

Means Analysis

Learning Objectives

After the completion of this chapter, you will be able to

- Explain the concept of means procedure
- Discuss the statistics related to means procedure
- Demonstrate the steps used in SPSS to execute means procedure
- Explain how to analyse and interpret the means statistics
- Discuss how to compute means by incorporating one and two additional variables
- Explain the importance of one-way ANOVA statistics during application of means procedure and how to interpret the results
- Discuss the concept and interpretation of Eta and Eta squared
- Explain the advantages and disadvantages of Eta squared

18.1 INTRODUCTION

The previous chapters focused on frequency analysis through graphical displays such as bar chart, line chart, pie chart, histogram and cross tabulation. Each graph displays a particular pattern and is applied for a specific purpose. Hence, it is required to be explicit in describing the terminology germane for each display. In this chapter, we extend our discussion to compute means of subgroups of sample data. The means analysis is applicable on metric (interval and ratio scale) data with respect to each category of explanatory variable measured at nominal scale. At the most basic level, the means analysis breakdowns the sample into distinct categories of explanatory variable and shows the average of the dependent variable for each of the respective category. For instance, we can compute the mean of Internet data usage amount between male and female telecom subscribers for a certain period of time. The data usage amount is measured at continuous scale, and its mean could be computed for two distinct categories of gender to generate inferences.

In means procedure, the SPSS also produces certain statistics along with the mean calculation such as median, grouped median, SEM, sum, number of cases, minimum and maximum value, range, variable value of first and last categories of grouping variable, standard deviation, variance, kurtosis, skewness, standard error of kurtosis and skewness and percentage of total sum. ANOVA and Eta table options are also available for first layer (independent variable) analysis. The ANOVA is used to analyse the means difference among the subgroups of the independent variable at certain LoS. Eta is used to assess the effect size in independent groups

of ANOVA. Eta squared indicates the proportion of variance in dependent variable caused by the presence of the independent variable.

Using appropriate scale of variable(s) during the means procedure is an important part for the accuracy of results in the means analysis. The selection of the dependent and independent variables is to be decided as per the research objectives and hypotheses poised in the study. The dependent variable should be measured at metric scale, whereas independent variable on non-metric (categorical) scale. In means procedure, the sample data is divided with respect to distinct categories of the nominal-based independent variable and mean is computed for each category of the independent variable. Thus, the means procedure explores the data for detailed comprehension.

18.2 STATISTICS RELATED TO MEANS PROCEDURE

The means procedure computes various descriptive statistics aiming to an in-depth understanding of data for various subgroups. One of the greatest advantages for computing statistics from mean procedure is to derive inferences for distinct categories of the independent variable. This could not be possible by analysing merely the frequency or descriptive statistics as discussed in the previous chapters. The various statistics associated with means procedure are as follows:

- **Median:** The median resembles the 50th percentile and indicates the value at above and below half of the cases that appear. It is computed by averaging two middle cases by arranging the cases in ascending or descending order (when the cases are even). The median is not affected by the extreme values in data set.
- **Standard error of mean:** The standard deviation of mean statistic derived from the sampling distribution is known as SEM. It measures the accuracy of the sample and indicates whether it represents the population. In other words, we can analyse the difference between the observed mean and hypothesized mean if the ratio of the difference to the standard error is less than -2 or greater than $+2$.
- **Sum:** The sum or total of the values for all the cases.
- **Minimum:** The smallest value of the variable based on the metric data.
- **Maximum:** The largest value of the variable based on the metric data.
- **Range:** The difference between the maximum and minimum value of the variable.
- **Variance:** Variance indicates the measure of dispersion around the mean and equals to the sum of squared deviation from the mean divided by one less than the number of cases for the respective sample.
- **Kurtosis:** A measure to indicate how much data is far or closer to the mean value. The shape of distribution will become more pick or flat in case of data close or away from the mean, respectively. The value of kurtosis is zero for the normal distribution. High positive kurtosis indicates more pickiness with shorter tails, whereas negative kurtosis creates flatness with shorter tails in data distribution.
- **Standard error of kurtosis:** It is used to measure the normality in data. If ratio of kurtosis to its standard error is less than -2 or greater than $+2$, the data will deviate from the normality.
- **Skewness:** It is a measure of lack of symmetry in distribution and indicates spreading of values from the mean in distribution. It also measures asymmetry in distribution. In

negative and positive skewed distribution, majority of values are spread above and below the mean, respectively.

- **Standard error of skewness:** It is also used to measure the normality. If the ratio of skewness to its standard error is less than -2 or greater than $+2$, the data will deviate from the normality.
- **Per cent of total sum:** Percentage of the total sum of each category for concerned variable.
- **N:** The total number of cases or objects in sample or population.
- **Per cent of total N:** Percentage of total number of cases for each category for the concerned variable.
- **ANOVA table and Eta squared (η^2):** The analysis of variance is used to test the hypothesis for measuring the mean difference among the various categories of the grouping variable. Eta squared is also known as correlation ratio or R^2 . It is computed by dividing the sum of squares for treatment effect (SS_{between}) by total sum of squares (SS_{total}). It presents the proportion of total variability in dependent variable accounted for by the variation in the independent variable. The partial Eta squared is the ratio of sum of squares effect (SS_{between}) to sum of squares effect plus error (SS_{within}).

$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{within}}}$$

- **Test of linearity:** It measures the strength of relationship among the groups and their means in ANOVA test.

18.3 EXECUTING MEANS PROCEDURE BY USING SPSS

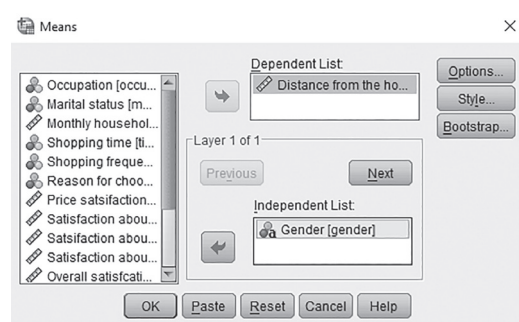
The mean or arithmetic average of the dependent variable for respective categories of the independent variable is computed with SPSS. The steps mentioned in Exhibit 18.1 are used to compute mean of *distance* for each category of *gender*.

Exhibit 18.1. Use retail.sav » Menu bar » Analyse » Compare means » Means » Select distance and transfer to the Dependent List » Select gender and transfer to the Independent List » Press OK

The dialog box appears on the screen as shown in Figure 18.1.

The output viewer from SPSS displays the mean of *distance* for distinct categories of gender as follows.

Figure 18.1 Main Dialog Box of Means



Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

Table 18.1 Means of *distance*

<i>Gender</i>	<i>Distance</i>	<i>Gender</i>	<i>Distance</i>
Male	6.0	Female	5.5

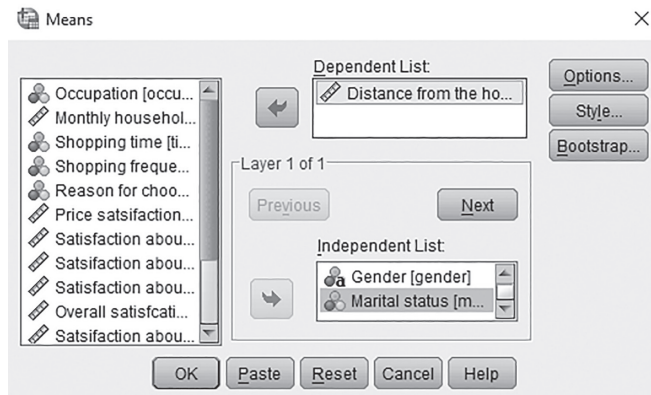
The output (Table 18.1) shows almost equal mean ratings of *distance* for both the categories of gender. Nevertheless, the male respondents covered more distance to reach the departmental store than the females. The number of means in the output is equal to the categories assigned for independent variable in the data set. Hence, interpretation can be made by comparing the dependent variable for every subgroup of population. The missing value, response error or non-existence of data for any subgroup in the population is also required to be considered for valid analysis.

The same dialogue box (Figure 18.1) could be used to create many tables at one time. SPSS produces various tables simultaneously based on the number of variables entered in the dependent and independent list during the execution process. By transferring two independent variables in the *Independent List* box, the means of the respective dependent variable is shown in distinct tables with corresponding to the distinct categories of each independent variable separately. The steps used to compute the means of *distance* belonging to the categories of *gender* and *marital status* are shown in Exhibit 18.2.

Exhibit 18.2. Use retail.sav » Menu bar » Analyse » Compare means » Means » Select distance and transfer to the Dependent List » Select gender and marital, and transfer to the Independent List » Press OK

The dialog box appears on the screen as shown in Figure 18.2.

Figure 18.2 Main Dialog Box of Means: Two Independent Variables



Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

Table 18.2a Means of *distance* for *gender*

Gender	Distance Covered (km)
Male	6.0
Female	5.5
Total	5.8

Table 18.2b Means of *distance* for *marital*

Marital Status	Distance Covered (km)
Married	6.1
Single	5.2
Total	5.8

The output viewer from SPSS displays means of *distance* for different categories of *gender* and *marital* status as shown in Tables 18.2a and 18.2b.

The outputs generated by the steps mentioned in Exhibit 18.2 show the following:

- The average distances covered by male and female respondents are 6 and 5.5 km, respectively. Similarly, the average distances covered by married and single are 6.1 and 5.2 km, respectively.
- All the cases covered the average distance of 5.8 km.

18.4 COMPUTING MEANS WITH ONE ADDITIONAL LAYERING VARIABLE

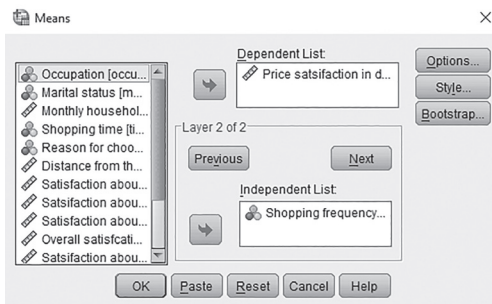
The layering variable is used as a control variable aiming to ascertain the relationship between two or more variables in cross-tabulation. Due to addition of one layering variable along with the independent variable, the parent table will be subdivided into certain panels as per the levels of the controlling variable. This option is used to measure the moderating effect (if any) of controlling variable on the pre-existing relationship between dependent and independent variables. For instance, we would measure the effect of *shopping frequency* as a controlling variable on the relationship between *gender* (independent variable) and *price* satisfaction (dependent variable) by using the steps mentioned in Exhibit 18.3.

Exhibit 18.3. Use retail.sav » Menu bar » Analyse » Compare means » Means » Select *price* and transfer to the Dependent List » Select *gender* and transfer to the Independent List » Click Next » Select *frequency* and transfer to the Independent List » Press OK

The dialog box appears on the screen as shown in Figure 18.3.

The output viewer from SPSS displays means of *price* satisfaction for different categories of *gender* by considering *shopping frequency* as the controlling variable (Table 18.3).

Figure 18.3 Main Dialog Box of Means: One Layering Variable



Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

Table 18.3 Mean Rating with One Additional Layering Variable

<i>Category-1</i>	<i>Shopping frequency</i>	<i>Mean</i>	<i>Std Deviation</i>	<i>Category-1</i>	<i>Mean</i>	<i>Std Deviation</i>
Male (Price satisfaction = 5.3)	First time	6.1	2.3	Female (Price satisfaction = 6.1)	4.0	0.81
	Once per month	4.5	2.7		6.1	2.3
	Once per week	5.0	2.6		6.6	1.9
	More than once in week	5.1	2.9		6.3	3.3
	Not fixed	6.1	1.9		6.0	0.81
	Total	5.3	2.6		6.1	2.4

The rating of price satisfaction is almost same for male (5.3) and female (6.1) customers. In this situation, if we divide the cases into distinct categories of shopping frequency by considering as layering variable, the behavioural differences of price satisfaction can be observed due to the interaction effect of shopping frequency. The female respondents belonging to 'first time' shopping frequency are least satisfied (4.0) with price as compared to males (6.1). Likewise, the price satisfaction is also different between male and female customers for the shopping frequency of 'once per month', 'once per week' and 'more than once in week'. Hence, we could interpret the change in main effect of one variable (price satisfaction) due to interaction effect of another variable (shopping frequency) in the analysis as shown in Table 18.3.

18.5 COMPUTING MEANS WITH TWO ADDITIONAL LAYERING VARIABLES

With addition of one more layering variable as controlling variable, the parent table will be subdivided into certain panels as per the combinations of each category of the first layering variable corresponding to each category of second layering variable until all possible combinations appear. Both layering variables should be measured at nominal scales. For instance, if we add *marital* status as second layering variable in the previously discussed scenario, the distinct combinations of second layering variable will be generated by software for each category of the shopping frequency (first layering variable) corresponding to each gender. In this procedure, the combinations will become more complex and incorporating controlling variables in each proceeding step leads to reduction in the observed frequency.

18.6 DESCRIPTIVE STATISTICS WITH MEAN PROCEDURE

Descriptive statistics such as median, SEM, sum, number of cases in mean, minimum, maximum, range, standard deviation, variance, kurtosis, skewness, standard error of kurtosis and skewness, and percentage of total sum are computed by SPSS to gain better understanding of data summary. These descriptors are also compared among different categories of the independent variable. The steps used to ascertain the descriptors of *price* for various categories of income groups are shown in Exhibit 18.4.

Figure 18.4a Main Dialog Box: Means Procedure

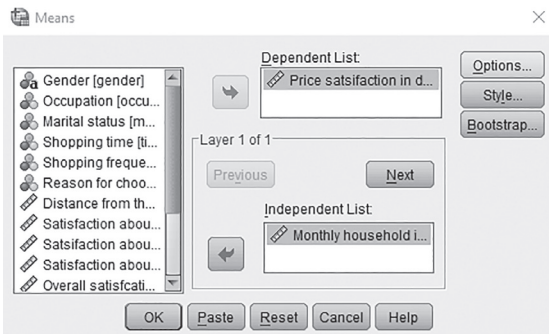
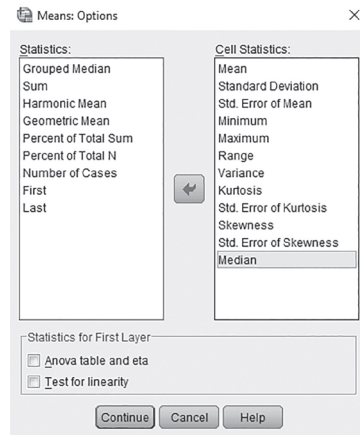


Figure 18.4b Means Procedure: Options



Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

Exhibit 18.4. Use retail.sav » Menu bar » Analyse » Compare means » Means » Select price and transfer to the Dependent List » Select income and transfer to the Independent List » Click » Option and transfer desired descriptive statistics into Cell Statistics » Continue » Press OK

The dialog boxes appear on the screen as shown in Figures 18.4a and 18.4b.

The output viewer from SPSS displays the means of *price* satisfaction for different categories of *income* as shown in Table 18.4.

Table 18.4 Descriptive Statistics with Mean Procedure

Descriptors	Monthly Household Income Categories (INR)					
	<10,000	10,001–20,000	20,001–30,000	30,001–40,000	>40,000	Total
Mean	5.6	6.1	5.4	5.7	5.4	5.6
Std deviation	2.4	2.2	2.4	3.2	2.04	2.5
SEM	0.47	0.54	0.51	0.58	0.43	0.23
Minimum	1.00	4.00	1.00	1.00	1.00	1.00
Maximum	10.00	10.00	10.00	10.00	9.00	10.00
Range	9.00	6.00	9.00	9.00	8.00	9.00
Variance	6.08	5.11	6.07	10.46	4.16	6.56
Kurtosis	-0.586	-1.108	-0.810	-1.333	0.438	-0.737
Std error of kurtosis	0.872	1.063	0.935	0.821	0.953	0.438
Skewness	-0.102	0.681	-0.034	-0.068	-0.723	-0.042
Std error of skewness	0.448	0.550	0.481	0.421	0.491	0.221
Median	5.0	5.0	6.0	6.0	6.0	6.0

The descriptors as mentioned in Table 18.4 are used to make the following interpretations:

- Mean rating of price satisfaction is almost same for all the income groups.
- Maximum variation occurs in the income group 30,001–40,000 INR due to the highest standard deviation (3.2) and variance (10.46).
- The lowest SEM (0.43) belongs to the income group of more than 40,000 INR, thus indicates more stability in data as compared to the rest of the groups.
- All income groups are negative kurtosis except the last category of income (>40,000). Data distribution of income group 30,001–40,000 INR is more flat as compared to other categories.
- All income categories are negatively skewed, except the income group 10,001–20,000 that is positively skewed (0.681).

18.7 ANOVA AND ETA

The means procedure also examines the ANOVA and Eta for the independent variable in the first-layer scenario. One-way ANOVA is used to measure whether the mean difference is significant across the categories of the first-layer independent variable. The p -value corresponding to between-groups of the independent variable appears in one-way ANOVA table. Other measures like Eta and Eta squared are the most common reported estimates of effect sized in case of ANOVA results. SPSS actually computes Eta squared. The Eta squared indicates the percentage by which the dependent variables are influenced by independent variable. Higher the value of Eta squared, higher will be the influence of the independent variable. Higher the F -statistics, lower will the probabilistic (p -value), and the Eta value will become close to 1.

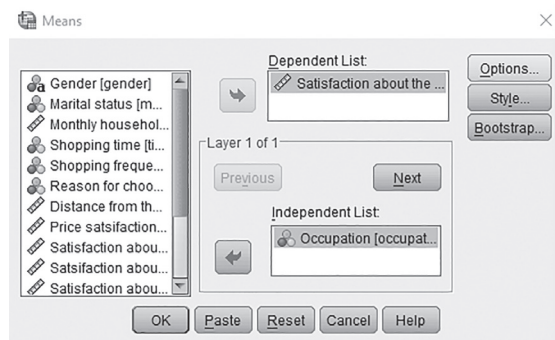
The steps mentioned in Exhibit 18.5 are used to analyse the variance among various categories of different *occupation* for *parking* as dependent variable. We would select ANOVA table and Eta under the statistics for the first layer.

Exhibit 18.5. Use retail.sav » Menu bar » Analyse » Compare means » Means » Select parking and transfer to the Dependent List » Select occupation and transfer to the Independent List » Click Option » Select Anova table and eta and Test for linearity in the area of Statistics for First Layer » Continue » Press OK

The dialog boxes appear on the screen as shown in Figures 18.5a and 18.5b.

The output viewer from SPSS shows the following results of ANOVA table and Eta as shown in Tables 18.5a–18.5c.

Figure 18.5a Main Dialog Box: Means Procedure



Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

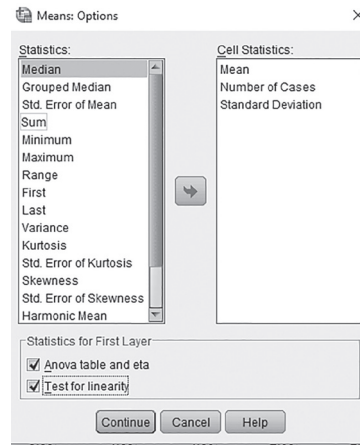
18.7.1 Descriptors of Sample

Table 18.5a reveals simply mean score, standard deviation and frequency of different occupation. The differences in means among all the categories are easily interpreted as shown below.

18.7.2 ANOVA Statistics

Table 18.5b presents the result of one-way ANOVA indicating *F*-statistics. The *F*-statistic is the ratio of variance between and within the sample. The sum of square between groups is calculated by computing sum of the squared deviations between the grand mean and each group mean, multiplied by the number of subjects in each group and sum of square within groups, is calculated by computing sum of the squared deviations between the mean for each category (occupation) and the observed value of each subject within that group. The degree of freedom between-groups is the number of groups minus one ($k-1$, i.e., $5-1 = 4$) and degree of freedom within-groups is the number of subjects or cases minus number of groups ($n-k$, i.e., $120-4 = 116$). The mean sum of squares for between-groups and within-groups are calculated by dividing sum of squares with respective degree of freedom. Finally, the *F*-statistics is calculated by the ratio of between-groups mean square divided by within-groups mean square. The last column of the table indicates the significance of probability about the differences in means that occur due to chance only. The lower *p*-value from 0.05 ($p < 0.05$, 0.000) at 5 per cent LoS indicates that a significant difference exists among the mean scores for categories of occupation.

Figure 18.5b Means Procedure: ANOVA



Reprint Courtesy: IBM.

Table 18.5a Means: Satisfaction for Parking

Occupation	N	Mean	Std Deviation	Occupation	N	Mean	Std Deviation
Business	36	1.4	0.69	Students	16	2.0	1.2
Government employee	22	1.5	0.74	Housewife	15	4.1	0.63
Private employee	31	3.2	1.34	Total	120	2.3	1.39

Table 18.5b ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
Parking × Occupation	Between-groups (Combined)	120.6	4	30.1	31.50	0.000
	Within-groups	110.0	115	0.957		
	Total	230.6	119			

18.7.3 Measures of Association

The last portion of this output, Table 18.5c, indicates the correlation between the two (parking × occupation) variables based on Eta. Eta measures the correlation between one nominal and another ratio or interval-based variable. The value 0.723 indicates strong correlation between the variables. The ETA squared indicates the proportion of the variance in dependent variable explained by the independent variable (Fritz, Morris, and Richler 2012). The value 0.523 indicates that 52% of variance in *parking* satisfaction is explained due to five different categories of occupation in the target population.

Table 18.5c Measures of Association

	<i>Eta</i>	<i>Eta Squared</i>
Parking × Occupation	0.723	0.523

Eta squared (η^2) is also known as the correlation ratio or R^2 and computed as dividing sum of square between the groups by total sum of squares (Levine and Hullett 2002). The formulae of η^2 is:

$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}}, \text{ thus } \eta^2 = \frac{120.6}{230.6} = 0.522.$$

18.8 ADVANTAGES OF ETA SQUARED

Eta squared is an important measure with a number of properties to explain effect size in statistical inferences; nevertheless, careful interpretation should be made by researchers due to certain disadvantages also. Following are the main advantages of Eta squared:

- It measures the association between two variables when one variable is on non-metric scale.
- Eta squared is often interpreted in terms of the percentage of variance accounted for by explanatory variable in the model and assumed to be a useful way for finding effect size.
- As Eta squared is always either equal or smaller to the partial-eta squared, it is considered more conservative method used by many researchers during interpretation.

SUMMARY

- **Concept of means procedure:** Means analysis is applicable on metric (interval or ratio scale) data with respect to each category of nominal-based independent variable. At the most basic level, means analysis breaks down the sample into distinct categories of the explanatory variable and shows the average of the dependent variable for each category.

- **Statistics related to means procedure:** Various statistics such as median, grouped median, SEM, sum, number of cases, mean, minimum, maximum, range, variable value of first and last category of the grouping variable, standard deviation, variance, kurtosis, standard error of kurtosis and skewness, and percentage of total sum can also be computed through means procedure. Descriptive statistics are calculated for metric variable corresponding to different categories of the independent variable.
- **Executing means procedure by using SPSS:** The following steps are used in SPSS to execute means procedure:

Computing mean for one independent variable containing distinctive categories »
 Computing mean for more than one independent variable » Computing mean by
 incorporating one or two layering variables » Descriptive statistics of dependent variable
 » ANOVA table and Eta squared » Linearity

- **Means for one and two additional variables:** The option layering variables are used to add more than one controlling variable to measure moderating effect on the relationship that exists between dependent and independent variables.
- **Interpretation of one-way ANOVA:** One-way ANOVA in means procedure is used to find whether means of different categories of the first layer independent variable significantly differ at certain LoS. Eta squared is used to measure the strength of association between the two variables when one is based on nominal and another is on ratio or interval scale.
- **Concept of Eta and Eta squared:** These measures are the most common estimates of effect size for ANOVA test. They indicate the percentage by which a dependent variable is influenced by an independent variable. Higher the value of Eta squared, higher will be the influence of the independent variable.

KEY TERMS

Controlling variable
 Correlation
 Degree of freedom
 Descriptive statistics
 Effect size
 Eta and Eta squared
F-statistics
 Interaction effect
 Layering variable

Means analysis
 Means squares between- and within-groups
 Measures of association
 Moderating variable
 One-way ANOVA
 Partial-eta squared
 Sum of square between-groups
 Sum of square within-groups
 Uncertainty coefficient

Solved Example

Q1. As a research analyst, you conduct a survey among 15 people of three different age (younger, middle-aged and older) groups aiming to measure their preference about using plastic money during routine transactions. Following data presents the measurement of preference on 5-point rating scale (from 1 = least preferred to 5 = highly preferred) along with respective gender of the respondents.

<i>Case</i>	<i>Age Group</i>	<i>Gender</i>	<i>Rating</i>	<i>Case</i>	<i>Age Group</i>	<i>Gender</i>	<i>Rating</i>
1	1	1	5	9	3	1	4
2	2	2	4	10	1	2	4
3	3	1	2	11	2	1	1
4	1	2	5	12	3	2	3
5	2	1	1	13	1	1	4
6	3	2	4	14	2	2	2
7	1	1	5	15	3	1	5
8	2	2	3				

Use code 1 = younger, 2 = medium age and 3 = older; and gender 1 = male, 2 = female.

Answer the following questions:

- Prepare data file in SPSS by using appropriate variables and scales.
- Comment on the preference level due to main effect of *gender* and *age*.
- Is there any effect of *age* by considering as controlling variable?
- Analyse the results by using ANOVA statistics and Eta?

Solution

- Following steps are used to prepare data file:

Open SPSS Windows » Select variable view » Type first variable *age* (in extreme left corner) » Use string » Click value column » Type 1, 2 and 3 in value option and mention younger, medium and older in label option by using *add* button one-by-one » Type second variable *gender* » Use string » Type 1 and 2 in value option and mention male and female in label option one-by-one » Type third variable *rating* and use numeric » Select data view » Assign rating preference for each case.

The variable and data views generated based on the above commands are shown in Figures 18.6a and 18.6b.

Figure 18.6a Creating Variables

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	age	String	8	0		{1, younger}...	None	8	Left	Nominal	Input
2	gender	String	8	0		{1, male}...	None	8	Left	Nominal	Input
3	rating	Numeric	8	2		None	None	8	Right	Unknown	Input
4											

Figure 18.6b Data Entry for Variables

	age	gender	rating				
				9	3	1	4
1	1	1	5	10	1	2	4
2	2	2	4	11	2	1	1
3	3	1	2	12	3	2	3
4	1	2	5	13	1	1	4
5	2	1	1	14	2	2	2
6	3	2	4	15	3	1	5
7	1	1	5				
8	2	2	3				

Source: Reprint Courtesy of International Business Machines Corporation, © International Business Machines Corporation.

- b. In order to find out the main effects of *gender* and *age*, the steps as mentioned in Exhibit 18.2 are followed and the output viewer of SPSS produced the following tables.

Table 18.6 Descriptive Statistics: Solution

	Age			Gender	
	Mean	Std Deviation		Mean	Std Deviation
Younger	4.60	0.548	Male	3.38	1.768
Medium	2.20	1.304	Female	3.57	0.976
Older	3.60	1.140			

It can be seen that male and female have an almost equal mean ratings of preference, whereas this difference is quite observable among three different age groups. Younger people having highest preference for using plastic money (4.6) followed by older (3.6) and medium age (2.2) groups.

- c. The steps as mentioned in Exhibit 18.3 are used in order to measure the effect of *age* as a controlling variable. Following output is created by SPSS by considering *age* as first layering variable.

Table 18.7 Effects of Controlling Variable

<i>Gender</i>	<i>Age</i>	<i>Mean</i>	<i>Std Deviation</i>	<i>Gender</i>	<i>Age</i>	<i>Mean</i>	<i>Std Deviation</i>
Male	Younger	4.67	0.577	Female	Younger	4.50	0.707
	Medium	1.00	0.000		Medium	3.00	1.000
	Older	3.67	1.528		Older	3.50	0.707

The above-mentioned results indicate the change in main effect of *gender* due to interaction of another variable included as *age* in the situation. Hence, genderwise mean difference (2.0) could be easily observed between male and female consumers belonging to the medium age group.

- d. We would apply ANOVA statistics and Eta to examine the mean difference among three different age groups by using the steps as mentioned in Exhibit 18.4 and the following results are created.

Table 18.8 AVONA Table: Solution

		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
rating * age	Between-groups (Combined)	14.533	2	7.267	6.606	0.012
	Within-groups	13.200	12	1.100		
	Total	27.733	14			

The last column of the table indicates the *p*-value associated to *F*-statistic. The lower *p*-value from 0.05 ($p < 0.05$, 0.012) at 5% LoS indicates the significant difference in mean scores of preference for using plastic money among the three different categories of age groups.

Table 18.9 Correlation: Nominal by Nominal

<i>Measures of Association</i>		
	<i>Eta</i>	<i>Eta Squared</i>
rating * age	0.72	0.524

ETA measures the correlation between one nominal (*age*) and other interval (*preference* rating) based variable. The value 0.73 indicates a strong association between the variables. The value 0.524 Eta squared indicates, that 52 per cent of variance in *preference* is explained due to three different age categories of the respondents in the given scenario.

Hands-on Practice

- Q1.** In an open-door competition of 6th standard students, the marks obtained in math and science have been recorded for randomly selected 10 students belonging to Sections A–C. The data is presented as follows:

<i>Section A</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Math	13	15	17	12	11	14	10	9	14	12
Science	16	17	6	13	12	3	15	12	11	10
<i>Section B</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Math	13	14	11	12	11	15	12	11	5	6
Science	14	14	5	16	11	18	8	16	12	18
<i>Section C</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Math	13	5	14	16	18	11	15	5	13	16
Science	13	14	14	15	11	16	9	15	11	19

Use Code 1 = Section A, 2 = Section B and 3 = Section C; and Code 1 = Math and Code 2 = Science.

Answer the following questions:

- a. Prepare data set in SPSS by incorporating three-variable section based on nominal scale, whereas math and science as ratio scale.
- b. Explore mean procedure for the scores of math and science among all three different sections and interpret the results.

Problem Based on Real Data

- Q1.** A Web link at [https://www.kaggle.com/yersever/500-person-gender-height-weight-body-massindex/version/2#_ =](https://www.kaggle.com/yersever/500-person-gender-height-weight-body-massindex/version/2#_=) displays a dataset of 500 persons' body information from open data source Kaggle Inc. (2019). Dataset include the variables gender (male/female), height (cm), weight (kg) and index (from 0 = extremely weak, 1 = weak, 3 = normal, 4 = obesity and 5 = extreme obesity). Do the following tasks based on the existing dataset.
- a. Compute the mean height and weight for male and female genders.
 - b. Compute the mean height and weight for 0 to 5 index. Identify the index displaying the highest mean height and weight.
 - c. Compute the mean height for male and female genders with respect to index from 0 to 1.
 - d. Use ANOVA and Eta commands and find out the difference in mean height and weight among various index (from 0 to 5). Interpret the results based on the *F*-statistics and Eta.

REFERENCES

- Kaggle Inc. 2019. *500 Persons Gender-Height-Weight-Body Mass Index*. Available at: <https://www.kaggle.com/yervever/500-person-gender-height-weight-bodymassindex> (accessed on 10 December 2018).
- Levine, T. R., and C. R. Hullett 2002. 'Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research'. *Human Communication Research* 28 (4): 612–625.
- Fritz, C. O., P. E. Morris, and J. J. Richler. 2012. 'Effect Size Estimates: Current Use, Calculations, and Interpretation'. *Journal of Experimental Psychology: General* 141 (1): 2–18.