# Chapter 9
# Analysis of Covariance: Increasing Precision in Comparison by Controlling Covariate

**Learning Objectives**

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

- Learn the concept of analysis of covariance.
- Know the application of analysis of covariance in different situation.
- Describe the concept of covariate and neutralize its effect from the treatment effect.
- Know the model involved in the analysis of covariance.
- Understand the concept of adjusting treatment means for covariate using linear regression.
- Understand the analysis of covariance graphically.
- Learn the method of using analysis of covariance.
- Understand as to why the analysis of covariance is efficient design in comparison to one-way analysis of variance.
- To be able to formulate the hypotheses in analysis of covariance.
- Understand the assumptions used in analysis of covariance.
- Know the method of preparing data file for analysis in SPSS.
- Learn the steps involved in using SPSS for analysis of covariance.
- Interpret the output obtained in analysis of covariance.
- Learn the model way of writing the results of analysis.

## Introduction

To compare the effectiveness of two or more treatments on certain criterion variable, we use one-way analysis of variance technique. This technique has been discussed in Chap. 7. In testing the comparative effectiveness of different treatments, the subjects are selected in each experimental group by using the

J.P. Verma, *Data Analysis in Management with SPSS Software*,
DOI 10.1007/978-81-322-0786-3_9, © Springer India 2013

principle of randomization. In a situation if the randomization is not possible, groups are equated on the basis of one or more known parameters. The randomization or matching is done in order to have the similar initial conditions so that whatever the changes in criterion variable occurs in the treatment groups can be attributed due to the treatments only. But in many situations, randomization of subjects or experimental units may not be possible as the experimenter may be forced to take the two or more intact samples from different locations due to administrative or financial constraints. For example, consider an experiment where it is desired to compare the effect of different types of tariff incentives on the mobile recharge revenue. In this situation, an experimenter does not have any choice to select the subjects randomly in different experimental groups. Samples will have to be drawn from the predefined clientele sets of different mobile companies. In such situations, groups are not homogeneous initially. These subjects in intact groups may differ in so many ways which might affect their behavior pattern. Thus, statistical control or indirect procedure is necessary to reduce the experimental error which causes due to such initial differences in the groups.

In experimental research, the individual variations that appear within the measures on the criterion variable are potentially correlated with some extraneous variable. If the criterion variable is a measure of how well subjects learn English speaking under one or the other of the two methods of instructions, the potential correlates are likely to include parameters such as prior knowledge of grammar, motivation level, aptitude, age, and intelligence. These potential correlates are known as covariates. Analysis of covariance can be used to compare the effectiveness of these instructional methods on learning English speaking after removing the effect of the identified covariates.

## Introductory Concepts of ANCOVA

Analysis of covariance (ANCOVA) is a statistical technique that may be considered as an extension of analysis of variance (ANOVA). Analysis of covariance combines features of one-way analysis of variance with simple linear regression. It is so because the treatment groups are compared like the way we do in analysis of variance and we adjust the measurement on criterion variable on the basis of covariate by using the concept of regression analysis.

In ANCOVA, we try to minimize the error variance by controlling the concomitant variable which varies along with the criterion variable in all the experimental groups. These concomitant variables are known as covariate. Typically, a covariate is highly correlated with a criterion variable; that is, the covariate contains information about the criterion variable and therefore possibly also explains the difference in the treatment groups. The purpose in the ANCOVA design is to isolate the variability component due to covariate so that group difference if any may be solely attributed to the treatments only. Analysis of covariance is used in a situation where there is at least one categorical factor and one or more continuous covariates.

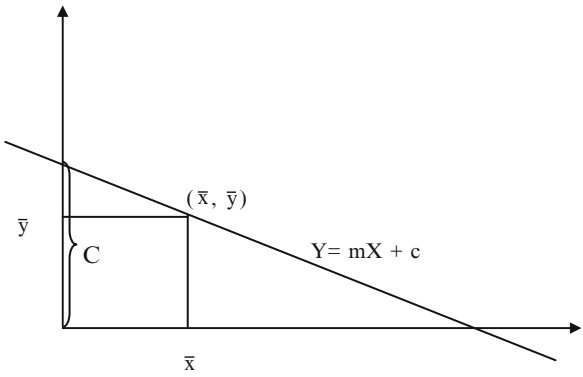


**Fig. 9.1** Regression equation of $Y$ on $X$

Since ANCOVA is the combination of ANOVA and regression analysis, this design can be used with any ANOVA model. One can do a completely randomized design, randomized block design, a Latin square design, or any other design if a covariate is put on it. All we need is one measurement for the covariate to go with every observation. In this chapter, only completely randomized design shall be discussed as an ANOVA model.

## Graphical Explanation of Analysis of Covariance

In order to understand the ANCOVA model, let us first refresh our concept of representing the line in the slope intercept form. You may recall that this line used to be represented by

$$\mathrm{Y} = \mathrm{mX} + \mathrm{c} \tag{9.1}$$

where $m$ is the slope and $c$ is the intercept of the line on $y$-axis. Graphically this equation can be represented by the Fig. 9.1.

Equation of line in any form may be converted to this slope intercept form for comparing their slopes and intercepts.

The line shown in Fig. 9.1 is a regression line for estimating the value of $Y$ if the value of $X$ is known. Now if you look at the vertical line over $\bar{x}$, it intersects the regression line at ($\bar{x}$ and $\bar{y}$). In other words, the point ($\bar{x}$, $\bar{y}$) actually lies on the regression line. This concept shall be used to explain the analysis of covariance.

To understand ANCOVA, let us consider $A$ and $B$ represent the two treatments. Further, $Y_A$ and $Y_B$ represent the value of criterion variable, whereas $X_A$ and $X_B$ represent the value of covariate in the two treatment groups A and B respectively. These two treatments are represented by the lines $A$ and $B$ in Fig. 9.2. If higher value

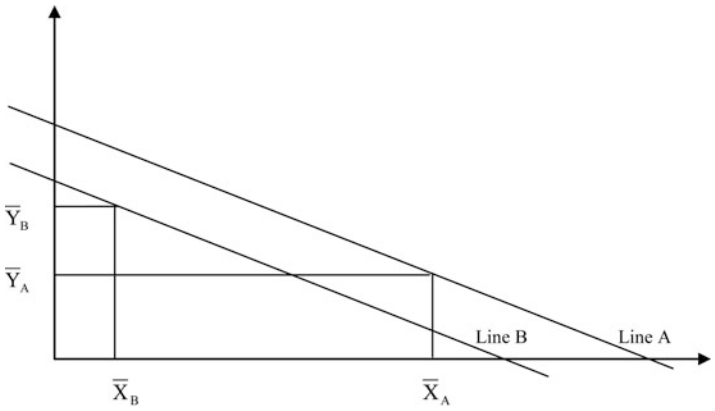


**Fig. 9.2** Geometrical representation of two treatments ($Y's$) with their covariates ($X's$)

of $Y$ indicates better performance, then obviously treatment $A$ would be better treatment because the line $A$ is higher everywhere than line $B$.

But sometimes corresponding to very low value of $\overline{X}_B$ and very high value of $\overline{X}_A$, the values of $\overline{Y}_B$ is higher than $\overline{Y}_A$. This can mislead the researcher. If line $B$ is always higher than line $A$, then the sample means end up reversed. The difference observed in $\overline{Y}'s$ is not because of the treatments but due to the covariate means ($\overline{X}'s$) which are so apart.

In analysis of covariance, what we do is to compare the values of $\overline{Y}_A$ and $\overline{Y}_B$ at $\bar{X}$ (overall grand mean of $\overline{X}_A$ and $\overline{X}_B$). Now the distance between the two curves is the same because the lines are parallel. What we do here is to adjust both means so that we are evaluating the points of the same $X$ on the curves. In this way, we get a more balanced comparison.

If treatments are compared without using ANCOVA, what may happen? Suppose the above-mentioned lines $A$ and $B$ are exactly the same, still $\overline{Y}_i$'s may be different. This may be because the effect of one treatment is observed at higher average covariate mean and the other treatment effect is measured with lower average covariate mean. This fact may not be revealed if the analysis of covariance is not done. Thus, in analysis of covariance, we compare the effect of treatments mean ($\overline{Y}_i$'s) by adjusting them with the average covariate means ($\bar{X}$).

## Analysis of Covariance Model

If $Y_{ij}$ represents the $j$th score of the criterion variable in the $i$th treatment group and $X_{ij}$ represents the $j$th score of the covariate in the $i$th treatment group, then the one-way ANCOVA model is represented as follows:

$$Y_{ij} = \mu + \beta(X_{ij} - \bar{X}) + \varepsilon_{ij} \tag{9.2}$$

where

$\mu$ is the overall population mean (on criterion variable)
$\beta$ is slope of the lines

$\bar{X}$ is combined mean of the covariate in all the treatment groups

$\varepsilon_{ij}$ is unexplained error terms which are independent and normally distributed with mean 0 and variance 1

One-way analysis of covariance fits a straight line to each treatment group of $X$-$Y$ data, such that the slopes of the lines are all equal. This fitted model may then be used to test the following null hypothesis:

$H_0$: The intercepts for each line are equal.

This hypothesis tests as to whether all the treatment group means are equal or not after making the adjustment for $X$ (covariate). Here, we assume that the slopes are equal. It is so because there is no point of comparing the treatments effect if one of the treatments produces positive effect whereas other induces negative effect.

Let us see how the treatment means are adjusted for covariate and are computed for comparison. Adding both sides of Eq. (9.2) for $j$ and dividing by $n$ (number of scores in each treatment group), we get

$$\frac{1}{n}\sum_j Y_{ij} = \mu + \beta\left(\frac{1}{n}\sum_j X_{ij} - \bar{X}\right) + \frac{1}{n}\sum_j \varepsilon_{ij}$$

$$\Rightarrow \quad \overline{Y}_i = \mu + \beta(\overline{X}_i - \bar{X}) + \bar{\varepsilon}_i$$

Since mean of $\varepsilon_i$ is zero, the equation becomes

$$\overline{Y}_i = \mu + \beta(\overline{X}_i - \bar{X}) \tag{9.3}$$

where

$\overline{Y}_i$ is mean of the criterion variable in the $i$th treatment group
$\mu$ is the overall population mean (on criterion variable)
$\overline{X}_i$ is mean of the covariate ($X$ data) in the $i$th treatment group

Other symbols have their usual meanings. If one-way ANCOVA model has two treatment groups, then the model (9.3) will give rise to two straight lines as shown in Fig. 9.2. Thus, by testing the hypothesis $H_0$ in ANCOVA, we actually compare the two treatments $\overline{Y}_A$ and $\overline{Y}_B$ after adjusting it for the covariates.

**Remark**

1. If the slope of line A and B is equal it indicates that the effect of both the treatments are in one direction only. Both the treatments will induce either positive or negative effect.
2. By comparing the intercepts of the lines, we try to compare whether the effect of all the treatments on the criterion variable is same or not.

## What We Do in Analysis of Covariance?

In analysis of covariance, the purpose of the analysis is to compare the posttreatment means of the groups by adjusting the initial variations in the grouping. The statistical control is achieved by including measures on supplementary or concomitant variate ($X$) in addition to the variate of primary interest ($Y$) after implementing the treatments. The concomitant variate that may not be of experimental interest is called covariate and designated as $X$. Let us designate the variable which is of interest in the experiment as $Y$, also known as criterion variable. Thus, in ANCOVA, we have two observations ($X$ and $Y$) from each subject. Measurements on $X$ (covariate) are obtained prior to the administration of treatments and are primarily to adjust the measurements on $Y$ (criterion). Covariate is the variable which is assumed to be associated with the criterion variable. When $X$ and $Y$ are associated, a part of the variability of $Y$ is due to the variation in $X$. If the value of covariate $X$ is constant over the experimental units, there would be corresponding reduction in the variance of $Y$.

Let us consider an example in which the analysis of covariance can be applied to reduce the estimate of experimental error. Suppose an experiment is conducted to study the effect of three different types of advertising campaign on the sale of a product. Further, the experimenter is forced to use three intact groups of outlets from three different states. However, there is a freedom to assign groups randomly to the treatment conditions (types of advertisement). Out of three groups, one may serve as control. Since the subjects cannot be assigned randomly to the treatment groups, the possibility of initial differences (before administration of treatment) among the groups always exist. Thus, one may decide to record the sale of each outlet ($X$) before applying the treatments, which serves as a measure of covariate. This measure of covariate is used to adjust the post advertisement sale figure ($Y$) in different outlets. The $Y$ is the sale of the product which is obtained after the implementation of the treatments (advertisements).

Thus, the variates $X$ and $Y$ can be defined as follows:

$X =$ sale of the product on 15 consecutive days in each experimental groups before treatments (advertisements).

$Y =$ sale of the product on 15 consecutive days in each experimental groups after treatments (advertisements).

In general, if $Y$ measures (criterion) are substantially correlated with $X$ measures (covariate), the analysis of covariance will result in similar estimate of experimental error than would be obtained from the analysis of variance.

Thus, in ANCOVA, the following null hypothesis is tested

$$H_0 : \mu_{\text{Adj\_Post\_Adv\_1}} = \mu_{\text{Adj\_Post\_Adv\_2}} = \mu_{\text{Adj\_Post\_Control}}$$

against the possible alternatives that at least one group mean differs
where

Adj_Post_Adv_1 is adjusted mean sale in the first treatment group (where first advertisement campaign was used)
Adj_Post_Adv_2 is adjusted mean sale in the second treatment group (where second advertisement campaign was used)
Adj_Post_Adv_3 is adjusted mean sale in the third treatment group (where no advertisement campaign was used)

ANCOVA table generated in the SPSS output contains the value of $F$-statistics along with their significance value. Thus, if $F$-value is significant, the null hypothesis is rejected and the post hoc test is used to compare the adjusted posttreatment means of different groups in pairs.

## When to Use ANCOVA

The analysis of covariance is used to test the comparative effectiveness of two or more treatments on the criterion variable after adjusting for their initial differences due to covariate. The covariate should be identified before the experiment, and its value should be measured on each of the experimental units. In many situations, it is not possible to identify single covariate which affects the measure on criterion variable during experimentation. In that case, initial testing ($X$) on the criterion variable in each of the treatment group may be considered as covariate, and the measure on the criterion variable after the treatment ($Y$) in all the treatment groups is the one in which we are interested to investigate. The analysis of covariance design should be used if the following things happen:

(a) The response on the criterion variable is continuous.
(b) There are one or more classification variables (treatment groups).
(c) There are one or more continuous independent variables (covariate).

## Assumptions in ANCOVA

In using the analysis of covariance design following assumptions are made:

1. The criterion variable must have the same variance in each of the treatment groups.
2. The data on criterion variable must have been obtained randomly.
3. The interaction between the criterion variable and covariate is negligible. The adjusted mean of the criterion variable in each of the treatment groups is computed owing to this assumption. If this assumption is violated, then the adjustment of the criterion variable to a common value of the covariate will be misleading.
4. Since ANCOVA uses the concept of linear regression, the assumption of linearity between independent and dependent variable must hold true.
5. The regression coefficients (slope) for each treatment groups must be homogeneous. If this assumption is violated, then the ANCOVA results will be misleading.

## Efficiency in Using ANCOVA over ANOVA

The ANCOVA design is more efficient in comparison to one-way ANOVA. It is because of the fact that in ANCOVA, a part of the variability due to error component is defined by the covariate and, hence, the error variance reduces comprehensively in comparison to one-way ANOVA design. In one-way ANOVA, the total variability is split into two components, that is, between groups and within groups. Here, the variability due to covariate is confounded into error component, and, hence, this design is inferior to ANCOVA in a situation where the covariate effects the measurement on criterion variable. In fact, one-way ANOVA should be used only in a situation where it is known that all the treatment groups are homogenous in all respect and perfect control is observed during the entire period of experimentation.

## Solved Example of ANCOVA Using SPSS

**Example 9.1** A study was planned to investigate the effect of different doses of vitamin C in curing the cold. Forty five subjects who were suffering from cold symptoms were divided into three groups. The first two groups were given a low dose and high dose of vitamin C every day whereas the third group was given a placebo. The number of days these subjects were suffering from cold before starting the treatment was taken as the covariate whereas the curing time in each treatment group was recorded as a dependent variable. The data so obtained on the subjects are shown in the Table 9.1.

**Table 9.1** Data on cold duration before and during implementation of vitamin C in different groups

| S.N. | Contents of vitamin C | | | | | |
|---|---|---|---|---|---|---|
| | High dose | | Low dose | | Placebo | |
| | Pre days | Post days | Pre days | Post days | Pre days | Post days |
| 1 | 0 | 2 | 14 | 12 | 1 | 10 |
| 2 | 10 | 3 | 16 | 13 | 10 | 8 |
| 3 | 11 | 5 | 5 | 8 | 5 | 14 |
| 4 | 15 | 9 | 12 | 10 | 6 | 9 |
| 5 | 6 | 3 | 0 | 1 | 10 | 13 |
| 6 | 12 | 8 | 8 | 4 | 5 | 11 |
| 7 | 9 | 7 | 12 | 9 | 12 | 15 |
| 8 | 13 | 7 | 5 | 10 | 13 | 15 |
| 9 | 1 | 6 | 19 | 10 | 6 | 10 |
| 10 | 8 | 13 | 14 | 8 | 19 | 20 |
| 11 | 7 | 12 | 6 | 11 | 8 | 12 |
| 12 | 6 | 10 | 8 | 11 | 8 | 14 |
| 13 | 4 | 3 | 5 | 8 | 6 | 12 |
| 14 | 3 | 2 | 2 | 6 | 5 | 9 |
| 15 | 4 | 3 | 4 | 6 | 8 | 14 |

Pre days: Cold duration before treatment
Post days: Cold duration during treatment

Apply analysis of covariance to see as to which dose of vitamin C is more effective in controlling cold. Test your hypothesis at 5% level of significance.

*Solution* In this example, the variables are as follows:

*Treatments:* The three treatments are as follows:
Treatment *A*: Administering high dose of vitamin C
Treatment *B*: Administering low dose of vitamin C
Treatment *C*: Administering placebo

*Covariate*: The number of days having cold symptoms before treatment.
*Dependent Variable*: *Curing time of cold symptoms*

Here, it is required to compare the average curing time among the three treatment groups, that is, high dose of vitamin C, low dose of vitamin C, and placebo, after adjusting for the covariate (average number of days having cold symptoms before treatments).

Thus, the following null hypothesis needs to be tested

$$H_0 : \mu_{\text{Adj\_Days\_in\_Treatment\_}A} = \mu_{\text{Adj\_Days\_in\_Treatment\_}B} = \mu_{\text{Adj\_Days\_in\_Treatment\_}C}$$

against the alternative hypothesis that at least one group mean (adjusted) is different

where

$\mu_{\text{Adj\_Days\_in\_Treatment\_}A}$ is adjusted mean curing time in treatment group $A$

$\mu_{\text{Adj\_Days\_in\_Treatment\_}B}$ is adjusted mean curing time in treatment group $B$

$\mu_{\text{Adj\_Days\_in\_Treatment\_}C}$ is adjusted mean curing time in treatment group $C$

The SPSS output provides ANCOVA table along with pairwise comparison of adjusted post means of different treatment groups. The pairwise comparison of means is done only when the $F$-ratio is significant.

The analysis of covariance table generated in the SPSS output looks similar to the one- way ANOVA table as only adjusted post means is compared here. In the ANCOVA table, $F$-value is shown along with its significance value ($p$-value). The $F$-value would be significant if its corresponding $p$-value is less than .05, and in that case null hypothesis would be rejected. Once the $F$-value is found to be significant, then a post hoc test is used to compare the paired means. SPSS provides the choice of post hoc test to be used in the 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. The SPSS output provides the significance value ($p$-value) for each pair of difference of group means. Thus, by looking at the values of means, the best treatment may be identified.

## *Computations in ANCOVA Using SPSS*

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

Before using SPSS commands to solve the problem of analysis of covariance, a data file needs to be prepared. The following steps will help you prepare the data file:

(i) *Starting the SPSS*: Follow the below-mentioned 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 variables in the study.

(ii) *Defining variables*

In this example, three variables, vitamin dose, cold duration before treatment and cold duration during treatment need to be defined. The procedure of defining these variables along with their characteristics is as follows:

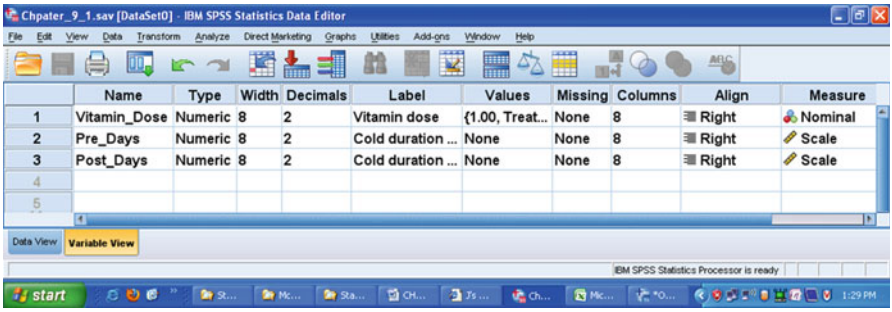


**Fig. 9.3** Defining variables and their characteristics

1. Click the **Variable View** to define the variables and their properties.
2. Write short name of the variables as *Vitamin_Dose*, *Pre_Days* and *Post_Days* under the column heading **Name**.
3. Under the column heading **Label**, define full name of these variables as *Vitamin dose*, *Cold duration before treatment*, and *Cold duration during treatment*. Other names may also be chosen for describing these variables.
4. Under the column heading **Measure,** select the option "Nominal" for the variable *Vitamin dose* and "Scale" for the variables *Cold duration before treatment* and *Cold duration during treatment*.
5. For the variable *Vitamin dose*, double-click the cell under the column **Values** and add the following values to different labels:

| Value | Label |
|---|---|
| 1 | Treatment *A* |
| 2 | Treatment *B* |
| 3 | Treatment *C* |

6. Use default entries in rest of the columns.

After defining the variables in variable view, the screen shall look as shown in Fig. 9.3.

(iii) *Entering data:* After defining all the variables in the **Variable View**, click **Data View** on the left corner in the bottom of the screen shown in Fig. 9.3 to open the format for entering the data column wise. After entering the data, the screen will look like Fig. 9.4. Save the data file in the desired location before further processing.

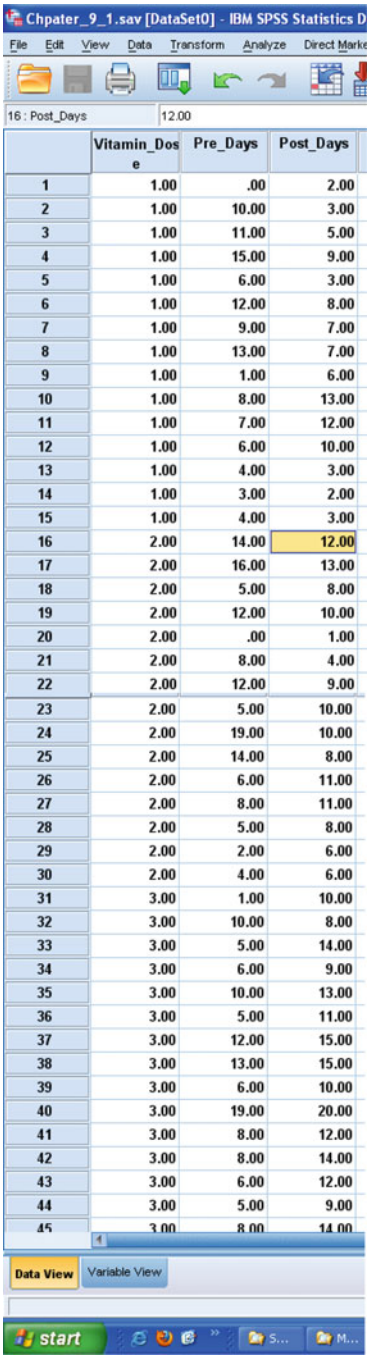


**Fig. 9.4** Screen showing entered data for all the variables in the data view

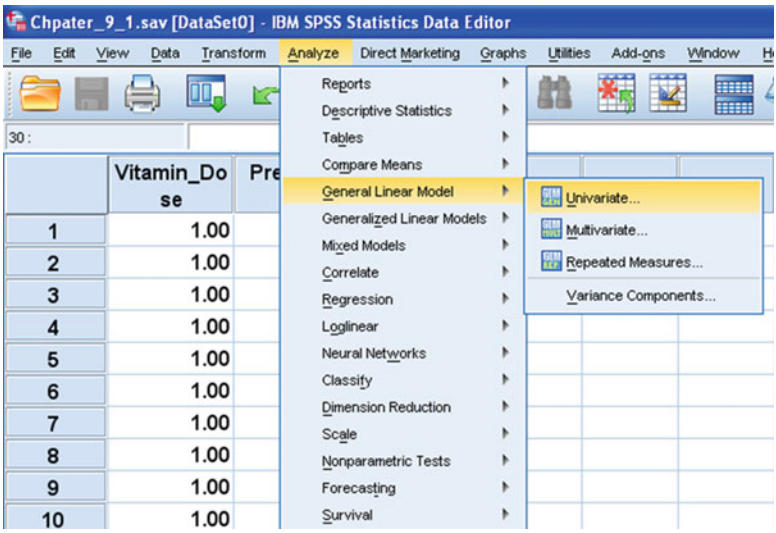


**Fig. 9.5** Sequence of SPSS commands for analysis of covariance

(b) ***SPSS commands for ANCOVA***

After preparing the data file, do the following steps for analysis of covariance:

(i) *Initiating the SPSS commands for ANCOVA:* In data view, click the following commands in sequence:

**Analyze ⇒ General Linear Model ⇒ Univariate**

The screen shall look like Fig. 9.5.

(ii) *Selecting variables for ANCOVA:* After clicking the **Univariate** option, you will be taken to the next screen as shown in Fig. 9.6 for selecting variables. Select the variables as follows:

- *Cold duration during treatment* from left panel to the "Dependent variable" section of the right panel.
- *Vitamin dose* from left panel to the "Fixed Factor(s)" section of the right panel.
- *Cold duration before treatment* from the left panel to the "Covariate(s)" section of the right panel.

(iii) *Selecting the options for computation:* After selecting the variables, different options need to be defined for generating the output in ANCOVA. This shall be done as follows:

- Click the tag **Model** on the screen shown in Fig. 9.6 and do the following:
  - Select the sum of squares option as "Type I."
  - Press ***Continue***.

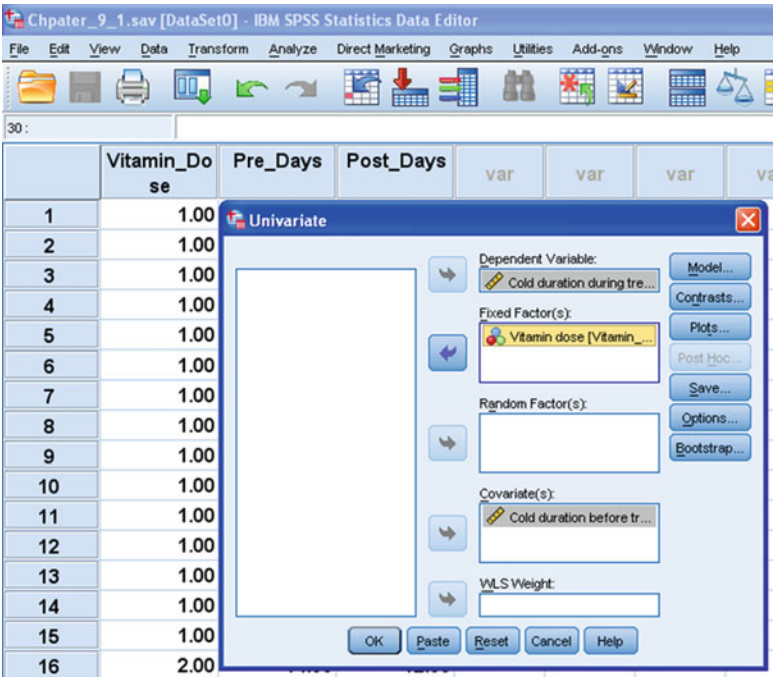


**Fig. 9.6** Selecting variables for ANCOVA

The screen will look like Fig. 9.7.

– Click the tag **Options** in the screen shown in Fig. 9.6 and do the following:

  – Select the variables *Overall* and *Vitamin_Dose* from the left panel to the "Display Means for" section of the right panel.
  – Check the option "Compare main effects."
  – Check the option "Descriptive statistics."
  – Ensure "Significance level" as .05. This value is written by default; however, you may write some other level of significance as .01 or .10, etc.
  – Click ***Continue***.

  The screen will look like Fig. 9.8.

  – Click **OK** on the screen shown in Fig. 9.6.

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

Clicking the option **OK** on the screen shown in Fig. 9.6 will take you to the output window. Relevant outputs can be selected by using the right click of the

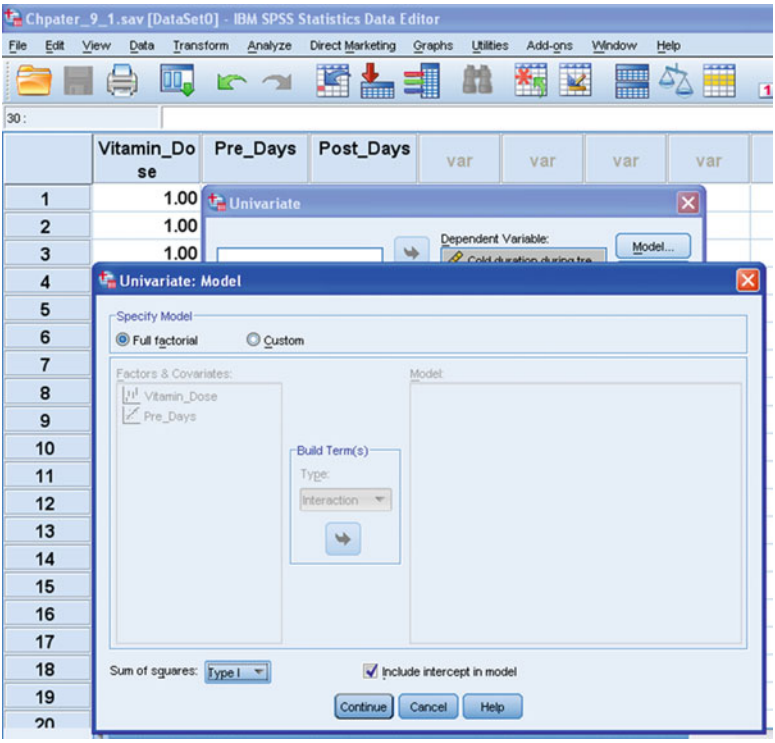


**Fig. 9.7** Screen showing option for choosing sum of squares and model type

mouse and may be copied in the word file. The identified outputs shall be rearranged for interpreting the findings. The details have been shown under the heading **Model Way of Writing the Results**.

(d) ***SPSS output***

The readers should note the kind of outputs to be selected from the output window of SPSS for explaining the findings. The following four outputs have been selected for discussing the results of ANCOVA:

1. Descriptive statistics
2. Adjusted estimates of the dependent variable
3. ANCOVA table
4. Post hoc comparison table

These outputs have been shown in Tables 9.2, 9.3, 9.4, and 9.5.

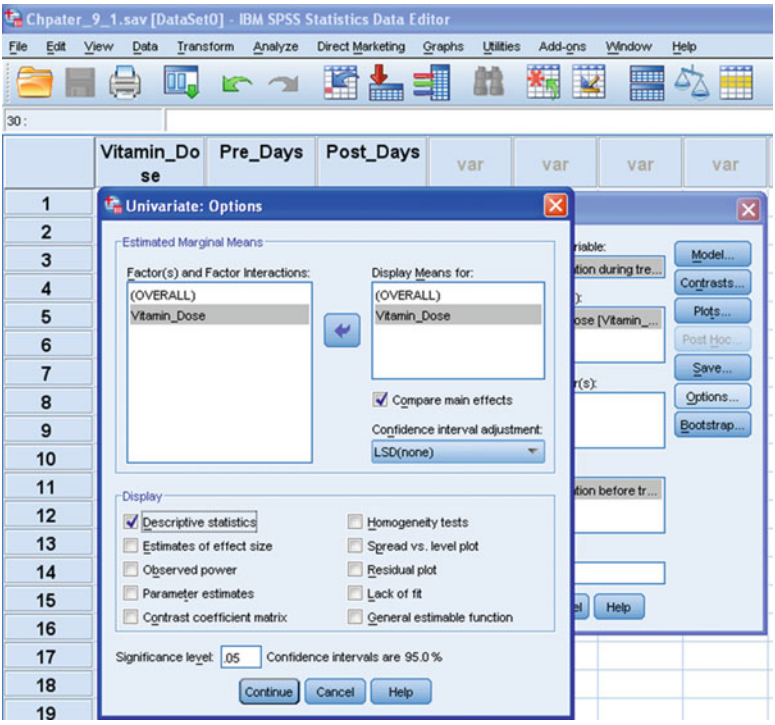


**Fig. 9.8** Selecting options for ANCOVA output

**Table 9.2** Descriptive statistics

| Vitamin dose | Mean | Std. deviation | *N* |
|---|---|---|---|
| Treatment *A* | 6.2000 | 3.62925 | 15 |
| Treatment *B* | 8.4667 | 3.18179 | 15 |
| Treatment *C* | 12.4000 | 3.11219 | 15 |
| Total | 9.0222 | 4.14778 | 45 |

Dependent variable: Cold duration during treatment

**Table 9.3** Adjusted estimates

| | | | 95% Confidence interval | |
|---|---|---|---|---|
| Vitamin dose | Mean | Std. error | Lower bound | Upper bound |
| Treatment *A* | 6.508[a] | .704 | 5.086 | 7.930 |
| Treatment *B* | 8.204[a] | .703 | 6.784 | 9.625 |
| Treatment *C* | 12.355[a] | .701 | 10.939 | 13.771 |

Dependent variable: Cold duration during treatment

[a]Covariates appearing in the model are evaluated at the following values: Cold duration before treatment = 8.0222

**Table 9.4** Tests "between-subjects" effects

| Source | Type I sum of squares | df | Mean square | *F* | Sig. |
|---|---|---|---|---|---|
| Corrected model | 454.761[a] | 3 | 151.587 | 20.565 | .000 |
| Intercept | 3,663.022 | 1 | 3,663.022 | 496.941 | .000 |
| Pre_Days | 183.993 | 1 | 183.993 | 24.961 | .000 |
| Vitamin_Dose | 270.768 | 2 | 135.384 | 18.367 | .000 |
| Error | 302.217 | 41 | 7.371 | | |
| Total | 4,420.000 | 45 | | | |
| Corrected total | 756.978 | 44 | | | |

Dependent variable: Cold duration during treatment
[a]R squared = .601 (adjusted R squared = .572)

**Table 9.5** Pairwise comparisons

| | | | | | 95% Confidence interval for difference[a] | |
|---|---|---|---|---|---|---|
| (I) Vitamin dose | (J) Vitamin dose | Mean diff. (I–J) | SE | Sig.[a] | Lower bound | Upper bound |
| Treatment *A* | Treatment *B* | −1.697 | .999 | .097 | −3.714 | .321 |
| | Treatment *C* | −5.847* | .994 | .000 | −7.855 | −3.839 |
| Treatment *B* | Treatment *A* | 1.697 | .999 | .097 | −.321 | 3.714 |
| | Treatment *C* | −4.151* | .992 | .000 | −6.155 | −2.146 |
| Treatment *C* | Treatment *A* | 5.847* | .994 | .000 | 3.839 | 7.855 |
| | Treatment *B* | 4.151* | .992 | .000 | 2.146 | 6.155 |

Dependent variable: Cold duration during treatment
Based on estimated marginal means
*The mean difference is significant at the .05 level
[a]Adjustment for multiple comparisons: Least significant difference (equivalent to no adjustments)

## Model Way of Writing the Results of ANCOVA and Their Interpretations

The above output generated by the SPSS can be shown in a much more user-friendly format by modifying the relevant contents of the Tables 9.2, 9.3, 9.4, and 9.5. The below-mentioned edited outputs can directly be shown in the project, dissertation, or thesis. These modified outputs shall be used to discuss the findings of ANCOVA.

(a) ***Descriptive Statistics of the Data Obtained on the Criterion Variable***
The mean and standard deviation of the criterion variable in different treatment groups have been shown in Table 9.6. Entries in this table have been copied from Table 9.2. If you are interested in computing different descriptive statistics for the covariate (Number of days having cold symptoms before treatment) also, the same be computed by using the procedure discussed in Chap. 2. However, the SPSS does not generate these statistics during ANCOVA analysis.

Look at the table heading which can be used in writing the final results in your study.

**Table 9.6** Mean and standard deviation of cold duration in different groups during treatment

| Vitamin dose | Mean | Std. deviation | *N* |
|---|---|---|---|
| Treatment *A* | 6.2 | 3.6 | 15 |
| Treatment *B* | 8.5 | 3.2 | 15 |
| Treatment *C* | 12.4 | 3.1 | 15 |
| Total | 9.0222 | 4.14778 | 45 |

Values have been rounded off

**Table 9.7** Adjusted mean and standard error for the data on cold duration in different groups during treatment

| | | | 95% Confidence interval | |
|---|---|---|---|---|
| Vitamin dose | Mean | Std. error | Lower bound | Upper bound |
| Treatment *A* | 6.5[a] | .70 | 5.09 | 7.93 |
| Treatment *B* | 8.2[a] | .70 | 6.78 | 9.63 |
| Treatment *C* | 12.4[a] | .70 | 10.94 | 13.77 |

[a]Covariates appearing in the model are evaluated at the following values:
Cold duration before treatment = 8.0222
Values have been rounded off

From Table 9.6, it can be seen that average time taken to cure the cold symptoms is highest in treatment group *C* whereas the least time is in treatment group *A*. Treatment C signifies the placebo, whereas treatment *A* is the high dose of vitamin C. The next question is to see whether this difference is significant or not after adjusting for the covariate (number of days having cold symptoms before treatment).

(b) ***Descriptive Statistics of the Data Obtained on the Criterion Variable after Adjusting for Covariate***

The adjusted mean and standard error of the criterion variable in different treatment groups have been shown in Table 9.7. The mean of criterion variable has been obtained in all the three treatment groups after adjusting for the covariate (Number of days having cold symptoms before treatment). These data have been taken from Table 9.3. Readers may note that these values are different from that of the unadjusted values shown in Table 9.6. The advantage of using the ANCOVA is that the differences in the posttesting means are compensated for the initial differences in the scores. In other words, it may be said that the effect of covariate is eliminated in comparing the effectiveness of treatments on the criterion variable.

Kindly note the heading of the table which may be used for writing the final results of ANCOVA.

(c) ***ANCOVA Table for the Data on Criterion Variable (Number of Days Having Cold Symptoms During Treatment)***

The main ANCOVA table may be reproduced by deleting some of the unwanted details of Table 9.4. The final results of ANCOVA have been shown in Table 9.8. The "significance" (Sig.) value has been named as *p*-value. In most of the scientific literature, *p*-value is used instead of term significance value.

**Table 9.8** ANCOVA table for the data on cold duration in different groups during treatment

| Source | Sum of squares | df | Mean square | *F* | (*p*-value) Sig. |
|---|---|---|---|---|---|
| Pre_Days | 183.993 | 1 | 183.993 | 24.961 | .000 |
| Vitamin_Dose | 270.768 | 2 | 135.384 | 18.367 | .000 |
| Error | 302.217 | 41 | 7.371 | | |
| Corrected total | 756.978 | 44 | | | |

**Table 9.9** Pairwise comparisons

| (I) Vitamin dose | (J) Vitamin dose | Mean diff. (I–J) | (*p*-value) Sig.[a] |
|---|---|---|---|
| Treatment *A* | Treatment *B* | −1.697 | .097 |
| | Treatment *C* | −5.847* | .000 |
| Treatment *B* | Treatment *A* | 1.697 | .097 |
| | Treatment *C* | −4.151* | .000 |
| Treatment *C* | Treatment *A* | 5.847* | .000 |
| | Treatment *B* | 4.151* | .000 |

Dependent variable: Cold duration during treatment
Based on estimated marginal means
*The mean difference is significant at the .05 level
[a]Adjustment for multiple comparisons: Least significant difference (equivalent to no adjustments)

Table 9.8 shows the *F*-value for comparing the adjusted means of the criterion variable in three Vitamin_Dose groups (treatment *A*, treatment *B*, and treatment *C*). You can note that *F*-statistic computed for Vitamin_Dose is significant because *p* -value associated with it is .000 which is less than .05. Thus, the null hypothesis of no difference among the adjusted means for the data on criterion variable (number of days having cold symptoms during treatment) in three treatment groups may be rejected at 5% level.

*Remark:* You can see that the *F*-value for Pre_Days (covariate) is also significant. It shows that the initial conditions of the experimental groups are not same, and that is why we are applying ANCOVA after adjusting mean values of the criterion variable for the covariate.

(d) ***Post Hoc Comparison for the Group Means in Post-measurement Adjusted with the Initial Differences***

Since *F*-statistic is significant, post hoc comparison has been made for the adjusted means of the three treatment groups, which is shown in Table 9.9. This, table has been obtained by deleting some of the information from Table 9.5. It may be noted here that *p*-value for the mean difference between treatments *A* and *C* as well between treatments *B* and *C* is .000. Since *p* value is less than .05, both these mean differences are significant at 5% level. Thus, the following conclusions can be drawn:

(i) There is a significant difference between the adjusted means of criterion variable (Number of days having cold symptoms during treatment) in treatment *A* (High vitamin C dose) and treatment *C* (Placebo).

**Table 9.10** Post hoc comparison of adjusted means in different groups for the data on cold duration during treatment with graphics

| Treatment *C* | Treatment *B* | Treatment *A* |
|---|---|---|
| 12.4 | 8.2 | 6.5 |

"____" represents no significant difference between the means

Treatment *A*: Administering high dose of vitamin C
Treatment *B*: Administering low dose of vitamin C
Treatment *C*: Administering placebo

(ii) There is a significant difference between the adjusted means of criterion variable (Number of days having cold symptoms during treatment) in treatment *B* (Low vitamin C dose) and treatment *C* (Placebo).
(iii) There is no significant difference between the adjusted means of criterion variable (Number of days having cold symptoms during treatment) in treatment *A* (High vitamin C dose) and treatment *B* (Low vitamin dose).

In order to find as to which treatment is the best, one can see the adjusted mean values of criterion variable in different treatment groups given in Table 9.7. Clubbing these adjusted means with the three conclusions mentioned above, one may get the answer. However, this task becomes much easier if Table 9.10 is developed. This table can be created by using the values of different adjusted group means from Table 9.7 and using *p*-values of mean differences from Table 9.9. In this table, the adjusted means of the criterion variable in different treatment groups have been shown in the descending order. If the difference between any two group means is significant (which can be seen from Table 9.10), nothing is done, and if the mean difference is not significant, an underline is put below both the group means.

Thus, it may be concluded that the average curing time in high and low vitamin groups was same. Further, the average curing time in both these groups was significantly less than that of placebo group.

Hence, it may be inferred that high vitamin dose as well as low vitamin dose are equally effective in curing the cold symptoms in comparison to that of placebo.

## Summary of the SPSS Commands

(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 *Vitamin_Dose* as nominal variable and *Pre_Days* and *Post_Days* as scale variables.
(iii) Under the column heading **Values** against the variable *Vitamin_Dose,* define "1" for Treatment *A*, "2" for Treatment *B*, and "3" for Treatment *C*.
(iv) After defining the variables, type the data for these variables by clicking **Data View**.
(v) In the data view, follow the below-mentioned command sequence for ANCOVA:

**Analyze ⇒ General Linear Model ⇒ Univariate**

(vi) Select the variables *Cold duration during treatment*, *Vitamin dose,* and *Cold duration before treatment from* left panel to the "Dependent variable" section, "Fixed Factor(s)" section, and "Covariate(s)" section of the right panel, respectively.
(vii) Click the tag **Model** and select the Sum of Squares option as "Type I." Press ***Continue***.
(viii) Click the tag **Options** and select the variables *Overall* and *Vitamin_Dose* from the left panel to the "Display Means for" section of the right panel. Check the option "Compare main effects" and "Descriptive statistics." Ensure the value of significance as .05 or .01 as the case may be. Press ***Continue***.
(ix) Click **OK** for output.

## Exercise

*Short Answer Questions*

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

Q1. What do you mean by the covariate? How it is controlled in ANCOVA? Give a specific example.
Q2. Describe an experimental situation where ANCOVA can be applied. Construct null hypothesis and all possible alternative hypotheses.
Q3. Thirty boys were selected for direct marketing of a vacuum cleaner in three similar cities. In each of the city, 10 boys were sent for direct marketing for a month. Three different kinds of incentives, namely, conveyance allowance, two percent bonus, and gifts were offered to these sales agents in these three cities on completing the target. To compare the effectiveness of three different incentives on sale, which statistical technique should be used?
Q4. If two treatment groups are to be compared on some criterion variable, how do you interpret if the slopes of the two regression lines are same? Further, if the intercepts are equal, what it conveys? Explain by means of graphical representation.

Q5. Explain the statement "the analysis of covariance is a mix of one-way ANOVA and linear regression."
Q6. Why the observed mean of criterion variable is adjusted in ANCOVA? How this adjustment is done?
Q7. What are the various assumptions used in analysis of covariance?
Q8. Which design is more efficient and why among one-way ANOVA and ANCOVA?

*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 designing an experiment, if the randomization is not possible, control is observed by matching the groups. This matching is done on the variable which is

   (a) Independent
   (b) Extraneous
   (c) Dependent
   (d) Any variable found suitable

2. Covariate is a variable which is supposed to be correlated with

   (a) Criterion variable
   (b) Independent variable
   (c) Dependent variable
   (d) None of the above

3. In ANCOVA, while doing post hoc analysis, which group means are compared?

   (a) Pretest group means
   (b) Posttest group means
   (c) Pretest adjusted group means
   (d) Posttest adjusted group means

4. In ANCOVA, if the slopes of the regression lines in different treatment groups are same, one can infer that

   (a) Some of the treatments will show the improvement where the other treatments may show the deterioration.
   (b) All the treatments will show either deterioration or improvement but with varying degrees.
   (c) One cannot tell about the improvement or deterioration due to different treatments.
   (d) All treatments will have the same amount of improvement in the criterion variable.

5. In ANCOVA, if intercepts of the regression lines in the two treatment groups are same, then it may be inferred that

(a) One treatment is better than other.
(b) One cannot say which treatment is more effective.
(c) Both the treatments are equally effective.
(d) No conclusion can be drawn.

6. In ANCOVA model, the error component is independently and identically normally distributed with

(a) Mean 0 and variance 1
(b) Mean 1 and variance 0
(c) Equal mean and variance 1
(d) Mean 0 and equal variance

7. In ANCOVA, the adjusted mean $\mu$ in the $i$th treatment group is obtained from the formula

(a) $\mu = \overline{Y}_i + \beta(\overline{X}_i - \bar{X})$
(b) $\mu = \overline{Y}_i - \beta(\overline{X}_i - \bar{X})$
(c) $\mu = \overline{Y}_i - \beta(\bar{X} - \overline{X}_i)$
(d) $\mu = \overline{Y}_i + \beta + (\overline{X}_i - \bar{X})$

8. In analysis of covariance, the criterion variable should be

(a) Continuous
(b) Nominal
(c) Ordinal
(d) Dichotomous always

9. One of the assumptions in using ANCOVA is

(a) The data on criterion variable must have been obtained by stratified sampling.
(b) The regression coefficients for each treatment groups must be heterogeneous.
(c) The interaction between the criterion variable and covariate is significant.
(d) The criterion variable must have the same variance in each of the treatment groups.

10. Choose the correct statement.

(a) The ANCOVA is more efficient than ANOVA because part of the error variance is explained by the covariate.
(b) ANOVA is more efficient than ANCOVA if the initial conditions are not same.
(c) ANOVA and ANCOVA are equally effective, and it is the matter of choice as to which analysis is to be used.
(d) All the above statements are correct.

11. In order to compare the effectiveness of three training programs on financial knowledge, an experiment was planned. Three groups of employees were

tested for their financial knowledge before and after the training program. While using SPSS for ANCOVA, three variables, namely, Pre_Knowledge Drib, Post_Knowledge, and Treatment_Group, need to be defined. Choose the correct types of each variable.

(a) Pre_Knowledge and Post_Knowledge are Scale and Treatment_Group is Ordinal.
(b) Pre_Knowledge and Post_Knowledge are Nominal and Treatment_Group is Scale.
(c) Pre_Knowledge and Treatment_Group are Scale and Post_Knowledge is Nominal.
(d) Pre_Knowledge and Post_Knowledge are Scale and Treatment_Group is Nominal.

12. While using SPSS for ANCOVA, the three variables, namely, Pre_Test, Post_-Test, and Treatment_Group, are classified as

(a) Post_Test as Dependent variable whereas Pre_Test and Treatment_Group as Fixed Factor(s)
(b) Post_Test as Dependent variable, Pre_Test as Covariate, and Treatment_Group as Fixed Factor
(c) Treatment_Group as Dependent variable, Pre_Test and Post_Test as Fixed Factor(s)
(d) Treatment_Group as Dependent variable, Post_Test as Covariate, and Pre_Test as Fixed Factor

13. Choose the correct sequence of commands in SPSS for starting ANCOVA.

(a) Analyze → Univariate → General Linear Model
(b) Analyze → General Linear Model → Multivariate
(c) Analyze → General Linear Model → Univariate
(d) Analyze → General Linear Model → Repeated Measures

*Assignments*

1. In a psychological experiment 60, subjects were randomly divided into three equal groups. These groups were taught with audiovisual aid, traditional method, and need-based methods. Prior to the treatments, learning motivation of all the subjects was assessed. After 4 weeks, improvement in academic achievements was noted. The data so obtained on academic achievements is shown in the Table A-1.
   Apply analysis of covariance to see as to which methodology of teaching is more effective for academic achievement. Test your hypothesis at .05 as well as .01 level of significance.
2. A study was conducted to know the impact of gender on life optimism. Since age is considered as factor effecting life optimism, it was considered as covariate.

**Table A-1** Scores on academic achievements and learning motivation in three types of teaching methods

| S.N. | Audiovisual group | | Traditional group | | Need-based group | |
|---|---|---|---|---|---|---|
| | Motivation | Achievement | Motivation | Achievement | Motivation | Achievement |
| 1 | 2 | 5 | 2 | 3 | 1 | 12 |
| 2 | 1 | 10 | 3 | 3 | 3 | 18 |
| 3 | 3 | 12 | 1 | 9 | 2 | 11 |
| 4 | 0 | 14 | 1 | 13 | 6 | 25 |
| 5 | 1 | 14 | 4 | 13 | 3 | 9 |
| 6 | 5 | 16 | 5 | 13 | 5 | 18 |
| 7 | 3 | 18 | 6 | 13 | 3 | 12 |
| 8 | 4 | 18 | 1 | 15 | 4 | 10 |
| 9 | 4 | 18 | 2 | 15 | 3 | 11 |
| 10 | 5 | 18 | 4 | 17 | 6 | 16 |
| 11 | 2 | 22 | 6 | 17 | 7 | 18 |
| 12 | 3 | 22 | 5 | 21 | 4 | 14 |
| 13 | 7 | 28 | 2 | 22 | 3 | 17 |
| 14 | 4 | 24 | 5 | 22 | 2 | 10 |
| 15 | 6 | 24 | 5 | 22 | 5 | 19 |
| 16 | 3 | 26 | 5 | 22 | 3 | 14 |
| 17 | 4 | 26 | 5 | 22 | 2 | 16 |
| 18 | 4 | 26 | 6 | 20 | 3 | 15 |
| 19 | 8 | 26 | 9 | 22 | 4 | 16 |
| 20 | 3 | 29 | 4 | 24 | 5 | 15 |

**Table A-2** Data on age and life optimism of the male and female

| | Male | | Female | |
|---|---|---|---|---|
| S.N. | Age | Life optimism | Age | Life optimism |
| 1 | 53 | 18 | 20 | 26 |
| 2 | 38 | 26 | 24 | 15 |
| 3 | 18 | 20 | 18 | 16 |
| 4 | 26 | 27 | 38 | 17 |
| 5 | 39 | 19 | 35 | 16 |
| 6 | 38 | 29 | 25 | 24 |
| 7 | 30 | 15 | 17 | 10 |
| 8 | 60 | 23 | 19 | 18 |
| 9 | 22 | 22 | 21 | 19 |
| 10 | 31 | 14 | 19 | 19 |
| 11 | 21 | 14 | 18 | 21 |
| 12 | 25 | 23 | 38 | 13 |
| 13 | 22 | 23 | 37 | 15 |
| 14 | 20 | 19 | 21 | 11 |
| 15 | 27 | 23 | 20 | 11 |
| 16 | 24 | 20 | 20 | 14 |
| 17 | 29 | 18 | 41 | 16 |
| 18 | 27 | 19 | 40 | 17 |
| 19 | 32 | 13 | 28 | 17 |
| 20 | 17 | 14 | 99 | 8 |

A questionnaire was administered on 20 male and 20 female subjects to know their life optimism. Their age was also noted. The data so obtained are listed in the Table A-2.

Apply analysis of covariance and discuss your findings to compare the life optimism among male and female adjusted for their age. Test your hypothesis at 5% as well as at 1% level.

*Answers to Multiple-Choice Questions*

| | | | | | |
|---|---|---|---|---|---|
| Q.1 | b | Q.2 | a | Q.3 | d |
| Q.4 | b | Q.5 | c | Q.6 | d |
| Q.7 | b | Q.8 | a | Q.9 | d |
| Q.10 | a | Q.11 | d | Q.12 | b |
| Q.13 | c | | | | |