Chapter 7
One-Way ANOVA: Comparing Means of More than Two Samples

Learning Objectives

After completing this chapter, you should be able to do the following:

- Understand the basics of one-way analysis of variance (ANOVA).
- Learn to interpret the model involved in one-way analysis of variance.
- Learn the different designs of ANOVA.
- Describe the situations in which one-way analysis of variance should be used.
- Learn the manual procedure of applying one-way ANOVA in testing of hypothesis.
- Construct the null and research hypotheses to be tested in the research study.
- Learn what happens if multiple t-tests are used instead of one-way ANOVA.
- Understand the steps involved in one-way analysis of variance in equal and unequal sample sizes.
- Interpret the significance of F-statistic using the concept of p value.
- Know the procedure of making data file for analysis in SPSS.
- Understand the steps involved in using SPSS for solving the problems of one-way analysis of variance.
- Describe the output of one-way analysis of variance obtained in SPSS.

Introduction

One-way analysis of variance is a statistical technique used for comparing means of more than two groups. It tests the null hypothesis that samples in different groups have been drawn from the same population. It is abbreviated as one-way ANOVA. This technique can be used in a situation where the data is measured either on interval or ratio scale. In one-way ANOVA, group means are compared by comparing the variability between groups with that of variability within the groups. This is done by computing an F-statistic. The F-value is computed by dividing the mean sum of squares between the groups by the mean sum of squares within the groups.

J.P. Verma, *Data Analysis in Management with SPSS Software*,
DOI 10.1007/978-81-322-0786-3_7,

As per the central limit theorem, if the groups are drawn from the same population, the variance between the group means should be lower than the variance within the groups. Thus, a higher ratio (F-value) indicates that the samples have been drawn from different populations.

There are varieties of situations in which one-way analysis of variance can be used to compare the means of more than two groups. Consider a study in which it is required to compare the responses of the students belonging to north, south, west and east regions towards liking of mess food in the university. If the quality of mess food is rated on a scale of 1–10 (1 = "I hate the food," 10 = "Best food ever"), then the responses of the students belonging to different regions can be obtained in the form of the interval scores. Here the independent variable would be the student's region having four different levels namely north, south, east and west whereas the response of the students shall be the dependent variable. To achieve the objective of the study the null hypothesis of no difference among the mean responses of the four groups may be tested against the alternative hypothesis that at least one group mean differs. If the null hypothesis is rejected, a post hoc test is used to get the correct picture as to which group's liking is the best.

Similarly a human resource manager may wish to determine whether the achievement motivation differs among the employees in three different age categories (<25, 26–35, and >35 years) after attending a training program. Here, the independent variable is the employee's age category, whereas the achievement motivation is the dependent variable. In this case, it is desired to test whether the data provide sufficient evidence to indicate that the mean achievement motivation of any age category differs from other. The one-way ANOVA can be used to answer this question.

Principles of ANOVA Experiment

There are three basic principles of design of experiments, that is, randomization, replication, and local control. Out of these three, only randomization and replication need to be satisfied by the one-way ANOVA experiments. Randomization refers to the random allocation of the treatment to experimental units. On the other hand, replication refers to the application of each individual level of the factor to multiple subjects. In other words, the experiment must be replicated in more than one subject. In the above example several employees in each age group should be selected in a random fashion in order to satisfy the principles of randomization and replication. This facilitates in drawing the representative sample.

One-Way ANOVA

It is used to compare the means of more than two independent groups. In one-way ANOVA, the effect of different levels of only one factor on the dependent variable is investigated. Usually one-way ANOVA is used for more than two groups because two groups may be compared using t-test. In comparing two group means, the t and F are related as $F = t^2$. In using one-way ANOVA, the experimenter is often interested in investigating the effect of different treatments on some subjects. Which may be people, animals, or plants, etc. For instance, obesity can be compared among the employees of three different departments: marketing, production, and human resource of an organization. Similarly anxiety of the employees can be compared in three different units of an organization. Thus, one-way ANOVA has a wide application in management sciences, humanities, and social sciences.

Factorial ANOVA

A factorial design is the one in which the effect of two factors on the dependent variable is investigated. Here each factor may have several levels and each combination becomes a treatment. Usually factorial ANOVA is used to compare the main effect of each factor as well as their interaction effects across the levels of other factor on the criterion variable. But the situation may arise where each combination of levels in two factors is treated as a single treatment and it is required to compare the effect of these treatments on the dependent variable. In such situation one-way ANOVA can be used to test the required hypothesis. Consider a situation where the effect of different combination of duration and time on learning efficiency is to be investigated. The duration of interest is 30 and 60 minutes and the subjects are given training in the morning and evening sessions for a learning task. The four combinations of treatments would be morning time with 30 minutes duration, morning time with 60 minutes duration, evening time with 30 minutes duration and evening time with 60 minutes duration. In this case neither the main effect nor the interaction effects are of interest to the investigator rather just the combinations of these levels form four levels of the independent treatment.

If the number of factors and their levels are large, then lots of experimental groups need to be created which is practically not possible, and in that case fractional factorial design is used. In this design, only important combinations are studied.

Repeated Measure ANOVA

Repeated measure ANOVA is used when same subjects are given different treatments at different time interval. In this design, same criterion variable is measured many times on each subject. This design is known as repeated measure design because repeated measures are taken at different time in order to see the impact of time on changes in criterion variable. In some studies of repeated measure design, same criterion variable is compared under two or more different conditions. For example, in order to see the impact of temperature on memory retention, a subject's memory might be tested once in an air-conditioned atmosphere and another time in a normal room temperature.

The experimenter must ensure that the carryover effect does not exist in administering different treatments on the same subjects. The studies in repeated measure design are also known as longitudinal studies.

Multivariate ANOVA

Multivariate ANOVA is used when there are two or more dependent variables. It provides solution to test the three hypotheses, namely, (a) whether changes in independent variables have significant impact in dependent variables, (b) whether interaction among independent variables is significant, and (c) whether interaction among dependent variables is significant. Multivariate analysis of variance is also known as MANOVA. In this design, the dependent variables must be loosely related with each other. They should neither be highly correlated nor totally uncorrelated among themselves. Multivariate ANOVA is used to compare the effects of two or more treatments on a group of dependent variables. The dependent variables should be such so that together it conveys some meaning. Consider an experiment where the impact of educational background on three personality traits honesty, courtesy, and responsibility is to be studied in an organization. The subjects may be classified on the basis of their educational qualification; high school, graduation or post-graduation. Here the independent variable is the Education with three different levels: high school, graduation, and postgraduation, whereas the dependent variables are the three personality traits namely honesty, courtesy, and responsibility. The one-way MANOVA facilitates us to compare the effect of education on the personality as a whole of an individual.

One-Way ANOVA Model and Hypotheses Testing

Let us suppose that there are r groups of scores where first group has n_1 scores, second has n_2 scores, and so on, and rth group has n_r scores. If X_{ij} represents the jth score in the ith group ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, n_i$), then these scores can be shown as follows:

						Total	Mean
	1	X_{11}	X_{12} …..	X_{1j} ….	X_{1n_1}	R_1	$\overline{X}_1$
	2	X_{21}	X_{22} …..	X_{2j} ….	X_{2n_2}	R_2	$\overline{X}_2$
	.	.	.	.	.		
	.	.	.	.	.		
Samples	i	X_{i1}	X_{i2} …..	X_{ij} ….	X_{in_i}	R_i	$\overline{X}_i$
	.	.	.	.	.		
	.	.	.	.	.		
	r	X_{r1}	X_{r2}…..	X_{rj} ….	X_{rn_r}	R_r	$\overline{X}_r$
						$G = R_1 + R_2 + \ldots R_r$	

Here,

$N = n_1 + n_2 + \ldots n_r$, the total of all the scores
G is the grand total of all N scores
R_i is the total of all the scores in ith group

The total variability among the above-mentioned N scores can be attributed due to the variability between groups and variability within groups. Thus, the total variability can be broken into the following two components:

$$\begin{aligned} &\text{Total variability} = \text{Variability between groups} + \text{Variability within Groups} \\ &\text{or} \qquad \text{TSS} = \text{SS}_\text{b} + \text{SS}_\text{w} \end{aligned} \tag{7.1}$$

This is known as one-way ANOVA model where it is assumed that the variability among the scores may be due to the groups. After developing the model, the significance of the group variability is tested by comparing the variability between groups with that of variability within groups by using the *F*-test.

The null hypothesis which is being tested in this case is that whether variability between groups (SS_b) and variability within the groups (SS_w) are the same or not. If the null hypothesis is rejected, it is concluded that the variability due to groups is significant, and it is inferred that means of all the groups are not same. On the other hand, if the null hypothesis is not rejected, one may draw the inference that group means do not differ significantly. Thus, if r groups are required to be compared on some criterion variable, then the null hypothesis can be tested by following the below mentioned steps:

(a) *Hypothesis construction:* The following null hypothesis is tested

$$\text{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_r$$

against the alternative hypothesis that at least one mean differs.

(b) *Level of significance:* The level of significance may be chosen beforehand. Usually it is taken as .05 or .01.

(c) Statistical test: The *F*-test is used to test the above mentioned hypothesis. If *F*-value is significant, it indicates that the variability between groups is significantly higher than the variability within groups; in that case, the null hypothesis of no difference between the group means is rejected. *F*-value is obtained by computing the total sum of squares (TSS), sum of squares between groups (SS_b), and sum of squares within groups (SS_w):

   (i) *Total sum of squares (TSS):* It indicates the variation present in the whole data set around its mean value and is obtained by adding the sum of squares due to "between groups" and sum of squares due to "within groups." The total sum of squares can be defined as the sum of squared deviations of all the scores from their mean value. It is usually denoted by TSS and is given by

$$\text{TSS} = \sum_i \sum_j \left(X_{ij} - \frac{G}{N} \right)^2$$

after solving

$$= \sum_i \sum_j X_{ij}^2 - \frac{G^2}{N} \tag{7.2}$$

Here G is the grand total of all the scores. The degrees of freedom for total sum of squares is $N - 1$, and, therefore, mean sum of squares is computed by dividing TSS by $N - 1$.

(ii) *Sum of squares between groups (SS_b)*: The sum of squares between groups can be defined as the variation of group around the grand mean of the data set. In other words, it is the measure of variation between the group means and is usually denoted by SS_b. This is also known as the variation due to assignable causes. The sum of squares between groups is computed as

$$\text{SS}_\text{b} = \sum_i \frac{R_i^2}{n_i} - \frac{G^2}{N} \tag{7.3}$$

Since r samples are involved in one-way ANOVA, the degrees of freedom for between groups is $r - 1$. Thus, mean sum of squares for between groups (MSS_b) is obtained by dividing SS_b by its degrees of freedom $r - 1$.

(iii) *Sum of squares within groups* (SS_w): The sum of squares within groups is the residual variation and is referred as variation due to non-assignable causes. It is the average variation within the groups and is usually denoted by SS_w:

$$\text{SS}_\text{w} = TSS - SS_\text{b} \tag{7.4}$$

The degrees of freedom for the sum of squares within groups is given by $N - r$, and, therefore, mean sum of squares for within groups (MSS_w) is obtained by dividing SS_w by $N - r$.

(iv) *ANOVA table:* After computing all sum of squares, these values are used in the analysis of variance(ANOVA) table for computing F-value as shown below.

ANOVA table

Sources of variation	SS	df	MSS	F-value
Between groups	SS_b	$r - 1$	$\text{MSS}_\text{b} = \frac{\text{SS}_\text{b}}{r-1}$	$F = \frac{\text{MSS}_\text{b}}{\text{MSS}_\text{w}}$
Within groups	SS_w	$N - r$	$\text{MSS}_\text{w} = \frac{\text{SS}_\text{w}}{N-r}$	
Total	TSS	$N - 1$		

Remark: Sum of squares are additive in nature, but mean sum of squares are not

(v) *F-statistic*: Under the normality assumptions, the F-value obtained in the above table, that is,

$$F = \frac{\mathrm{MSS_b}}{\mathrm{MSS_w}} \tag{7.5}$$

follows an F-distribution with $(r-1, N-r)$ degrees of freedom.

This test statistic F is used to test the null hypothesis of no difference among the group means.

(d) *Decision criteria*: The tabulated value of F at .05 and .01 level of significance with $(r-1, N-r)$ degrees of freedom may be obtained from Tables A.4 and A.5, respectively, in the Appendix. If calculated value of F is greater than tabulated F, the null hypothesis is rejected. And in that case it is concluded that at least one of the means will be different. Since ANOVA does not tell us where the difference lies, post hoc test is used to get the clear picture. There are several post hoc tests which can be used, but least significant difference (LSD) test is generally used in equal sample sizes, whereas Scheffe's test is most often used in unequal sample sizes.

In all the post hoc tests, a critical difference is calculated at a particular level of significance, and if the difference of any pair of observed means is higher than the critical difference, it is inferred that one mean is higher than the other; otherwise, group means are equal. By comparing all pair of means, conclusion is drawn as to which group mean is the highest. The procedure of such comparison can be seen in the solved Example 7.1.

LSD test provides the critical difference (CD) which is used for comparing differences in all the pair of means. The CD is computed as follows:

$$\text{Critical difference} = t_{.05}(N-r) \times \sqrt{\frac{2}{n}(\mathrm{MSS})_{\mathrm{w}}} \tag{7.6}$$

where the symbols have their usual meanings.

Since Scheffe's test is used in case of unequal sample size hence it also provides different critical difference for comparing different pair of group means. Here the critical difference (CD) is calculated as follows:

$$\text{CD for comparing } i^{\text{th}} \text{ and } j^{\text{th}} \text{ group means}$$
$$= \sqrt{(r-1)F_{.05}(r-1, N-r)} \times \sqrt{\mathrm{MSS_w}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{7.7}$$

where n_i and n_j represent the sample sizes of ith and jth groups, respectively, and other symbols have their usual meanings.

The SPSS output provides p value (significant value) for each pair of means to test the significance of difference between them. If p value for any pair of means is less than .05, it is concluded that means are significantly different otherwise not. SPSS provides various options for post hoc tests. One may choose one or more options for analysis while using SPSS.

Assumptions in Using One-Way ANOVA

While applying one-way ANOVA for comparing means of different groups, the following assumptions are made:

1. The data must be measured either on interval or ratio scale.
2. The samples must be independent.
3. The dependent variable must be normally distributed.
4. The population from which the samples have been drawn must be normally distributed.
5. The variances of the population must be equal.
6. The errors are independent and normally distributed.

Remarks

1. ANOVA is a relatively robust procedure in case of violations of the normality assumption.
2. In case the data is ordinal, a nonparametric alternative such as Kruskal-Wallis one-way analysis of variance should be used instead of parametric one-way ANOVA.

Effect of Using Several *t*-tests Instead of ANOVA

Many times a researcher argues that what if I use three t-tests rather than using one-way ANOVA in comparing the means of three groups. One of the logics is that, why to use three times t-test if equality of means can be tested by using one-way ANOVA once. If the number of groups are more, then one needs to apply large number of t-tests. For example, in case of six groups, one needs to apply ${}^6C_2 = 15$, t-tests instead of one-time one-way ANOVA. This may be one of the arguments of the researcher in favor of using one-way ANOVA, but the main problem in using multiple t-tests instead of one-way ANOVA is that the type I error gets inflated.

If the level of significance has been chosen as $p_{1,}$ then Fisher has showed that the type I error rate expands from p_1 to some larger value as the number of tests between paired means increases. The error rate expansion is constant and predictable which can be computed by the following equation:

$$p = 1 - (1 - p_1)^r \tag{7.8}$$

where p is the new level of significance and r is the number of t-tests used for comparing all the pair of group means.

For example, in comparing three group means, if t-tests are used instead of one-way ANOVA and if the level of significance is chosen as .05, then the total number of paired comparison would be ${}^{3}C_2 = 3$.

Here, $p_1 = 0.05$ and $r = 3$, and, therefore, the actual level of significance becomes

$$\begin{aligned} p &= 1 - (1 - p_1)^r \\ &= 1 - (1 - 0.5)^3 = 1 - 0.95^3 = 1 - 0.8574 \\ &= 0.143 \end{aligned}$$

Thus, in comparing three group means instead of using one-way ANOVA, if three t-tests are applied, then level of significance shall inflate from .05 to 0.143.

Application of One-Way ANOVA

One-way ANOVA is used when more than two group means are compared. Such situations are very frequent in management research where a researcher may like to compare more than two group means. For instance, one may like to compare the mood state of the employees working in three different plants or to compare the occupational stress among three different age categories of employees in an organization.

Consider an experiment where a market analyst of a company is interested to know the effect of three different types of incentives on the sale of a particular brand of shampoo. Shampoo is sold to the customers with three schemes. In the first scheme 20% extra is offered in the same price, in the second scheme shampoo is sold with free bath soap, whereas in the third scheme it is sold to the customers with a free ladies' perfume. These three schemes are offered to the customers in the same outlet for 3 months. During the second month, sales of the shampoo are recorded in all three schemes for 20 days. In this situation, scheme is the independent variable having three different levels: 20% extra shampoo, shampoo with a bath soap, and shampoo with a ladies' perfume whereas, the sales figure is the dependent variable. Here the null hypothesis which is required to be tested would be

H_0 : Average sale of shampoo in all three incentive groups are same against the alternative hypothesis.
H_1 : At least one group mean is different.

The one-way ANOVA may be applied to compute F-value. If F-statistic is significant, the null hypothesis may be rejected, and in that case, a post hoc test may be applied to find as to which incentive is the most attractive in improving the sale of the shampoo. On the other hand, if F-value is not significant, one fails to reject the null hypothesis, and in that case, there would be no reason to believe that any one incentive is better than others to enhance the sale.

Example 7.1 An audio company predicts that students learn more effectively with a constant low-tune melodious music in background, as opposed to an irregular loud orchestra or no music at all. To verify this hypothesis, a study was planned by dividing 30 students into three groups of ten each. Students were assigned to these three groups in a random fashion, and all of them were given a comprehension to read for 20 min. Students in group 1 were asked to study the comprehension with low-tune melodious music at a constant volume in the background. Whereas the students in group 2 were exposed to loud orchestra and group 3 to no music at all while reading the comprehension. After reading the comprehension, they were asked to solve few questions. The marks obtained are shown in the Table 7.1.

Do these data confirm that learning is more effective in particular background music? Test your hypothesis at 5% level.

Solution Following steps shall be taken to test the required hypothesis:

(a) *Hypotheses construction:* The researcher is interested in testing the following null hypothesis:

$$\mathrm{H_0}: \mu_{\text{Music}} = \mu_{\text{Orchestra}} = \mu_{\text{Without\_Music}}$$

against the alternative hypothesis that at least one mean is different.

(b) *Level of significance:* 0.05

(c) *Statistical test:* One-way ANOVA shall be used to test the null hypothesis. In order to complete the ANOVA table, first, all the sum of squares are computed. Here,

Number of groups $= r = 3$
Sample size in each group $= n = 10$
Total number of scores $= nr = 30$

The computation of group total, group means, and grand total has to be computed first which is shown in Table 7.2.

(i) $$\text{Correction factor(CF)} = \frac{G^2}{N} = \frac{135^2}{30} = 607.5$$

(ii) $$\begin{aligned}\text{Raw sum of squares(RSS)} &= \sum_{\mathrm{i}}\sum_{\mathrm{j}} \mathrm{X}_{\mathrm{ij}}^2 \\ &= \left(8^2+4^2+8^2+\ldots 9^2+6^2\right) \\ &\quad + \left(4^2+6^2+3^2+\ldots 4^2+3^2\right) \\ &\quad + \left(3^2+4^2+6^2+\ldots 1^2+2^2\right) \\ &= 440+188+127=755\end{aligned}$$

Table 7.1 Comprehension scores of the students under three different environments

Music	Orchestra	Without music
8	4	3
4	6	4
8	3	6
6	4	2
6	3	1
7	8	2
3	3	6
7	2	4
9	4	1
6	3	2

Table 7.2 Computation of group total, group means, and grand total

	Music	Orchestra	Without music	
	8	4	3	
	4	6	4	
	8	3	6	
	6	4	2	
	6	3	1	
	7	8	2	
	3	3	6	
	7	2	4	
	9	4	1	
	6	3	2	
Group total	$R_1 = 64$	$R_2 = 40$	$R_3 = 31$	$G = R_1 + R_2 + R_3 = 135$
Group mean	6.4	4	3.1	

(iii) $$\text{Total sum of square(TSS)} = \text{RSS} - \text{CF} = \sum_i \sum_j X_{ij}^2 - \frac{G^2}{N}$$
$$= 755 - 607.5 = 147.5$$

(iv) $$\text{Sum of squares between groups } (\text{SS}_\text{b}) = \sum_i \frac{R_i^2}{n_i} - \frac{G^2}{N}$$
$$= \frac{64^2 + 40^2 + 31^2}{10} - 607.5$$
$$= 665.7 - 607.5 = 58.2$$

(v) $$\text{Sum of square within groups}(\text{SS}_\text{w}) = \text{TSS} - \text{SS}_\text{b}$$
$$= 147.5 - 58.2 = 89.3$$

Table 7.3 ANOVA table for the data on comprehension test

Sources of variation	SS	df	MSS	F-value
Between groups	58.2	$r-1=2$	$\frac{58.2}{2}=29.1$	
				8.79
Within groups	89.3	$N-r=27$	$\frac{89.3}{27}=3.31$	
Total	147.5	$N-1=29$		

Table 7.4 Group means and their comparison

Music	Orchestra	Without music	CD at 5% level
6.4	4	3.1	**1.67**

"____" represents no significant difference between the means at 5% level

(d) *Decision criteria*

From Table A.4 in the Appendix, $F_{.05}(2,27)=4.22$.

Since calculated $F(=8.79) > F_{.05}(2,27)$, the null hypothesis may be rejected. It is therefore concluded that learning efficiency in all the three experimental groups is not same. In order to find as to which group's learning efficiency is best, the least significance difference (LSD) test shall be applied. The critical difference in LSD test is given by

$$\begin{aligned} \text{CD} &= t_{.05}(27) \times \sqrt{\frac{2 \times \text{MSS}_\text{w}}{n}} \\ &= 2.052 \times \sqrt{\frac{2 \times 3.31}{10}} = 1.67 \end{aligned}$$

(e) *Results*

The group means may be compared by arranging them in descending order as shown in the Table 7.4

It is clear from Table 7.4 that the mean difference between "music" and "orchestra" groups as well as "music" and "without music" groups is greater than the critical difference. Since the mean difference between orchestra and without music groups is significant hence it is shown by clubbing their means by the line as shown in the Table 7.4.

(f) *Inference*

From the results, it is clear that the mean learning performance in music group is significantly higher than that of orchestra as well as nonmusic groups, whereas the mean learning of orchestra group is equal to that of nonmusic group. It is therefore concluded that melodious music improves the learning efficiency.

Table 7.5 Data on psychological health

S.N.	Banking	Insurance	Retail
1	45	41	58
2	41	38	54
3	47	43	49
4	59	53	65
5	48	43	51
6	45	42	56
7	38	40	41
8	48	42	51
9	39	32	45
10	42	39	53
11	38	36	37
12	36	32	42
13	45	40	44
14	38	39	32
15	42	40	50

Solved Example of One-Way ANOVA with Equal Sample Size Using SPSS

Example 7.2 The data in the following table indicates the psychological health ratings of corporate executives in banking, insurance, and retail sectors. Apply one-way ANOVA to test whether the executives of any particular sector are healthier in their psychological health in comparison to other sectors. Test your hypothesis at 5% as well as 1% level (Table 7.5).

Solution

In this problem, it is required to test the following null hypothesis

$$H_0 : \mu_{\text{Banking}} = \mu_{\text{Insurance}} = \mu_{\text{Retail}}$$

against the alternative hypothesis that at least one mean differs.

The SPSS output provides F-value along with its significance value (p value). The F-value would be significant if the p value is less than .05. If F-value becomes significant, a post hoc test shall be used to compare the paired means. SPSS provides facility to choose one or more post hoc test for analysis.

In this example, since the sample sizes are equal, LSD test shall be used as a post hoc test for comparing the group means. However, one can choose other post hoc tests as well. The SPSS output provides the p value for testing the significance of the difference between each pair of group means. Thus, by looking to the results of post hoc test, one can determine as to which group mean is higher. The procedure has been discussed while interpreting the output.

Computations in One-Way ANOVA with Equal Sample Size

(a) ***Preparing data file***

A data file needs to be prepared before using the SPSS commands for one-way ANOVA with equal samples size. The following steps will help you prepare the data file:

(i) *Starting the SPSS:* Use the following command sequence to start SPSS:

Start → Programs → IBM SPSS Statistics → IBM SPSS Statistics 20

After clicking the **Type in Data,** you will be taken to the **Variable View** option for defining the variables in the study.

(ii) *Defining variables:* There are two variables in this example which need to be defined along with their properties while preparing the data file. These variables are psychological health and sector. The psychological health is defined as scale variable, whereas sector is defined as nominal variable as they are measured on interval as well as nominal scales, respectively. The procedure of defining these variables in the SPSS is as follows:

1. Click the **Variable View** to define the variables and their properties.
2. Write short name of the variables as *Psy_Health* and *Sector* under the column heading **Name**.
3. Under the column heading **Label**, full names of these variables have been defined as *Psychological health rating* and *Different sector,* respectively. You may choose some other names of these variables as well.
4. For the variable *Sector,* double-click the cell under the column heading **Values** and add the following values to different levels:

Value	Label
1	Banking
2	Insurance
3	Retail

5. Under the column heading **Measure**, select the option “Scale” for the *Psy_Health* variable and “Nominal” for the *Sector* variable.
6. Use default entries in rest of the columns. The screen shall look like Fig. 7.1.

Remark: Many variables can be defined in the variable view simultaneously if ANOVA is to be applied for more than one variable.

(iii) *Entering data*

After defining both the variables in **Variable View,** click **Data View** on the left corner in the bottom of the screen as shown in Fig. 7.1 to open the

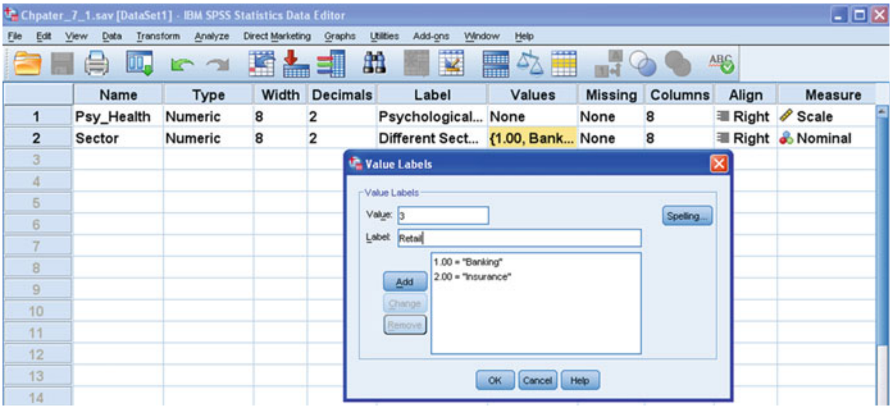

Fig. 7.1 Defining variables along with their characteristics

data entry format column wise. After entering the data, the screen will look like as shown in Fig. 7.2. Since the data is large, only a portion of data is shown in the figure. Save the data file in the desired location before further processing.

(b) **SPSS *commands for one-way ANOVA***
After entering all the data in the data view, follow the below-mentioned steps for one-way analysis of variance:

(i) *Initiating the SPSS commands for one-way ANOVA:* In data view, click the following commands in sequence:

Analyze ⇨ Compare Means ⇨ One-Way ANOVA

The screen shall look like Fig. 7.3.

(ii) *Selecting variables for one-way ANOVA:* After clicking the **One-Way ANOVA** option, you will be taken to the next screen for selecting variables. Select the variables *Psychological health rating* and *Different sector* from left panel to the "Dependent list" section and "Factor" section of the right panel, respectively. The screen will look like Fig. 7.4.

(iii) *Selecting the options for computation:* After selecting the variables, option needs to be defined for generating the output in one-way ANOVA. Take the following steps:

– Click the tag **Post Hoc** in the screen shown in Fig. 7.4.

– Check the option "LSD." LSD test is selected because the sample sizes are equal. You may choose any other post hoc test if you so desire.
– Write "Significance level" as .05. By default, it is selected. However, you may select any other significance level like .01 or .10 as well.
– Click ***Continue***.

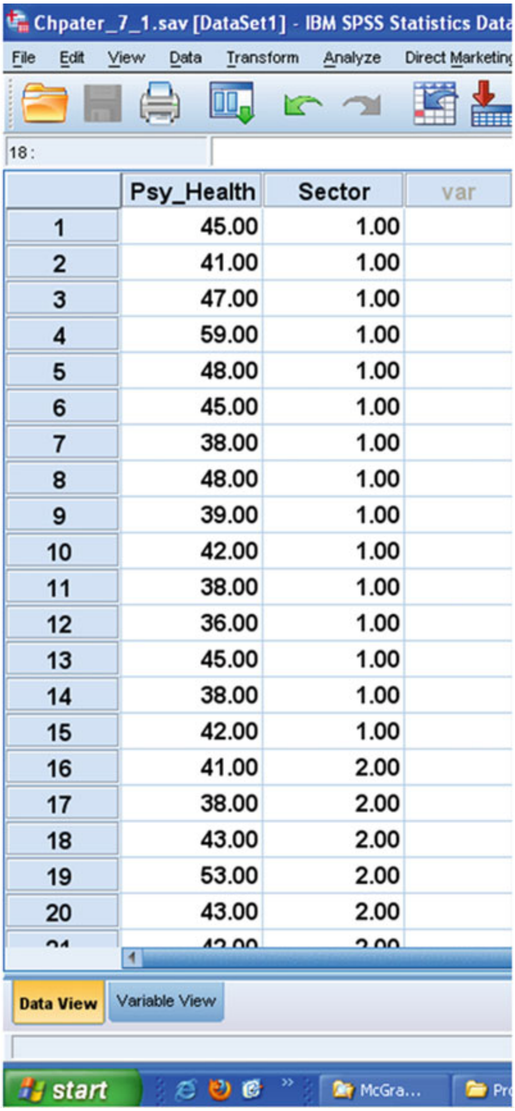

Fig. 7.2 Screen showing entered data for the psychological health and sector in the data view

The screen will look like Fig. 7.5.

- Click the tag **Options** in the screen shown in Fig. 7.4 and then check 'Descriptive'.
- Click ***Continue***.

The screen for this option shall look like Fig. 7.6.

- After selecting the options, the screen shown in Fig. 7.4 shall be restored.
- Click **OK**.

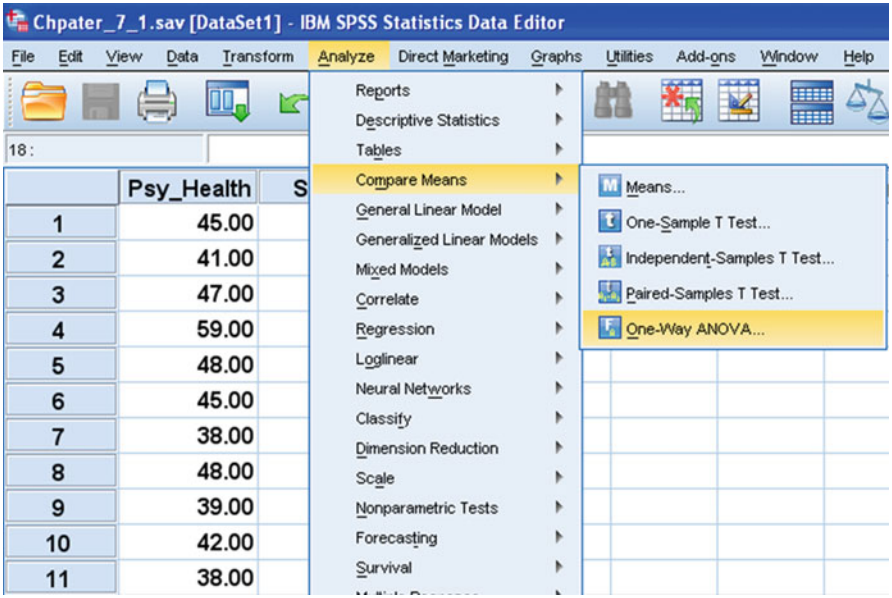

Fig. 7.3 Screen showing SPSS commands for one-way ANOVA

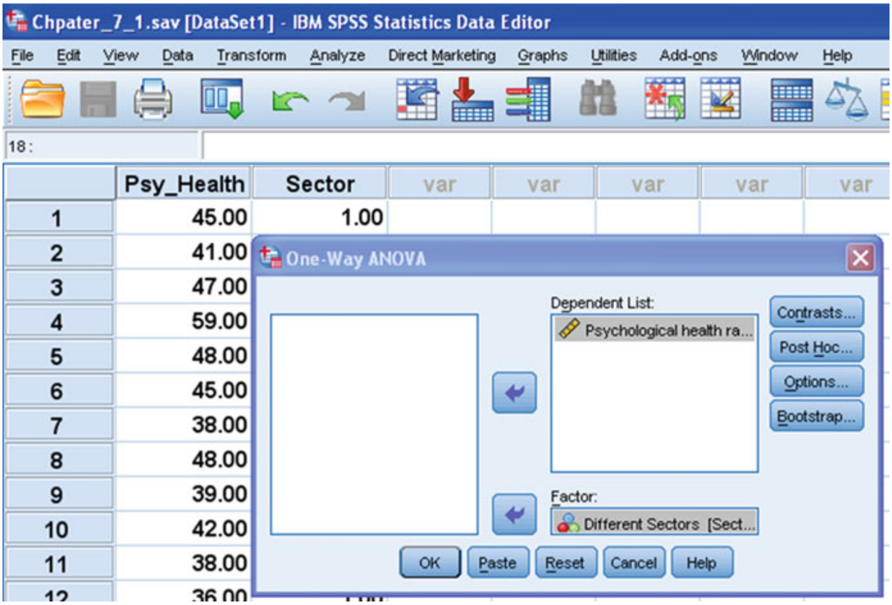

Fig. 7.4 Screen showing selection of variables for one-way ANOVA

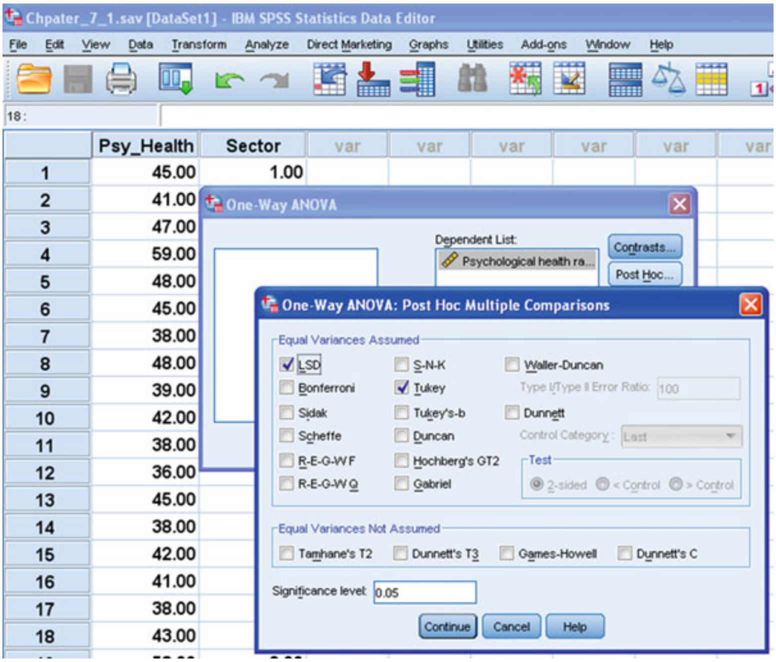

Fig. 7.5 Screen showing options for post hoc test and significance level

(c) ***Getting the output***

After clicking **OK** in the screen shown in Fig. 7.4, the output shall be generated in the output window. The relevant outputs may be selected by using right click of the mouse and may be copied in the word file. Here, the following outputs shall be generated:

1. Descriptive statistics
2. ANOVA table
3. Post hoc comparison table

In this example, all the outputs so generated by the SPSS will look like as shown in Tables 7.6, 7.7, and 7.8.

Interpretations of the Outputs

Different descriptive statistics have been shown in Table 7.6 which may be used to study the nature of the data. Further descriptive profiles of the psychological health rating for the corporate executives in different sectors can be developed by using the values of mean, standard deviation, and minimum and maximum scores in each groups. The procedure of developing such profile has been discussed in Chap. 2 of this book.

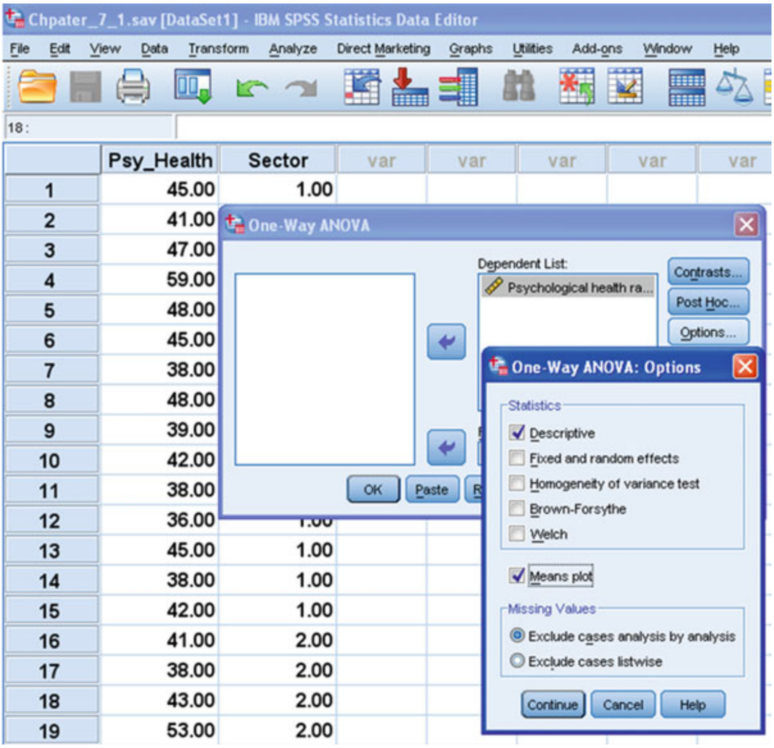

Fig. 7.6 Screen showing options for descriptive statistics

Table 7.6 Descriptive statistics for the data on psychological health among corporate executives in different sectors

					95% confidence interval for mean			
	N	Mean	SD	SE	Lower bound	Upper bound	Min.	Max.
Banking	15	43.40	5.84	1.51	40.17	46.63	36.00	59.00
Insurance	15	40.00	4.97	1.28	37.25	42.75	32.00	53.00
Retail	15	48.53	8.53	2.20	43.81	53.26	32.00	65.00
Total	45	43.98	7.38	1.10	41.76	46.20	32.00	65.00

Note: Values have been rounded off nearest to the two decimal places

Table 7.7 ANOVA table for the data on psychological health

	Sum of squares	df	Mean square	*F*	Sig. (*p* value)
Between groups	553.64	2	276.82	6.31	.004
Within groups	1,843.33	42	43.89		
Total	2,396.98	44			

Note: Values have been rounded off nearest to the two decimal places

Table 7.8 Post hoc comparison of means using LSD test

(*I*) Different sectors	(*J*) Different sectors	Mean difference ($I - J$)	Std. error	Sig. (*p* value)
Banking	Insurance	3.40	2.42	.167
	Retail	−5.13*	2.42	.040
Insurance	Banking	−3.40	2.42	.167
	Retail	−8.53**	2.42	.001
Retail	Banking	5.13*	2.42	.040
	Insurance	8.53**	2.42	.001

Note: The values of lower bound and upper bound have been omitted from the original output. The values have been rounded off nearest to the two decimal places
*The mean difference is significant at 5% level
**The mean difference is significant at 1% level

Table 7.9 Mean scores on psychological health in different groups

Retail	Banking	Insurance
48.53	43.40	40.00

"____" represents no significant difference between the means

The mean of different groups in Table 7.6 and the results of Table 7.8 have been used to prepare the graphics shown in Table 7.9 which can be used to draw conclusions about post hoc comparison of means.

The *F*-value in Table 7.7 is significant at 5% level because its *p* value (=.004) is less than .05. Thus, the null hypothesis of no difference among the means of the three groups may be rejected at 5% level. Since the *p* value is also less than .01, the null hypothesis may be rejected at 1% level also.

Here, the *F*-value is significant; hence, the post hoc test needs to be applied for testing the significance of mean difference between different pairs of groups. Table 7.8 provides such comparison. It can be seen from this table that the difference between banking and retail groups on their psychological health rating is significant at 5% level because the *p* value for this mean difference is .04 which is less than .05.

Similarly, the difference between insurance and retail groups on their psychological health is also significant at 5% as well as 1% level because the *p* value attached to this mean difference is .001 which is less than .05 as well as .01.

There is no significant difference between the banking and insurance groups on their psychological health rating because the *p* value attached to this group is .167 which is more than .05.

All the above-mentioned three findings can be very easily understood by looking to the graphics in Table 7.9. From this table, it is clear that the mean psychological health rating score is highest among the executives in the retail sector in comparison to that of banking and insurance sectors. It may thus be concluded that the psychological health of the executives in the retail sector is best in comparison to that of banking and insurance sectors.

Solved Example of One-Way ANOVA with Unequal Sample

Example 7.3 A human resource department of an organization conducted a study to know the status of occupational stress among their employees in different age categories. A questionnaire was used to assess the stress level of the employees in three different age categories: <40, 40–55, and >55 years. The stress scores so obtained are shown in Table 7.10.
Apply one-way analysis of variance to test whether mean stress score of the employees in any two age categories are different. Test your hypothesis at 5% level.

Solution Solving problems of one-way ANOVA with equal and unequal samples through SPSS are almost similar. In case of unequal sample size, one should be careful in feeding the data. The procedure of feeding the data in this case shall be discussed below. Here, the SPSS procedure shall be discussed in brief as it is exactly similar to the one discussed in Example 7.2. Readers are advised to refer to the procedure mentioned in Example 7.2 in case of doubt in solving this problem of unequal sample size.

Here, the null hypothesis which needs to be tested is

$$H_0 : \mu_A = \mu_B = \mu_C$$

against the alternative hypothesis that at least one group mean differs.

If the null hypothesis is rejected, post hoc test will be used for comparing group means. Since the sample sizes are different, the Scheffe's test has been used for post hoc analysis.

Table 7.10 Occupational stress scores among the employees in different age categories

Group A (<40 years)	Group B (40–55 years)	Group C (>55 years)
54	75	55
48	68	51
47	68	59
54	71	64
56	79	52
62	86	48
56	81	65
45	79	48
51	72	56
54	78	49
48	69	
52		

Computations in One-Way ANOVA with Unequal Sample Size

(a) ***Preparing data file:***

(i) *Starting the SPSS:* Start the SPSS the way it has been done in the above-mentioned example and click the **Type in Data** option. You will be taken to the **Variable View** option for defining the variables in the study.

(ii) *Defining variables:* There are two variables in this example that need to be defined along with their properties while preparing the data file. The two variables are stress scores and age group. The stress score is defined as scale variable, whereas age group is defined as nominal variable as they are measured on interval as well as nominal scales, respectively. The procedure of defining these variables in the SPSS is as follows:

1. Click **Variable View** to define variables and their properties.
2. Write short name of the variables as *Stress* and *Age_Gp* under the column heading **Name**.
3. Under the column heading **Label**, full names of these variables may be defined as *Stress scores* and *Age group,* respectively. You may choose some other names of these variables as well.
4. For the variable *Age group,* double-click the cell under the column heading **Values** and add the following values to different levels:

Value	Label
1	Group A (<40 years)
2	Group B (40–55 years)
3	Group C (>55 years)

5. Under the column heading **Measure,** select the option "Scale" for the *Stress* variable and "Nominal" for the *Age_Gp* variable.
6. Use default entries in rest of the columns.

After defining all the variables in variable view, the screen shall look like Fig. 7.7.

Remark: More than one variable can be defined in the variable view for doing ANOVA for many variables simultaneously.

(iii) *Entering the data*: After defining the variables in the **Variable View**, enter the data, column-wise in **Data View**. The data feeding shall be done as follows:

Format of data feeding in Data View

	S.N.	Stress	Age_Gp
Group A $n_1 = 12$	1	54	1
	2	48	1
	3	47	1
	4	54	1
	5	56	1
	6	62	1
	7	56	1
	8	45	1
	9	51	1
	10	54	1
	11	48	1
	12	52	1
Group B $n_2 = 11$	13	75	2
	14	68	2
	15	68	2
	16	71	2
	17	79	2
	18	86	2
	19	81	2
	20	79	2
	21	72	2
	22	78	2
	23	69	2
Group C $n_3 = 10$	24	55	3
	25	51	3
	26	59	3
	27	64	3
	28	52	3
	29	48	3
	30	65	3
	31	48	3
	32	56	3
	33	49	3

After feeding the data as mentioned above, the final screen shall look like Fig. 7.8.

(b) **SPSS *commands for one-way ANOVA for unequal sample size***
After entering all the data in data view, save the data file in the desired location before further processing.

(i) *Initiating the SPSS commands for one-way ANOVA:* In data view, go to the following commands in sequence:

Analyze ⇨ Compare Means ⇨ One-Way ANOVA

(ii) *Selecting variables for analysis:* After clicking the **One-Way ANOVA** option, you will be taken to the next screen for selecting variables. Select

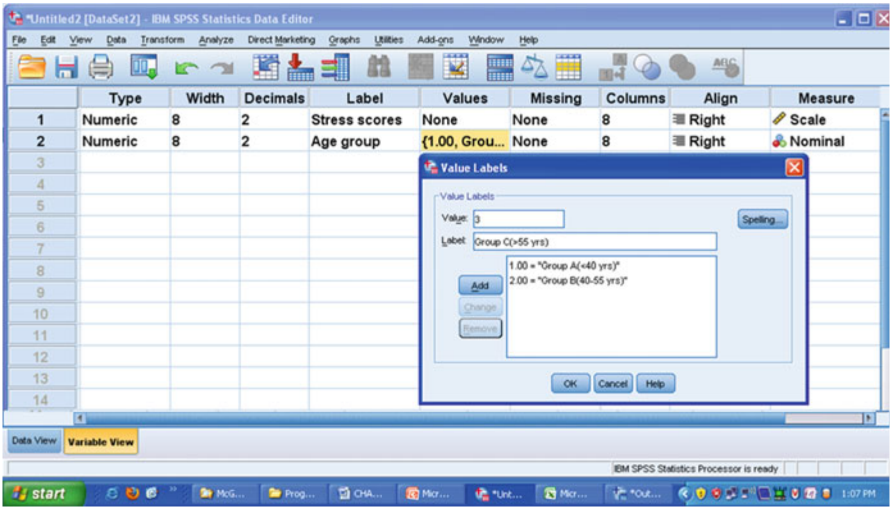

Fig. 7.7 Defining variables along with their characteristics

the variables *Stress scores* and *Age group* from left panel to the “Dependent list” section and “Factor” section of the right panel, respectively. The screen shall look like Fig. 7.9.

(iii) *Selecting options for computation:* After variable selection, option needs to be defined for generating outputs in one-way ANOVA. This shall be done as follows:

- Click the tag **Post Hoc** in the screen shown in Fig. 7.9.
  - Check the option “Scheffe.” This test is selected because the sample sizes are unequal; however, you can choose any other test if you so desire.
  - If graph needs to be prepared, select the option “Means plot.”
  - Write “Significance level” as .05. Usually this is written by default; however, you may write any other significance level like .01 or .10 as well.
  - Click ***Continue***.
- Click the tag **Options** and then check “Descriptive.” Click ***Continue***.
- After selecting the options, click **OK**.

(c) ***Getting the output***

After clicking **OK** on the screen as shown in Fig. 7.9, the output shall be generated in the output window. The relevant outputs can be selected by using right click of the mouse and may be copied in the word file. The following output shall be generated in this example:

(a) *Descriptive statistics*

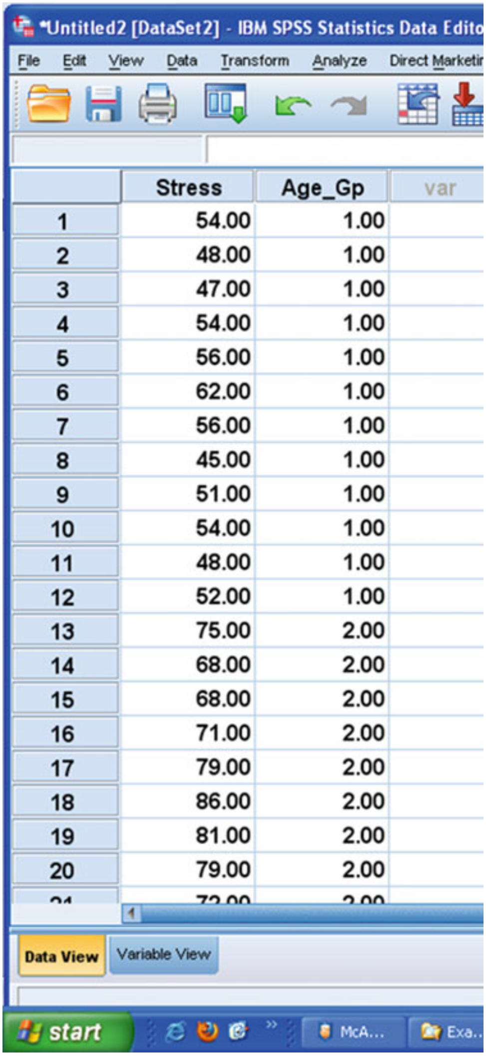

Fig. 7.8 Showing data entry of stress scores for the employees in different age categories in data view

(b) *ANOVA table*
(c) *Post hoc comparison table*
(d) *Graph for means plot*

These outputs are shown in Tables 7.11, 7.12, and 7.13 and in Fig. 7.10. The Table 7.14 has been developed by using the descriptive statistics from the Table 7.11 and inputs from Table 7.13.

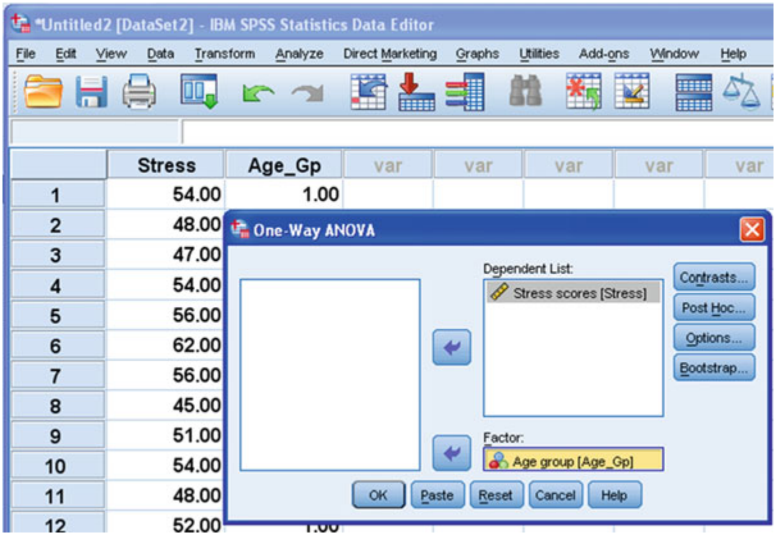

Fig. 7.9 Screen showing selection of variables

Table 7.11 Descriptive statistics for the data on occupational stress of employees in different age categories

					95% confidence interval for mean			
	N	Mean	SD	SE	Lower bound	Upper bound	Min.	Max.
Group A (<40 years)	12	52.25	4.77	1.38	49.23	55.28	45.00	62.00
Group B (40–55 years)	11	75.09	5.97	1.80	71.08	79.10	68.00	86.00
Group C (>55 years)	10	54.70	6.29	1.99	50.20	59.20	48.00	65.00
Total	33	60.61	11.80	2.05	56.42	64.79	45.00	86.00

Table 7.12 ANOVA table for the data on occupational stress

	Sum of squares	df	Mean square	*F*	Sig.
Between groups	3494.620	2	1747.310	54.42	.000
Within groups	963.259	30	32.109		
Total	4457.879	32			

Interpretation of the Outputs

Table 7.11 shows the descriptive statistics of the data on occupational stress of employees in different age categories. These statistics can be used to develop a graphic profile of the employee's occupational stress in different age categories.

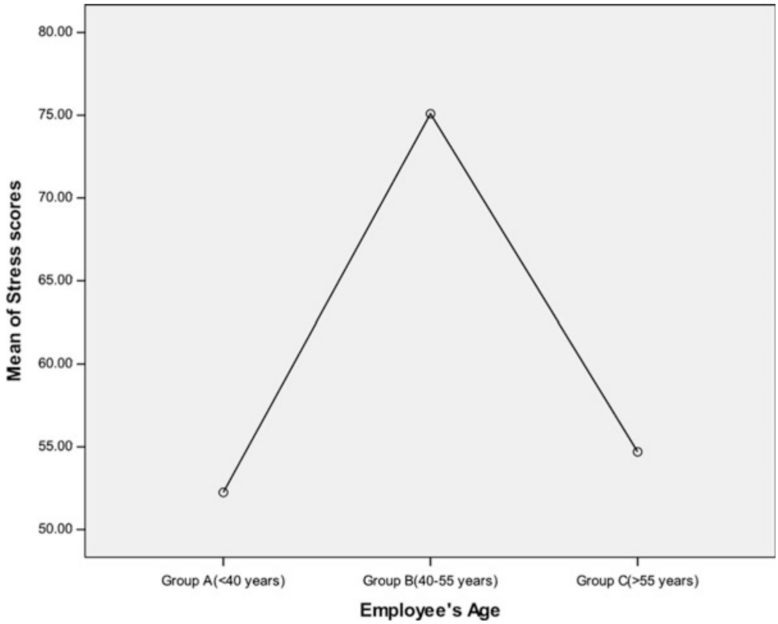

Fig. 7.10 Graphical presentation of mean scores of occupational stress in three different age categories

The procedure of developing such profile has been discussed in detail in Chap. 2 of this book. Further, these descriptive statistics can be used to discuss the nature of data in different age categories.

Table 7.12 gives the value of calculated F. The p value attached with the F is .000 which is less than .05 as well as .01; hence, it is significant at 5% as well as 1% levels. Since the F-value is significant, the null hypothesis of no difference in the occupational stress among the employees in all the three age categories is rejected. The post hoc test is now used to compare the means in different pairs.

SPSS provides the option of choosing the post hoc test, and, therefore, one may choose any one or more test for post hoc analysis. In this example, the Scheffe's test was chosen to compare the means in different pairs. Table 7.13 provides such comparisons.

It can be seen that the difference between occupational stress of the employees in group A (<40 years) and group B (40–55 years) is significant at 5% as well as at 1% level both as the p value for this mean difference is .000 which is less than .05 as well as .01. Similarly, the mean difference between occupational stress of the employees in group B (40–55 years) and group C (>55 years) is also significant at 5% as well as 1% level both as the p value for this mean difference is .000 which is also less than .05 and .01. However, there is no significant difference between the occupational stress of the employees in group A (<40 years) and group C (>55 years) because the p value is .606.

Table 7.13 Post hoc comparison of group means using Scheffe's test

(I) Age group	(J) Age group	Mean diff. (I − J)	SE	Sig. (*p* value)
Group A (<40 years)	Group B (40–55 years)	−22.84091*	2.36531	.000
	Group C (>55 years)	−2.45000	2.42623	.606
Group B (40–55 years)	Group A (<40 years)	22.84091*	2.36531	.000
	Group C (>55 years)	20.39091*	2.47585	.000
Group C (>55 years)	Group A (<40 years)	2.45000	2.42623	.606
	Group B (40–55 years)	−20.39091*	2.47585	.000

Note: The values of lower bound and upper bound have been omitted from the original output
*The mean difference is significant at 5% as well as 1% levels

Table 7.14 Mean scores on occupational stress in different groups

Group B (40–55 years)	Group C (>55 years)	Group A (<40 years)
75.09	54.70	52.25

"____" represents no significant difference between the means

The above results can be easily understood by looking to the graphics in Table 7.14. This table has been obtained by combining the results of Tables 7.11 and 7.13.

Table 7.14 reveals that the mean occupational stress is highest among the employees in group B (40–55 years). Further, mean occupation stress is similar in group C (>55 years) and group A (<40 years). Since the option for mean plot was selected in the SPSS, Fig. 7.10 has been generated in the output which shows the mean plots of all the groups. The graph provides the conclusion at a glance.

Inference: On the basis of the results obtained above, it may be inferred that the occupational stress among the employees in the age category 40–55 years is maximum. The researcher may write their own reasons for these findings after studying the lifestyle and working environment of the employees in this age category. The results in the study provide an opportunity to the researcher to write their own reasoning or develop their theoretical concepts supported by the review of literatures.

Summary of the SPSS Commands for One-Way ANOVA (Example 7.2)

(i) Start the SPSS by using the following commands:

Start → Programs → IBM SPSS Statistics → IBM SPSS Statistics 20

(ii) Click **Variable View** tag and define the variables *Psy_Health* and *Sector* as scale and nominal variables, respectively.

(iii) Under the column heading **Values**, define "1" for banking, "2" for insurance, and "3" for retail.

(iv) After defining variables, type the data for these variables by clicking **Data View**.

(v) In the data view, follow the below-mentioned command sequence for the computation involved in one-way analysis of variance:

Analyze ⇨ Compare Means ⇨ One-Way ANOVA

(vi) Select the variables *Psychological health rating* and *Different sector* from left panel to the "Dependent list" section and "Factor" section of the right panel, respectively.

(vii) Click the tag **Post Hoc** and check the option "LSD" and ensure that the value of "Significance level" is written as .05. Click ***Continue***.

(viii) Click the tag **Options** and then check "Descriptive." Press ***Continue***.

(ix) Press **OK** for output.

Exercise

Short Answer Questions

Note: Write answer to each of the following questions in not more than 200 words.

Q.1. In an experiment, it is desired to compare the time taken to complete a task by the employees in three age groups, namely, 20–30, 31–40, and 41–50 years. Write the null hypothesis as well as all possible types of alternative hypotheses.

Q.2. Explain a situation where one-way analysis of variance can be applied. Which variances are compared in one-way ANOVA?

Q.3. Define principles of ANOVA. What impact it will have if these principles are not met?

Q.4. In what situations factorial experiments are planned? Discuss a specific situation where it can be used.

Q.5. What is repeated measure design? What precaution one must take in framing such an experiment?

Q.6. Discuss the procedure of one-way ANOVA in testing of hypotheses.

Q.7. Write a short note on post hoc tests.

Q.8. What do you mean by different sum of squares? Which sum of square you would like to increase and decrease in your experiment and why?

Q.9. What are the assumptions in applying one-way ANOVA?

Q.10. If you use multiple t-tests instead of one-way ANOVA, what impact it will have on results?

Q.11. Analysis of variance is used for comparing means of different groups, but in doing so F-test is applied, which is a test of significance for comparing the variances of two groups. Discuss this anomaly.

Q.12. What do you mean by the post hoc test? Differentiate between LSD and Scheffe's test.

Q.13. What is p value? In what context it is used?

Multiple Choice Questions

Note: For each of the question, there are four alternative answers. Tick mark the one that you consider the closest to the correct answer.

1. In one-way ANOVA experiment, which of the following is a randomization assumption that must be true?

   (a) The treatment must be randomly assigned to the subjects.
   (b) Groups must be chosen randomly.
   (c) The type of data can be randomly chosen to either categorical or quantitative.
   (d) The treatments must be randomly assigned to the groups.

2. Choose the correct statement.

   (a) Total sum of square is additive in nature.
   (b) Total mean sum of square is additive in nature.
   (c) Total sum of square is nonadditive.
   (d) None of the above is correct.

3. In one-way ANOVA, X_{ij} represents

   (a) The sample mean of the criterion variable for the ith group
   (b) The criterion variable value for the ith subject in the jth group
   (c) The number of observations in the jth group
   (d) The criterion variable value for the jth subject in the ith group

4. In one-way ANOVA, TSS measures

   (a) The variability within groups
   (b) The variability between groups
   (c) The overall variability in the data
   (d) The variability of the criterion variable in any group.

5. In an experiment, three unequal groups are compared with total number of observations in all the groups as 31 (with some items missing). Calculate the test statistic for one-way ANOVA F-test.

Source	df	SS	MS	F
Between groups		7.5	3.75	?
Within groups				
Total		20.8		

(a) 17.89
(b) 789
(c) 7.89
(d) 78.9

6. Choose the correct statement.

(a) LSD may be used for unequal sample size.
(b) Scheffe's test may be used for unequal sample size.
(c) Scheffe's test may be used for comparing more than ten groups.
(d) None of the above is correct.

7. If two groups having 10 observations in each are compared by using one-way ANOVA and if $SS_w = 140$, then what will be the value of MSS_w?

(a) 50
(b) 5
(c) 0.5
(d) 50.5

8. In a one-way ANOVA, if the level of significance is fixed at .05 and if p value associated with F-statistics is 0.062, then what should you do?

(a) Reject H_0, and it is concluded that the group population means are not all equal.
(b) Reject H_0, and it may be concluded that it is reasonable that the group population means are all equal.
(c) Fail to reject H_0, and it may be concluded that the group population means are not all equal.
(d) Fail to reject H_0, and it may be concluded that there is no reason to believe that the population means differ.

9. Choose the correct statement.

(a) If F-statistic is significant at .05 level, it will also be significant at .01 level.
(b) If F-statistic is significant at .01 level, it may not be significant at .05 level.
(c) If F-statistic is significant at .01 level, it will necessarily be significant at .05 level.
(d) If F-statistic is not significant at .01 level, it will not be significant at .05 level.

10. Choose the correct statement.

(a) If p value is 0.02, F-statistic shall be significant at 5% level.
(b) If p value is 0.02, F-statistic shall not be significant at 5% level.
(c) If p value is 0.02, F-statistic shall be significant at 1% level.
(d) None of the above is correct.

11. In comparing the IQ among three classes using one-way ANOVA in SPSS, choose the correct statement about the variable types.

(a) IQ is a nominal variable and class is a scale variable.
(b) Both IQ and class are the scale variables.
(c) IQ is a scale variable and class is a nominal variable.
(d) Both IQ and class are the nominal variables.

12. If product sales are to be compared in three outlets, then choose the valid variable names in SPSS.

(a) Product_Sale and Outlet
(b) Product-Sale and Outlet
(c) Product_Sale and 3Outlet
(d) Product-Sale and 3_Outlet

13. If three groups of students are compared on their work efficiency and in each group there are 12 subjects, what would be the degrees of freedom for the within group in one-way ANOVA?

(a) 30
(b) 31
(c) 32
(d) 33

14. Choose the correct model in one-way ANOVA.

(a) $TSS = (SS)_b + (SS)_w$
(b) $TSS = (SS)_b - (SS)_w$
(c) $TSS = (SS)_b \times (SS)_w$
(d) $TSS = (SS)_b/(SS)_w$

15. In one-way ANOVA F-test, if SS_w decreases (other sums of squares and degrees of freedom remain the same), then which of the following is true?

(a) The value of the test statistic increases.
(b) The p value increases.
(c) Both (a) and (b).
(d) Neither (a) nor (b).

16. In a one-way ANOVA, the p value associated with F-test is 0.100. If the level of significance is taken as .05, what would you do?

(a) Reject H_0, and it is concluded that some of the group population means may differ.
(b) Reject H_0, and it is reasonable to assume that all the group population means are equal.
(c) Fail to reject H_0, and it is concluded that some of the group population means differ.
(d) Fail to reject H_0, and it is reasonable to assume that all the group population means are equal.

17. In one-way ANOVA, four groups were compared for their memory retention power. These four groups had 8, 12, 10, and 11 subjects, respectively. What shall be the degree of freedom of between groups?

    (a) 41
    (b) 37
    (c) 3
    (d) 40

18. If motivation has to be compared among the employees of three different units using one-way ANOVA, then the variables Motivation and Units need to be selected in SPSS. Choose the correct selection strategy.

    (a) Motivation in “Factor” section and Plant in “Dependent list” section.
    (b) Motivation in “Dependent list” section and Plant in “Factor” section.
    (c) Both Motivation and Plant in “Dependent list” section.
    (d) Both Motivation and Plant in “Factor” section.

Assignments

1. A CFL company was interested to know the impact of weather on the life of the bulb. The bulb was lit continuously in hot humid and cold environmental conditions till it was fused. The following are the number of hours it lasted in different conditions:

Life of bulbs (in hours) in different environmental conditions

S.N.	Humid	Hot	Cold
1	400	450	520
2	425	460	522
3	423	480	529
4	465	490	521
5	422	540	529
6	435	580	540
7	444	598	579
8	437	589	595
9	437	540	510
10	480	598	530
11	475	578	567
12	430	549	529
13	431	542	523
14	428	530	510
15	412	532	570

   Apply one-way analysis of variance and test whether the average life of bulbs are same in all the weather conditions. Test your hypothesis at 5% level of significance as well as 1% level of significance.

2. It was experienced by a researcher that the housewives read local news with more interests in comparison to the news containing health information and read

health news with more interest in comparison to that of international news. To test this hypothesis, ten housewives were selected at random in each of the three groups. First group was given an article containing local news for reading, the second group read an article about health, whereas the third group was given an article related with international news. After an hour, subjects in each of these groups were tested for a recall measure test where they were asked true-false questions about the news story they read. The scores so obtained on the recall measure test in all the three groups are shown below:

Data on recall measure in three groups

Local news	Health news	International news
15	14	10
16	12	8
15	14	12
18	16	11
12	11	13
16	14	16
17	14	9
15	12	8
16	13	12
15	13	12

Apply one-way ANOVA and discuss your findings at 5% as well as 1% levels.

Answers to Multiple-Choice Questions

Q.1	a	Q.2	a	Q.3	d	Q.4.	c
Q.5	c	Q.6	b	Q.7	b	Q.8.	d
Q.9	c	Q.10	a	Q.11	c	Q.12.	a
Q.13	d	Q.14	a	Q.15	a	Q.16.	d
Q.17	c	Q.18	b				