

Chapter 449

Repeated Measures Analysis of Variance

Introduction

This procedure performs an analysis of variance on repeated measures (within-subject) designs using the general linear models approach. The experimental design may include up to three between-subject terms as well as three within-subject terms. Box's M and Mauchly's tests of the assumptions about the within-subject covariance matrices are provided. Geisser-Greenhouse, Box, and Huynh-Feldt corrected probability levels on the within-subject F-tests are given along with the associated test power.

Repeated measures designs are popular because they allow a subject to serve as their own control. This improves the precision of the experiment by reducing the size of the error variance on many of the F-tests, but additional assumptions concerning the structure of the error variance must be made.

This procedure uses the general linear model (GLM) framework to perform its calculations. Identical results can be achieved by using the GLM ANOVA program. The user input of this procedure is simply the GLM panel modified to allow a more direct specification of a repeated-measures model. We refer you to the GLM ANOVA chapter for details on the calculations and interpretations of analysis of variance. We will concentrate here on providing information specific to repeated measures analysis.

An Example

This section will give an example of a repeated-measures experiment. An experiment was conducted to study the effects of four drugs upon reaction time to a set of tasks using five subjects.

Subject	Drug 1	Drug 2	Drug 3	Drug 4
1	30	28	16	34
2	14	18	10	22
3	24	20	18	30
4	38	34	20	44
5	26	28	14	30

Discussion

One way of categorizing experimental designs is as *between subject* or *within subject*. Examples of between-subject designs are the common factorial designs in which the experimental units (the subjects) are assigned to separate treatment conditions. Usually, this assignment is done at random. The experimenter wants to know if the variability from subject to subject is smaller than the variability from treatment to treatment. The basic assumption is that the subjects are independent from one another.

Within-subject designs are those in which multiple measurements are made on the same individual. Because the response to stimuli usually varies less within an individual than between individuals, the within-subject variability is usually less than (or at most equal to) the between-subject variability. Reducing the underlying variability reduces the sample size which reduces cost.

Disadvantages of Within-Subjects Designs

The main advantage of within-subjects designs is in the reduced variability that is achieved by controlling from differences from one subject to the next. There are several disadvantages to this type of design:

1. Practice effect. In some experiments, subjects systematically improve as they practice the task being studied. In other cases, subjects may systematically get worse as they get fatigued or bored with the experimental task. Note that only the treatment administered first is immune to practice effects. Hence, experimenters often make some effort to balance the number of subjects receiving each treatment first.
2. Carryover effect. In many drug studies, it is important to “wash out” one drug completely before the next drug is administered. Otherwise, the influence of the first drug carries over into the response to the second drug. Note that practice effects refer to a general change in response because the task is repeated, but carryover effects refer to specific, lasting effects of a particular treatment.
3. Statistical analysis. The statistical model that justifies the analysis is very restrictive since the individual responses must have certain mathematical properties. Also, missing responses are much more difficult to deal with in this case.
4. Generalizability. Experimenters assume that differences between treatments are design independent. That is, if a completely random design was constructed, the same treatment differences would be observed. This is not always the case.

Even in the face of all these disadvantages, repeated measures (within-subject) designs are popular in many areas of research. It is important that you recognize these problems going in, rather than learning of them later after the experiment has been conducted.

Assumptions

The following assumptions are made when using the F test to analyze a factorial experimental design.

1. The response variable is continuous.
2. The residuals follow the normal probability distribution with mean equal to zero and constant variance.
3. The subjects are independent. Since in a within-subject design, responses coming from the same subject are not usually independent, assumption three must be modified for responses within a subject. The independence between subjects is still assumed.
4. The within-subject covariance matrices are equal for all between-subject groups. In this type of experiment, the repeated measurements on a subject may be thought of as a multivariate response vector having a certain covariance structure. This assumption states that these covariance matrices are constant from group to group. This assumption is tested by Box's M test. Of course, this assumption unnecessary in the single-group design.
5. All of the within-subject covariance matrices are circular. One way of defining circularity is that the variances of differences between any two measurements within a subject are constant. Since responses that are close together in time often have a higher correlation than those that are far apart, it is common for this assumption to be violated. This assumption is tested by Mauchly's test and by studying the values of epsilon (defined below). The circularity assumption is not necessary when only two repeated measures are made.

Repeated Measures Analysis of Variance

The program provides formal tests of these assumptions. However, these tests have their own assumptions which also may be violated, so a more common strategy is to assume that the circularity is violated and take appropriate action. **NCSS** does this for you automatically.

Technical Details

Other than reformatting the input panel, the main difference between this procedure and the GLM procedure is the inclusion of the Geisser-Greenhouse correction and associated tests of assumptions. Because of this, we will present only those results here. You can obtain a more general overview of analysis of variance in the One-Way Analysis of Variance and General Linear Models chapters.

Covariance Matrix Assumptions

The covariance matrix for a design with m subjects and k measurements per subject may be represented as

$$\Sigma = [\sigma_{ij}]$$

Valid F tests in a repeated-measures design require that the covariance matrix is a type H matrix. A type H matrix has the *circularity* property that

$$\Sigma = A + A' + \lambda I_k$$

where I_k is the identity matrix of order k and λ is a constant.

This property may also be defined as

$$\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} = 2\lambda$$

One type of matrix that has this property is one which has *compound symmetry*. A matrix with this property has all elements on the main diagonal equal and all elements off the main diagonal equal. An example of a covariance matrix with compound symmetry is

$$\begin{bmatrix} 9 & 2 & 2 & 2 \\ 2 & 9 & 2 & 2 \\ 2 & 2 & 9 & 2 \\ 2 & 2 & 2 & 9 \end{bmatrix}$$

An example of a type H matrix which does not have compound symmetry is

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 & 5 \\ 3 & 6 & 5 & 6 \\ 4 & 5 & 8 & 7 \\ 5 & 6 & 7 & 10 \end{bmatrix}$$

Note that if the diagonal elements are equal, which implies that the variation within each subject is constant, a type H matrix must have compound symmetry.

Epsilon

Epsilon is a measure of the extent to which a covariance matrix departs from circularity. It was developed by Box (see Winer(1991) or Kirk (1982)) and is estimated by

$$\hat{\varepsilon} = \frac{k^2 \left(\sum_{i=1}^k \frac{S_{ii}}{k} - \sum_{i=1}^k \sum_{j=1}^k \frac{S_{ij}}{k^2} \right)^2}{(k-1) \left[\sum_{i=1}^k \sum_{j=1}^k S_{ij}^2 - 2k \sum_{j=1}^k \left(\sum_{i=1}^k \frac{S_{ji}}{k} \right)^2 + k^2 \left(\sum_{i=1}^k \sum_{j=1}^k \frac{S_{ij}}{k^2} \right)^2 \right]}$$

where the estimated covariance matrix is given by

$$\hat{\Sigma} = [s_{ij}]$$

and k is the number of levels of the within subject factor.

For two- and three-way interaction terms, epsilon is estimated by

$$\hat{\varepsilon} = \frac{(\sum_{i=1}^r z_{ii})^2}{r \sum_{i=1}^r \sum_{j=1}^r z_{ij}^2}$$

where $Z = CSC$ and C is a contrast matrix appropriate for testing that interaction.

This estimate of epsilon is biased, especially for large values of epsilon. To correct for this bias, Huynh and Feldt developed another estimate of epsilon, which is calculated as follows

$$\tilde{\varepsilon} = \text{Min} \left[\frac{N(k-1)\hat{\varepsilon} - 2}{(k-1)[N - g - (k-1)\hat{\varepsilon}]}, 1 \right]$$

where N is the total number of subjects and g is the number of levels of the between factors.

The range of epsilon is

$$\frac{1}{k-1} \leq \varepsilon \leq 1$$

When $\varepsilon = 1$, the matrix is circular. When $\varepsilon = 1/k - 1$, the matrix differs maximally from circularity.

Box's estimator tends to underestimate epsilon and the Huynh-Feldt estimate tends to overestimate it. Simulation studies have found Box's estimate to be the one that should be used to adjust the F tests.

Geisser-Greenhouse Adjustment

All F ratios of within subject factors and interactions require the assumption that the covariance matrix is of type H in order for the F ratio to follow the F distribution with degrees of freedom $df1$ and $df2$. When the covariance matrix is not of type H , Geisser and Greenhouse suggested that the distribution of the F ratio be approximated by an F distribution with degrees of freedom $\varepsilon(df1)$ and $\varepsilon(df2)$ where ε is set at its minimum, that is, $\varepsilon = 1/k - 1$.

Box suggested that rather than use the minimum to adjust the degrees of freedom, ε should be set at the Geisser-Greenhouse value, $\hat{\varepsilon}$. Realizing the $\hat{\varepsilon}$ is biased, Huynh and Feldt suggested that $\tilde{\varepsilon}$ be used.

Repeated Measures Analysis of Variance

Simulation studies have shown that using Box's adjustment consistently gives the most accurate significance levels.

Mauchly's Test of Compound Symmetry

Mauchly (1940) developed a test to determine if a sample covariance matrix has compound symmetry. The formulas for Mauchly's test statistic W , given in Huynh and Feldt (1970), are as follows

$$W = |CSC'| / (\text{trace}(CSC')/p)^p$$

$$\chi_{p(p+1)/2-1}^2 = -(N-g) \left(1 - \frac{2p^2 + p + 2}{6p(N-g)} \right) \ln(W)$$

where g is the number of groups, N is the number of subjects, C is a contrast matrix with p rows suitable for testing a main effect or interaction, S is a k -by- k matrix of the pooled group covariances. Note that usually, p equals the degrees of freedom of the corresponding term.

Data Structure

The data must be entered in a format that puts the response values in one variable and the values of each of the factors in other variables. We will first present an example of a single-group repeated measures design followed by an example of a design with one between factor and one within factor.

Single-Group Repeated Measures Design Example – Reaction Dataset

The experiment described in this example was conducted to study the effects of four drugs upon reaction time to a set of tasks. The five subjects were given extensive training in the tasks prior to the experiment so that there would be no carryover (learning) effect from one trial to the next. The five subjects were chosen at random from the population of interest.

The order in which the drugs were administered was randomized separately for each subject. A sufficient time was allowed between trials to wash out the effect of the previous drug before administering the next drug. The results of this experiment are recorded in the Reaction dataset. This design is often referred to as a *randomized block design*.

Reaction Dataset

Row	Person	Drug	Test
1	1	1	30
2	1	2	28
3	1	3	16
4	1	4	34
5	2	1	14
6	2	2	18
7	2	3	10
8	2	4	22
9	3	1	24
10	3	2	20
11	3	3	18
12	3	4	30
13	4	1	38
14	4	2	34
15	4	3	20
16	4	4	44
17	5	1	26
18	5	2	28
19	5	3	14
20	5	4	30

Repeated Measures Analysis of Variance

Heart Rate Data - Exercise Dataset

The following dataset is an example of a one between-factor and one within-factor repeated measures design. An experiment was conducted to study the effects of exercise on heart rate. The subjects were randomly divided into three groups of six subjects each. The first group did not have a regular exercise plan. The second group exercised once a week. The third group exercised daily. Each subject's heart rate was recorded when the experiment began, at the end of ten weeks, and at the end of twenty weeks. These data are stored in a database called EXERCISE.

Exercise Dataset

Row	Exercise	Subj	Time	Heart Rate
1	0 - None	1	0	87
2	0 - None	1	10	77
3	0 - None	1	20	84
4	0 - None	2	0	67
5	0 - None	2	10	65
6	0 - None	2	20	62
7	0 - None	3	0	55
8	0 - None	3	10	52
9	0 - None	3	20	58
10	0 - None	4	0	66
11	0 - None	4	10	70
12	0 - None	4	20	65
13	0 - None	5	0	88
14	0 - None	5	10	82
15	0 - None	5	20	85
16	0 - None	6	0	75
17	0 - None	6	10	72
18	0 - None	6	20	79
19	1 - Weekly	7	0	84
20	1 - Weekly	7	10	78
21	1 - Weekly	7	20	74
22	1 - Weekly	8	0	78
23	1 - Weekly	8	10	72
24	1 - Weekly	8	20	68
25	1 - Weekly	9	0	64
26	1 - Weekly	9	10	53
27	1 - Weekly	9	20	54
28	1 - Weekly	10	0	73
29	1 - Weekly	10	10	68
30	1 - Weekly	10	20	63
31	1 - Weekly	11	0	84
32	1 - Weekly	11	10	77
33	1 - Weekly	11	20	74
34	1 - Weekly	12	0	55
35	1 - Weekly	12	10	53
36	1 - Weekly	12	20	52
37	2 - Daily	13	0	72
38	2 - Daily	13	10	55
39	2 - Daily	13	20	53
40	2 - Daily	14	0	83
41	2 - Daily	14	10	72
42	2 - Daily	14	20	69
43	2 - Daily	15	0	75
44	2 - Daily	15	10	63
45	2 - Daily	15	20	65
46	2 - Daily	16	0	55
47	2 - Daily	16	10	49
48	2 - Daily	16	20	51
49	2 - Daily	17	0	83
50	2 - Daily	17	10	76
51	2 - Daily	17	20	72
52	2 - Daily	18	0	63
53	2 - Daily	18	10	54
54	2 - Daily	18	20	55

Missing Values

There are two kinds of unbalance that can occur in repeated-measures designs. First, in multi-group designs, there may be a different number of subjects in each group. This type of unbalance causes no problems in the F-tests. Second, some individuals may not have had all measurements. When this occurs, the program makes the additional assumption that the within-subject sample effects sum to zero. Every effort should be made to avoid missing values because of the additional assumptions that must be made. However, even when data are missing, meaningful conclusions can be drawn.

Example 1 – Running Repeated Measures ANOVA

This section presents an example of how to run an analysis of a typical repeated measures design with one between factor and one within factor. These data are contained in the Exercise dataset.

Setup

To run this example, complete the following steps:

1 Open the Exercise example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Exercise** and click **OK**.

2 Specify the Repeated Measures Analysis of Variance procedure options

- Find and open the **Repeated Measures Analysis of Variance** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable(s)**Heartrate**
 Subject Variable.....**Subject**
 Between Factor Variable 1**Exercise**
 Comparisons.....**Each with First**
 .(This will generate the test of the no exercise group with the weekly exercise group and the no exercise group with the daily exercise group.)
 Within Factor Variable 1**Time**

Reports Tab

All Reports**Checked**
 Compute Power**Checked**
 Tukey-Kramer Test.....**Checked**
 Show Notes**Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Expected Mean Squares

Expected Mean Squares

Model Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A: Exercise	2	Yes	B(A)	$S+csB+bcsA$
B(A): Subject	15	No	S(ABC)	$S+csB$
C: Time	2	Yes	BC(A)	$S+sBC+absC$
AC	4	Yes	BC(A)	$S+sBC+bsAC$
BC(A)	30	No	S(ABC)	$S+sBC$
S(ABC)	0	No	S	S

Note: Expected Mean Squares are for the balanced cell-frequency case.

The expected mean square expressions are provided to show the appropriate error term for each factor. The correct error term for a factor is that term that is identical except for the factor being tested.

Note that in the repeated measures model, there are two error terms that are used: the between error labeled B(A) and the within error labeled BC(A).

Model Term

The source of variation or term in the model.

DF

The degrees of freedom, which is the number of observations used by this term.

Term Fixed?

Indicates whether the term is fixed or random.

Denominator Term

Indicates the term used as the denominator in the F-ratio. This is the error term for this term.

Expected Mean Square

This expression represents the expected value of the corresponding mean square if the design was completely balanced. S represents the expected value of the mean square error (sigma). The uppercase letters represent either the adjusted sum of squared treatment means if the factor is fixed, or the variance component if the factor is random. The lowercase letter represents the number of levels for that factor, and s represents the number of replications of the experimental layout.

These EMS expressions are provided to determine the appropriate error term for each factor. The correct error term for a factor is that term whose EMS is identical except for the factor being tested.

Analysis of Variance Table

Analysis of Variance Table

Model Term	DF	Sum of Squares	Mean Square	F-Ratio	P-Value	Power†
A: Exercise	2	427.4445	213.7222	0.6124	0.55504	0.13334
B(A): Subject	15	5234.556	348.9704			
C: Time	2	547.4445	273.7222	36.9156	0.00000*	1.00000
AC	4	191.4444	47.86111	6.4548	0.00072*	0.97763
BC(A)	30	222.4444	7.414815			
S	0					
Total (Adjusted)	53	6623.333				
Total	54					

* Term significant at $\alpha = 0.05$

† Power was calculated using the observed F-Ratio as the population effect size with a significance level of $\alpha = 0.05$.

Model Term

The source of variation, which is the term in the model.

DF

The degrees of freedom, which is the number of observations used by the corresponding model term.

Sum of Squares

This is the sum of squares for this term. It is usually included in the ANOVA table for completeness, not for direct interpretation.

The sums of squares are calculated as follows. First, the sum of squares of each term that does not involve the subject factor is computed using the difference between two reduced models. For example, the sum of squares for A is computed as the difference between the sum of squares for the model A+C+AC and the sum of squares for the model C+AC. The sum of squares for C and AC is computed similarly.

Next, the sum of squares of the subject factor is computed by treating the subjects as a one-way design, computing the subject sum of squares, and subtracting the sum of squares of all terms that occur before it in the model—in this case, the sum of squares of factor A.

Next, the sum of squares of the BC(A) interaction is computed by treating this term as a one-way design, computing its sum of squares, and subtracting the sum of squares of all terms that occur before it in the model—in this case, the sum of squares for A, B(A), C, and AC.

The computations are carried out in this manner to give reasonable tests in the cases when there are unequal numbers of subjects per group or some subjects have missing measurements. The results are similar to the Type III sum of squares computations given by SAS.

Mean Square

An estimate of the variation accounted for by this term. It is the sum of squares divided by the degrees of freedom.

Repeated Measures Analysis of Variance

F-Ratio

The ratio of the mean square for this term and the mean square of its corresponding error term. This is also called the F-test value.

P-Value

The significance level of the above F-ratio, or the probability of an F-ratio larger than that obtained by this analysis. For example, to test at an alpha of 0.05, this probability would have to be less than 0.05 to make the F-ratio significant. Note that if the value is significant at the specified value of alpha, a star is placed to the right of the F-Ratio.

This F-ratio is only valid if all the assumptions are valid. You should study the results of the preliminary tests to determine if the assumptions hold.

Power

Power is the probability of rejecting the hypothesis that the means are equal when they are in fact not equal. Power is one minus the probability of type II error (β). The power of the test depends on the sample size, the magnitudes of the variances, the alpha level, and the actual differences among the population means.

The power value calculated here assumes that the population standard deviation is equal to the observed standard deviation and that the differences among the population means are exactly equal to the differences among the sample means.

High power is desirable. High power means that there is a high probability of rejecting the null hypothesis when the null hypothesis is false. This is a critical measure of precision in hypothesis testing.

Generally, you would consider the power of the test when you accept the null hypothesis. The power will give you some idea of what actions you might take to make your results significant. If you accept the null hypothesis with high power, there is not much left to do. At least you know that the means are not different. However, if you accept the null hypothesis with low power, you can take one or more of the following actions:

1. Increase your alpha level. Perhaps you should be testing at $\alpha = 0.05$ instead of $\alpha = 0.01$. Increasing the alpha level will increase the power.
2. Increasing your sample size will increase the power of your test if you have low power. If you have high power, an increase in sample size will have little effect.
3. Decrease the magnitude of the variance. Perhaps you can redesign your study so that measurements are more precise and extraneous sources of variation are removed.

P-Values for F-Tests with Geisser-Greenhouse Adjustments

P-Values for F-Tests with Geisser-Greenhouse Adjustments

Model Term	DF	F-Ratio	P-Value			
			Regular	Lower-Bound Epsilon	Geisser-Greenhouse Epsilon	Huynh-Feldt Epsilon
A: Exercise	2	0.6124	0.55504			
B(A): Subject	15					
C: Time	2	36.9156	0.00000*	0.00002*	0.00000*	0.00000*
AC	4	6.4548	0.00072*	0.00950*	0.00075*	0.00072*
BC(A)	30					
S	0					

* Term significant at $\alpha = 0.05$

This table presents the F ratios from the analysis of variance table with probability levels computed using the three Geisser-Greenhouse adjustments. These are explained in detail below. Note that no adjustments are made to between-subjects terms (A in this example). Also note that in designs involving two or three within factors, different adjustment factors are computed for each term. The values of epsilon are shown in the Covariance Matrix Circularity report.

Model Term

The source of variation, which is the term in the model.

F-Ratio

The F-ratio is repeated from Analysis of Variance Table.

Regular P-Value

The probability level is repeated from Analysis of Variance Table.

Lower-Bound Epsilon P-Value

This is the probability level of the corresponding F-ratio using the minimum epsilon. This correction involves multiplying both the numerator and denominator degrees of freedom by the minimum epsilon and then calculating the probability level. Since this epsilon is a value between zero and one, the impact of this adjustment is to reduce the degrees of freedom.

This adjustment is made to correct for a non-circular covariance matrix. Simulation studies have shown these probability levels to be too conservative and so we do not recommend its use. Usually, the Geisser-Greenhouse epsilon is used instead.

Geisser-Greenhouse Epsilon P-Value

This is the probability level of the corresponding F-ratio using the Geisser-Greenhouse epsilon. This adjustment involves multiplying both the numerator and denominator degrees of freedom by the Geisser-Greenhouse epsilon and then calculating the probability level. Since this epsilon is a value between zero and one, the impact of this adjustment is to reduce the degrees of freedom.

This adjustment is made to correct for non-circularity in the covariance matrix. Box suggested that rather than using the theoretical minimum value of the Geisser-Greenhouse epsilon, you should use the value estimated by the data.

Simulation studies have shown this adjustment to give very accurate probability levels. We recommend its use.

Huynh-Feldt Epsilon P-Value

This is the probability level of the corresponding F-ratio using the Huynh-Feldt version of the Geisser-Greenhouse correction. This correction involves multiplying both the numerator and denominator degrees of freedom by their epsilon and then calculating the probability level. Since this epsilon is a value between zero and one, the impact of this adjustment is to reduce the degrees of freedom.

This adjustment is made to correct for non-circularity in the covariance matrix. Huynh and Feldt showed that Geisser-Greenhouse estimate of epsilon was biased so they developed a less biased version. When this estimate is greater than one, it is set equal to one.

Simulation studies have shown this adjustment to give accurate probability levels, but not as accurate as Geisser-Greenhouse correction. Hence, we recommend the Geisser-Greenhouse correction.

Strategy for the Geisser-Greenhouse Adjustment

Kirk (1982) recommends the following three step testing strategy:

1. Check the Regular Prob Level. If this probability level is not significant (if it is not lower than 0.05, say), stop and declare the F not significant. If this F is significant, proceed to step 2.
2. Check the Lower-Bound Prob Level. If this probability is significant (less than 0.05, say), stop and declare the F significant. If this F is not significant, proceed to step 3.
3. Check the Geisser-Greenhouse Prob Level. If this probability is significant, stop and declare the F significant. If this probability level is not significant, declare the F as not significant.

Power of F-Tests with Geisser-Greenhouse Adjustments

Power of F-Tests with Geisser-Greenhouse Adjustments

Model Term	DF	F-Ratio	Power†			
			Regular	Lower-Bound Epsilon	Geisser-Greenhouse Epsilon	Huynh-Feldt Epsilon
A: Exercise	2	0.6124	0.13334			
B(A): Subject	15					
C: Time	2	36.9156	1.00000	0.99989	1.00000	1.00000
AC	4	6.4548	0.97763	0.83457	0.97663	0.97763
BC(A)	30					
S	0					

† Power was calculated using the observed F-Ratio as the population effect size with a significance level of $\alpha = 0.05$.

This table presents the F ratios from the analysis of variance table with the associated power values. The definition of power is discussed above in the Analysis of Variance section. This table lets you compare the statistical power of the four tests.

Note how the power decreases as the more conservative tests are used. Since the Geisser-Greenhouse is the most conservative test, it has the lowest power.

Model Term

The source of variation, which is the term in the model.

F-Ratio

The F-ratio is repeated from Analysis of Variance Table.

Regular Power

This gives the power. The definition of power and how it is calculating was provided in the Analysis of Variance Table section.

Lower-Bound Epsilon Power

This gives the power when the Lower-Bound correction is used.

Geisser-Greenhouse Epsilon Power

This gives the power when the Geisser-Greenhouse correction is used.

Huynh Feldt Epsilon Power

This gives the power when the Huynh Feldt correction is used.

Box's M Test for Equality of Between-Group Covariance Matrices

Box's M Test for Equality of Between-Group Covariance Matrices

Model Term	Box's M Statistic	DF1	DF2	Approximate F-Test		Approximate Chi ² -Test		Reject Equal Covariance Matrices at $\alpha = 0.2$?
				F-Ratio	P-Value	Chi ²	P-Value	
BC(A)	16.9396	12	1090.38	0.9883	0.45785	12.0459	0.44200	No

This section presents the results of a preliminary test to determine if the data meet the assumption of equal covariance matrices across groups. This test is discussed in detail in the Equality of Covariance Matrices chapter. Since the test depends heavily on the assumption of multivariate normality, when the data fail to pass the test, it may or may not be because of the covariances matrices are unequal.

When your data fail this test, one remedy is to transform the response variable by taking the square root, the logarithm, or the inverse. Often, a power transformation such as these will correct both non-normality and unequal variance. Of course, after applying such a variance stabilizing transformation, you have to discuss your results in the transformed metric—you cannot discuss the means in the original (untransformed) metric.

Note that this test requires the number of subjects per group to be greater than the number of levels of the within-subject factor(s).

Model Term

This is the term whose covariance matrices are being tested. The factor in parentheses represents the term(s) forming the groups, the first factor listed (B in this example) is the subject factor, and the rest of the factors are used to form the multivariate response. In this example, factor C, which has three levels, becomes the multivariate response vector. If more than one factor is listed here, they are combined into one single factor to form the multivariate response vector.

Box's M Statistic

This is the value of Box's M statistic used to test the equality of the covariance matrices.

DF1

The numerator degrees of freedom of the approximate F-distribution used to evaluate Box's M statistic. This value need not be an integer. This value is also the degrees of freedom of the approximate Chi-square statistic.

DF2

The denominator degrees of freedom of the approximate F-distribution used to evaluate Box's M statistic. This value need not be an integer.

Approximate F-Test F-Ratio

The value of the approximate F-test used to evaluate Box's M statistic.

Repeated Measures Analysis of Variance

Approximate F-Test P-Value

The probability level of the Box's M statistic based on an F-distribution approximation. If this value is less than a chosen significance value, such as 0.10, you must assume that the covariance matrices are not equal and take appropriate action.

Approximate Chi²-Test Chi² Value

The value of the approximate Chi-square test used to evaluate Box's M statistic.

Approximate Chi²-Test P-Value

The probability level of the Box's M statistic based on a Chi-square approximation. If this value is less than a chosen significance value, such as 0.10, you must assume that the covariance matrices are not equal and take appropriate action.

Reject Equal Covariance Matrices at $\alpha = 0.2$?

Using the value of the Assumption Alpha contained on the Reports tab panel, this provides the result of the test.

Covariance Matrix Circularity**Covariance Matrix Circularity**

Model Term	Epsilon			Mauchly's Test of Compound Symmetry*				Reject Covariance Matrix Circularity at $\alpha = 0.2$?
	Lower-Bound	Geisser-Greenhouse	Huynh-Feldt	Test Statistic	Chi ²	DF	P-Value	
BC(A)	0.5	0.9896294	1	0.9895	0.1475	2	0.92891	No

* Mauchly's Test actually tests the more restrictive assumption that the pooled covariance matrix has compound symmetry.

This section provides an analysis of the circularity (sometimes called the sphericity) assumption that is required for all of the within-subject F tests. The formulas are given in the Technical Details at the beginning of the chapter. You can often correct circularity problems by taking the logarithm of the responses.

Some statisticians believe you should ignore this test since it relies heavily on the multivariate normality of your data. They suggest that you routinely use Box's Geisser-Greenhouse correction which corrects for this problem.

Model Term

This is the term whose covariance matrix is being tested for circularity. The factor in parentheses represents the term(s) forming the groups, the first factor listed (B in this example) is the subject factor, and the rest of the factors are used to form the multivariate response. In this example, factor C, which has three levels, becomes the multivariate response vector. If more than one factor is listed, they are combined into one single factor to form the multivariate response vector.

Repeated Measures Analysis of Variance

Lower-Bound Epsilon

This is the minimum value of epsilon. The maximum value is one. This value is used to adjust the F-test by multiplying it times both the numerator and denominator degrees of freedom when calculating the probability levels.

Geisser-Greenhouse Epsilon

This is the estimate of epsilon that was suggested by Box. It serves as an index of the severity of non-circularity. Values of epsilon near one indicate that the covariance matrix is circular. Values of epsilon near the minimum (the Lower Bound Epsilon) indicate that the covariance matrix assumption is violated.

This value is used to adjust the F-test by multiplying it times both the numerator and denominator degrees of freedom when calculating the probability levels.

Huynh-Feldt Epsilon

This is an estimate of epsilon that was suggested by Huynh and Feldt to correct for bias found in the Geisser-Greenhouse estimate. This estimate is always greater than or equal to the Geisser-Greenhouse estimate. It is possible for this value to be greater than one. When this happens, the value is set equal to one.

Epsilon serves as an index of the severity of non-circularity. Values near one indicate that the covariance matrix is circular. Values near the minimum (the Lower Bound Epsilon) indicate that the covariance matrix assumption is violated.

This value is used to adjust the F-test by multiplying it times both the numerator and denominator degrees of freedom when calculating the probability levels.

Mauchly's Test Statistic

This is the value of Mauchly's test statistic. It tests the assumption that the pooled covariance matrix has compound symmetry. Compound symmetry is slightly more restrictive than circularity. The value of this statistic ranges from zero to one.

Chi²

This chi-square value is used to test the significance of the Mauchly test statistic.

DF

This is the degrees of freedom of the chi-square approximation of Mauchly's test statistic.

P-Value

This is the significance level of the chi-square test. When this value is small (0.10 or less), the data fail Mauchly's test for compound symmetry.

Reject Covariance Matrix Circularity at $\alpha = 0.2$?

This field indicates whether the data passed or failed Mauchly's test.

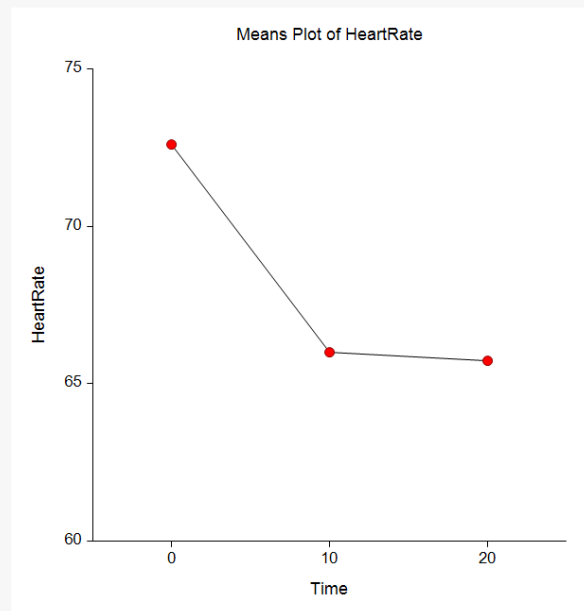
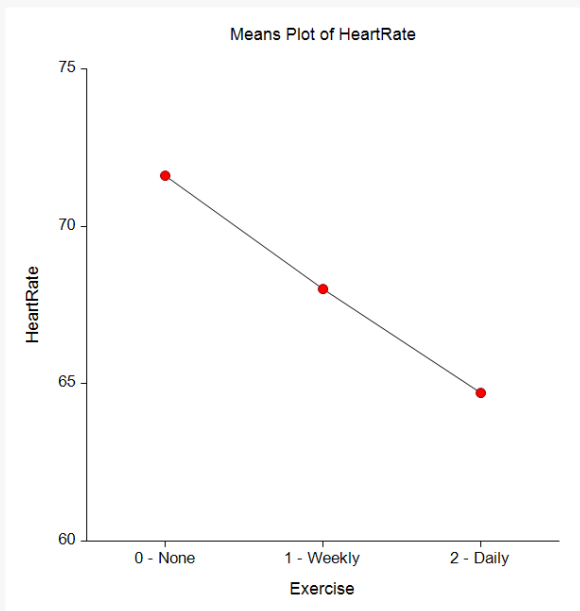
Repeated Measures Analysis of Variance

Means and Standard Errors and Plots

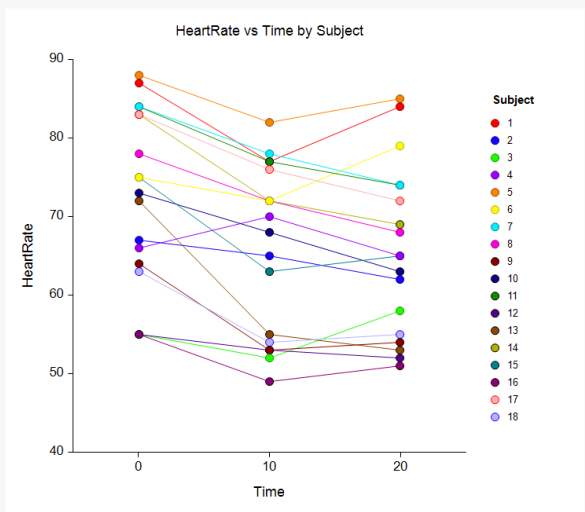
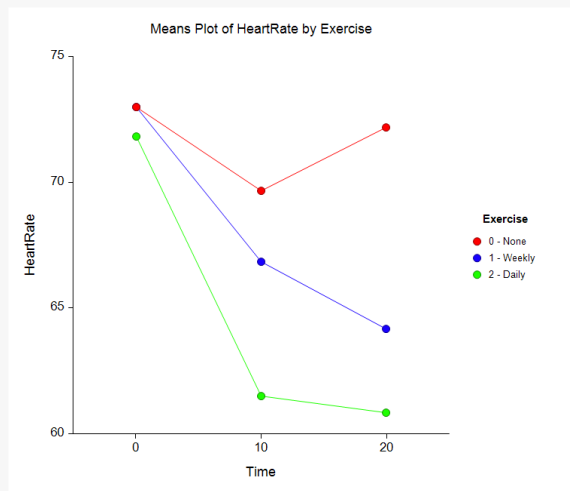
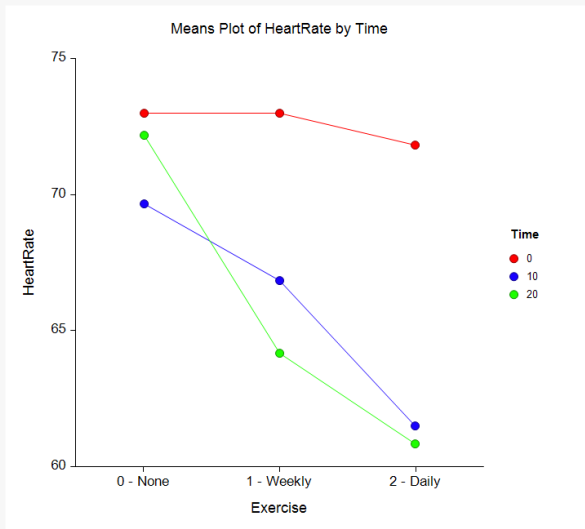
Means and Standard Errors

Term	Count	Mean	Standard Error
All	54	68.11111	
A: Exercise			
0 - None	18	71.61111	4.403095
1 - Weekly	18	68	4.403095
2 - Daily	18	64.72222	4.403095
C: Time			
0	18	72.61111	0.641821
10	18	66	0.641821
20	18	65.72222	0.641821
AC: Exercise, Time			
0 - None, 0	6	73	1.111667
0 - None, 10	6	69.66666	1.111667
0 - None, 20	6	72.16666	1.111667
1 - Weekly, 0	6	73	1.111667
1 - Weekly, 10	6	66.83334	1.111667
1 - Weekly, 20	6	64.16666	1.111667
2 - Daily, 0	6	71.83334	1.111667
2 - Daily, 10	6	61.5	1.111667
2 - Daily, 20	6	60.83333	1.111667

Plots



Repeated Measures Analysis of Variance



Term

The label for this line of the report.

Count

The number of observations in the mean.

Mean

The value of the sample mean.

Standard Error

The standard error of the mean. Note that these standard errors are the square root of the mean square of the error term for this term divided by the count. These standard errors are not the same as the simple standard errors calculated separately for each group. The standard errors reported here are those appropriate for testing multiple comparisons.

Repeated Measures Analysis of Variance

Plots

These plots display the means for each factor and two-way interaction. Note how easily you can see patterns in the plots.

Multiple Comparison Reports

Tukey-Kramer Multiple Comparison Test

Comparison Term: A: Exercise

Group	Count	Mean	Different From Groups
0 - None	18	71.61111	
1 - Weekly	18	68	
2 - Daily	18	64.72222	

$\alpha = 0.05$, Error Term = B(A), DF = 15, MSE = 348.9704, Critical Value = 3.6732

Notes:

This report provides multiple comparison tests for all pairwise differences between the means.

Tukey-Kramer Multiple Comparison Test

Comparison Term: C: Time

Group	Count	Mean	Different From Groups
0	18	72.61111	10, 20
10	18	66	0
20	18	65.72222	0

$\alpha = 0.05$, Error Term = BC(A), DF = 30, MSE = 7.414815, Critical Value = 3.4865

Notes:

This report provides multiple comparison tests for all pairwise differences between the means.

These sections present the results of the multiple-comparison procedures selected. These reports all use a uniform format that will be described by considering Tukey-Kramer Multiple-Comparison Test. The reports for the other procedures are similar. For more information on the interpretation of the various multiple-comparison procedures, turn to the section by that name in the One-Way ANOVA chapter.

Group

The label for this group.

Count

The number of observations in the mean.

Mean

The value of the sample mean.

Repeated Measures Analysis of Variance

Different from Groups

A list of those groups that are significantly different from this group according to this multiple-comparison procedure. All groups not listed are not significantly different from this group.

Alpha

The level of significance that you selected.

Error Term

The term in the ANOVA model that is used as the error term.

DF

The degrees of freedom for the error term.

MSE

The value of the mean square error.

Critical Value

The value of the test statistic that is “just significant” at the given value of alpha. This value depends on which multiple-comparison procedure you are using. It is based on the t-distribution or the studentized range distribution. It is the value of t, F, or q in the corresponding formulas.

Planned Comparison Reports

This section presents the results of any planned comparisons that were selected.

Planned Comparison: 0 - None vs. 1 - Weekly

Comparison Term: A: Exercise

Test Results

Comparison Value	Standard Error	T-Statistic	Two-Sided P-Value	Reject H0 at $\alpha = 0.05?$	95% Confidence Interval Limits for the Comparison Value	
					Lower	Upper
-3.611111	6.226916	0.5799	0.57058	No	-16.88347	9.661246

$\alpha = 0.05$, Error Term = B(A), DF = 15, MSE = 348.9704

Group Details

Group	Comparison Coefficient	Count	Mean
0 - None	-1	18	71.61111
1 - Weekly	1	18	68
2 - Daily	0	18	64.72222

Repeated Measures Analysis of Variance

Planned Comparison: 0 - None vs. 2 - Daily

Comparison Term: A: Exercise

Test Results

Comparison Value	Standard Error	T-Statistic	Two-Sided P-Value	Reject H0 at $\alpha = 0.05$?	95% Confidence Interval Limits for the Comparison Value	
					Lower	Upper
-6.888889	6.226916	1.1063	0.28602	No	-20.16125	6.383469

$\alpha = 0.05$, Error Term = B(A), DF = 15, MSE = 348.9704

Group Details

Group	Comparison Coefficient	Count	Mean
0 - None	-1	18	71.61111
1 - Weekly	0	18	68
2 - Daily	1	18	64.72222

Notes:

This section presents the planned comparison testing whether each group is significantly different from the first group. This would be useful when the first group is a control group.

Comparison Value

The value of the comparison. This is formed by multiplying the Comparison Coefficient times the Mean for each group and summing.

Standard Error

This is the standard error of the estimated comparison value. It is the denominator of the T-Value (above).

T-Statistic

The t-test used to test whether the above Comparison Value is significantly different from zero.

$$t_f = \frac{\sum_{i=1}^k c_i M_i}{\sqrt{MSE \sum_{i=1}^k \frac{c_i^2}{n_i}}}$$

where **MSE** is the mean square error, **f** is the degrees of freedom associated with **MSE**, **k** is the number of groups, **c_i** is the comparison coefficient for the *i*th group, **M_i** is the mean of the *i*th group, and **n_i** is the sample size of the *i*th group.

Two-Sided P-Value

The significance level of the above T-Value. The Comparison is statistically significant if this value is less than the specified alpha.

Repeated Measures Analysis of Variance

Reject H0 at $\alpha = 0.05$?

The decision based on the specified value of the multiple comparison alpha.

Group

The label for this group.

Comparison Coefficient

The coefficient (weight) used for this group.

Count

The number of observations in the mean.

Mean

The value of the sample mean.

Alpha

The level of significance that you selected.

Error Term

The term in the ANOVA model that is used as the error term.

DF

The degrees of freedom of the error term.

MSE

The value of the mean square error.

Example 2 – Single-Group Repeated Measures Design

This section presents an example of how to analyze a single-group repeated measures design. The dataset was given at the beginning of the chapter and is contained in the Reaction dataset.

Setup

To run this example, complete the following steps:

1 Open the Reaction example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Reaction** and click **OK**.

2 Specify the Repeated Measures Analysis of Variance procedure options

- Find and open the **Repeated Measures Analysis of Variance** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable(s) **Test**
 Subject Variable..... **Person**
 Between Factor Variable 1 **<Empty>**
 Within Factor Variable 1 **Drug**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Single-Group Repeated Measures Output

Expected Mean Squares

Model Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A: Person	4	No	S(AB)	S+bsA
B: Drug	3	Yes	AB	S+sAB+asB
AB	12	No	S(AB)	S+sAB
S(AB)	0	No		S

Note: Expected Mean Squares are for the balanced cell-frequency case.

Repeated Measures Analysis of Variance

Analysis of Variance Table

Model Term	DF	Sum of Squares	Mean Square	F-Ratio	P-Value
A: Person	4	680.8	170.2		
B: Drug	3	698.2	232.7333	24.7589	0.00002*
AB	12	112.8	9.4		
S	0				
Total (Adjusted)	19	1491.8			
Total	20				

* Term significant at $\alpha = 0.05$

P-Values for F-Tests with Geisser-Greenhouse Adjustments

Model Term	DF	F-Ratio	P-Value			
			Regular	Lower-Bound Epsilon	Geisser-Greenhouse Epsilon	Huynh-Feldt Epsilon
A: Person	4					
B: Drug	3	24.7589	0.00002*	0.00762*	0.00065*	0.00002*
AB	12					
S	0					

* Term significant at $\alpha = 0.05$

Covariance Matrix Circularity

Model Term	Epsilon			Mauchly's Test of Compound Symmetry*				Reject Covariance Matrix Circularity at $\alpha = 0.2$?
	Lower-Bound	Geisser-Greenhouse	Huynh-Feldt	Test Statistic	Chi ²	DF	P-Value	
AB	0.3333333	0.604874	1	0.1865	4.5716	5	0.47037	No

* Mauchly's Test actually tests the more restrictive assumption that the pooled covariance matrix has compound symmetry.

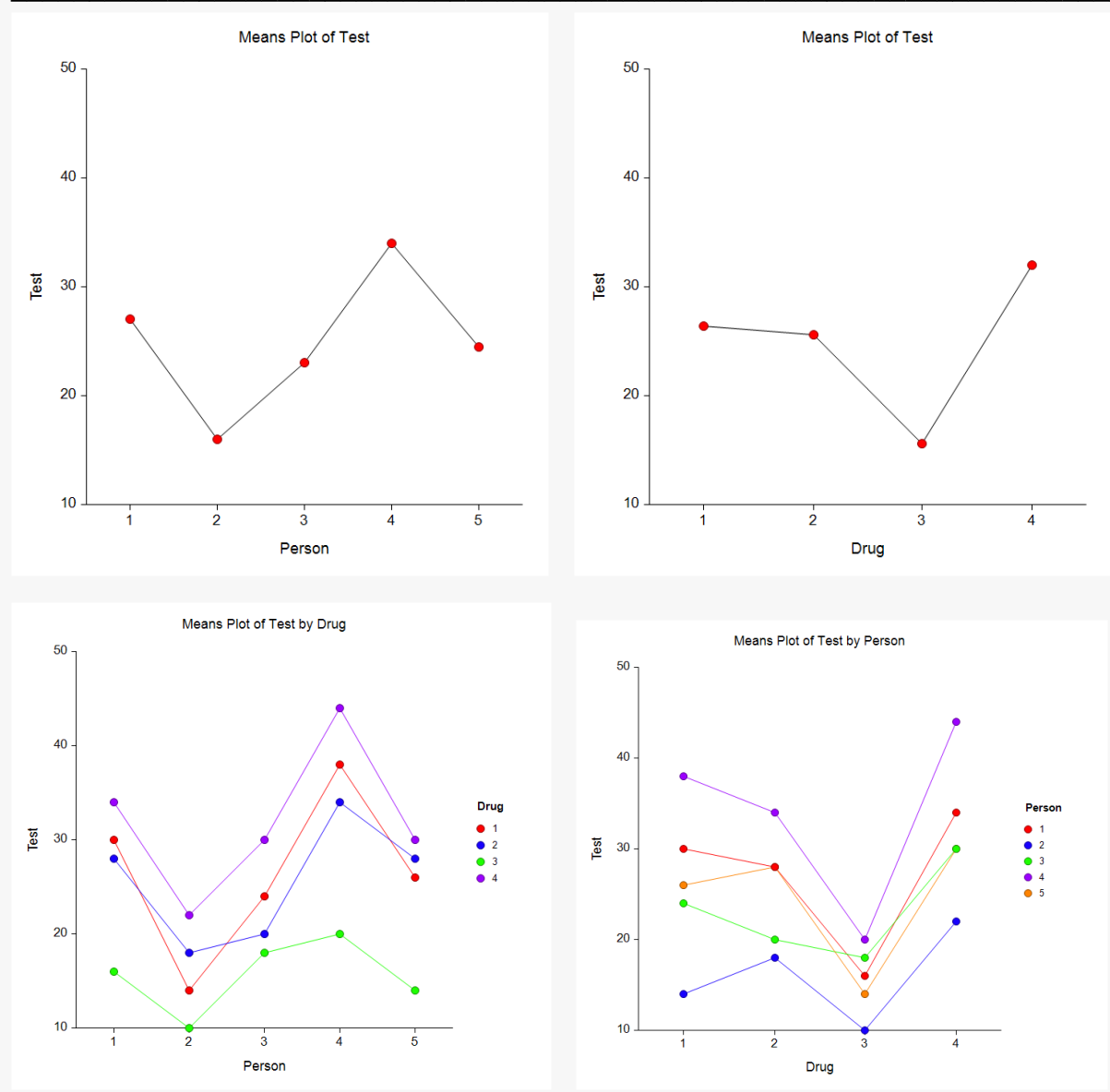
Repeated Measures Analysis of Variance

Means and Standard Errors

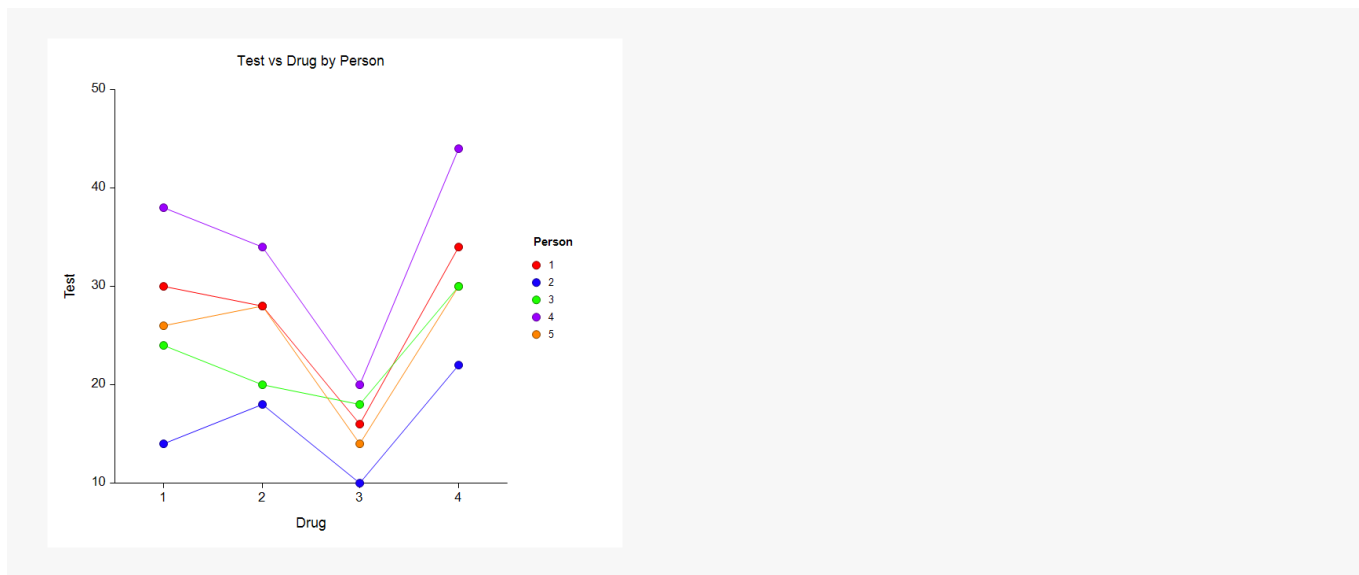
Term	Count	Mean	Standard Error
All	20	24.9	
A: Person			
1	4	27	0
2	4	16	0
3	4	23	0
4	4	34	0
5	4	24.5	0
B: Drug			
1	5	26.4	1.371131
2	5	25.6	1.371131
3	5	15.6	1.371131
4	5	32	1.371131
AB: Person, Drug			
1, 1	1	30	0
1, 2	1	28	0
1, 3	1	16	0
1, 4	1	34	0
2, 1	1	14	0
2, 2	1	18	0
2, 3	1	10	0
2, 4	1	22	0
3, 1	1	24	0
3, 2	1	20	0
3, 3	1	18	0
3, 4	1	30	0
4, 1	1	38	0
4, 2	1	34	0
4, 3	1	20	0
4, 4	1	44	0
5, 1	1	26	0
5, 2	1	28	0
5, 3	1	14	0
5, 4	1	30	0

Repeated Measures Analysis of Variance

Plots



Repeated Measures Analysis of Variance



Our only comment about this output is to note that the Box's M test section was omitted because there is only one group.