INTRODUCTION TO STATISTICAL PACKAGES: SPSS

Structure

DMS,KU

4.1 Introduction

SPSS (*Statistical Package for Social Sciences)* is a powerful, user-friendly software package for the manipulation and statistical analysis of data. The package is particularly useful for students and researchers in Social and behavioral sciences. This unit gives a brief and straightforward description of how to conduct a range of statistical analyses using SPSS. It concentrates largely on how to use SPSS to get results and on how to correctly interpret these results. The material is not intended in any way to be an introduction to statistics and, it is assumed that most readers will have some know how of basic statistical concepts such as *linear regression, correlation, significance tests,* and simple *analysis of variance*.

Most SPSS users prefer to use its Windows graphic interface that is, pointing with the mouse and clicking on the options they want. Nonetheless, SPSS provides a way to not only type commands but also switch between the command mode and the Windows point-and-click method.
SPSS is the package most widely used by social and behavioral scientists because of the following reasons:
1. Of the major packages, it seems to be the easiest to use for the most widely used statistical techniques
2. One can use it with either a Windows point-and-click approach or through syntax (SPSS commands). Each has its own advantages, and the user can switch between the approaches.
3. Many of the widely used social science data sets come with an easy method to translate them into SPSS; this significantly reduces the preliminary work needed to explore new data.
Overall, SPSS is a good first statistical package for people wanting to perform quantitative research in Social, Statistical and Management sciences because it is easy to use and because it can be a good starting point to learn more advanced statistical packages.

4.2 Objectives

After going through this unit, students will be able to

- Understand the purpose and use of statistical packages
- Define data analysis
- Convert raw data into SPSS form
- Manipulate data in various ways
- Analyze data for research purposes
- Present analyzed data using charts
- Learn how taxes are paid online

4.3 Getting started with SPSS

In a typical session with SPSS, you are likely to work with three kinds of windows and have the opportunity to save the contents of each. The **Data Editor Window**, which is the first window you encounter, is used to define and enter your data and to perform statistical procedures. The results of the statistical tests appear in the **Output Window**. The **Syntax Window** can be used to keep a record of the operations that you perform on your data. This window will automatically open when you click a **Paste** function. So, for example, when you select certain cases

for analysis or transform your data or compute a correlation, clicking the **Paste** function will store a record of what was performed (in command language). Beyond serving as a log for your operations, it is possible to run commands from the **Syntax** window; we will not discuss the syntax window in this script; to find out how to do this, you can make use of the **Help** commands.

Getting Help

Online help is provided from the Help menu or via context menus or Help buttons on dialogue boxes. Here the general help facility is described. The required menu is available from any window and provides three major help facilities:

Help — Statistics Coach helps users unfamiliar with SPSS or the statistical procedures available in SPSS to get started. This facility prompts the user with simple questions in non technical language about the purpose of the statistical analysis and provides visual examples of basic statistical and charting features in SPSS.

Help — Tutorial provides access to an introductory SPSS tutorial, including a comprehensive overview of SPSS basics. It is designed to provide a step-by-step guide for carrying out a statistical analysis in SPSS.

Help — Topics opens the Help Topics: SPSS for Windows box, which provides access to Contents, Index, and Find tabs. Under the Contents tab, double-clicking items with a book symbol expands or collapses their contents. The Index tab provides an alphabetical list of topics. Once a topic is selected, or the first few letters of the word are typed in, the Display button provides a description. The Find tab allows for searching the help files for specific words and phrases.

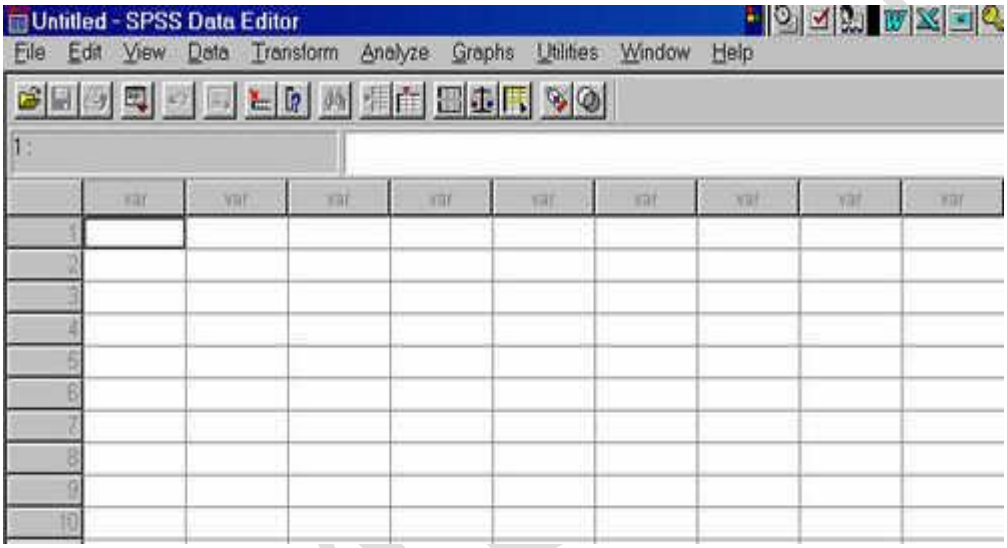
You can save any of the windows. To save a window, make that particular window active. You can activate a window by using the **Window** menu. Once you have activated the desired window, click on **File** and then **Save**. The contents of each of these windows can also be printed by activating a window and clicking on the **Printer** icon (or you can open the **File** menu and then click on **Print**).
SPSS automatically adds a three-letter suffix to the end of the file name (".sav" for data editor files, ".spo" for output files, and ".sps" for syntax files). Thus, with a particular data set, it is useful to keep a single name for the prefix of all three files and let the suffix distinguish among the different file types.
Basic information on these files is summarized below.

Windows	File Extension	Function
Editor	.sav	Used to define, enter and edit data and to run statistical tests.
Output	.spo	Contains the results of the statistical procedures.
Syntax	.sps	This window is activated when we click on the paste function and records a record of the operations that are pasted. SPSS commands can be run from this window.

4.3.1 Entering, Transforming and Selecting Data

The first step in working with SPSS is to enter your data and to create an SPSS data file. Although we will assume that you are typing in your data for the first time, you should be aware that SPSS can also read existing files from other programs such as Excel, Quattro and Lotus 1-2-3.

When you first open up SPSS for Windows, you will be prompted to make a decision about running the tutorial, typing in new data, opening up an existing file, etc. Because you will be creating a new data file, click on **Type in data** and then click **OK**. This should bring you to the initial screen which is the data editor shown below.



Remember that the key to typing in your data is to realize that the data from each case or participant must be typed on a separate line.

Let's understand this using an example. Assume that you are interested in developing a profile of people who use a Momo (A Chinese Dish with soup) kitchen in your city. To do this you collected the following information from a random sample of 50 users of the momo kitchen.

Person	Gender	Age	Number of Siblings	Health Score	Personality Score	Activity Score
1	Male	76	0	16.64	15	-4
2	Female	28	3	60.83	22	4
3	Male	39	0	44.25	18	2
4	Male	47	1	49.13	36	0
5	Female	56	0	30.67	25	-1
6	Female	61	0	29.37	20	-3
..						
..						
50	Male	59	0	35.92	31	2

As you type in your data, you will need to do each of the following.

1. Create a separate line for each case, which in this particular example is each person.
2. Create a column for each variable of interest. In this example, we will use seven columns, one for each of the following variables (person, gender, age, number of siblings, health score, personality score, and activity score). Note that it may not be essential to create a column for a person identification number, but we will do so simply to help us keep track of the data.
3. Develop a numerical code for the gender variable. In this case, we will assign the value of 1 to females and the value of 2 to males.

4.3.2 Creating the Data File

In this section, we will describe the step-by-step procedure for creating your data file.

1. Notice on the bottom left hand corner of the screen that you can access either the **Data View** or **Variable View** windows. To view one window or the other, simply click on it. For now, in establishing the variables and their characteristics, you should work in the **Variable View** window.

There are 10 characteristics to be specified under the columns of the Variable View:
a) Name— the chosen variable name. This can be up to eight alphanumeric characters but must begin with a letter.

b) Type— the type of data. SPSS provides a default variable type once variable values have been entered in a column of the Data View. The type can be changed by highlighting the respective entry in the second column of the Variable View and clicking the three-periods symbol (…) appearing on the right-hand side of the cell. This results in the Variable Type box being opened, which offers a number of types of data including various formats for numerical data, dates, or currencies.

c) Width— the width of the actual data entries. The default width of numerical variable entries is eight which can be increased or decreased.

d) Decimals— the number of digits to the right of the decimal place to be displayed for data entries. This is not relevant for string data and for such variables the entry under the fourth column is given as a greyed-out zero.

DMS.KU

e) Label— a label attached to the variable name. It is generally a good idea to assign variable labels. They are helpful for reminding users of the meaning of variables.

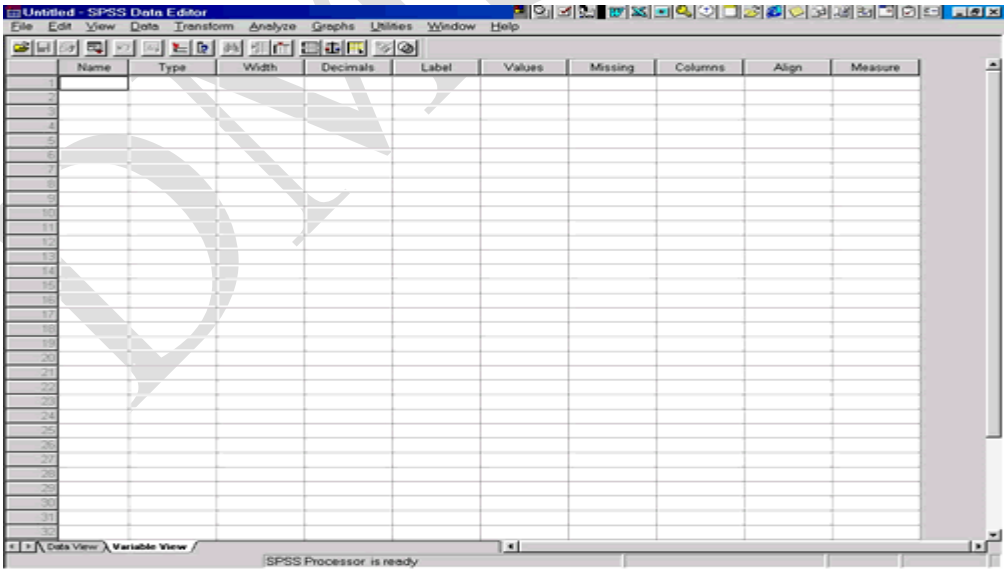
f) Values— labels attached to category codes. For categorical variables, an integer code should be assigned to each category and the variable defined to be of type "numeric."

g) Missing— missing value codes. SPSS recognizes the period symbol as indicating a missing value. If other codes have been used (e.g., 99, 999) these have to be declared to represent missing values.

h) Columns — width of the variable column in the Data View. The default cell width for numerical variables is eight. The cell width can be changed in the same way as the width of the data entries or simply by dragging the relevant column boundary.

i) Align— alignment of variable entries. The SPSS default is to align numerical variables to the right-hand side of a cell and string variables to the left.

j) Measure— measurement scale of the variable. The default chosen by SPSS depends on the data type. For example, for variables of type "numeric," the default measurement scale is a continuous or interval scale (referred to by SPSS as "scale"). For variables of type "string," the default is a nominal scale.



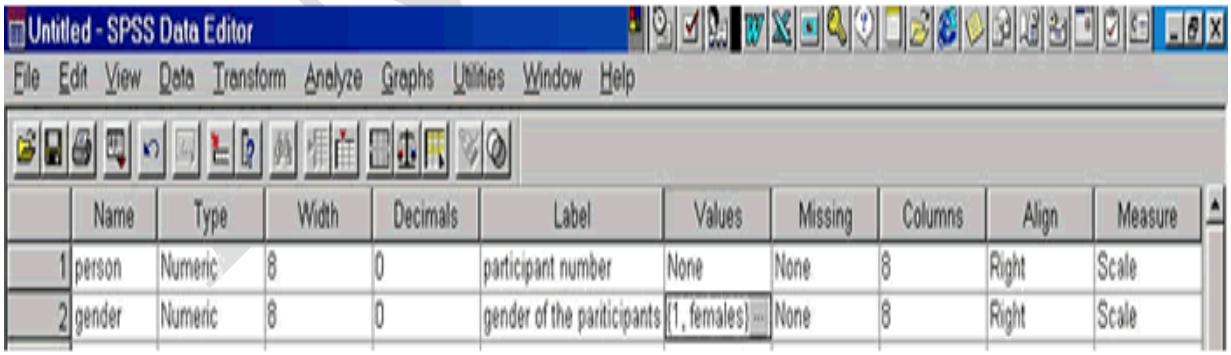
Variable View Window

2. In the space for **Name**, type the desired variable name, which can be of maximum 8 characters in length. The first character must be alphabetic; the remaining characters can be alphabetic and/or numeric, and no spaces can appear in the name. We will type the name "person" in our example.

3. Next press the Tab key and you will notice that SPSS assigns default values to all of the settings. You should go through the relevant column for that particular variable. Clicking on a column will often expose a shaded area. You can click on this to pursue your options.

In our example, because there are no decimal points in our person variable, click on the gray portion following **Numeric** and change the **Decimal Places** to 0. Also, click on **Label** and type in a label like "participant number" in the **Variable Label** slot. Note that we could have exercised other options such as modifying the column width and whether the numbers appear left-justified, right-justified, or centered.

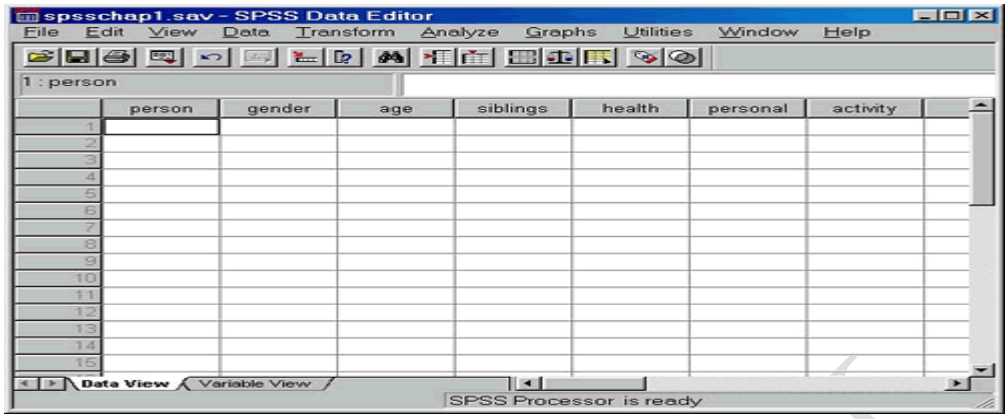
4. Now, you should set up the next variable. Click on line 2 of the **Name** column, on the next **var** column and type "gender". Because we have decided to use the codes of 1 and 2 to represent females and males, respectively, you can click on **Decimals** to change the decimal places to 0. Next, click on **Label** to provide a label for your variable--for example, you might type "gender of the participants." In this case, because we have specified numeric codes for the different values for our variable and we are likely to forget these over time, we should specify **Values**. In the field for **Value**, type a "1," and in the field for **Label**, type a label such as "females." Then create the label of "males" for a value "2." At this point, your window should look like the one below.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	person	Numeric	8	0	participant number	None	None	8	Right	Scale
2	gender	Numeric	8	0	gender of the participants	{1, females}...	None	8	Right	Scale

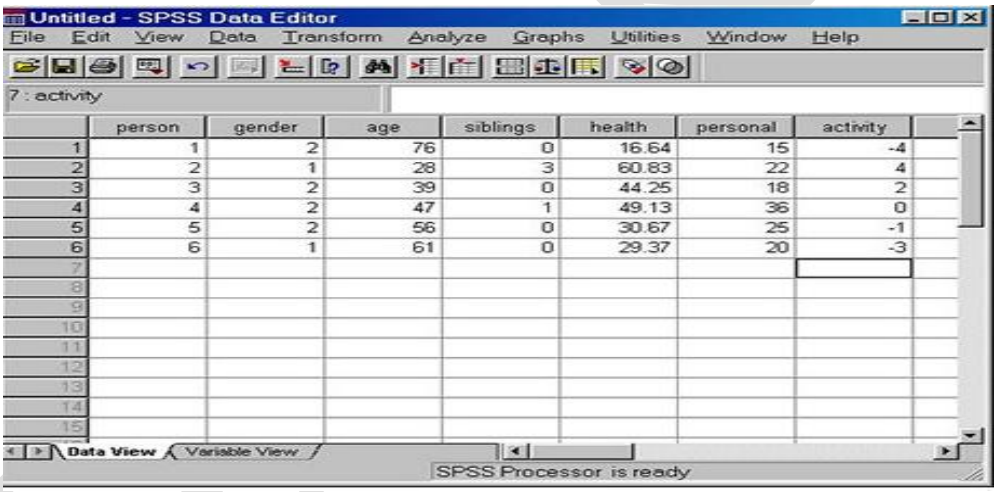
Current Variable View Window

5. You should now define each of the remaining four variables. After you do this, click on the **Data View** window (lower left hand corner). You should see all of the variable names that you have entered like in the figure below.



Data File (Data View)

6. Now, type in the data for the first six persons in our data sheet. Start by clicking in the left-most column of the first line and type the person's number (i.e., "1"), then, press the Tab Key or the right arrow key and type the first person's gender (i.e., "2"). Continue to do this until the data are typed in. Below is listed a copy of the data file that was created while preparing the script.

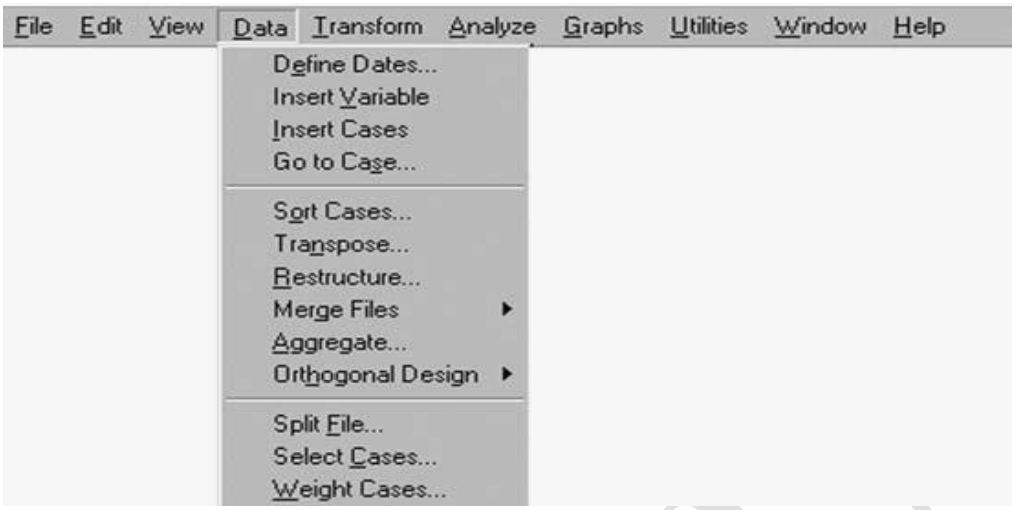


	person	gender	age	siblings	health	personal	activity
1	1	2	76	0	16.64	15	-4
2	2	1	28	3	60.83	22	4
3	3	2	39	0	44.25	18	2
4	4	2	47	1	49.13	36	0
5	5	2	56	0	30.67	25	-1
6	6	1	61	0	29.37	20	-3

Data file Containing the First Six Lines of the Data

7. When you are satisfied with your data file, you should save it. Click on **File** and **Save** and type in a file name (e.g., "momo"). Note that SPSS automatically adds the ". sav" suffix to your file name. This is the SPSS suffix that is used to designate data files.

The data file as displayed in the Data View spreadsheet is not always organized in the appropriate format for a particular use. The Data dropdown menu provides procedures for reorganizing the structure of a data file (see figure below).



The first four command options from the Data drop-down menu are concerned with editing or moving within the Data View spreadsheet. Date formats can be defined or variables or cases inserted. The following set of procedures allows the format of a data file to be changed:

1. **Sort Cases…** opens a dialogue box that allows sorting of cases (rows) in the spreadsheet according to the values of one or more variables. Cases can be arranged in ascending or descending order.

2. **Transpose…** opens a dialogue for swapping the rows and columns of the data spreadsheet. The Variable(s) list contains the columns to be transposed into rows and an ID variable can be supplied as the Name Variable to name the columns of the new transposed spreadsheet.

3. **Restructure…** calls the **Restructure Data Wizard**, a series of dialogue boxes for converting data spreadsheets between what are known as "long" and "wide" formats. These formats are relevant in the analysis of repeated measures.

4. **Merge files** allow either **Add Cases…** or **Add Variables…** to an existing data file. A dialogue box appears that allows opening a second data file. This file can either contain the same variables but different cases (to add cases) or different variables but the same cases (to add variables).

5. **Aggregate…** combines groups of rows (cases) into single summary rows and creates a new aggregated data file. The grouping variables are supplied under the Break Variable(s) list of the Aggregate Data dialogue box and the variables to be aggregated under the Aggregate Variable(s) list. The Function… sub-dialogue box allows for the aggregation function of each

variable to be chosen from a number of possibilities (mean, median, value of first case, number of cases, etc.).

Finally, the **Split File…**, **Select Cases…**, and **Weight Cases…** procedures allow using the data file in a particular format without changing its appearance in the Data View spreadsheet.

4.3.3 Computing Means and Standard Deviations

Once you have typed in your data, performing statistical procedures is relatively simple. To give you a sense of how to do so, as well as to expose you to some of the powerful SPSS tools for analyzing data, several examples will be dealt with. In the first example, you will compute the mean and standard deviations for each of the variables from our momo kitchen study. In the second, you will do this separately for males and females. Finally, in a third example, you will compute a new variable that is a composite of the Health Activity and Personality Scores and compute the mean and standard deviation for this measure.

4.3.3.1 Computing the Mean and Standard Deviation for All Scores

1. Click on **Analyze** then **Descriptive Statistics**, then **Descriptives**.

2. Highlight each of the variables for which you are interested in computing descriptive statistics (e.g., Age, Number of Siblings, Health Score, Personality Score, and Activity Score) and move them into the **Variable(s)** column. Note that you can move all of these variables over at one time by clicking and dragging over the items that you want to select.

3. Click on **Options** and select the desired statistics. At minimum, you should select the **Mean** and **Std. deviation**. Now, click on **Continue** and then **OK**.

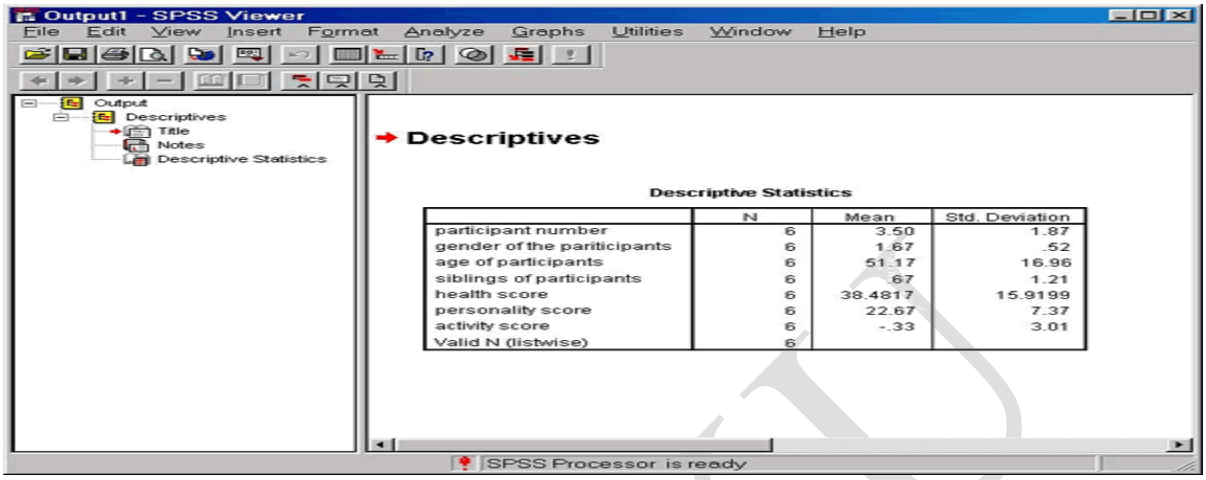
4. Your output should look like the one shown below. You may have fewer or more statistics depending on your selection in the **Options** menu. Note that the variable labels appear on your printout.

5. If you would like a hard copy of this output, you can print it by clicking on the print icon on the tool bar. Also, you can save the output by clicking on the disk

DMS.KU

icon. Note that SPSS automatically adds the ".spo" suffix to your output file name.

You can also make use of Menus or shortcut keys to perform such functions.



Output from the Analysis

4.3.3.2Computing Means and Standard Deviations for Males and Females Separately

From the output screen, you cannot get to all of the options that you might be interested in. If the option that you are interested in is not available, click on **Window**, and then **SPSS Data Editor** to return to the **Data Editor** window. You will need to do this to do the exercises like splitting data.

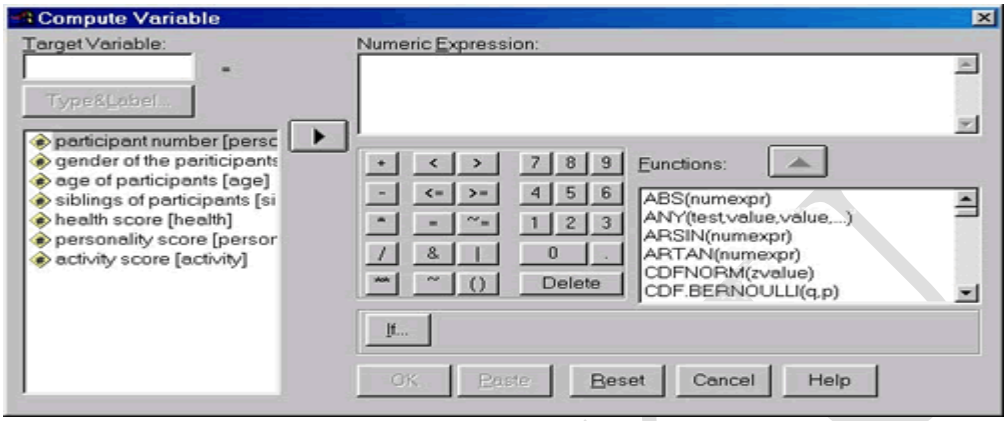
1. First click on **Data** and then **Split File**. This allows you to split the file according to a particular variable and conduct separate analyses for each level of the variable.
2. Next you should select **Organize output by groups** and move **Gender** from the variable list to the **Groups Based on** list. When you have done this, click on **OK**.
3. Now, click on **Analyze**, **Descriptive Statistics**, and **Descriptives**. Select the desired variable(s) and **Options**.
4. Note in your output that two sets of data summaries are presented--one for females and one for males.

4.3.3.3 Computing a New Variable and Then Performing Descriptive Statistics

Assume that in addition to our interest in calculating the means and standard deviations for the collected variables, we wish to compute a composite score that roughly represents some overall measure of physical and mental health. Specifically, assume that we wish to compute the mean and standard deviation for a new variable called overall health that represents the average of each person's

Health Score and Personality Score. You can do this by performing the following steps.

1. Click on **Transform** and then **Compute**. This should produce the screen shown below.



Compute Variable Screen

2. Type the name of the variable that you wish to create ("ovhealth" here) in the **Target Variable** field. Now, you need to type in the computation that you wish to have performed in the **Numeric Expression** field. You can use all the operations listed on the bottom of this screen, and it is important to realize that operations within parentheses are performed first. Thus, if we wish to compute the mean of the Health and Personality Scores, we need to add these together before we divide by 2. To do so, simply type (or move over) variable names for the Health and Personality Scores and enclose these within parentheses. Next, click on the / button and follow this with the number 2. Your target variable and your numeric expression fields might look like the following:

Target Variable	**Numeric Expression**
ovhealth =	(health + personal)/2

3. Now you should click on the **Type&Label** field. This will enable you to create a longer label for your variable and modify the type and width of the variable. Once you have done this, press **Continue**. Then press **OK**.

4. Once you have done this, notice that the new variable appears in your data file. Now you are ready to calculate the mean and standard deviation for this new variable by using the procedures outlined above.

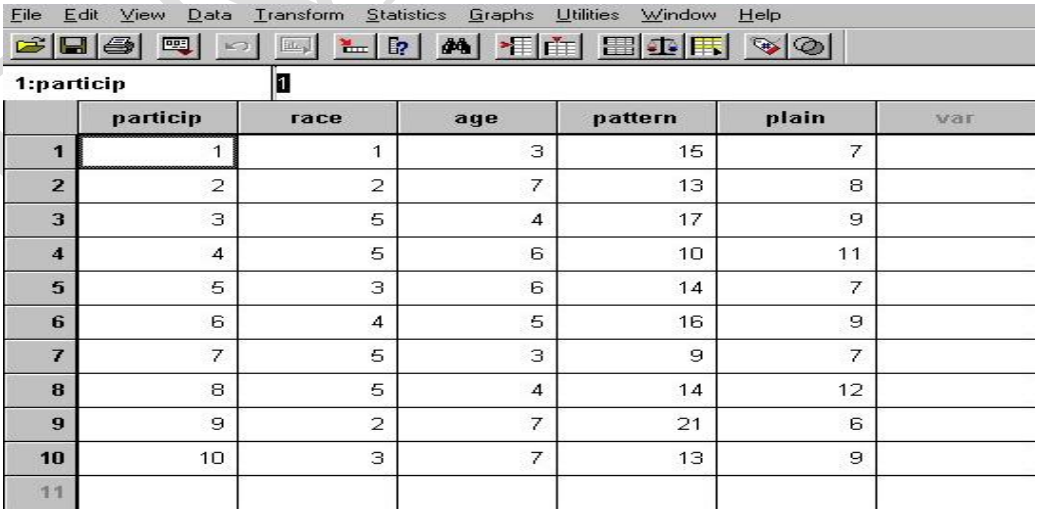
4.3.3.4 Another Example of Creating a Data File for a Data Set

To gain more confidence in creating a data file and computing descriptive statistics, work through the following example. Assume that you have conducted

an experiment to determine if newborn infants prefer patterned or plain stimuli. In the study, 10 infants were presented with both plain and patterned figures. The amount of time spent in looking at each type of stimulus was measured. In addition, the race of the infants and the age (in days) were recorded. The data sheet for this experiment appears below.

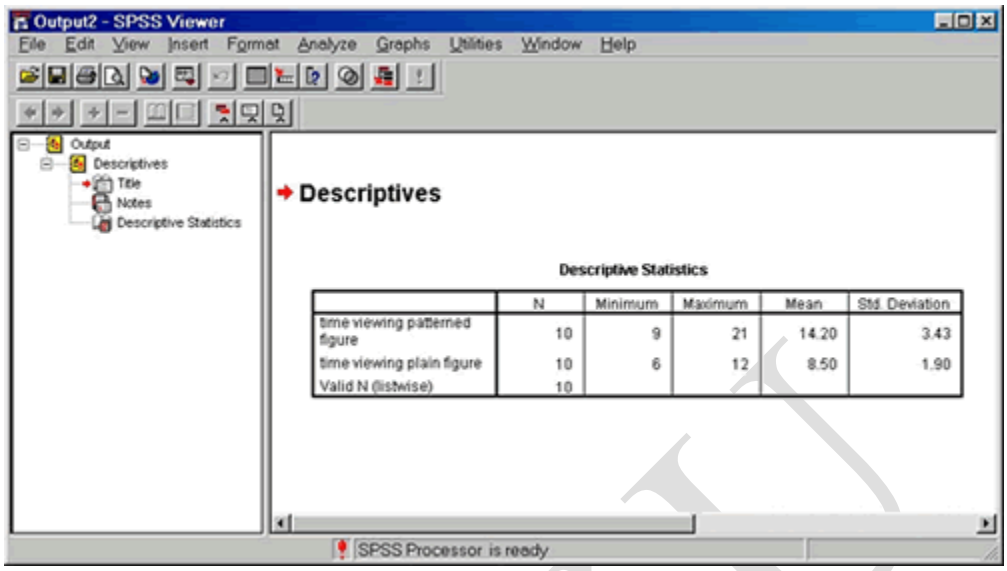
Participant	Race	Age	Time Viewing Patterned Figure	Time Viewing Plain Figure
1	Am. Indian	3	15	7
2	Asian	7	13	8
3	White	4	17	9
4	White	6	10	11
5	Afr. Amer.	6	14	7
6	Hispanic	5	16	9
7	White	3	9	7
8	White	4	14	12
9	Asian	7	21	6
10	Afr. Amer.	7	13	9

Below is one possible way to create your data file. Note that you have to create a numeric code for the race variable and in this case we used 1 = American Indian, 2 = Asian, 3 = African American, 4 = Hispanic, and 5 = White. Once you have typed the data file, compare it with the one below. Then, try to compute means and standard deviations for the time viewing the patterned figure and the time viewing the plain figure variables. The results should look like those in the output file below.



	particip	race	age	pattern	plain	var
1	1	1	3	15	7	
2	2	2	7	13	8	
3	3	5	4	17	9	
4	4	5	6	10	11	
5	5	3	6	14	7	
6	6	4	5	16	9	
7	7	5	3	9	7	
8	8	5	4	14	12	
9	9	2	7	21	6	
10	10	3	7	13	9	
11						

Data File for the Data



Output file

4.4 Computing a T-Test

T-Test is used to examine the effects of one independent variable on one or more dependent variables and is restricted to comparisons of two conditions or groups (two levels of the independent variable). The results of this test enable you to determine if two means differ significantly. Two basic experimental designs, **between-subjects** and **within-subjects** designs can be analyzed with a t-test.

4.4.1 T-Test for Between-Subjects Designs

This section will describe how to analyze the results of between-subjects designs. It is important to know that the **Independent Samples T-Test** is used with between-subjects designs and the **Paired Samples T-Test** is used with within-subjects designs. So you will be using the Independent Sample T-Test here.

A two-group between-subjects design is one in which participants have been randomly assigned to the two levels of the independent variable. In this design, each participant is assigned to only one group, and consequently, the two groups are independent of one another. For example, assume that you are interested in studying the effects of two types of drugs (X, Y) on reaction time. If you randomly assign some participants to the Drug X group and other participants to the Drug Y group, then you are using a between-subjects design. (Remember that in a within-subjects design, all participants would receive both levels of the drug.)

An Example: Parental Involvement Experiment

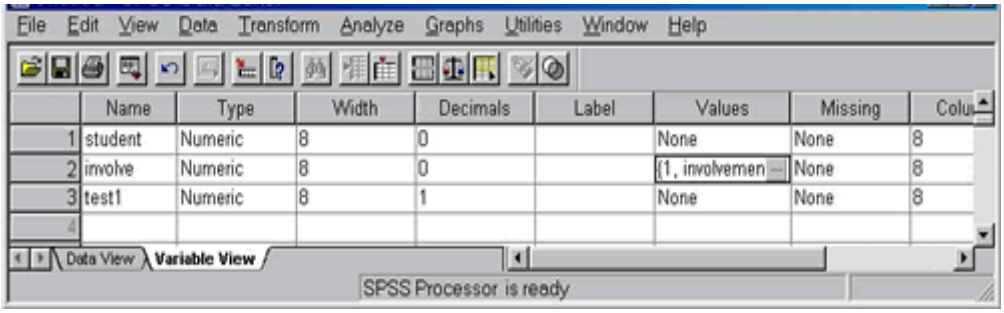
Assume that you studied the effects of parental involvement (independent variable) on students' grades (dependent variable). Half of the students in a third grade class were randomly assigned to the parental involvement group. The teacher contacted the parents of these children throughout the year and told them

about the educational objectives of the class. Further, the teacher gave the parents specific methods for encouraging their children's educational activities. The other half of the students in the class were assigned to the no-parental involvement group. The scores on the first test were tabulated for all the children, and these are presented below.

Student	Parental Involvement Condition	Test 1
01	Involvement	78.6
02	Involvement	64.9
03	Involvement	100.0
04	Involvement	83.7
05	Involvement	94.0
06	Involvement	78.2
07	Involvement	76.9
08	Involvement	82.0
09	No involvement	81.0
10	No involvement	69.5
11	No involvement	73.8
12	No involvement	66.7
13	No involvement	54.8
14	No involvement	69.3
15	No involvement	73.5
16	No involvement	79.4

Creating Your Data File

There is a key point to keep in mind when creating a data file for an independent samples t-test. That is, that you must create a column for your independent variable condition. In this case, that is the parental involvement condition, and you should create a numeric code that allows SPSS to know the parental involvement condition that the score is in. So, the first part of your data file might look like the one below, with three variables--one for student number, one for parental involvement condition (using a code of "1" for involvement and "2" for no involvement), and score on Test1. Remember, that in creating the data file, you should create a variable **Label** for each variable and **Value** label for the parental involvement variable. Your variable view file should look something like the one below.



Variable View File for the Parental Involvement Experiment

After creating the Data file, you need to follow the below given steps to compute the T-Test for the data:

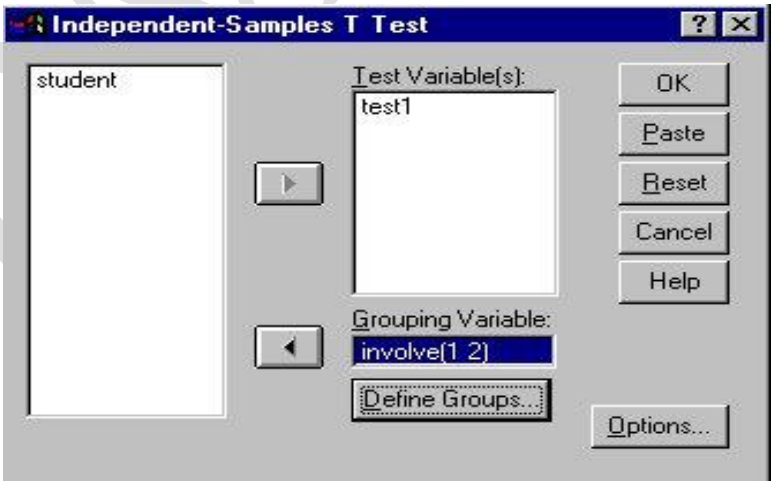
1. Click on **Analyze**, then **Compare Means**, then **Independent Samples T-Test**.

2. Now, move the dependent variable (in this case, labeled "test1") into the **Test Variable** field.

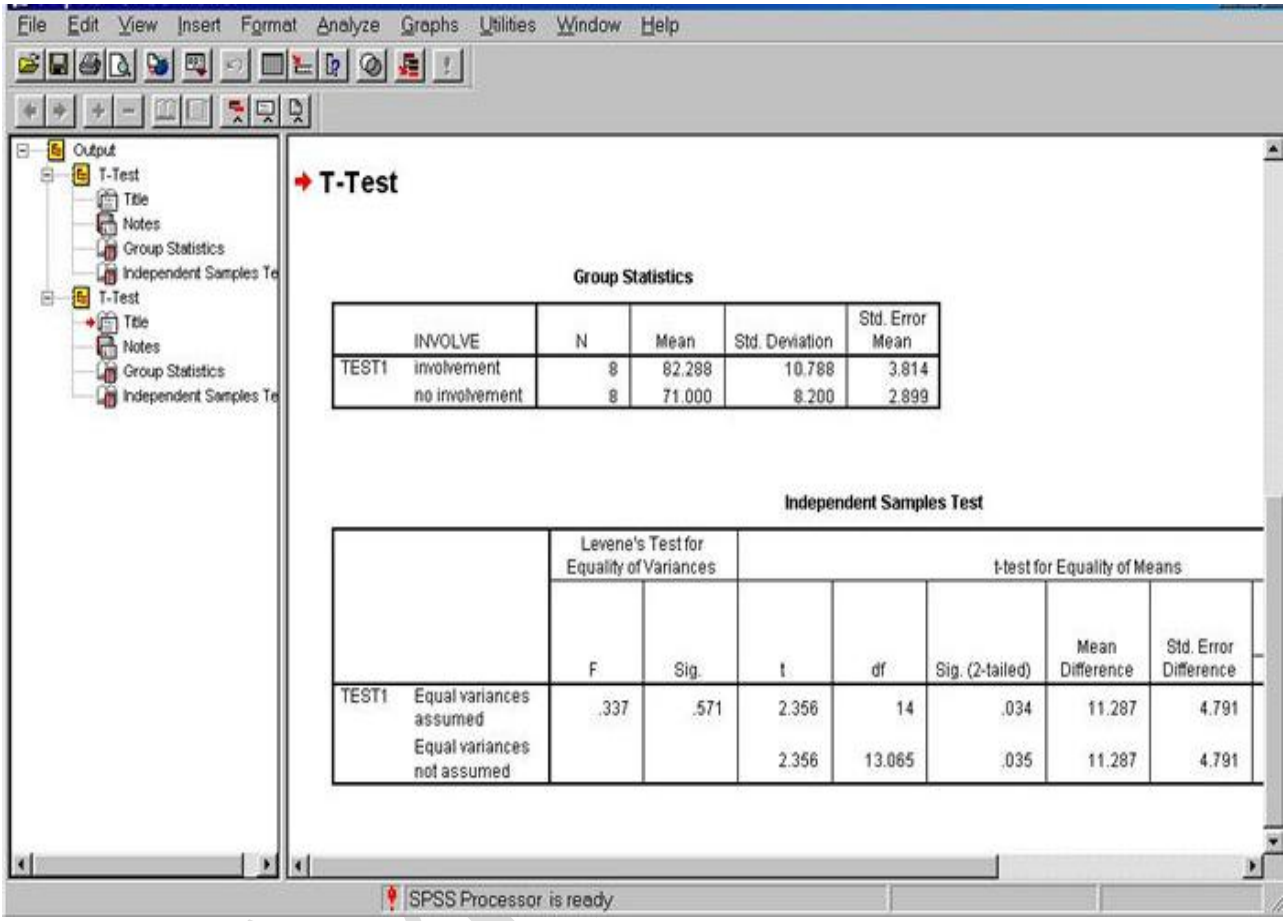
3. Move your independent variable (in this case, "involve") into the **Grouping Variable** field. You should be aware that **Grouping Variable** stands for your independent variable.

4. You will notice that there are question marks in the parentheses following your independent variable in the **Grouping Variable** field. This is because you need to define the particular groups that you want to compare.
To do so, click on **Define Groups,** and indicate the numeric values that each group represents. In this case, you will want to put a "1" in the field labeled **Group 1** and a "2" in the field labeled **Group 2**. Once you have done this, click on **Continue**. Your independent-samples t-test screen should look like that below.

5. Now click on **OK** to run the t-test. You may also want to click on **Paste** in order to create a record of what you have done.
The output from the above carried out t-test procedure is shown below. The output from the test is then described below the Screen.



Independent Samples t-test Figure



Output from Independent Samples t-test

1. The first table lists the number of participants (N), mean, standard deviation, and standard error of the mean for both the groups. Notice that the value labels are printed as well as the variable labels for your variables.

2. The second table initially presents you with an F-test (Levine's test for equality of variances) that evaluates the basic assumption of the t-test that the variances of the two groups are approximately equal (homogeneity of variance). If the F value reported here is very high and the significance level is very low--usually lower than .05 or .01), then the assumption of homogeneity of variance has been violated. If this is the case, you should use the t-test in the lower half of the table, whereas if you have not violated the homogeneity assumption, you should use the t-test in the upper half of the table.

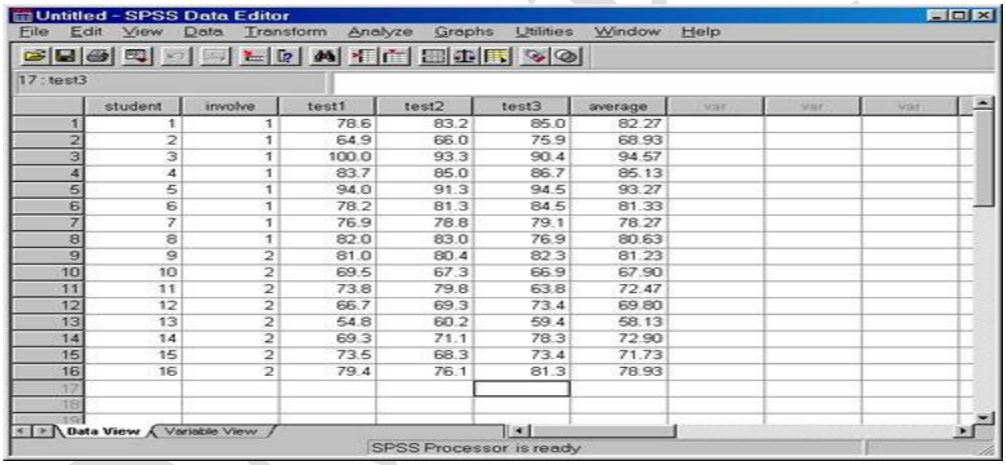
3. In this particular case, you can see that we have not violated the homogeneity assumption, and we should use the t of 2.356, degrees of freedom of 14, and the

significance level of .034. Thus, our data show that parental involvement has a significant effect on grades, $t(14) = 2.356$, $p < .05$.

Note: Students from Commerce Background, as in the case of this script can take examples from their own field as e.g. Effect of using IT on Stock Markets, Effect of using ATM's in Banks on Customer Satisfaction and the like.

4.4.2 Gaining Confidence: Extending the Parental Involvement Experiment

Assume that the course had three tests and you wished to examine the effects of parental involvement on all three tests as well as a final term average. So, to do this, assume that the test score you already have in your data file is the score from the first test. Add the Test 2 and Test 3 scores (shown below in the data file) for each of the 16 students. Once you have done this, try to get SPSS to perform four t-tests--one for each of the three term grades and one for the final grade average. (Note that you can get SPSS to calculate the term average with the **Transform Compute** menu.)



	student	involve	test1	test2	test3	average
1	1	1	78.6	83.2	85.0	82.27
2	2	1	64.9	66.0	75.9	68.93
3	3	1	100.0	93.3	90.4	94.57
4	4	1	83.7	85.0	86.7	85.13
5	5	1	94.0	91.3	94.5	93.27
6	6	1	78.2	81.3	84.5	81.33
7	7	1	76.9	78.8	79.1	78.27
8	8	1	82.0	83.0	76.9	80.63
9	9	2	81.0	80.4	82.3	81.23
10	10	2	69.5	67.3	66.9	67.90
11	11	2	73.8	79.8	63.8	72.47
12	12	2	66.7	69.3	73.4	69.80
13	13	2	54.8	60.2	59.4	58.13
14	14	2	69.3	71.1	78.3	72.90
15	15	2	73.5	68.3	73.4	71.73
16	16	2	79.4	76.1	81.3	78.93
17						
18						

If you have problems working with this example, you should look at the steps below.

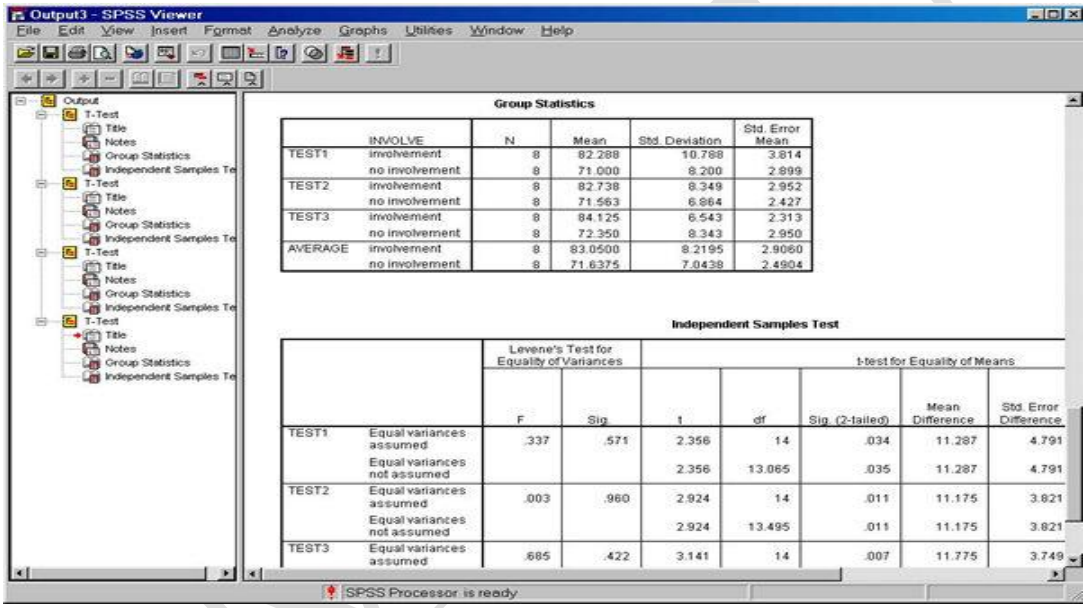
1. To create a new variable that represents the average of all three tests, click on **Transform**, then **Compute**. Type in the name of the variable that you wish to create (e.g., "average") in the **Target Variable** field. In the **Numeric Expression** field, type (or click on the appropriate characters) the expression that represents the average. In this case, you might type the following expression

Target Variable	**Numeric Expression**
average	(test1 + test2 + test3)/3

Note that the three test scores are included within parentheses. This is necessary because SPSS first performs operations that are within parentheses, and that we

want to add all the numbers before dividing. Make sure that you also create a label for your new variable. Once you have created the proper expression, click on **OK**, and this should take you to the SPSS data editor where you should see a new column that represents the average of the three test scores.

2. Now click on **Statistics**, then **Compare Means**, then **Independent Samples t-test**. You should then move the four dependent variables (test1, test2, test3, average) into the **Test Variable** field. Next, move the independent variable (i.e., involve) into the **Grouping Variable** field. Next, click on **Define Groups** and indicate the two levels of your involve variable. When you are finished, click on **OK**. The first portion of your output should look like that below.



Group Statistics

	INVOLVE	N	Mean	Std. Deviation	Std. Error Mean
TEST1	involvement	8	82.288	10.788	3.814
	no involvement	8	71.000	8.200	2.899
TEST2	involvement	8	82.738	8.349	2.952
	no involvement	8	71.563	6.864	2.427
TEST3	involvement	8	84.125	6.543	2.313
	no involvement	8	72.350	8.343	2.950
AVERAGE	involvement	8	83.0500	8.2195	2.9060
	no involvement	8	71.6375	7.0438	2.4904

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
TEST1	Equal variances assumed	.337	.571	2.356	14	.034	11.287	4.791
	Equal variances not assumed			2.356	13.065	.035	11.287	4.791
TEST2	Equal variances assumed	.003	.960	2.924	14	.011	11.175	3.821
	Equal variances not assumed			2.924	13.495	.011	11.175	3.821
TEST3	Equal variances assumed	.685	.422	3.141	14	.007	11.775	3.749

Output for the Average Variable

4.5 Analysis of Variance for One-Independent Variable Designs

This Section describes how to compute the analysis of variance for between-subjects designs that contain only one independent variable. This analysis is used to determine if two or more group Means differ significantly. In the case of two-group designs, it yields the same probability level as the T-Test procedure. An F ratio is the statistic that is calculated and a significant F indicates that at least two-group Means differ.

Example: Preparation Hours and Scholastic Aptitude Test (SAT) Performance

As an educational researcher, assume that you conducted an experiment to test whether preparation hours/courses (independent variable) affect Scholastic

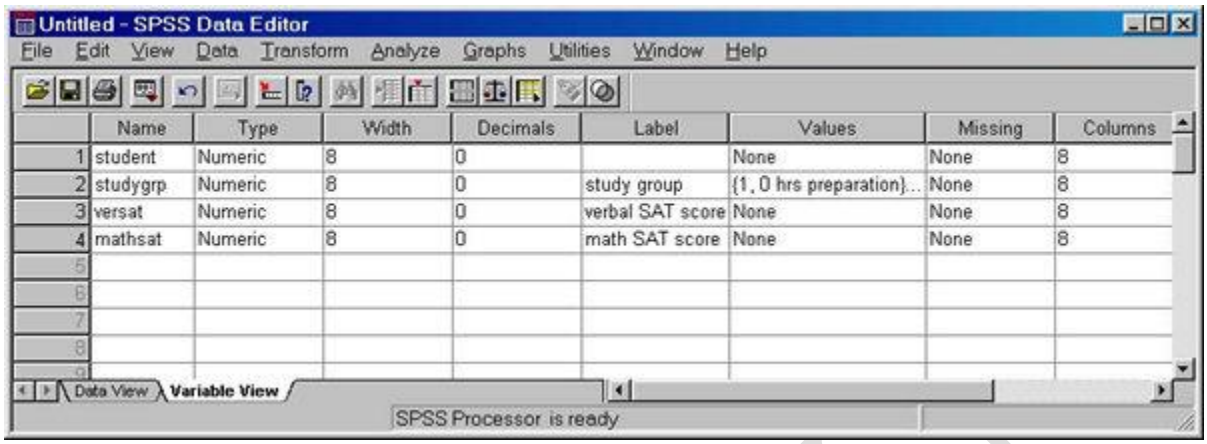
Aptitude Test (SAT) performance (dependent variable). Specifically, 5 high school students were randomly assigned to each of three study conditions:

1. 0 hours of coursework preparation.
2. 20 hours of coursework preparation
3. 40 hours of coursework preparation

After all the coursework was completed, all 15 students were given the SAT and their scores on the verbal and math sections are listed below.

Student	Study Group (hours)	Verbal SAT Score	Math SAT Score
1	0	350	500
2	0	480	520
3	0	395	450
4	0	440	510
5	0	385	470
6	20	500	560
7	20	490	540
8	20	560	580
9	20	495	590
10	20	495	550
11	40	550	610
12	40	590	630
13	40	580	620
14	40	550	590
15	40	620	580

You should now create a data file that represents these data. Remember that you need to put each participant's data on one line and that you need to create a variable for the study group. You might create a data file that looks like the one below. You will also need to create a numeric code for your independent variable. In this case, we can assign 1 to 0 hours of coursework, 2 to 20 hours of coursework, and 3 to 40 hours of coursework.



Variable View File for the SAT Experiment

	student	studygrp	verbsat	mathsat	var	var
1	1	1	350	500		
2	2	1	480	520		
3	3	1	395	450		
4	4	1	440	510		
5	5	1	385	470		
6	6	2	500	560		
7	7	2	490	540		
8	8	2	560	580		
9	9	2	495	590		
10	10	2	495	550		
11	11	3	550	610		
12	12	3	590	630		
13	13	3	580	620		
14	14	3	550	590		
15	15	3	620	580		

Data File for the SAT Experiment

Performing the Analysis

There are several ways to compute the analysis of variance with SPSS. We will use a procedure called **One-Way ANOVA.**

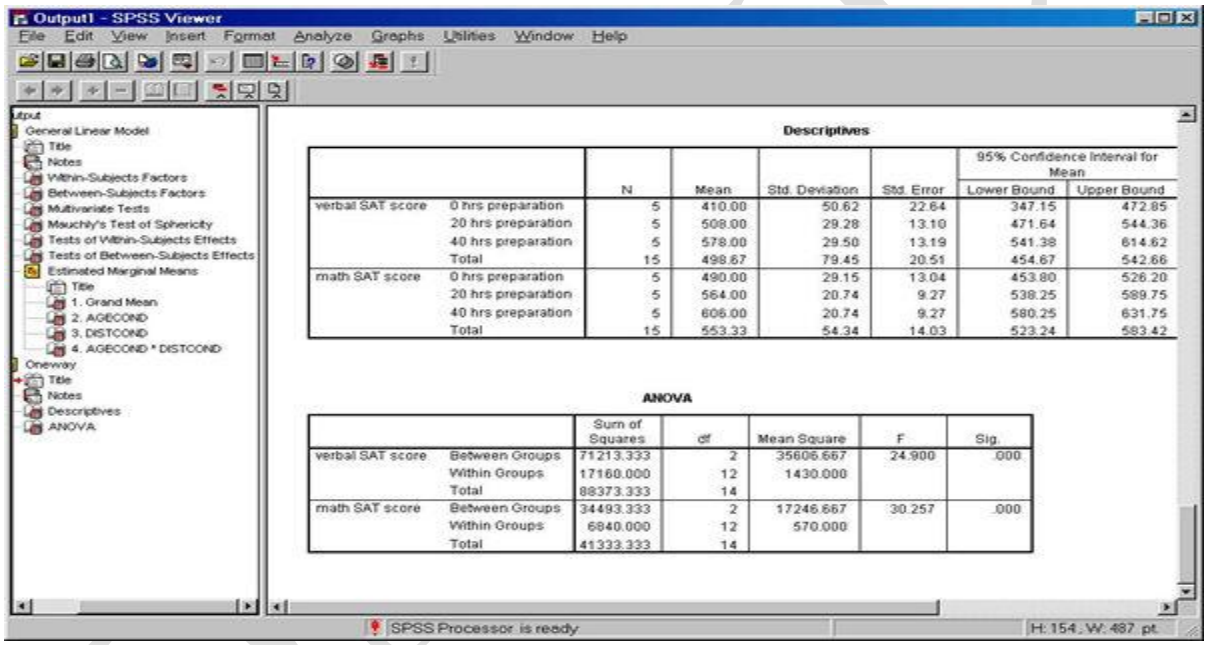
1. Click on **Analyze**, then **Compare Means**, then **One-Way ANOVA**.

2. Move the dependent measure that you wish to analyze into the **Dependent List** field. Let's analyze the data for the verbal SAT and math SAT, so move both "verbsat" and "mathsat" into the field. Next, you should move the independent variable into the **Factor** field. In our case, this is the "studygrp" variable. Now click on **Options** and select **Descriptive** statistics. When you are done, click on **Continue** and then on **OK**.

3. You may also want to compute a total SAT score (using the **Transform** and **Compute** menu) and then perform an analysis of variance on this total score.

Examining the Output

You will notice that SPSS provides you with several types of information in the output for the analysis of variance.



Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
						Lower Bound	Upper Bound
verbal SAT score	0 hrs preparation	5	410.00	50.62	22.64	347.15	472.85
	20 hrs preparation	5	508.00	29.28	13.10	471.64	544.36
	40 hrs preparation	5	578.00	29.50	13.19	541.38	614.62
	Total	15	498.67	79.45	20.51	454.67	542.66
math SAT score	0 hrs preparation	5	490.00	29.15	13.04	453.80	526.20
	20 hrs preparation	5	564.00	20.74	9.27	538.25	589.75
	40 hrs preparation	5	606.00	20.74	9.27	580.25	631.75
	Total	15	553.33	54.34	14.03	523.24	583.42

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
verbal SAT score	Between Groups	71213.333	2	35606.667	24.900	.000
	Within Groups	17160.000	12	1430.000		
	Total	88373.333	14			
math SAT score	Between Groups	34493.333	2	17246.667	30.257	.000
	Within Groups	6840.000	12	570.000		
	Total	41333.333	14			

ANOVA Output

1. First, you are presented with the F statistic for each analysis that you requested. In this case, you requested one analysis of variance for the verbal SAT measure and one for the math SAT measure. If your F is large and your significance level is low (usually less than .05 or .01), then you can conclude that your results were not due to chance. In this case, course preparation had a significant influence on verbal SAT scores, $F(2, 12) = 24.90$, $p < .05$. Note that another name for the Mean Square within Groups is the mean square of the Error Term.
2. Because we asked for descriptive statistics, the mean and number of valid participants in each condition is presented.

4.5.1 Creating Tables for Your Data

If you want SPSS to list descriptive information about your groups in tabular form, you can do this through the **Tables** command. Assume that we would like to have the mean, standard deviation, and standard error of the mean for verbal scores for each of our three groups.

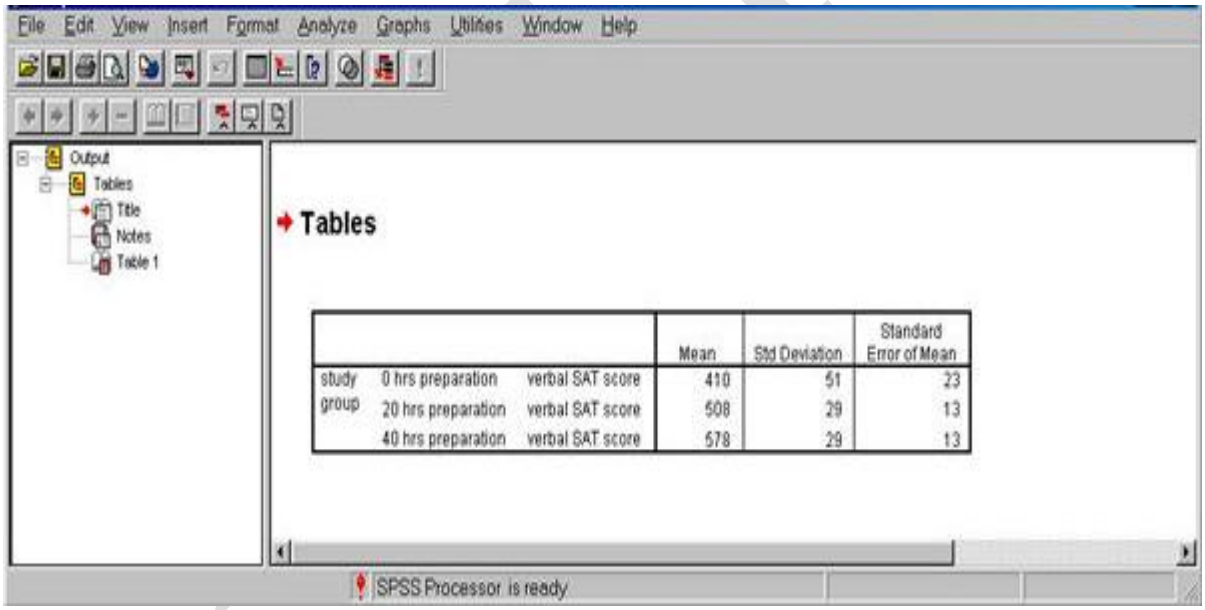
1. Click on **Analyze**, then **Custom Tables**, then **Basic Tables**.

2. Move the dependent variable measure ("verbsat") into the **Summaries** field.

3. Move the independent variable ("studygrp") into the **Subgroups Down** field.

4. Click on **Statistics** and **Add** the statistics that you would like to see. For this example, calculate the mean, standard deviation, and standard error of the mean (**S. E. Mean**), and any other statistics that you would like to see.

5. Click on **Continue** and then **OK**. The statistics should then appear in a table like the one shown below. Note that the value labels appear in your output (that is, if you specify the value label when entering the variables of data).



Showing a Table

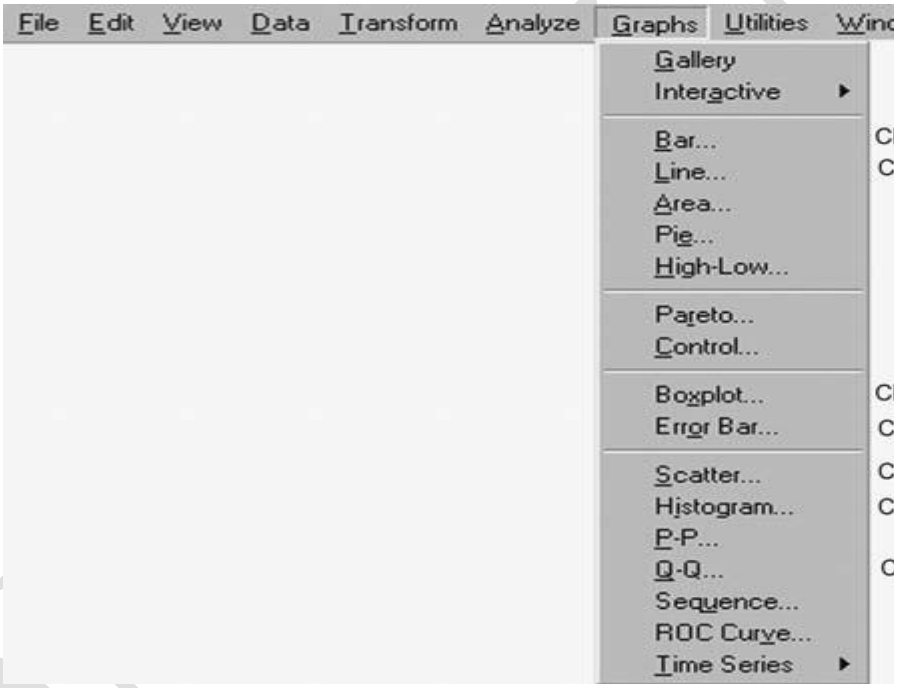
4.6 Charting Data in SPSS

The data that we analyze in SPSS can be charted in various ways to present in a better way. This section gives a brief introduction of how charts can be created in SPSS.

4.6.1 Creating a Simple Bar Graph

SPSS has a variety of graphs that you can use to display your data. You can make a simple bar graph of your data from this experiment by following the steps listed below.

1. Click on **Graphs**, then **Bar** as shown below.



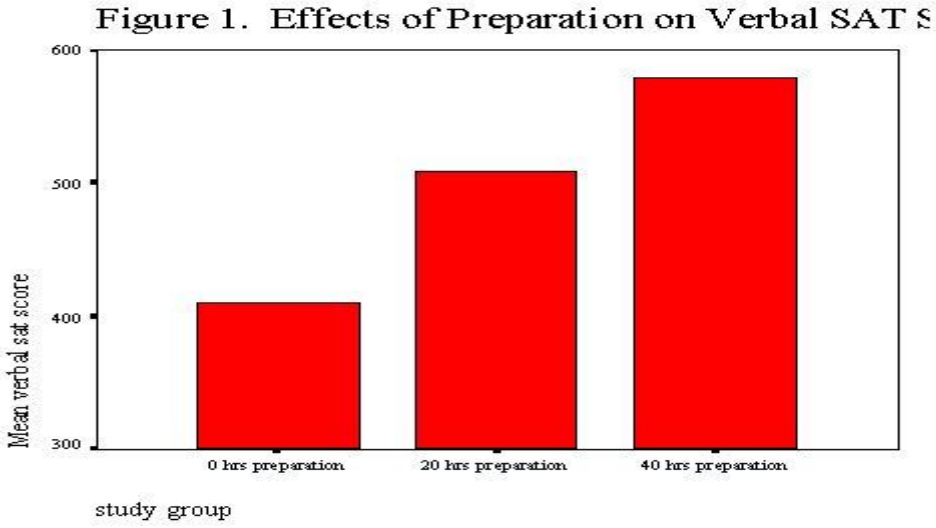
2. In the **Bar Charts** menu, click on **Simple** and then **Define**.

3. In the **Define Simple Bar** menu, first select **Other** summary function (this enables you to plot the means of your groups) and then select your dependent variable measure ("verbsat") and move it into the **Variable** field. The default will be to compute the mean for that variable, but you can modify this (to select the median, mode, or other summary statistics) by clicking **Change Summary**.

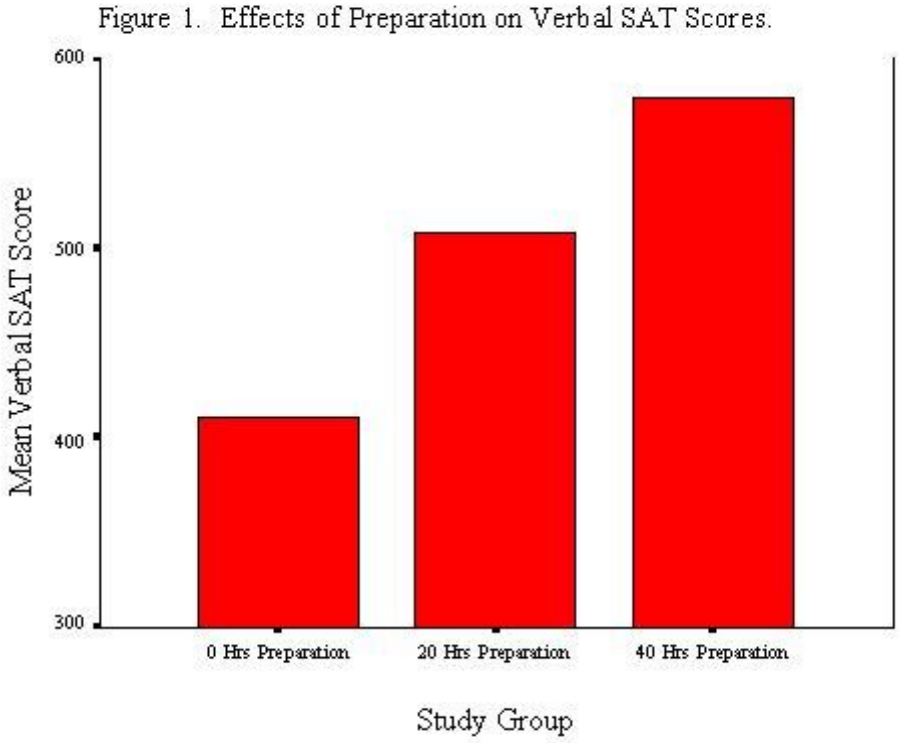
4. Now move the independent variable ("studygrp") into the **Category Axis** field.

5. If you wish to create a title for your figure, click on **Titles,** and type a title.

6. When you are ready, click on **OK**, and a simple Bar chart as shown will be created



7. SPSS uses default options in creating your graph and you can modify it by double clicking on your graph. This will access the **SPSS Chart Editor**. In the graph below, you can see that the title does not fit with the default font size. Also, you may want to change the font, size, and location of the X- and Y-axis labels. With a bit of editing, you can get it to look like the one below.



8. You should note that when you are in the **SPSS Chart Editor**, you can save the settings that you created in your graph by clicking on **File** and then **Save Chart Template**. Doing this will save your preferences for the various graph options (e.g., font size, centering vs. left or right justification of labels) and will enable you to apply these to future graphs.

Note: Similarly you can create other types of charts for your data as demanded by the situation and you should create a chart type that better suits your data.

4.7 Online Submission of Tax Returns

Internet has facilitated the life of people by offering great many services, one of the important being Electronic-Payments. E-Payment facilitates payment of direct taxes online by taxpayers.

The Service Providers

To avail of this facility the taxpayer is required to have a net-banking account with any of the banks listed below, which are the only banks offering this facility at present.

1. Axis Bank
2. State Bank of India
3. Punjab National Bank
4. Indian Overseas Bank
5. Canara Bank
6. Indian Bank
7. Bank of India
8. Corporation Bank
9. State Bank of Bikaner & Jaipur
10. State Bank of Travancore
11. State Bank of Indore
12. Vijaya Bank
13. HDFC Bank
14. Oriental Bank of Commerce

DMS,KU

15. State Bank of Patiala

16. Bank of Baroda

17. IDBI Bank

18. State Bank of Mysore

19. Bank of Maharashtra

20. State Bank of Hyderabad

21. Union Bank of India

22. Allahabad Bank

23. Dena Bank

24. Syndicate Bank

25. ICICI Bank

26. United Bank of India

27. UCO Bank

28. Central Bank of India

29. Andhra Bank

30. J and K Bank

4.7.2 Procedure for e-payment

1. To pay taxes online the taxpayer will select the relevant challan i.e. ITNS 280, ITNS 281, ITNS 282 or ITNS 283, as applicable.

2. Enter its PAN / TAN as applicable. There will be an online check on the validity of the PAN / TAN entered.

3. If PAN/ TAN is valid the taxpayer will be allowed to fill up other challan details like accounting head under which payment is made, name and address of TAN and also select the bank through which payment is to be made, etc.

4. On submission of data entered a confirmation screen will be displayed. If the taxpayer confirms the data entered in the challan, it will be directed to the net-banking site of the bank.

5. The taxpayer will login to the net-banking site with the user id/ password provided by the bank for net-banking purpose and enter payment details at the bank site.

DMS,KU

6. On successful payment a challan counterfoil will be displayed containing CIN, payment details and bank name through which e-payment has been made. This counterfoil is proof of payment being made.

4.7 Exercise

Q1. Fill in the Blanks

a) SPSS stands for_____

b) SPSS is mainly used for Data Analysis in Social and _____Sciences

c) SPSS has two modes _____and _____

d) In SPSS we come across _____ kinds of windows

e) The three windows that we use in SPSS are _____, _____and _____

f) The window used to define and enter user data is _____

g) The results of the statistical tests appear in the _____ Window

h) The window used to keep record of the operations on data is _____

i) The extension for output files is _____ and for syntax files is _____

j) The file name extension for data editor files is _____

Q2. Answer the following:

1. What is SPSS. What are its advantages?
2. Write down steps for starting SPSS in different ways?
3. Write functions of different windows used in SPSS?
4. How can you get help in SPSS if you need it?
5. What are the different types of Files in SPSS. How can we identify them?
6. Create a Data File for recording the use of Mobile Phones in different age groups. Assign appropriate codes where needed and then analyze the data and come with some conclusion?
7. Create a suitable Chart for the data in exercise 6 above?

DMS.KU

8. Compute a New variable in 6 above to show overall Mobile Phone usage in all the age groups?
9. Explore all menus provided in SPSS package?
10. Try to earn more and more statistical techniques and implement them in SPSS?

4.8 Suggested Readings

1. SPSS for Beginners, Vijay Gupta, VJbooks Inc.
2. A Handbook of Statistical Analysis Using SPSS, Landau & Everitt, Chapman and Hall/ CRC London
3. SPSS for Beginners: Advanced Methods, Vijay Gupta, VJbooks Inc.
4. SPSS user Manual

DMS,KU

Answers to Exercises:

Unit-I

Q1.

1. Integrated Circuit
2. Ultra Large Scale Integration
3. Firmware
4. Palmtop/PDA
5. PARAM
6. Control Unit, Arithmetic Logic Unit, Registers
7. Character Special
8. Memory/Registers
9. Seek Time
10. Digital, Analog
11. Communication
12. Intelligent
13. Packets
14. Gateway
15. Hub

Unit-II

Q1.

I. It is surrounded by a solid rectangular box

II. When the cell is Active

III. Clicking it with mouse, Use Cursor Keys, Use Goto, Home, End Keys

IV. 256 columns and 65536 rows

V. In the cell itself and the formula bar

VI. Press F5 (Goto) and then type V1200

VII. i) Saving- Ctrl+S, Clicking Shortcut icon in the tool bar, Using Save/Save As option in the File Menu

ii) Open- Ctrl+O, Clicking Shortcut icon in the tool bar, Using Open option in the File Menu

VIII. Column, Bar, Line, Pie, Scatter, area, Bubble etc

IX. By setting the Print Area in the File Menu under Print Area

X. Goto View Menu and Choose Header/Footer and fill up the necessary information

DMS,KU

Unit-III

Q1.

I. Accounting
II. Two main
III. F1 function key
IV. Ctrl+N
V. Alt+H
VI. F2 function key
VII. F12 function key
VIII. Sales
IX. Period (financial year)
X. Ledger

Unit-IV

Q1.

a. Statistical Package for Social Sciences
b. Behavioral
c. Command, Windows (GUI)
d. Three
e. Data Editor, Output and Syntax windows
f. Data Editor window
g. Output window
h. Syntax window
i. .spo, .sps
j. .sav

DMS,KU