



STATISTICAL FOUNDATION FOR DATA ANALYTICS CSHO531DAP

Record of Laboratory Experiments/ Programs



Submitted in Partial fulfillment of Laboratory Internal Assessment
for the course
STATISTICAL FOUNDATION FOR DATA ANALYTICS CSHO531DAP
in the department of Computer Science and Engineering

OCTOBER 23, 2022
ASHWIN A

Some Important Points

1. Entire document should be followed in Time New Roman Font & Size is 12
2. Student Name and Roll Number in every page; top right corner
3. Headings are as follows:
 - a. Problem Statement
 - b. Sequence of Execution steps with snapshots and descriptions
 - i. Step 1
 - ii. Step 2
 - iii. Step N
 - c. Conclusion
4. All the **snapshots** and the **descriptions** to be aligned left (Ctrl+L)

List of experiments with problem statements:

1. Introduction to IBM's SPSS [Statistical Product and Service Solutions] tool for Statistical Analysis
2. Basic operations performed in the tool
 - A. Define variables with data types and related parameters
 - B. Data set Variable view & Data view
 - C. Sort Variables, Transpose, Aggregate, Split file & Select cases
3. Introduction to Null Hypothesis
4. Basic operations on Statistical parameters
 - A. Mean, Median, Mode, Standard Deviation and Variance
 - B. Frequencies, Descriptives [Skewness & Kurtosis; Histogram with Normal curve]
5. Statistical tests and their significance value
 - A. One Way ANOVA
 - B. Independent Sample T-Test or Chi- Square Test

Experiment 1:

Problem Statement:

Introduction to IBM's SPSS (Statistical Product and Service Solutions) tool for Statistical Analysis.

Introduction:

SPSS is a Windows based program that can be used to perform data entry and analysis and to create tables and graphs. SPSS is capable of handling large amounts of data and can perform all of the analyses and much more.

Open SPSS on your computer, something that looks similar to the following screenshot is observed in Fig 1.

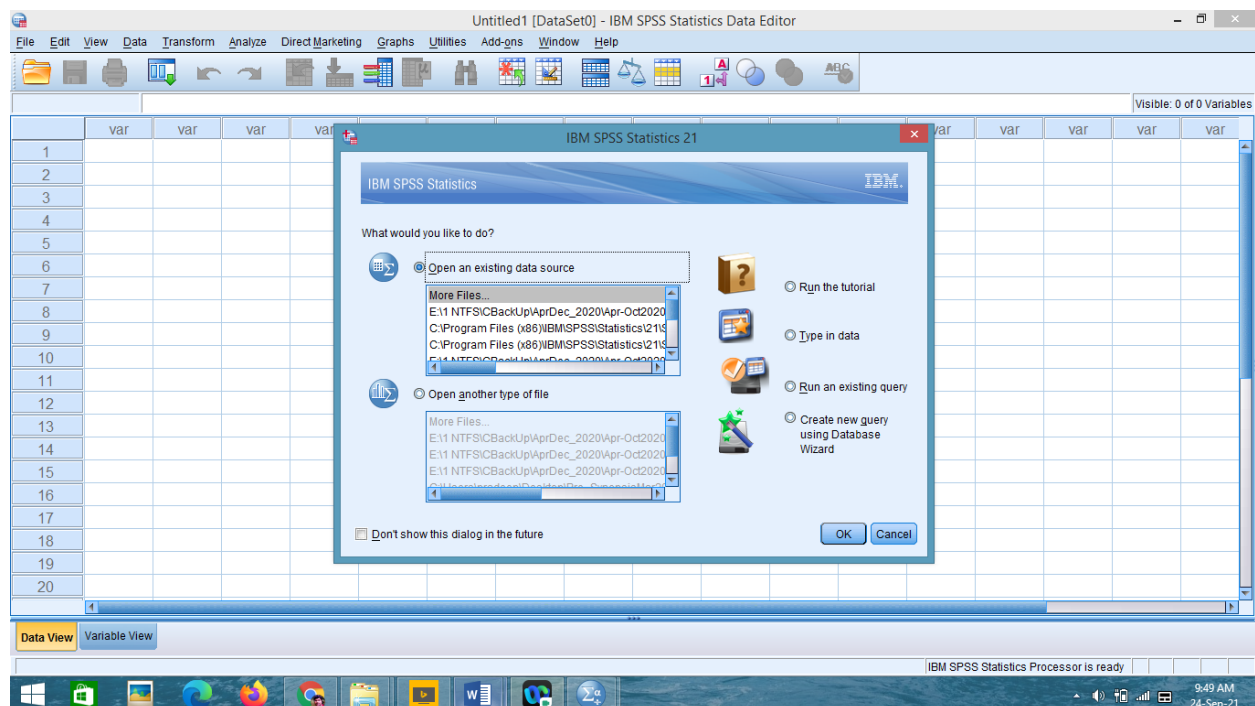
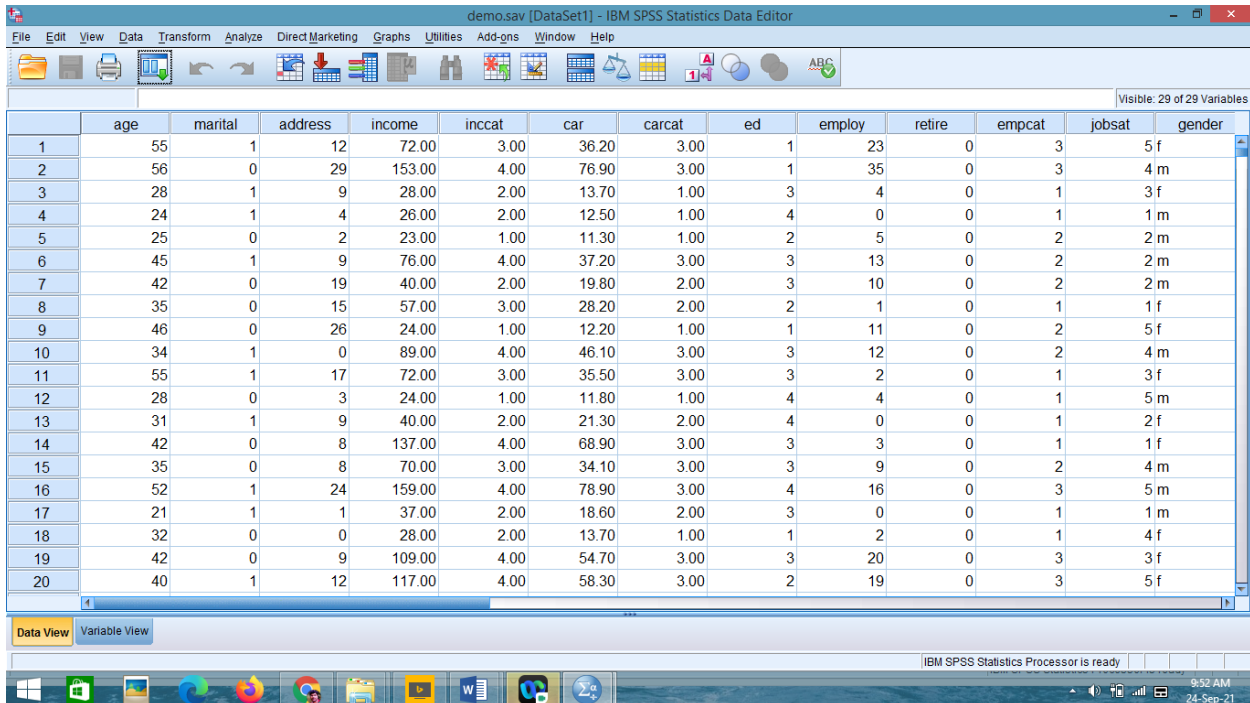


Fig 1: Basic layout of SPSS tool

SPSS automatically assumes that you want to open an existing file, and immediately opens a dialogue box to ask which file you'd like to open. It'll make it easier to navigate the interface and windows in SPSS if we open a file.

A demo.sav file is selected to open an existing dataset which can be seen in Fig 2.



	age	marital	address	income	inccat	car	carcat	ed	employ	retire	empcat	jobsat	gender
1	55	1	12	72.00	3.00	36.20	3.00	1	23	0	3	5	f
2	56	0	29	153.00	4.00	76.90	3.00	1	35	0	3	4	m
3	28	1	9	28.00	2.00	13.70	1.00	3	4	0	1	3	f
4	24	1	4	26.00	2.00	12.50	1.00	4	0	0	1	1	m
5	25	0	2	23.00	1.00	11.30	1.00	2	5	0	2	2	m
6	45	1	9	76.00	4.00	37.20	3.00	3	13	0	2	2	m
7	42	0	19	40.00	2.00	19.80	2.00	3	10	0	2	2	m
8	35	0	15	57.00	3.00	28.20	2.00	2	1	0	1	1	f
9	46	0	26	24.00	1.00	12.20	1.00	1	11	0	2	5	f
10	34	1	0	89.00	4.00	46.10	3.00	3	12	0	2	4	m
11	55	1	17	72.00	3.00	35.50	3.00	3	2	0	1	3	f
12	28	0	3	24.00	1.00	11.80	1.00	4	4	0	1	5	m
13	31	1	9	40.00	2.00	21.30	2.00	4	0	0	1	2	f
14	42	0	8	137.00	4.00	68.90	3.00	3	3	0	1	1	f
15	35	0	8	70.00	3.00	34.10	3.00	3	9	0	2	4	m
16	52	1	24	159.00	4.00	78.90	3.00	4	16	0	3	5	m
17	21	1	1	37.00	2.00	18.60	2.00	3	0	0	1	1	m
18	32	0	0	28.00	2.00	13.70	1.00	1	2	0	1	4	f
19	42	0	9	109.00	4.00	54.70	3.00	3	20	0	3	3	f
20	40	1	12	117.00	4.00	58.30	3.00	2	19	0	3	5	f

Fig 2: Demo.sav SPSS dataset/file

SPSS Menus and Icons

Review the options listed under each menu on the Menu Bar by clicking them one at a time. Follow along with the below descriptions.

File includes all of the options you typically use in other programs, such as open, save, exit. Notice that you can open or create new files of multiple types as illustrated to the right.

Edit includes the typical cut, copy, and paste commands, and allows you to specify various options for displaying data and output.

Click on **Options**, and you will see the dialog box to the left. You can use this to format the data, output, charts, etc. These choices are rather overwhelming, and you can simply take the default options for now. The author of your text (me) dumb to even know was too these options could easily be set.

View allows you to select which toolbars you want to show, select font size, add or remove the gridlines that separate each piece of data, and to select whether or not to display your raw data or the data labels.

Data allows you to select several options ranging from displaying data that is sorted by a specific variable to selecting certain cases for subsequent analyses.

Transform includes several options to change current variables. For example, you can change continuous variables to categorical variables, change scores into rank scores, add a constant to variables, etc.

Analyze includes all of the commands to carry out statistical analyses and to calculate descriptive statistics. Much of this book will focus on using commands located in this menu.

Graphs includes the commands to create various types of graphs including box plots, histograms, line graphs, and bar charts.

Utilities allows you to list file information which is a list of all variables, their labels, values, locations in the data file, and type.

Add-ons are programs that can be added to the base SPSS package. You probably do not have access to any of those. Window can be used to select which window you want to view (i.e., Data Editor, Output Viewer, or Syntax).

Help has many useful options including a link to the SPSS homepage, a statistics coach, and a syntax guide.

Using **topics**, you can use the index option to type in any key word and get a list of options, or you can view the categories and subcategories available under contents. This is an excellent tool and can be used to troubleshoot most problems.

The **Icons** directly under the Menu bar provide shortcuts to many common commands that are available in specific menus.

Conclusion:

The fundamental aspects and introduction of the statistical tool SPSS are briefly explained.

Experiment 2: (Part A)**Problem Statement:**

Define variables with data types and related parameters for a specific data set.

Describing Dataset (with list of variables and data types):

Dataset Name: Diamonds Dataset

This is a dataset of various dimensions, cut quality, color, country of origin and their corresponding prices.

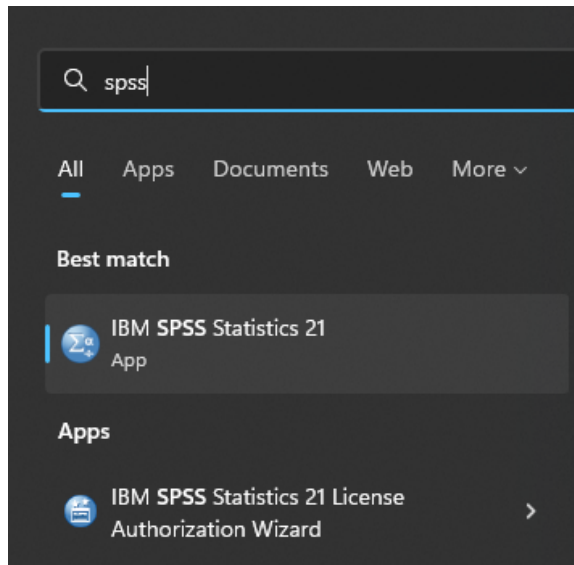
Dataset link: <https://www.kaggle.com/datasets/nancyalaswad90/diamonds-prices>

Variable	Data type
Carat	Numeric
Cut	Numeric
Color	String
Clarity	Numeric
Table	Numeric
Depth	Numeric
Country	Numeric
X	Numeric
Y	Numeric
Z	Numeric
Price	Numeric

Steps of Execution

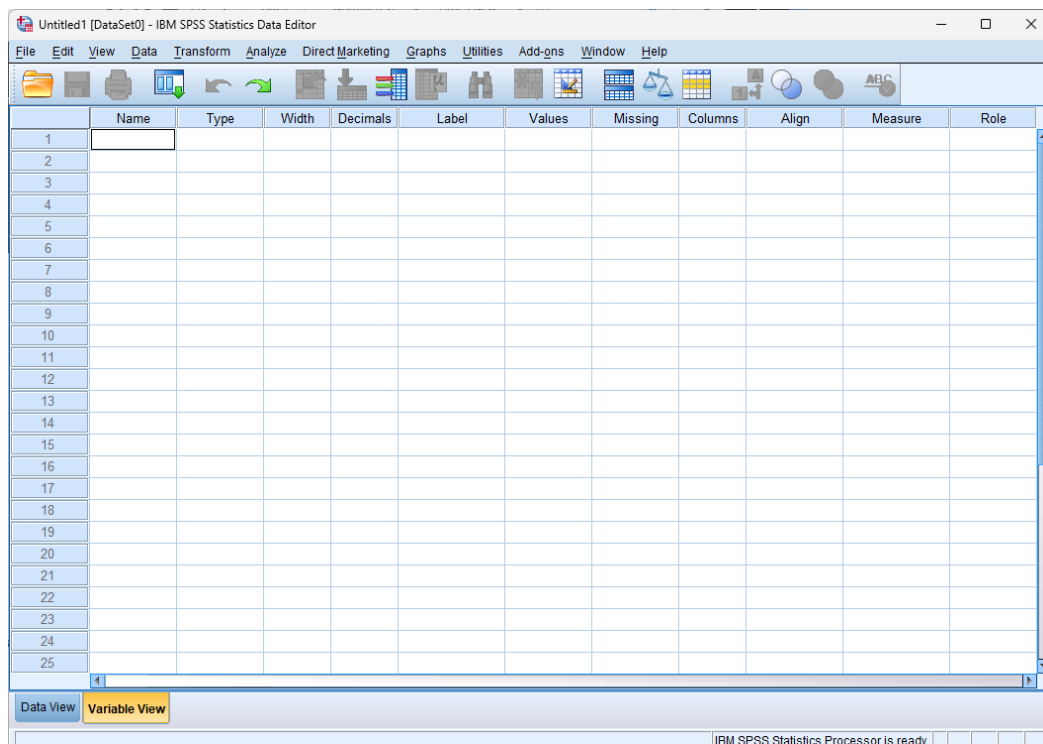
Step 1:

search and open SPSS



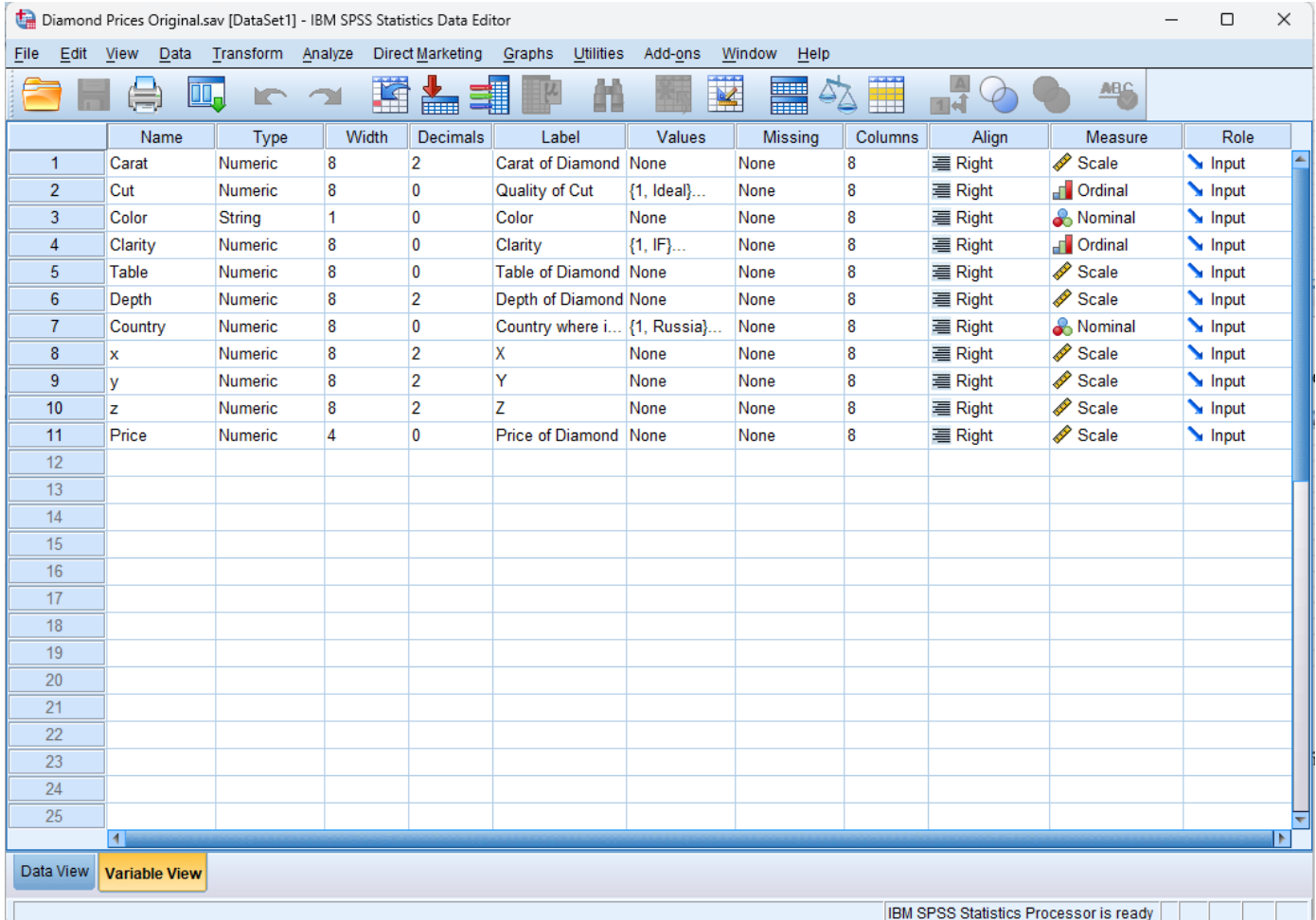
Step 2:

Enter variable View in SPSS editor



Step 3:

Enter all the attribute details and the corresponding data types and value



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Carat	Numeric	8	2	Carat of Diamond	None	None	8	Right	Scale	Input
2	Cut	Numeric	8	0	Quality of Cut	{1, Ideal}...	None	8	Right	Ordinal	Input
3	Color	String	1	0	Color	None	None	8	Right	Nominal	Input
4	Clarity	Numeric	8	0	Clarity	{1, IF}...	None	8	Right	Ordinal	Input
5	Table	Numeric	8	0	Table of Diamond	None	None	8	Right	Scale	Input
6	Depth	Numeric	8	2	Depth of Diamond	None	None	8	Right	Scale	Input
7	Country	Numeric	8	0	Country where i...	{1, Russia}...	None	8	Right	Nominal	Input
8	x	Numeric	8	2	X	None	None	8	Right	Scale	Input
9	y	Numeric	8	2	Y	None	None	8	Right	Scale	Input
10	z	Numeric	8	2	Z	None	None	8	Right	Scale	Input
11	Price	Numeric	4	0	Price of Diamond	None	None	8	Right	Scale	Input
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

Data View Variable View

IBM SPSS Statistics Processor is ready

Conclusion:

Variables and datatypes were defined and declared in the SPSS editor for a selected specific dataset.

Experiment 2: (Part B)

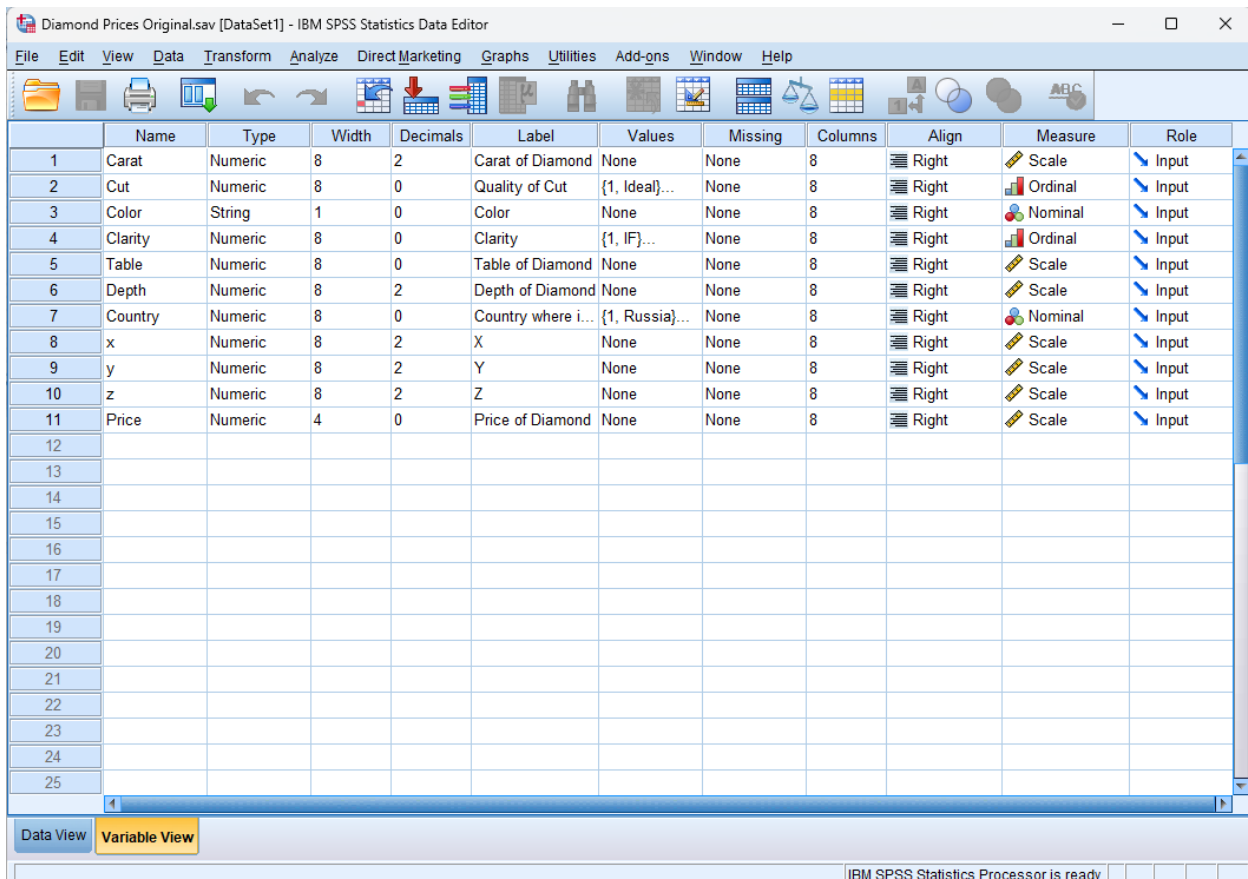
Problem Statement:

Data set Variable view & Data view

Steps of Execution

Step 1:

Open SPSS and Enter the Attribute details in the variable view



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Carat	Numeric	8	2	Carat of Diamond	None	None	8	Right	Scale	Input
2	Cut	Numeric	8	0	Quality of Cut	{1, Ideal}...	None	8	Right	Ordinal	Input
3	Color	String	1	0	Color	None	None	8	Right	Nominal	Input
4	Clarity	Numeric	8	0	Clarity	{1, IF}...	None	8	Right	Ordinal	Input
5	Table	Numeric	8	0	Table of Diamond	None	None	8	Right	Scale	Input
6	Depth	Numeric	8	2	Depth of Diamond	None	None	8	Right	Scale	Input
7	Country	Numeric	8	0	Country where i...	{1, Russia}...	None	8	Right	Nominal	Input
8	x	Numeric	8	2	X	None	None	8	Right	Scale	Input
9	y	Numeric	8	2	Y	None	None	8	Right	Scale	Input
10	z	Numeric	8	2	Z	None	None	8	Right	Scale	Input
11	Price	Numeric	4	0	Price of Diamond	None	None	8	Right	Scale	Input
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

IBM SPSS Statistics Processor is ready

Step 2:

Open the dataset using a relevant app(Microsoft Excel for csv) and copy the records from the dataset and paste the values into the corresponding column or attribute in SPSS dataview

Diamond Prices Original.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 11 of 11 Variables

	Carat	Cut	Color	Clarity	Table	Depth	Country	x	y	z	Price
1	.23	1	E	7	55	61.50	6	3.95	3.98	2.43	326
2	.21	2	E	6	61	59.80	2	3.89	3.84	2.31	326
3	.23	4	E	4	65	56.90	5	4.05	4.07	2.31	327
4	.29	2	I	5	58	62.40	7	4.20	4.23	2.63	334
5	.31	4	J	7	58	63.30	7	4.34	4.35	2.75	335
6	.24	3	J	3	57	62.80	9	3.94	3.96	2.48	336
7	.24	3	I	2	57	62.30	10	3.95	3.98	2.47	336
8	.26	3	H	6	55	61.90	8	4.07	4.11	2.53	337
9	.22	5	E	5	61	65.10	2	3.87	3.78	2.49	337
10	.23	3	H	4	61	59.40	8	4.00	4.05	2.39	338
11	.30	4	J	6	55	64.00	5	4.25	4.28	2.73	339
12	.23	1	J	4	56	62.80	5	3.93	3.90	2.46	340
13	.22	2	F	6	61	60.40	7	3.88	3.84	2.33	342
14	.31	1	J	7	54	62.20	5	4.35	4.37	2.71	344
15	.20	2	E	7	62	60.20	10	3.79	3.75	2.27	345
16	.32	2	E	8	58	60.90	4	4.38	4.42	2.68	345
17	.30	1	I	7	54	62.00	1	4.31	4.34	2.68	348
18	.30	4	J	6	54	63.40	8	4.23	4.29	2.70	351
19	.30	4	J	6	56	63.80	2	4.23	4.26	2.71	351
20	.30	3	J	6	59	62.70	1	4.21	4.27	2.66	351
21	.30	4	I	7	56	63.30	6	4.26	4.30	2.71	351
22	.23	3	E	5	55	63.80	2	3.85	3.92	2.48	352
23	.23	3	H	4	57	61.00	1	3.94	3.96	2.41	353
24	.31	3	J	6	62	59.40	5	4.39	4.43	2.62	353
25	.31	3	J	6	62	58.10	4	4.44	4.47	2.59	353

Data View Variable View

IBM SPSS Statistics Processor is ready

Conclusion:

The Attributes have been declared in the Variable view and values in the dataset have been entered in Dataview.

Experiment 2: (Part C)

Problem Statement:

Sort Variables, Transpose, Aggregate, Split file & Select cases

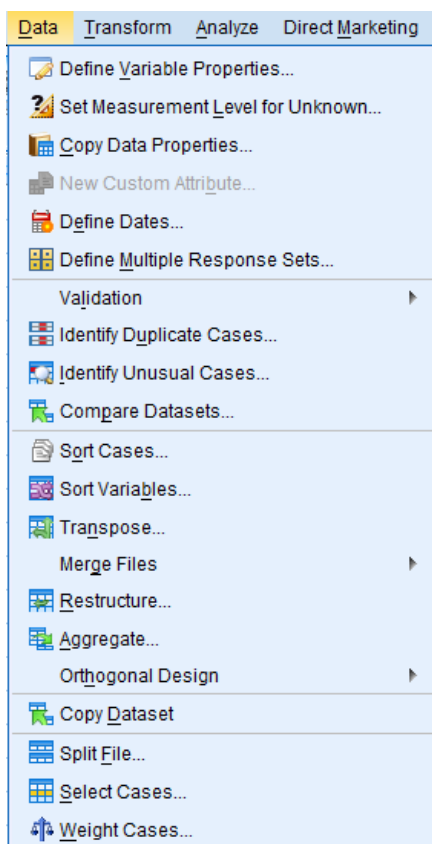
Steps of execution:

Sort Variables:

Arrange the attributes in ascending or descending order with respect to the given parameter

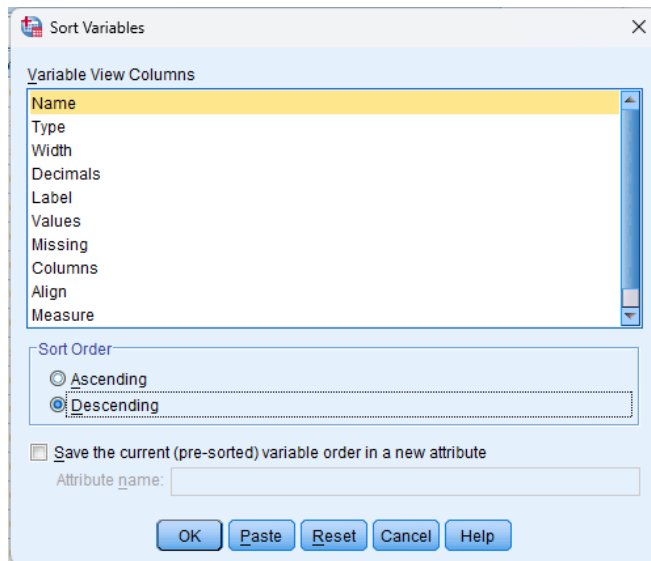
Step 1:

select Sort variables in Data→sort variables



Step 2:

Select the parameter and The order i.e Ascending or Descending order



Step 3:

The output is displayed

*Diamond Prices Original - Copy.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 11 of 11 Variables

	z	y	x	Table	Price	Depth	Cut	Country	Color	Clarity	Carat
1	2.43	3.98	3.95	55	326	61.50	1	6	E	7	.23
2	2.31	3.84	3.89	61	326	59.80	2	2	E	6	.21
3	2.31	4.07	4.05	65	327	56.90	4	5	E	4	.23
4	2.63	4.23	4.20	58	334	62.40	2	7	I	5	.29
5	2.75	4.35	4.34	58	335	63.30	4	7	J	7	.31
6	2.48	3.96	3.94	57	336	62.80	3	9	J	3	.24
7	2.47	3.98	3.95	57	336	62.30	3	10	I	2	.24
8	2.53	4.11	4.07	55	337	61.90	3	8	H	6	.26
9	2.49	3.78	3.87	61	337	65.10	5	2	E	5	.22
10	2.39	4.05	4.00	61	338	59.40	3	8	H	4	.23
11	2.73	4.28	4.25	55	339	64.00	4	5	J	6	.30
12	2.46	3.90	3.93	56	340	62.80	1	5	J	4	.23
13	2.33	3.84	3.88	61	342	60.40	2	7	F	6	.22
14	2.71	4.37	4.35	54	344	62.20	1	5	J	7	.31
15	2.27	3.75	3.79	62	345	60.20	2	10	E	7	.20
16	2.68	4.42	4.38	58	345	60.90	2	4	E	8	.32
17	2.68	4.34	4.31	54	348	62.00	1	1	I	7	.30
18	2.70	4.29	4.23	54	351	63.40	4	8	J	6	.30
19	2.71	4.26	4.23	56	351	63.80	4	2	J	6	.30
20	2.66	4.27	4.21	59	351	62.70	3	1	J	6	.30
21	2.71	4.30	4.26	56	351	63.30	4	6	I	7	.30
22	2.48	3.92	3.85	55	352	63.80	3	2	E	5	.23
23	2.41	3.96	3.94	57	353	61.00	3	1	H	4	.23
24	2.62	4.43	4.39	62	353	59.40	3	5	J	6	.31
25	2.59	4.47	4.44	62	353	58.10	3	4	J	6	.31

Data View Variable View

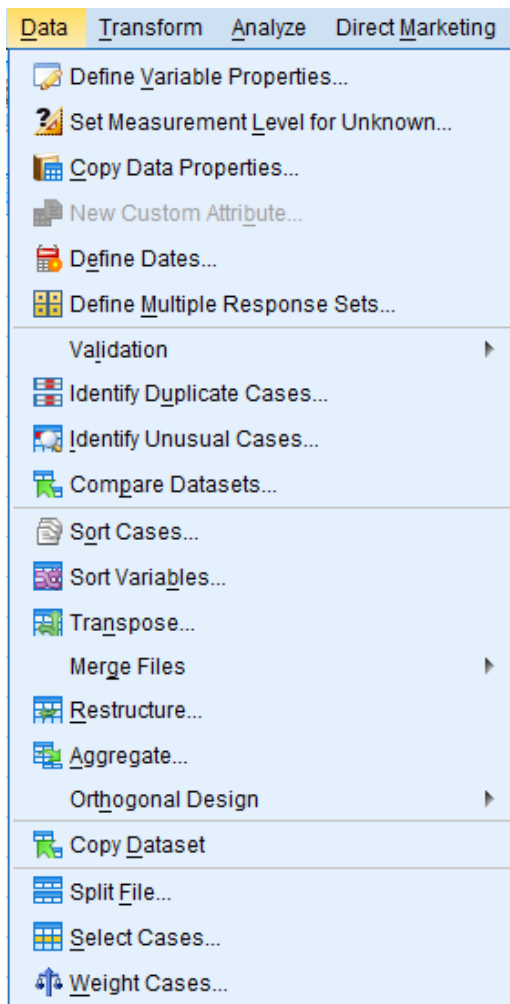
IBM SPSS Statistics Processor is ready

Aggregate:

Used to run aggregate functions(sum, mean, count, etc) on variables

Step 1:

Select Aggregate from Data → Aggregate



Step 2:

select the break variable and the summary variable and select the aggregate function from the functions tab.

Aggregate Data

Break Variable(s):
 Price of Diamond [Price]

Aggregated Variables:
 Summaries of Variable(s):
 Country_mean = MEAN(Country)

Function... Name & Label...
☐ Number of cases Name: N_BREAK

Save
☐ Add aggregated variables to active dataset
☒ Create a new dataset containing only the aggregated variables
 Dataset name: aa
☐ Write a new data file containing only the aggregated variables
 File... C:\Users\ashwin\Documents\aggr.sav

Options for Very Large Datasets
☐ File is already sorted on break variable(s)
☐ Sort file before aggregating

OK Paste Reset Cancel Help

Aggregate Data: Aggregate Function

Summary Statistics	Specific Values	Number of cases
<input checked="" type="radio"/> Mean	<input type="radio"/> First	<input type="radio"/> Weighted
<input type="radio"/> Median	<input type="radio"/> Last	<input type="radio"/> Weighted missing
<input type="radio"/> Sum	<input type="radio"/> Minimum	<input type="radio"/> Unweighted
<input type="radio"/> Standard Deviation	<input type="radio"/> Maximum	<input type="radio"/> Unweighted missing

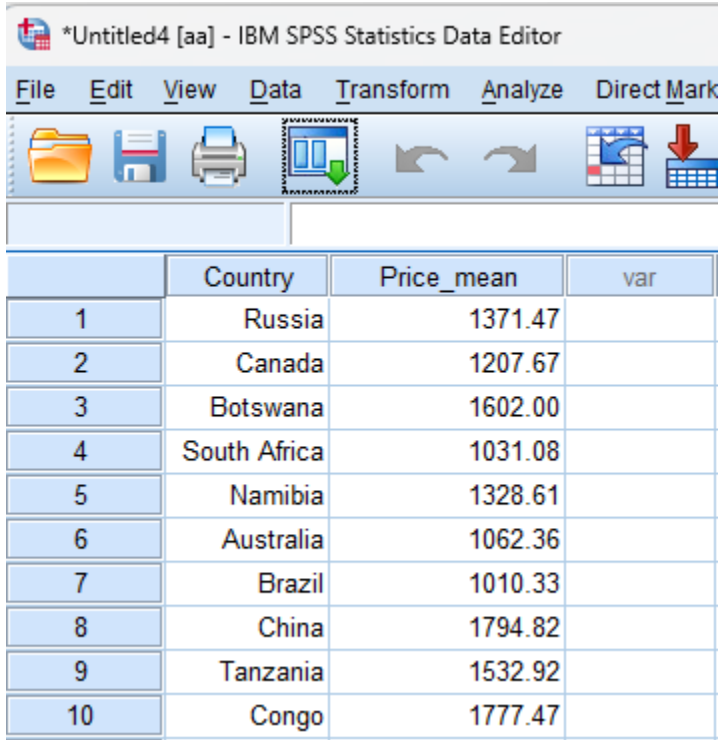
Percentages
☐ Above
☐ Below Value:
☐ Inside
☐ Outside Low: High:

Fractions
☐ Above
☐ Below Value:
☐ Inside
☐ Outside Low: High:

Continue Cancel Help

Step 3:

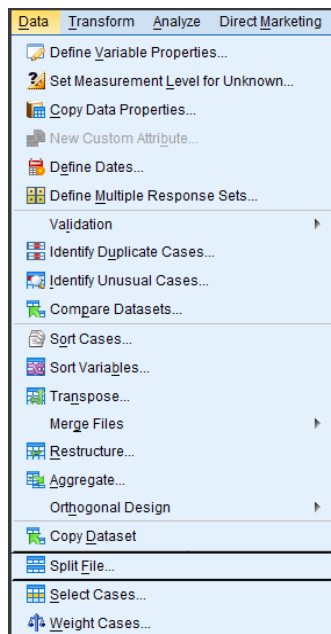
The output is displayed



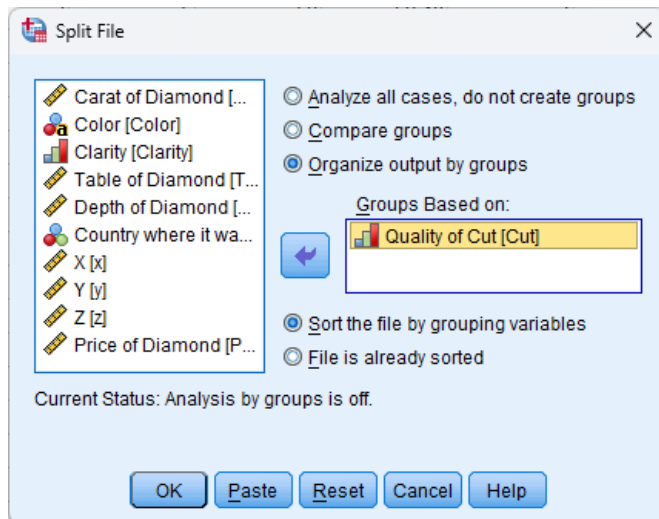
	Country	Price_mean	var
1	Russia	1371.47	
2	Canada	1207.67	
3	Botswana	1602.00	
4	South Africa	1031.08	
5	Namibia	1328.61	
6	Australia	1062.36	
7	Brazil	1010.33	
8	China	1794.82	
9	Tanzania	1532.92	
10	Congo	1777.47	

Split File:

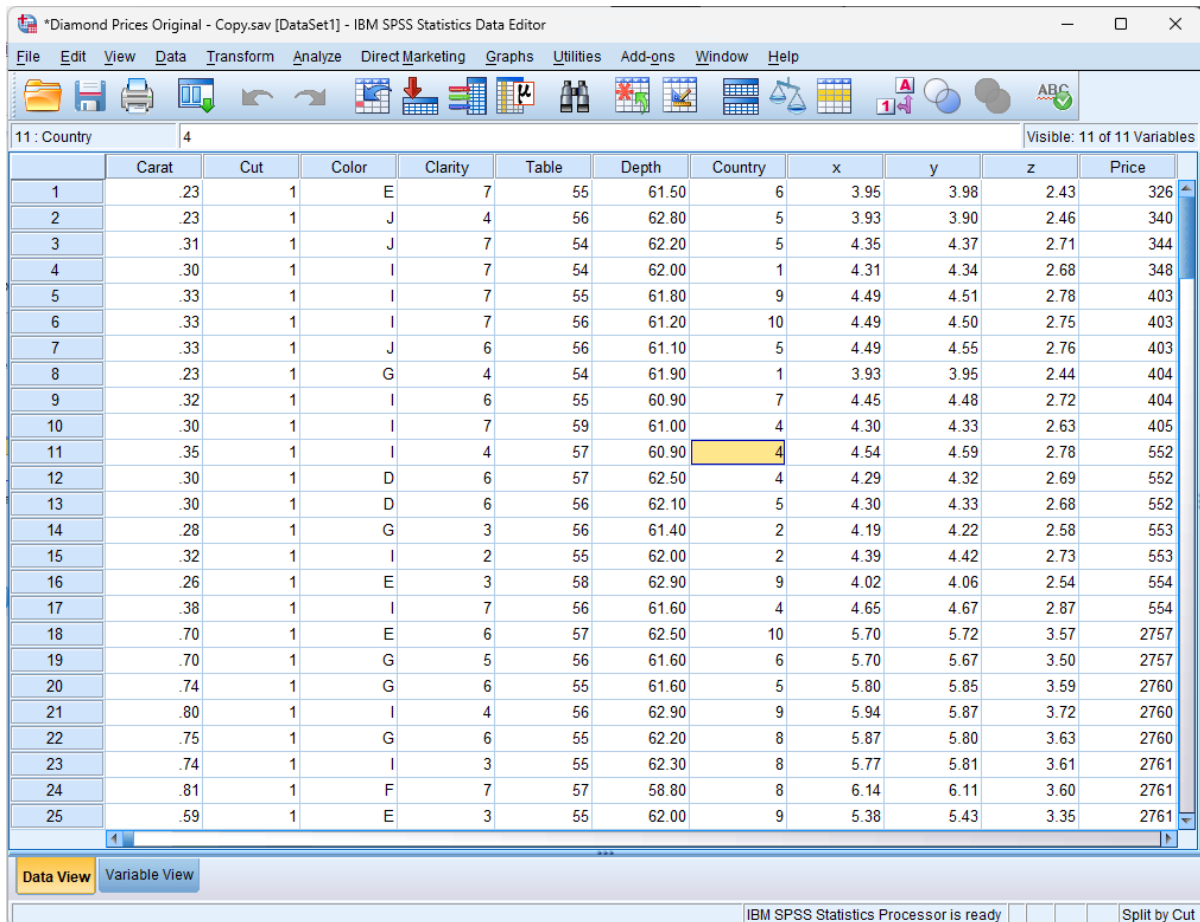
Step 1: select Split file Data tab Data→Split File



Step 2: Select the attribute to create groups based on



Step 3: Output is displayed in the variable view



*Diamond Prices Original - Copy.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

11 : Country 4 Visible: 11 of 11 Variables

	Carat	Cut	Color	Clarity	Table	Depth	Country	x	y	z	Price
1	.23	1	E	7	55	61.50	6	3.95	3.98	2.43	326
2	.23	1	J	4	56	62.80	5	3.93	3.90	2.46	340
3	.31	1	J	7	54	62.20	5	4.35	4.37	2.71	344
4	.30	1	I	7	54	62.00	1	4.31	4.34	2.68	348
5	.33	1	I	7	55	61.80	9	4.49	4.51	2.78	403
6	.33	1	I	7	56	61.20	10	4.49	4.50	2.75	403
7	.33	1	J	6	56	61.10	5	4.49	4.55	2.76	403
8	.23	1	G	4	54	61.90	1	3.93	3.95	2.44	404
9	.32	1	I	6	55	60.90	7	4.45	4.48	2.72	404
10	.30	1	I	7	59	61.00	4	4.30	4.33	2.63	405
11	.35	1	I	4	57	60.90	4	4.54	4.59	2.78	552
12	.30	1	D	6	57	62.50	4	4.29	4.32	2.69	552
13	.30	1	D	6	56	62.10	5	4.30	4.33	2.68	552
14	.28	1	G	3	56	61.40	2	4.19	4.22	2.58	553
15	.32	1	I	2	55	62.00	2	4.39	4.42	2.73	553
16	.26	1	E	3	58	62.90	9	4.02	4.06	2.54	554
17	.38	1	I	7	56	61.60	4	4.65	4.67	2.87	554
18	.70	1	E	6	57	62.50	10	5.70	5.72	3.57	2757
19	.70	1	G	5	56	61.60	6	5.70	5.67	3.50	2757
20	.74	1	G	6	55	61.60	5	5.80	5.85	3.59	2760
21	.80	1	I	4	56	62.90	9	5.94	5.87	3.72	2760
22	.75	1	G	6	55	62.20	8	5.87	5.80	3.63	2760
23	.74	1	I	3	55	62.30	8	5.77	5.81	3.61	2761
24	.81	1	F	7	57	58.80	8	6.14	6.11	3.60	2761
25	.59	1	E	3	55	62.00	9	5.38	5.43	3.35	2761

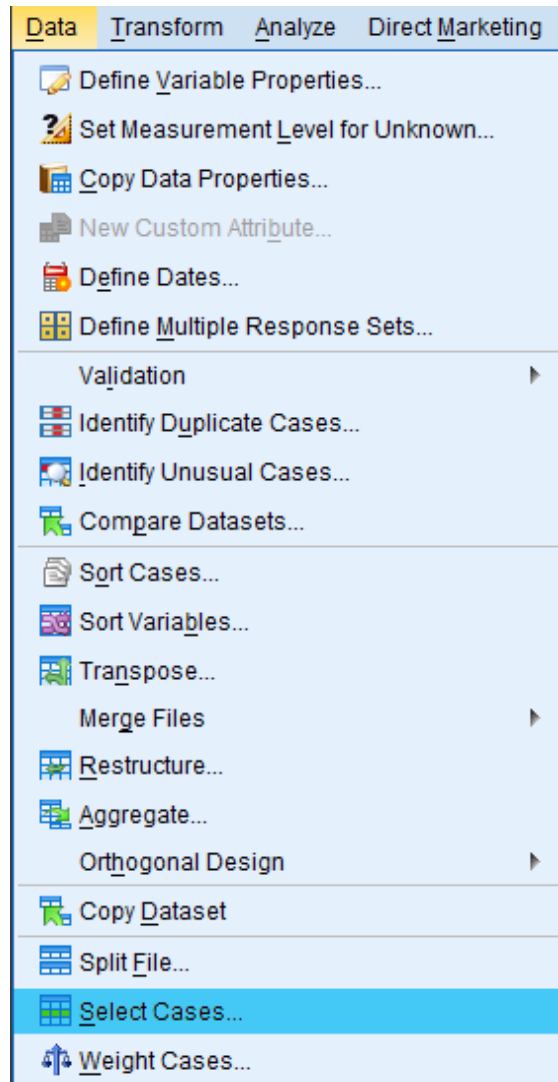
Data View Variable View

IBM SPSS Statistics Processor is ready Split by Cut

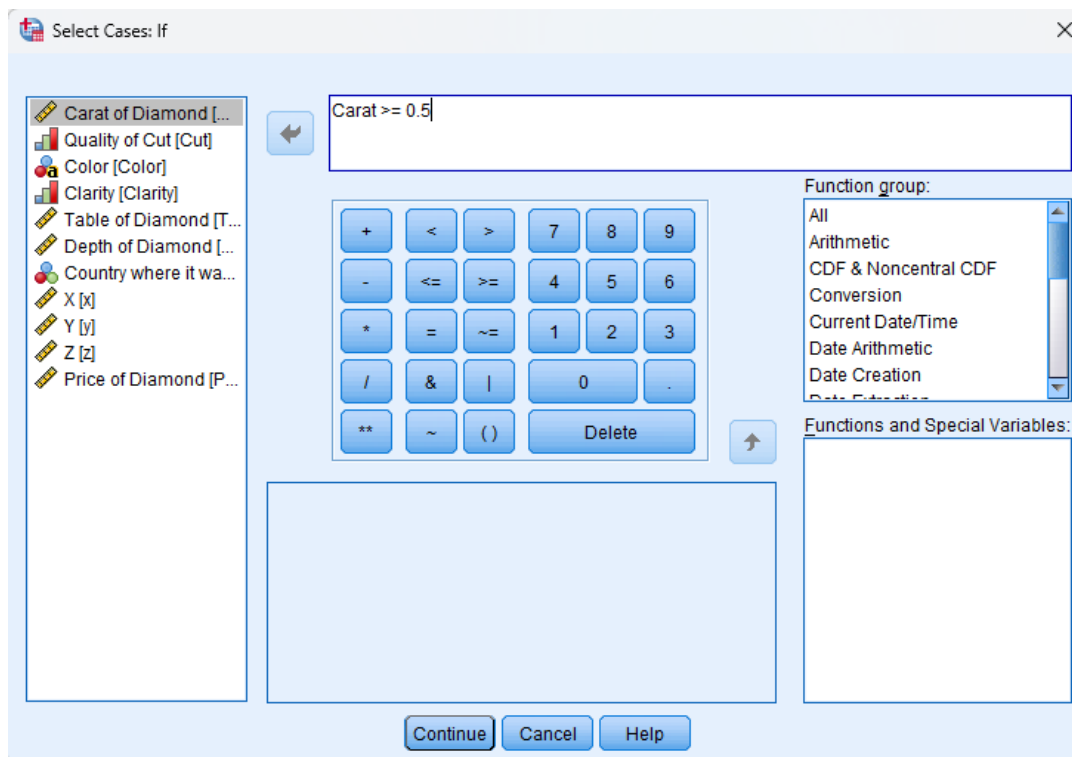
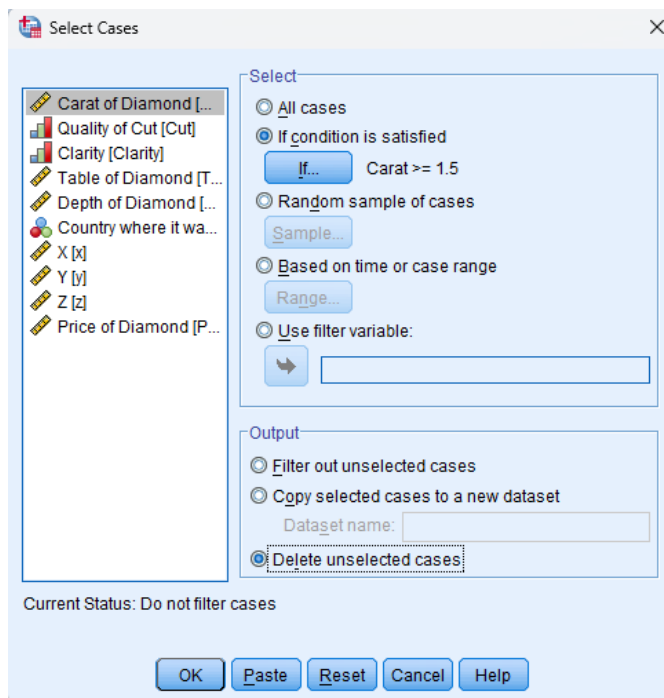
Select Cases

Step 1:

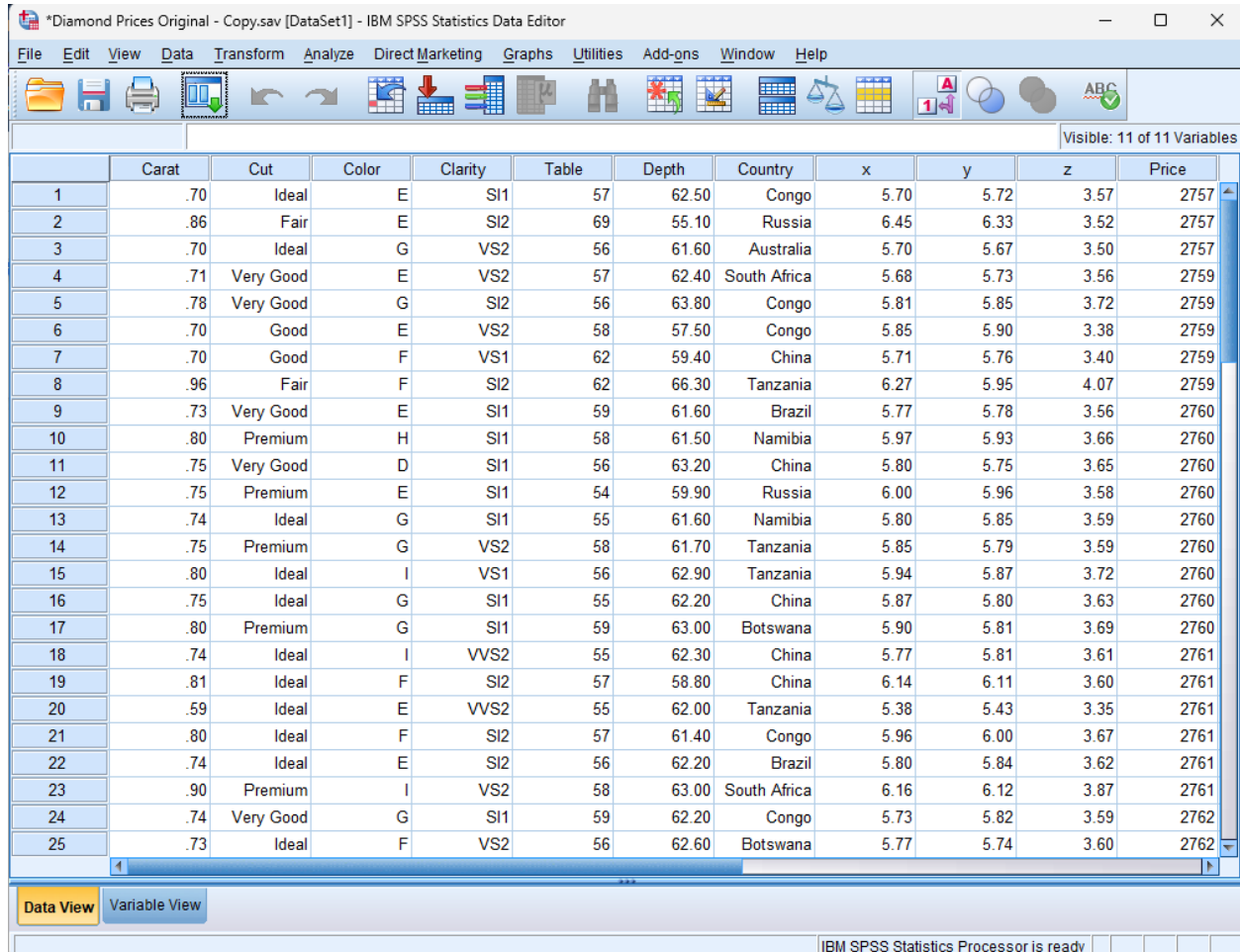
Select “Select Cases” from Data tab in Data→select Cases



Step 2: Select the condition to apply on the dataset



Step 3: output is displayed in the data view



	Carat	Cut	Color	Clarity	Table	Depth	Country	x	y	z	Price
1	.70	Ideal	E	SI1	57	62.50	Congo	5.70	5.72	3.57	2757
2	.86	Fair	E	SI2	69	55.10	Russia	6.45	6.33	3.52	2757
3	.70	Ideal	G	VS2	56	61.60	Australia	5.70	5.67	3.50	2757
4	.71	Very Good	E	VS2	57	62.40	South Africa	5.68	5.73	3.56	2759
5	.78	Very Good	G	SI2	56	63.80	Congo	5.81	5.85	3.72	2759
6	.70	Good	E	VS2	58	57.50	Congo	5.85	5.90	3.38	2759
7	.70	Good	F	VS1	62	59.40	China	5.71	5.76	3.40	2759
8	.96	Fair	F	SI2	62	66.30	Tanzania	6.27	5.95	4.07	2759
9	.73	Very Good	E	SI1	59	61.60	Brazil	5.77	5.78	3.56	2760
10	.80	Premium	H	SI1	58	61.50	Namibia	5.97	5.93	3.66	2760
11	.75	Very Good	D	SI1	56	63.20	China	5.80	5.75	3.65	2760
12	.75	Premium	E	SI1	54	59.90	Russia	6.00	5.96	3.58	2760
13	.74	Ideal	G	SI1	55	61.60	Namibia	5.80	5.85	3.59	2760
14	.75	Premium	G	VS2	58	61.70	Tanzania	5.85	5.79	3.59	2760
15	.80	Ideal	I	VS1	56	62.90	Tanzania	5.94	5.87	3.72	2760
16	.75	Ideal	G	SI1	55	62.20	China	5.87	5.80	3.63	2760
17	.80	Premium	G	SI1	59	63.00	Botswana	5.90	5.81	3.69	2760
18	.74	Ideal	I	VVS2	55	62.30	China	5.77	5.81	3.61	2761
19	.81	Ideal	F	SI2	57	58.80	China	6.14	6.11	3.60	2761
20	.59	Ideal	E	VVS2	55	62.00	Tanzania	5.38	5.43	3.35	2761
21	.80	Ideal	F	SI2	57	61.40	Congo	5.96	6.00	3.67	2761
22	.74	Ideal	E	SI2	56	62.20	Brazil	5.80	5.84	3.62	2761
23	.90	Premium	I	VS2	58	63.00	South Africa	6.16	6.12	3.87	2761
24	.74	Very Good	G	SI1	59	62.20	Congo	5.73	5.82	3.59	2762
25	.73	Ideal	F	VS2	56	62.60	Botswana	5.77	5.74	3.60	2762

Conclusion:

Sort Variables, Transpose, Aggregate, Split file, Select cases were successfully executed in SPSS

Experiment 3:

Problem Statement:

Introduction to Null Hypothesis

NULL HYPOTHESIS

The null hypothesis is a characteristic arithmetic theory, suggesting that no statistical relationship and significance exists in a set of given, single, observed variables between two sets of observed data and measured phenomena. Sometimes referred to as the ‘null’, it is represented as H_0 .

Commonly with analysis or testing, it is assumed there exists no significance or effect unless there is enough evidence to suggest otherwise, similar to how a trial presumes innocence.

When the sample contains sufficient evidence, the null can be rejected, and the effect be deemed statistically significant.

Conclusion:

The null hypothesis is elaborated upon, and can thus be applied in various situations, using different tests.

Experiment 4 (Part A):

Problem Statement

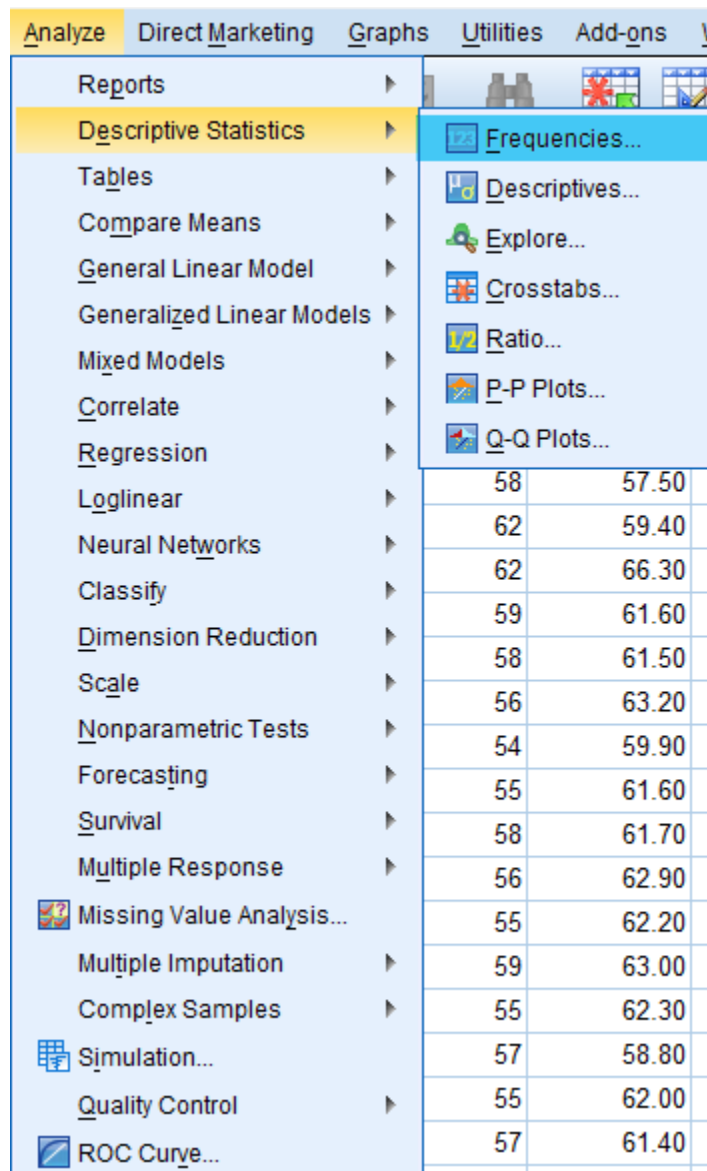
Basic operations on Statistical parameters

Mean, Median, Mode, Standard Deviation and Variance

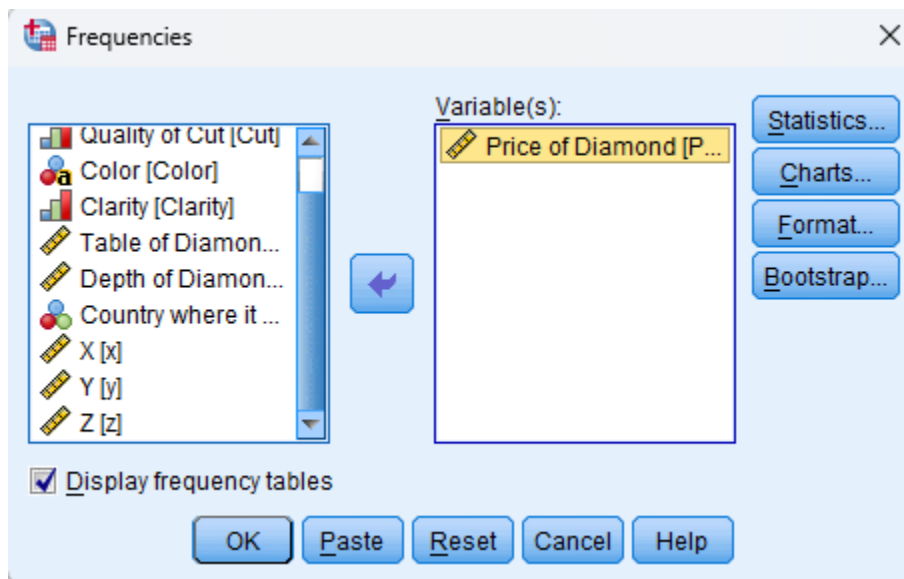
Steps of Execution:

Step 1:

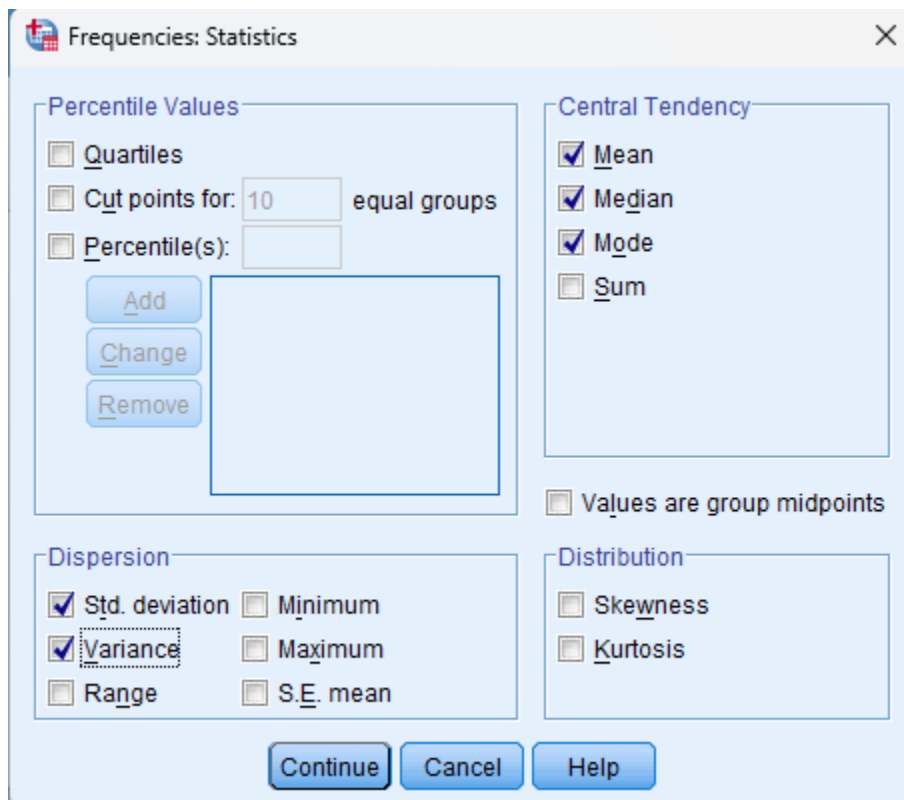
Select Frequencies from Analyze → Descriptive Statistics → Frequencies



Step 2: Select the variables on which the frequencies has to be applied



Step 3: Select the Frequencies that need to be displayed



Step 4: Output is displayed in the output viewer

➔ **Frequencies**

[DataSet1] C:\Users\ashwi\Documents\SPSSInc\Diamond Prices Original

Statistics

Price of Diamond

N	Valid	150
	Missing	0
Mean		1365.21
Median		554.00
Mode		554
Std. Deviation		1146.741
Variance		1315014.840

Conclusion:

Mean, Median, Mode, Standard deviation and Variance were successfully executed in SPSS

Experiment 4 (Part B):

Problem Statement

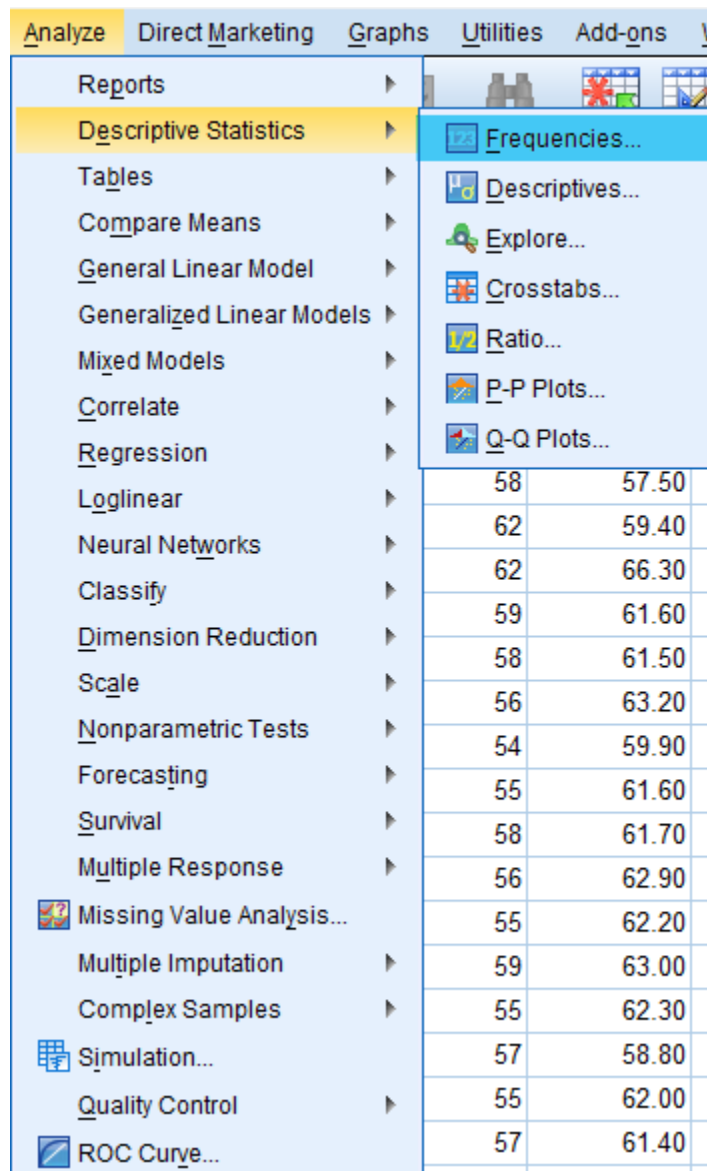
Basic operations on Statistical parameters

Frequencies, Descriptives [Skewness & Kurtosis; Histogram with Normal curve]

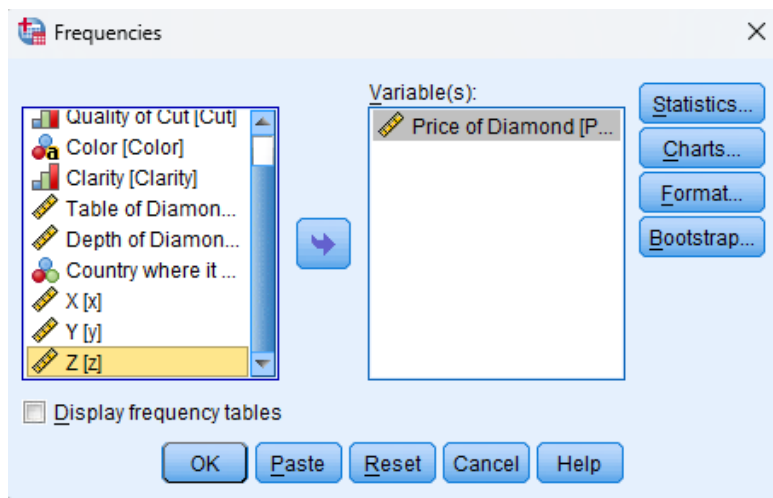
Steps of Execution:

Step 1:

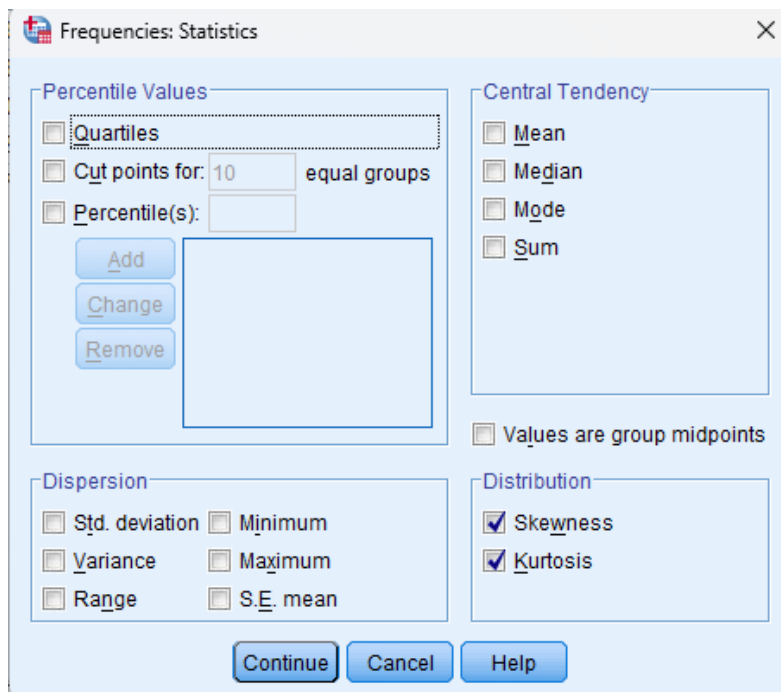
Select Frequencies from Analyze → Descriptive Statistics → Frequencies



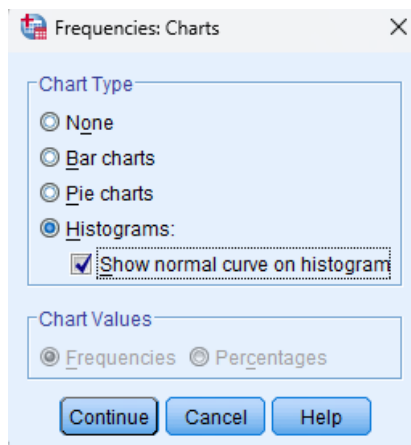
Step 2: Select the attribute to apply Skewness and kurtosis



Step 3: Select Skewness and Kurtosis in the statistics tab

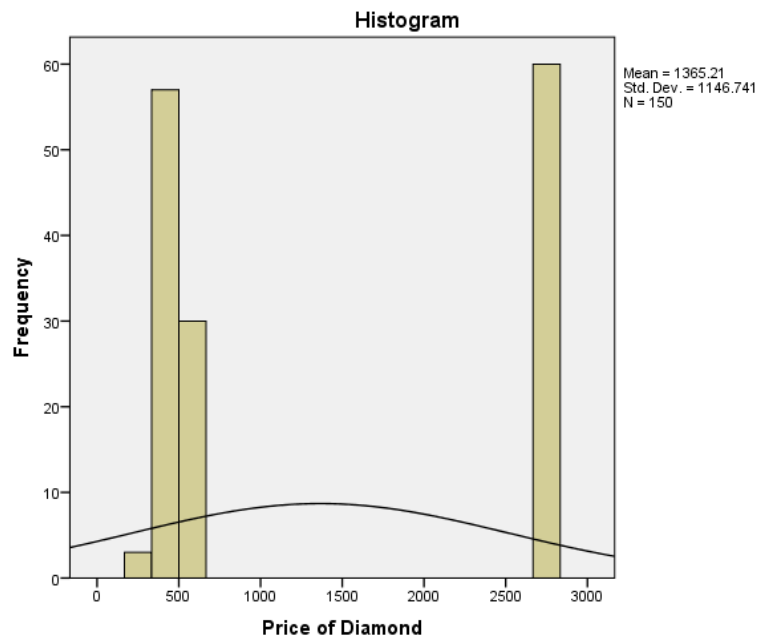


Step 4: Select Histogram chart in Charts tab



Step 5: Output is displayed in the Output viewer

Statistics		
Price of Diamond		
N	Valid	150
	Missing	0
Skewness		.401
Std. Error of Skewness		.198
Kurtosis		-1.849
Std. Error of Kurtosis		.394



Conclusion:

Skewness and Kurtosis was successfully applied to the dataset in SPSS

Experiment: 5 (Part A)

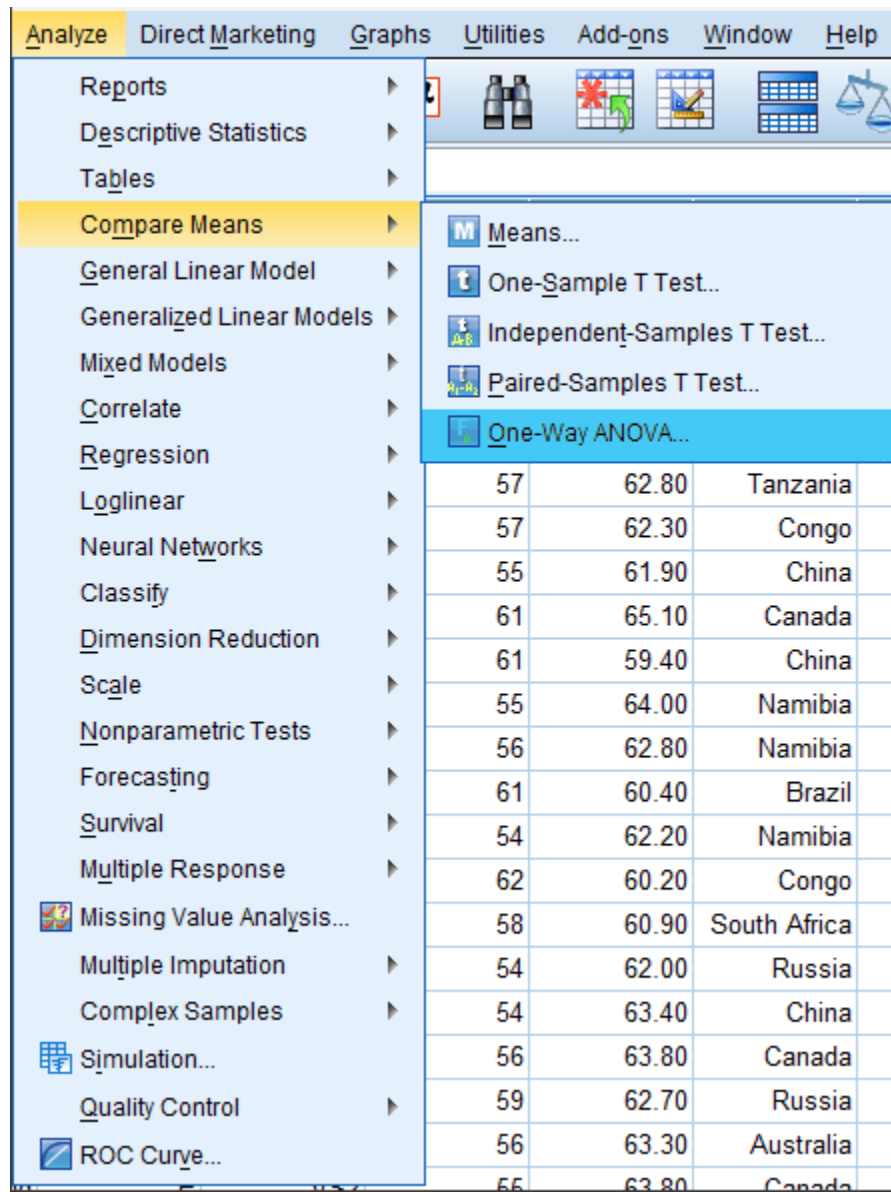
Problem Statement:

Statistical tests and their significance value

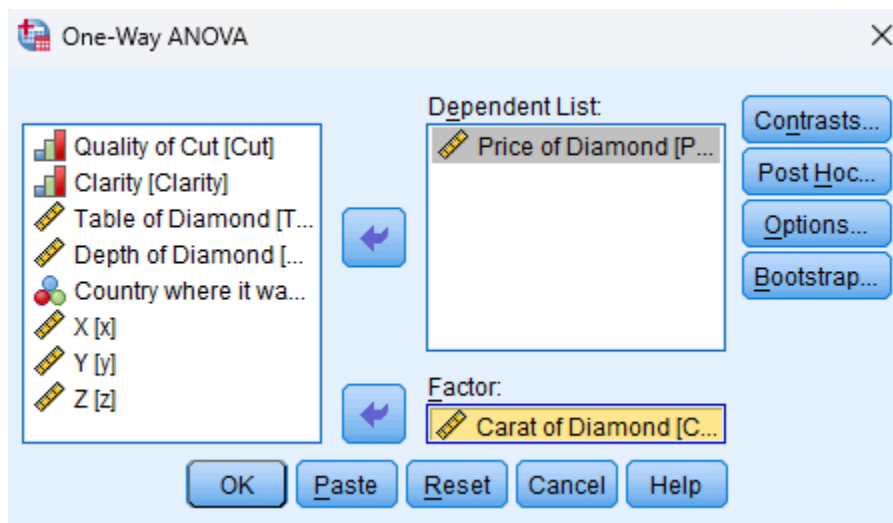
One Way ANOVA

Steps Of Execution:

Step 1: Open SPSS and Go to ANOVA in Analyze → Compare Means → One-way Anova



Step 2: Select the dependent attribute and the factor attribute and click OK



Step 3:

The output is displayed in the output viewer

➔ **Oneway**

[DataSet1] C:\Users\ashwi\Documents\SPSSInc\Diamond Prices Original

ANOVA					
Price of Diamond					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	195487646.1	33	5923868.064	1528.519	.000
Within Groups	449565.068	116	3875.561		
Total	195937211.2	149			

Conclusion:

One-Way ANOVA Test was executed and the significance value was found to be 0.00 hence there is a strong association between the two variables.

Experiment: 5 (Part B)

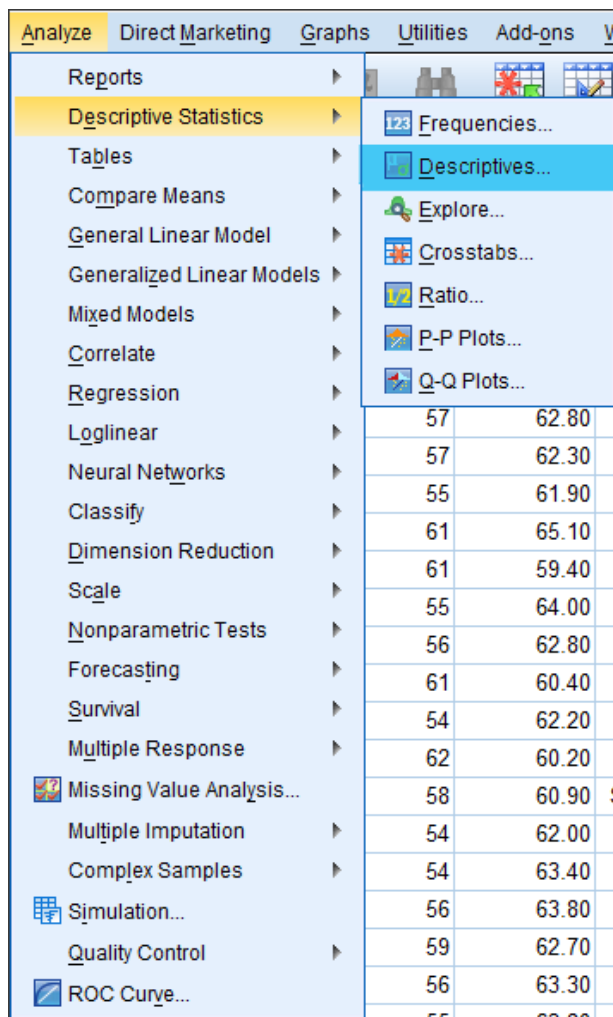
Problem Statement:

Statistical tests and their significance value

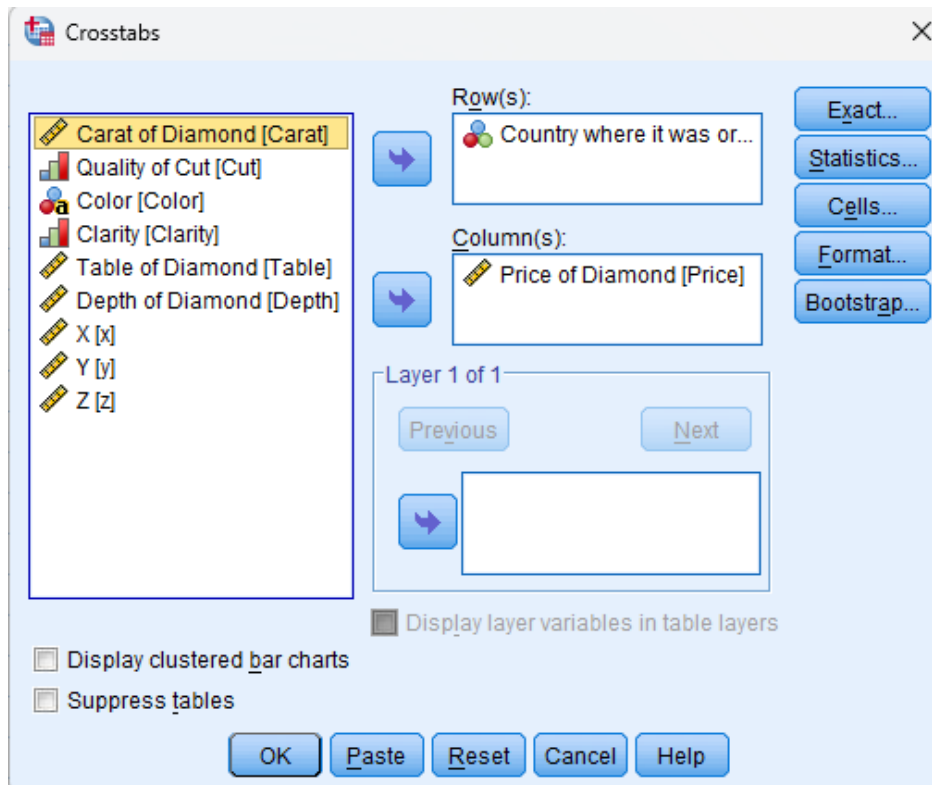
Chi- Square Test

Steps of Execution:

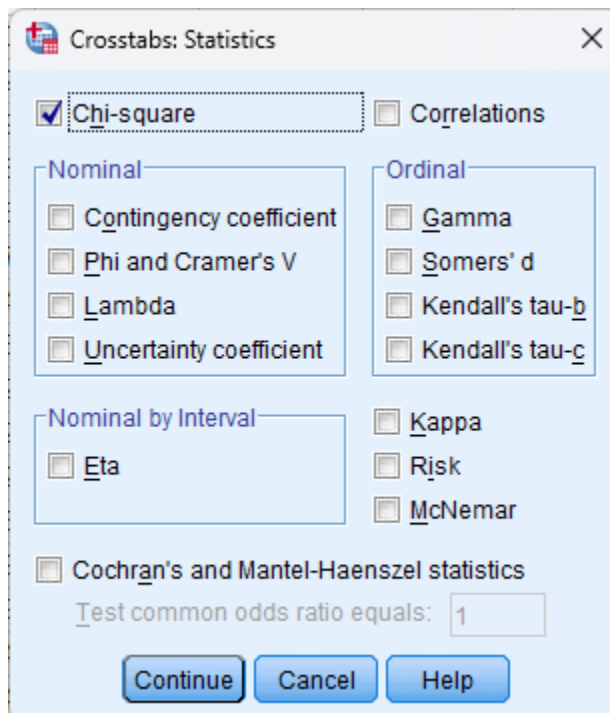
Step 1: select Crosstabs from Analyze → Descriptive Statistics → crosstabs



Step 2: Select attributes for Rows and Columns to check the significance value



Step 3: select Chi-Square test from the statistics tab



Step 4: The output is displayed in the output viewer

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	329.846 ^a	324	.400
Likelihood Ratio	276.206	324	.975
Linear-by-Linear Association	2.481	1	.115
N of Valid Cases	150		

a. 370 cells (100.0%) have expected count less than 5. The minimum expected count is .05.

Conclusion:

Chi-Square Test was conducted on the given datasets and the significance value came out to be 0.4 which means there is no significant relationship between the two attributes