis of DISTANCE with Residuals using the cleaned dataset;
run;
```

```
proc ttest data =cleaneddata_residuals;
var residuals;
run;
```

SAS® Studio

Server Files and Folders

- My Folders
 - sasuserv94
 - AAUP.txt
 - Assignment 1.sas
 - Assignment 1.sas-
 - Chapter1.sas**
 - Chapter1.sas-
 - FAA1.xls
 - FAA2.xls
 - Lecture 2.sas
 - Lecture 2.sas-
 - Lecture 3(2).sas
 - Lecture 3(2).sas-
 - Lecture 3(3).sas
 - Lecture 3(3).sas-
 - Lecture 3(3).sas-
 - Lecture 3.sas
 - Lecture 3.sas-
 - Lecture4.sas
 - Lecture4.sas-

RESULTS

Table of Contents

Regression Analysis of DISTANCE with Residuals using the cleaned dataset

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Number of Observations Read	830
Number of Observations Used	206
Number of Observations with Missing Values	630

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	P > F
Model	2	12502013	62501006	1027.01	<.0001
Error	203	12598437	62006		
Corrected Total	205	13760140			

Root MSE: 246.19387 R-Square: 0.9084
Dependent Mean: 277.05942 Adj R-Sq: 0.9075
Coeff Var: 8.97415

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	P > t
Intercept	Intercept	1	-5862.4123	189.45010	-30.94	<.0001
speed_Air	speed_Air	1	83.58000	1.80098	44.75	<.0001
height	height	1	12.81748	1.52011	7.02	<.0001

Regression Analysis of DISTANCE with Residuals using the cleaned dataset

The REG Procedure
Model: MODEL1
Dependent Variable: distance distance

Obs	Dependent Variable	Predicted Value	Std Error Predict	Residual	Std Residual	Student Residual	Cook's D
1	250.7						
2	1287.0						

Type here to search

SAS® Studio

Server Files and Folders

- My Folders
 - sasuserv94
 - AAUP.txt
 - Assignment 1.sas
 - Assignment 1.sas-
 - Chapter1.sas**
 - Chapter1.sas-
 - FAA1.xls
 - FAA2.xls
 - Lecture 2.sas
 - Lecture 2.sas-
 - Lecture 3(2).sas
 - Lecture 3(2).sas-
 - Lecture 3(3).sas
 - Lecture 3(3).sas-
 - Lecture 3(3).sas-
 - Lecture 3.sas
 - Lecture 3.sas-
 - Lecture4.sas
 - Lecture4.sas-

RESULTS

Table of Contents

Fit Diagnostics for distance

Sum of Residuals	8.87078E-10
Sum of Squared Residuals	12598437
Predicted Residual SS (PRESS)	13019082

Observations: 206
Parameters: 3
Error DF: 203
MSE: 62006
R-Square: 0.9084
Adj R-Square: 0.9075

Type here to search

SAS® Studio

Server Files and Folders

- Folder Shortcuts
- My Folders
 - sasuser.v94
 - AAUUP.txt
 - Assignment 1.sas
 - Assignment 1.sas~
 - Chapter1.sas
 - Chapter1.sas~
 - FAA1.xls
 - FAA2.xls
 - Lecture 2.sas
 - Lecture 2.sas~
 - Lecture 3(2).sas
 - Lecture 3(2).sas~
 - Lecture 3(3).sas
 - Lecture 3(3).sas~
 - Lecture 3.sas
 - Lecture 3.sas~
 - Lecture4.sas
 - Lecture4.sas~

CODE LOG RESULTS OUTPUT DATA

Table of Contents

Residual by Regressors for distance

Residual

speed_air

height

Regression Analysis of DISTANCE with Residuals using the cleaned dataset

The TTEST Procedure

Variable: residuals (Residual)

N	Mean	Std Dev	Std Err	Minimum	Maximum
205	4.95E-12	247.9	17.2729	-525.9	550.6

Mean	95% CL Mean	Std Dev	95% CL Std Dev
4.95E-12	-34.0553	34.0553	247.9

DF	t Value	Pr > t
205	0.00	1.0000

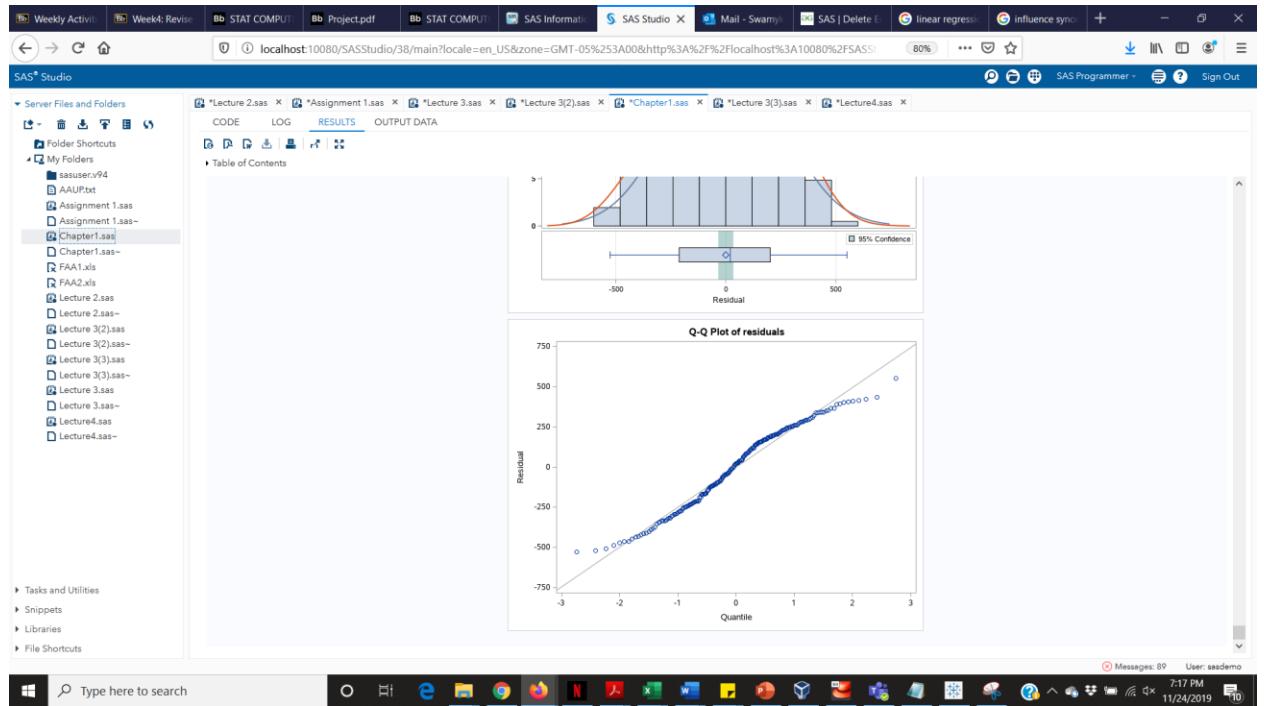
Distribution of residuals

With 95% Confidence Interval for Mean

Messages: 89 User: sasdemo

The screenshot shows the SAS Studio interface with the following details:

- Top Bar:** Shows tabs for "Weeky Activity", "Weeky: Revise", "STAT COMPUT", "Project.pdf", "STAT COMPUT", "SAS Informa", "SAS Studio", "Mail - Swamy", "SAS | Delete E", "linear regress", "influence sync", and a "+" button.
- Title Bar:** Displays the URL "localhost:10080/SASStudio/38/main?locale=en_US&zone=GMT-05%253A0&http%3A%2F%2flocalhost%3A10080%2FSAS%" and a progress bar at 80%.
- SAS Studio Sidebar:** Includes sections for "Server Files and Folders", "My Folders" (containing files like "sauserv94", "AAUP.txt", "Assignment 1.sas", "Assignment 1.sas~", "Chapter1.laa", "Chapter1.laa~", "FAA1.xls", "FAA2.xls", "Lecture 2.sas", "Lecture 2.sas~", "Lecture 3(2).sas", "Lecture 3(2).sas~", "Lecture 3(3).sas", "Lecture 3(3).sas~", "Lecture 3.sas", "Lecture 3.sas~", "Lecture4.sas", and "Lecture4.sas~"), "Tasks and Utilities", "Snippets", "Libraries", and "File Shortcuts".
- Central Area:** Features two plots:
 - Distribution of residuals:** A histogram titled "Distribution of residuals With 95% Confidence Interval for Mean". The x-axis is labeled "Residual" and ranges from -500 to 500. The y-axis is labeled "Percent" and ranges from 0 to 20. The histogram bars are blue, and a normal distribution curve (blue line) and a kernel density estimate (red line) are overlaid. A green box highlights the 95% confidence interval on the x-axis.
 - Q-Q Plot of residuals:** A scatter plot titled "Q-Q Plot of residuals". The x-axis is labeled "observed" and ranges from 0 to 750. The y-axis is labeled "residual" and ranges from 0 to 750. The data points form a roughly linear pattern, indicating approximate normality of the residuals.
- Bottom Taskbar:** Shows the Windows Start button, a search bar with the placeholder "Type here to search", and a row of icons for various applications including File Explorer, Edge, File Manager, and others.
- System Tray:** Shows icons for battery level, signal strength, and date/time (11/24/2019, 7:17 PM).
- Status Bar:** Displays "Messages: 89" and "User: sauserm".



From this we can see that speed of the air and height show significant values. Hence we can say that they are the influencing factors for the landing distance of the aircrafts.

CONCLUSION:

- Since the p-values of both speed_air and height are less than 0.05 they are the final parameters.
- Moreover, from the previous chapter it is concluded that speed_ground and speed_air are closely related. And from the above analysis we can see that speed_air is more significant than speed_ground. Hence we can drop speed_ground and retain speed_air for the final observations.
- From the residual analysis and T-test, we can see that the plot varies only slightly from the normal curve.

QUESTIONS :

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

The final model contains around 836 observations. All 950 flights are not taken, since there are a lot of duplicate, missing and abnormal values that don't validate the conditions. Hence, all these abnormal data are deleted and the data is cleaned and reduced to 836 observations.

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

The factors that impact the landing distance of a flight are air speed of the aircraft and the height when it is passing over the threshold of the runway.

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

There is no significant differences between the two aircraft makes.