

## Chapter 235

# Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority

---

## Introduction

This procedure analyzes data from a two-treatment, two-period (2x2) cross-over design where the goal is to demonstrate non-inferiority of a treatment versus a control or reference. The response is assumed to be a continuous random variable that follows the normal distribution. When the normality assumption is suspect, the nonparametric Mann-Whitney U (or Wilcoxon Rank-Sum) Test may be employed.

In the two-period cross-over design, subjects are randomly assigned to one of two groups. One group receives treatment *R* followed by treatment *T*. The other group receives treatment *T* followed by treatment *R*. Thus, the response is measured at least twice on each subject.

Cross-over designs are used when the treatments alleviate a condition, rather than effect a cure. After the response to one treatment is measured, the treatment is removed, and the subject is allowed to return to a baseline response level. Next, the response to a second treatment is measured. Hence, each subject is measured twice, once with each treatment.

Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both of these cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.

---

## Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment into the study may be easier because each patient will receive both treatments.

---

## Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. It may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

The design cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to effect a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. It is arguably simpler to measure a subject once than to obtain their measurement twice. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

## Technical Details

Suppose you want to evaluate the non-inferiority of a treatment,  $T$ , as compared to a control or reference,  $R$ , using data on subjects in a 2x2 cross-over design, where a period effect may be present. This procedure allows you to perform this type of analysis.

### Cross-Over Analysis

In the discussion that follows, we summarize the presentation of Chow and Liu (1999). We suggest that you review their book for a more detailed presentation.

The general linear model for the standard 2x2 cross-over design is

$$Y_{ijk} = \mu + S_{ik} + P_j + F_{(j,k)} + C_{(j-1,k)} + e_{ijk}$$

where  $i$  represents a subject (1 to  $n_k$ ),  $j$  represents the period (1 or 2), and  $k$  represents the sequence (1 or 2). The  $S_{ik}$  represent the random effects of the subjects. The  $P_j$  represent the effects of the two periods. The  $F_{(j,k)}$  represent the effects of the two formulations (treatments). In the case of the 2x2 cross-over design

$$F_{(j,k)} = \begin{cases} F_R & \text{if } k = j \\ F_T & \text{if } k \neq j \end{cases}$$

where the subscripts  $R$  and  $T$  represent the *reference* and *treatment* formulations, respectively.

The  $C_{(j-1,k)}$  represent the carry-over effects. In the case of the 2x2 cross-over design

$$C_{(j-1,k)} = \begin{cases} C_R & \text{if } j = 2, k = 1 \\ C_T & \text{if } j = 2, k = 2 \\ 0 & \text{otherwise} \end{cases}$$

where the subscripts  $R$  and  $T$  represent the *reference* and *treatment* formulations, respectively.

Assuming that the average effect of the subjects is zero, the four means from the 2x2 cross-over design can be summarized using the following table.

Sequence	Period 1	Period 2
1(RT)	$\mu_{11} = \mu + P_1 + F_R$	$\mu_{21} = \mu + P_2 + F_T + C_R$
2(TR)	$\mu_{12} = \mu + P_1 + F_T$	$\mu_{22} = \mu + P_2 + F_R + C_T$

where  $P_1 + P_2 = 0$ ,  $F_T + F_R = 0$ , and  $C_T + C_R = 0$ .

## Treatment Effect

### Two-Sample T-Test for Treatment Effect

The presence of a treatment (drug) effect can be studied by testing whether  $F_T - F_R = M$  using a  $t$  test. This test is calculated as follows

$$T_d = \frac{\hat{F} - M}{\hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$\hat{F} = \bar{d}_{.1} - \bar{d}_{.2}$$

$$\bar{d}_{.k} = \frac{1}{n_k} \sum_{i=1}^{n_k} d_{ik}$$

$$\hat{\sigma}_d^2 = \frac{1}{(n_1 + n_2 - 2)} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{.k})^2$$

$$d_{ik} = \frac{Y_{i2k} - Y_{i1k}}{2}$$

The null hypothesis of no drug effect is rejected at the  $\alpha$  significance level if

$$|T_d| > t_{\alpha/2, n_1 + n_2 - 2}$$

A  $100(1 - \alpha)\%$  confidence interval for  $F = F_T - F_R$  is given by

$$\hat{F} \pm (t_{\alpha/2, n_1 + n_2 - 2}) \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

### Mann-Whitney U (or Wilcoxon Rank-Sum) Test for Treatment Effect

Senn (2002) pages 113-114 describes Koch's adaptation of the Wilcoxon-Mann-Whitney rank sum test that tests treatment effects in the presence of period effects. The test is based on the period differences and assumes that there are no carryover effects. Koch's method calculates the ranks of the period differences for all subjects in the trial and then uses the Mann-Whitney U (or Wilcoxon Rank-Sum) Test to analyze these differences between the two sequence groups. The Mann-Whitney U (or Wilcoxon Rank-Sum) Test is described in detail in the Two-Sample T-Test chapter of the documentation.

## Carryover Effect

The 2x2 cross-over design should only be used when there is no carryover effect from one period to the next. The presence of a carryover effect can be studied by testing whether  $C_T = C_R = 0$  using a  $t$  test. This test is calculated as follows

$$T_c = \frac{\hat{C}}{\hat{\sigma}_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$\hat{C} = \bar{U}_{.2} - \bar{U}_{.1}$$

$$\bar{U}_{.k} = \frac{1}{n_k} \sum_{i=1}^{n_k} U_{ik}$$

$$\hat{\sigma}_u^2 = \frac{1}{(n_1 + n_2 - 2)} \sum_{k=1}^2 \sum_{i=1}^{n_k} (U_{ik} - \bar{U}_{.k})^2$$

$$U_{ik} = Y_{i1k} + Y_{i2k}$$

The null hypothesis of no carryover effect is rejected at the  $\alpha$  significance level if

$$|T_c| > t_{\alpha/2, n_1+n_2-2}.$$

A 100(1 -  $\alpha$ )% confidence interval for  $C = C_T - C_R$  is given by

$$\hat{C} \pm (t_{\alpha/2, n_1+n_2-2}) \hat{\sigma}_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

## Period Effect

The presence of a period effect can be studied by testing whether  $P_1 = P_2 = 0$  using a  $t$  test. This test is calculated as follows

$$T_P = \frac{\hat{P}}{\hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$\hat{P} = \bar{O}_{.1} - \bar{O}_{.2}$$

$$\bar{O}_{.1} = \bar{d}_1$$

$$\bar{O}_{.2} = -\bar{d}_2$$

## Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority

$$\hat{\sigma}_d^2 = \frac{1}{(n_1 - n_2 - 2)} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{.k})^2$$

$$d_{ik} = \frac{Y_{i2k} - Y_{i1k}}{2}$$

The null hypothesis of no drug effect is rejected at the  $\alpha$  significance level if

$$|T_P| > t_{\alpha/2, n_1+n_2-2}.$$

A  $100(1 - \alpha)\%$  confidence interval for  $P = P_2 - P_1$  is given by

$$\hat{P} \pm (t_{\alpha/2, n_1+n_2-2}) \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

---

## Non-Inferiority Test

This discussion is based on the book by Rothmann, Wiens, and Chan (2012) which discusses the two-independent sample case. Assume that higher values are better, that  $\mu_T$  and  $\mu_R$  represent the means of the two treatment groups, and that  $M$  is the positive *non-inferiority margin*. The null and alternative hypotheses when the **higher values are better** are

$$H_0: (\mu_T - \mu_R) \leq -M$$

$$H_1: (\mu_T - \mu_R) > -M$$

or

$$H_0: \mu_T \leq \mu_R - M$$

$$H_1: \mu_T > \mu_R - M$$

If, on the other hand, we assume that **higher values are worse**, then null and alternative hypotheses are

$$H_0: (\mu_T - \mu_R) \geq M$$

$$H_1: (\mu_T - \mu_R) < M$$

or

$$H_0: \mu_T \geq \mu_R + M$$

$$H_1: \mu_T < \mu_R + M$$

The two-sample t-test is usually employed to test that the treatment effect is zero. The non-inferiority test is a one-sided two-sample t-test that compares the treatment effect to a non-zero quantity,  $M$ . A one-sided edition of the Mann-Whitney U (or Wilcoxon Rank-Sum) nonparametric test is also optionally available.

## Data Structure

The data for a cross-over design is entered into three variables. The first variable contains the sequence number, the second variable contains the response in the first period, and the third variable contains the response in the second period. Note that each row of data represents the complete response for a single subject.

Chow and Liu (1999) give the following data on page 73. These data are contained in the dataset called ChowLiu73.

### ChowLiu73 Dataset

Sequence	Period 1	Period 2
1	74.675	73.675
1	96.400	93.250
1	101.950	102.125
1	79.050	69.450
1	79.050	69.025
1	85.950	68.700
1	69.725	59.425
1	86.275	76.125
1	112.675	114.875
1	99.525	116.250
1	89.425	64.175
1	55.175	74.575
2	74.825	37.350
2	86.875	51.925
2	81.675	72.175
2	92.700	77.500
2	50.450	71.875
2	66.125	94.025
2	122.450	124.975
2	99.075	85.225
2	86.350	95.925
2	49.925	67.100
2	42.700	59.425
2	91.725	114.05

## Example 1 – 2x2 Cross-Over Analysis for Non-Inferiority

This section presents an example of how to perform a non-inferiority test in an analysis of data from a 2x2 cross-over design. Chow and Liu (1999) page 73 provide an example of data from a 2x2 cross-over design. These data were shown in the Data Structure section earlier in this chapter. In this example, we'll assume that the test formulation will be deemed non-inferior if its mean is no more than 5 points lower than the standard or reference formulation.

### Setup

To run this example, complete the following steps:

#### 1 Open the ChowLiu73 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **ChowLiu73** and click **OK**.

#### 2 Specify the Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority procedure options

- Find and open the **Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority** 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

Sequence Group Variable .....	<b>Sequence</b>
Period 1 Variable .....	<b>Period1</b>
Period 2 Variable .....	<b>Period2</b>
Higher Values Are.....	<b>Better</b>
Non-Inferiority Margin .....	<b>5</b>

##### Reports Tab

Show Written Explanations.....	<b>Checked</b>
--------------------------------	----------------

#### 3 Run the procedure

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

## Cross-Over Effects and Means Summary

### Cross-Over Effects and Means Summary

Parameter	Count	Parameter Estimate	Standard Deviation	Standard Error	T*	Lower 95.0% CL	Upper 95.0% CL
Treatment Effect	24	-2.288	9.145	3.733	2.0739	-10.030	5.455
$\mu_1$ (or $\mu_R$ )	24	82.559		4.285			
$\mu_2$ (or $\mu_T$ )	24	80.272		4.395			
Period Effect	24	-1.731	9.145	3.733	2.0739	-9.474	6.011
$\mu (\text{Period}=1)$	24	82.281		4.043			
$\mu (\text{Period}=2)$	24	80.550		4.618			
Carryover Effect	24	-9.592	38.390	15.673	2.0739	-42.095	22.911
$(R+T) (Seq=1)$	12	167.627	33.234	9.594			
$(R+T) (Seq=2)$	12	158.035	42.930	12.393			

This report shows the estimated effects, means, standard deviations, standard errors, and confidence limits of various parameters and subgroups of the data. The least squares mean of treatment R is 82.559 and of treatment T is 80.272. The treatment effect  $[(\mu_2 - \mu_1)$  or  $(\mu_T - \mu_R)]$  is estimated to be -2.288. The period effect  $[(\mu|(\text{Period}=2)) - (\mu|(\text{Period}=1))]$  is estimated to be -1.731. The carryover effect  $[((R+T)|(Seq=2)) - ((R+T)|(Seq=1))]$  is estimated to be -9.592. Note that least squares means are created by taking the simple average of their component means, not by taking the average of the raw data. For example, if the mean of the 20 subjects in period 1 sequence 1 is 50.0 and the mean of the 10 subjects in period 2 sequence 2 is 40.0, the least squares mean is  $(50.0 + 40.0)/2 = 45.0$ . That is, no adjustment is made for the unequal sample sizes. Also note that the standard deviation of some of the subgroups is not calculated.

This report summarizes the estimated means and the treatment, period, and carryover effects.

### Parameter

These are the items displayed on the corresponding lines. Note that the *Treatment* line is the main focus of the analysis. The *Period* and *Carryover* information is used for preliminary tests of assumptions.

### Count

The count gives the number of non-missing values. This value is often referred to as the group sample size or  $n$ .

### Parameter Estimate

These are the estimated values of the corresponding parameters. Formulas for the three effects were given in the Technical Details section earlier in this chapter.

### Standard Deviation

The sample standard deviation is the square root of the sample variance. It is a measure of spread.

### Standard Error

These are the standard errors of each of the effects. They provide an estimate of the precision of the effect estimate. The formulas were given earlier in the Technical Details section of this chapter.

### T\*

This is the t-value used to construct the confidence interval. If you were constructing the interval manually, you would obtain this value from a table of the Student's t distribution with  $n - 1$  degrees of freedom.



## Lower and Upper Confidence Limits

These values provide a confidence interval for the estimated effect.

## Interpretation of the Above Report

This section provides a written interpretation of the above report.

## Cross-Over Analysis Detail

### Cross-Over Analysis Detail

Seq.	Period	Treatment	Count	Least Squares Mean	Standard Deviation	Standard Error
1	1	R	12	85.823	15.691	4.530
2	2	R	12	79.296	25.198	7.274
1	2	T	12	81.804	19.712	5.690
2	1	T	12	78.740	23.207	6.699
1	Difference	(T-R)/2	12	-2.009	6.423	1.854
2	Difference	(T-R)/2	12	0.278	11.225	3.240
1	Total	R+T	12	167.627	33.234	9.594
2	Total	R+T	12	158.035	42.930	12.393
.	.	R	24	82.559		4.285
.	.	T	24	80.272		4.395
1	.	.	24	83.814		
2	.	.	24	79.018		
.	1	.	24	82.281		4.043
.	2	.	24	80.550		4.618

This report shows the estimated effects, means, standard deviations, and standard errors of various subgroups of the data. The least squares mean of treatment R is 82.559 and of treatment T is 80.272. Note that least squares means are created by taking the simple average of their component means, not by taking the average of the raw data. For example, if the mean of the 20 subjects in period 1 sequence 1 is 50.0 and the mean of the 10 subjects in period 2 sequence 2 is 40.0, the least squares mean is  $(50.0 + 40.0)/2 = 45.0$ . That is, no adjustment is made for the unequal sample sizes. Also note that the standard deviation and standard error of some of the subgroups are not calculated.

This report provides the least squares means of various subgroups of the data.

### Seq.

This is the sequence number of the mean shown on the line. When the dot (period) appears in this line, the results displayed are created by taking the simple average of the appropriate means of the two sequences.

### Period

This is the period number of the mean shown on the line. When the dot (period) appears in this line, the results displayed are created by taking the simple average of the appropriate means of the two periods.

### Treatment

This is the treatment (or formulation) of the mean shown on the line. When the dot (period) appears in this line, the results displayed are created by taking the simple average of the appropriate means of the two treatments.

When the entry is  $(T-R)/2$ , the mean is computed on the quantities created by dividing the difference in each subject's two scores by 2. When the entry is  $R+T$ , the mean is computed on the sums of the subjects two scores.

## Count

The count is the number of subjects in the mean.

## Least Squares Mean

Least squares means are created by taking the simple average of their component means, not by taking a weighted average based on the sample size in each component. For example, if the mean of the 20 subjects in period 1 sequence 1 is 50.0 and the mean of the 10 subjects in period 2 sequence 2 is 40.0, the least squares mean is  $(50.0 + 40.0)/2 = 45.0$ . That is, no adjustment is made for the unequal sample sizes. Since least squares means are used in all subsequent calculations, these are the means that are reported.

## Standard Deviation

This is the estimated standard deviation of the subjects in the mean.

## Standard Error

This is the estimated standard error of the least squares mean.

---

## T-Test for Non-Inferiority

### T-Test for Non-Inferiority

Alternative Hypothesis	Mean Difference	Standard Error	T-Value	DF	Prob Level	Conclude Non-Inferiority at $\alpha = 0.050$ ?
$\mu_T > \mu_R - 5$	-2.288	3.733	0.7266	22	0.23757	No

The two treatment means in a 2x2 cross-over study are not significantly different based on a one-sided T-Test at the 0.05 significance level (the actual probability level was 0.23757). The design has 12 subjects in sequence 1 (RT) and 12 subjects in sequence 2 (TR). The average response to treatment R is 82.559, and the average response to treatment T is 80.272. The estimated treatment effect ( $\mu_T - \mu_R$ ) is -2.288.

This report presents the T-test for non-inferiority of the treatment. In this case, we cannot conclude non-inferiority with a P-value of 0.23757.

## Alternative Hypothesis

This states the alternative hypothesis of non-inferiority that is being tested.

## Mean Difference

This is the difference between the treatment means,  $\hat{\mu}_2 - \hat{\mu}_1$ . This is known as the treatment effect.

## Standard Error

This is the standard error of the treatment effect. It provides an estimate of the precision of the treatment effect estimate. The formula was given earlier in the Technical Details section of this chapter.

## T-Value

This is the test statistic calculated from the data that is used to test whether the effect is different from the null hypothesized value (0 in this case).

## DF

The *DF* is the value of the degrees of freedom. This is two less than the total number of subjects in the study.

## Prob Level

This is the probability level (*p*-value) of the test. If this value is less than the chosen significance level, then the corresponding effect is said to be significant. For example, if you are testing at a significance level of 0.05, then probabilities that are less than 0.05 are statistically significant. You should choose a value appropriate for your study.

## Conclude Non-Inferiority at $\alpha = 0.050$ ?

This column indicates whether or not the null hypothesis is rejected, in favor of the alternative hypothesis of non-inferiority, based on the *p*-value and chosen  $\alpha$ . A test in which the null hypothesis is rejected is sometimes called *significant*.

## T-Tests for Period and Carryover Effects (Two-Sided)

**T-Tests for Period and Carryover Effects (Two-Sided)**

Parameter	Estimate	Standard Error	T-Value	DF	Prob Level	Reject H0 at $\alpha = 0.050$ ?
Period Effect	-1.731	3.733	-0.4637	22	0.64739	No
Carryover Effect	-9.592	15.673	-0.6120	22	0.54681	No

A preliminary test failed to reject the assumption of equal period effects at the 0.05 significance level (the actual probability level was 0.64739). A preliminary test failed to reject the assumption of equal carryover effects at the 0.05 significance level (the actual probability level was 0.54681).

This report presents the T-tests for the period and carryover effects. In this case, both are not significantly different from zero.

## Parameter

These are the items being tested. The *Period* and *Carryover* lines are preliminary tests of assumptions.

## Estimate

These are the estimated values of the corresponding effects. Formulas for these effects were given in the Technical Details section earlier in this chapter.

## Standard Error

These are the standard errors of each of the effects. They provide an estimate of the precision of the effect estimate. The formulas were given earlier in the Technical Details section of this chapter.

## T-Value

These are the test statistics calculated from the data that are used to test whether the effect is different from zero.

## DF

The *DF* is the value of the degrees of freedom. This is two less than the total number of subjects in the study.

## Prob Level

This is the probability level (p-value) of the test. If this value is less than the chosen significance level, then the corresponding effect is said to be significant. Some authors recommend that the tests of assumptions (Period and Carryover) should be done at the 0.10 level of significance.

---

## Mann-Whitney U or Wilcoxon Rank-Sum Location Difference Test for Non-Inferiority

This test is the most common nonparametric substitute for the *t*-test when the assumption of normality is not valid. The test is based on the period differences and assumes that there are no carryover effects. This method calculates the ranks of the period differences for all subjects in the trial and then uses the Mann-Whitney U (or Wilcoxon Rank-Sum) Test to analyze these differences between the two sequence groups.

This test has good properties (asymptotic relative efficiency) for symmetric distributions. There are exact procedures for this test given small samples with no ties, and there are large sample approximations. When ties are present in the data, an approximate solution for dealing with ties is available. You can use the approximation provided but know that the exact results no longer hold.

The Mann-Whitney U (or Wilcoxon Rank-Sum) Test is described in detail in the Two-Sample T-Test chapter of the documentation.

---

### Mann-Whitney U or Wilcoxon Rank-Sum Location Difference Test for Non-Inferiority

---

#### Sequence Details

Sequence	Mann-Whitney U	Sum of Ranks (W)	Mean of W	Std Dev of W
R/T	68	146	150	17.32051
T/R	76	154	150	17.32051

Number of Sets of Ties = 0, Multiplicity Factor = 0

---

#### Test Results

Test Type	Alternative Hypothesis†	Z-Value	Prob Level	Conclude Non-Inferiority at $\alpha = 0.050$ ?
Exact*	LocT > LocR - 5		0.42142	No
Normal Approximation	LocT > LocR - 5	0.2309	0.40868	No
Normal Approx. with C.C.	LocT > LocR - 5	0.2021	0.41993	No

† "LocT" and "LocR" refer to the period difference location parameters of the two treatment groups distributions.

\* The Exact Test is provided only when there are no ties and the sample size is  $\leq 20$  in both groups.

---

The two location parameters in a 2x2 cross-over study are not significantly different based on a one-sided Exact Mann-Whitney U or Wilcoxon Rank-Sum Test at the 0.05 significance level (the actual probability level was 0.42142). The two location parameters in a 2x2 cross-over study are not significantly different based on a one-sided Mann-Whitney U or Wilcoxon Rank-Sum Test based on the Normal Approximation at the 0.05 significance level (the actual probability level was 0.40868). The two location parameters in a 2x2 cross-over study are not significantly different based on a one-sided Mann-Whitney U or Wilcoxon Rank-Sum Test based on the Normal Approximation with Continuity Correction at the 0.05 significance level (the actual probability level was 0.41993). The design had 12 subjects in sequence 1 (R/T) and 12 subjects in sequence 2 (T/R). The sum of ranks of the period differences in the first sequence, R/T, was 146. The sum of ranks of the period differences in the second sequence, T/R, was 154.

## Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority

This report presents the results of the Wilcoxon-Mann-Whitney test for non-inferiority of the treatment. The null and alternative hypotheses relate to the equality or non-equality of the central tendency of the two distributions of period differences. The software adds the null-hypothesized difference (i.e., non-inferiority margin) to each value of Group 2, and the test is run based on the original Group 1 values and the transformed Group 2 values.

The exact test is only provided when there are no ties and the sample size is less than or equal to 20 in both sequence groups.

## Tests of Assumptions Section

This section presents the results of tests for checking the normality assumption.

### Tests of the Normality Assumption for the Period Differences in Sequence 1

Normality Test	Test Statistic	Prob Level	Reject H0 of Normality at $\alpha = 0.050$ ?
Shapiro-Wilk	0.9418	0.52170	No
Skewness	-0.7849	0.43251	No
Kurtosis	0.3616	0.71767	No
Omnibus (Skewness or Kurtosis)	0.7468	0.68839	No

One of the underlying assumptions for the T-test is that the differences are normally distributed. This report presents the results of four different normality tests for the period differences within each sequence: Shapiro-Wilk Normality, Skewness Normality, Kurtosis Normality, Omnibus Normality.

### Tests of the Normality Assumption for the Period Differences in Sequence 2

Normality Test	Test Statistic	Prob Level	Reject H0 of Normality at $\alpha = 0.050$ ?
Shapiro-Wilk	0.9091	0.20784	No
Skewness	0.9127	0.36138	No
Kurtosis	-0.8364	0.40293	No
Omnibus (Skewness or Kurtosis)	1.5327	0.46472	No

One of the underlying assumptions for the T-test is that the differences are normally distributed. This report presents the results of four different normality tests for the period differences within each sequence: Shapiro-Wilk Normality, Skewness Normality, Kurtosis Normality, Omnibus Normality.

## Shapiro-Wilk Normality

This test for normality has been found to be the most powerful test in most situations. It is the ratio of two estimates of the variance of a normal distribution based on a random sample of  $n$  observations. The numerator is proportional to the square of the best linear estimator of the standard deviation. The denominator is the sum of squares of the observations about the sample mean. The test statistic  $W$  may be written as the square of the Pearson correlation coefficient between the ordered observations and a set of weights which are used to calculate the numerator. Since these weights are asymptotically proportional to the corresponding expected normal order statistics,  $W$  is roughly a measure of the straightness of the normal quantile-quantile plot. Hence, the closer  $W$  is to one, the more normal the sample is.

The probability values for  $W$  are valid for sample sizes greater than 3. The test was developed by Shapiro and Wilk (1965) for sample sizes up to 20. **NCSS** uses the approximations suggested by Royston (1992) and Royston (1995) which allow unlimited sample sizes. Note that Royston only checked the results for sample

## Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority

sizes up to 5000 but indicated that he saw no reason larger sample sizes should not work.  $W$  may not be as powerful as other tests when ties occur in your data.

### Skewness Normality

This is a skewness test reported by D'Agostino (1990). Skewness implies a lack of symmetry. One characteristic of the normal distribution is that it has no skewness. Hence, one type of non-normality is skewness.

The Value is the test statistic for skewness, while Prob Level is the  $p$ -value for a two-tailed test for a null hypothesis of normality. If this  $p$ -value is less than a chosen level of significance, there is evidence of non-normality. Under Decision ( $\alpha = 0.050$ ), the conclusion about skewness normality is given.

### Kurtosis Normality

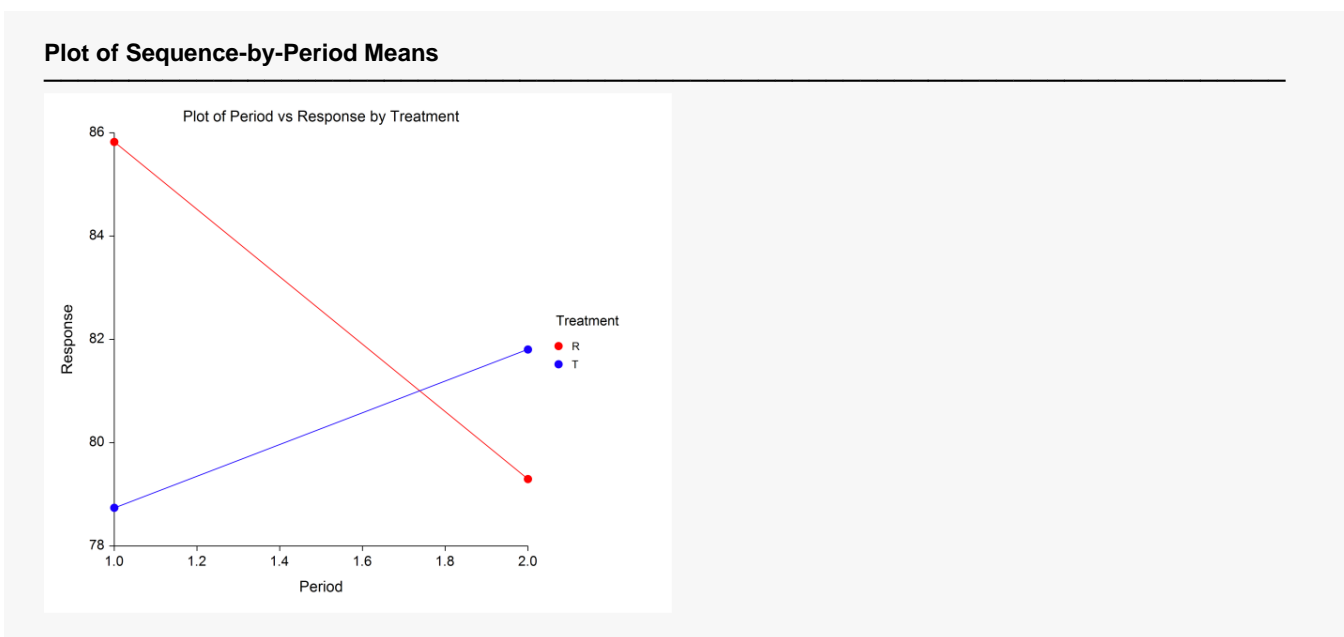
Kurtosis measures the heaviness of the tails of the distribution. D'Agostino (1990) reported a second normality test that examines kurtosis. The Value column gives the test statistic for kurtosis, and Prob Level is the  $p$ -value for a two-tail test for a null hypothesis of normality. If this  $p$ -value is less than a chosen level of significance, there is evidence of kurtosis non-normality. Under Decision ( $\alpha = 0.050$ ), the conclusion about normality is given.

### Omnibus Normality

This third normality test, also developed by D'Agostino (1990), combines the skewness and kurtosis tests into a single measure. Here, as well, the null hypothesis is that the underlying distribution is normally distributed. The definitions for Value, Prob Level, and Decision are the same as for the previous two normality tests.

---

## Plot of Sequence-by-Period Means

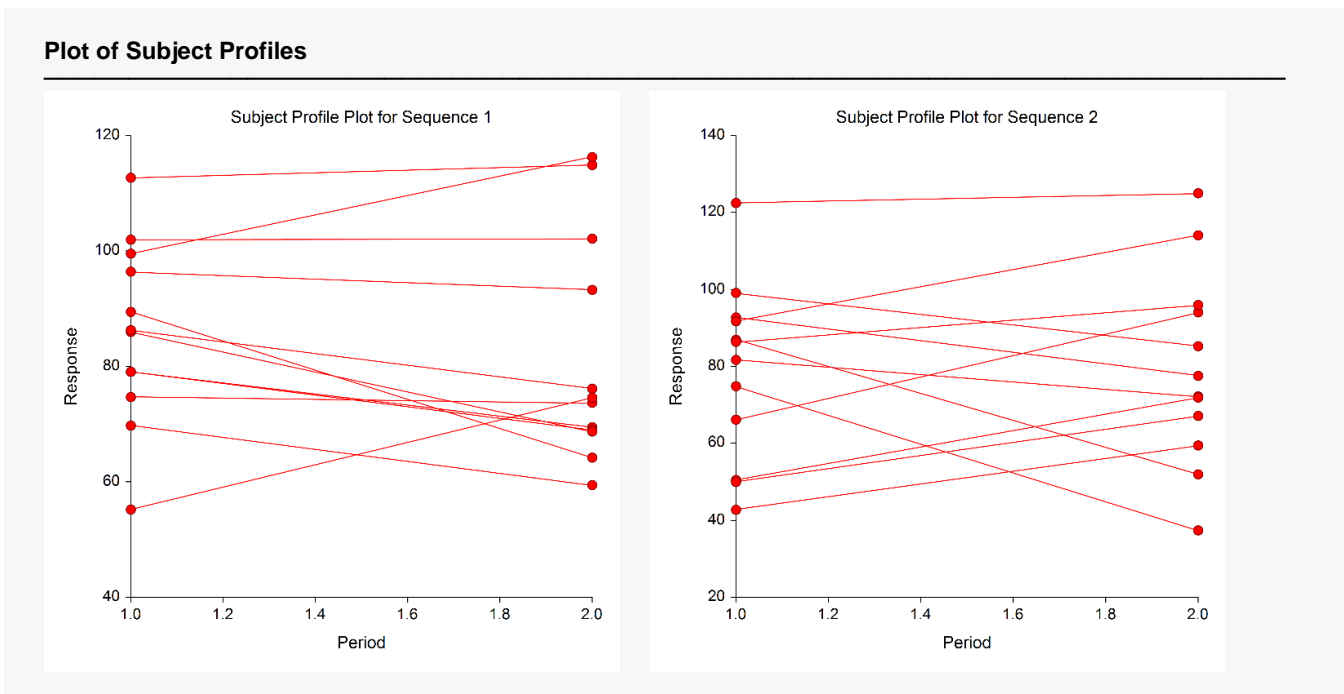


The sequence-by-period means plot shows the mean responses on the vertical axis and the periods on the horizontal axis. The lines connect like treatments. The distance between these lines represents the magnitude of the treatment effect.

## Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority

If there is no period, carryover, or interaction effects, two horizontal lines will be displayed. The tendency for both lines to slope up or down represents period and carryover effects. The tendency for the lines to cross represents period-by-treatment interaction. This is also a type of carryover effect.

## Plot of Subject Profiles

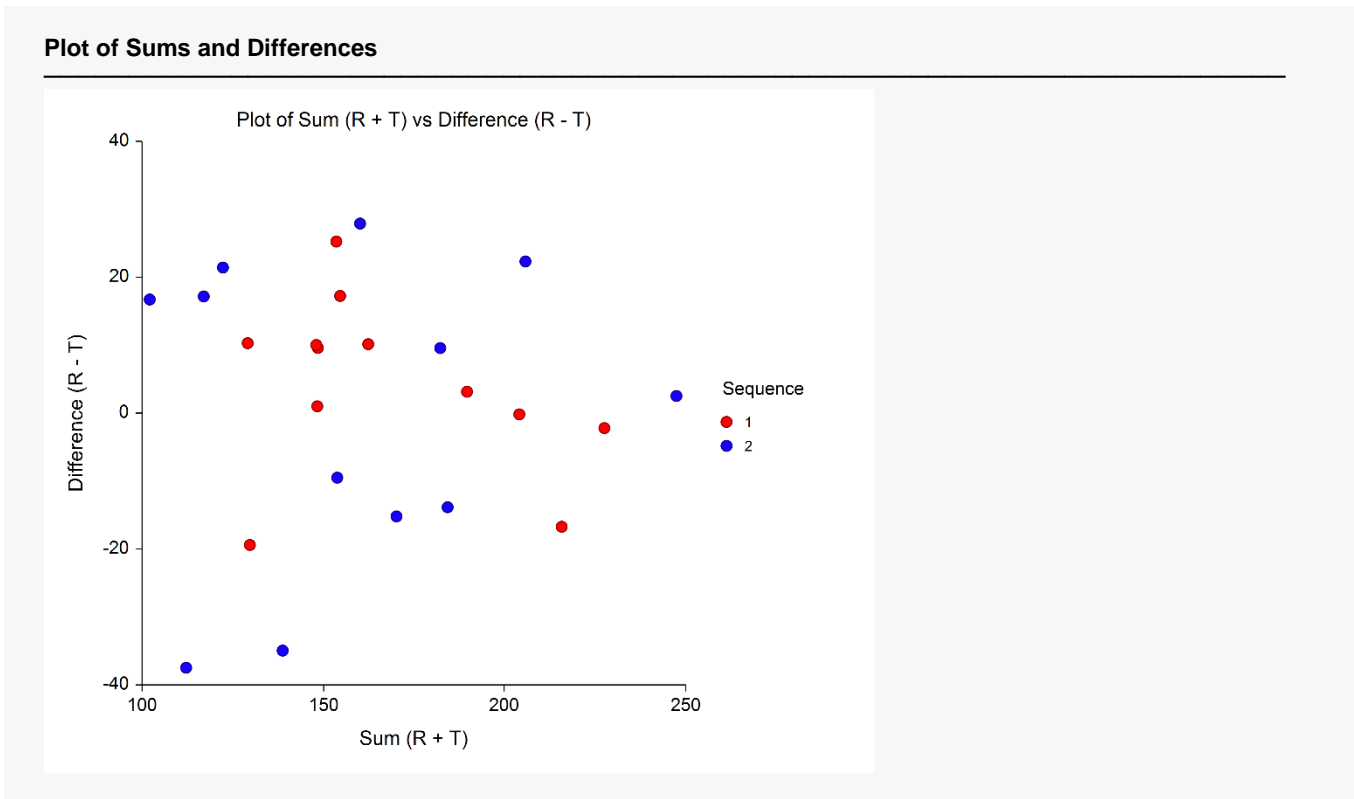


The profile plot displays the raw data for each subject. The response variable is shown along the vertical axis. The two sequences are shown along the horizontal axis. The data for each subject is depicted by two points connected by a line. The subject's response to the reference formulation is shown first followed by their response to the treatment formulation. Hence, for sequence 2, the results for the first period are shown on the right and for the second period on the left.

This plot is used to develop a feel for your data. You should view it first as a tool to check for outliers (points and subjects that are very different from the majority). Note that outliers should be removed from the analysis only if a reason can be found for their deletion. Of course, the first step in dealing with outliers is to double-check the data values to determine if a typing error might have caused them. Also, look for subjects whose lines exhibit a very different pattern from the rest of the subjects in that sequence. These might be a signal of some type of data-recording or data-entry error.

The profile plot allows you to assess the consistency of the responses to the two treatments across subjects. You may also be able to evaluate the degree to which the variation is equal in the two sequences.

## Plot of Sums and Differences



The sums and differences plot shows the sum of each subject's two responses on the horizontal axis and the difference between each subject's two responses on the vertical axis. Dot plots of the sums and differences have been added above and to the right, respectively.

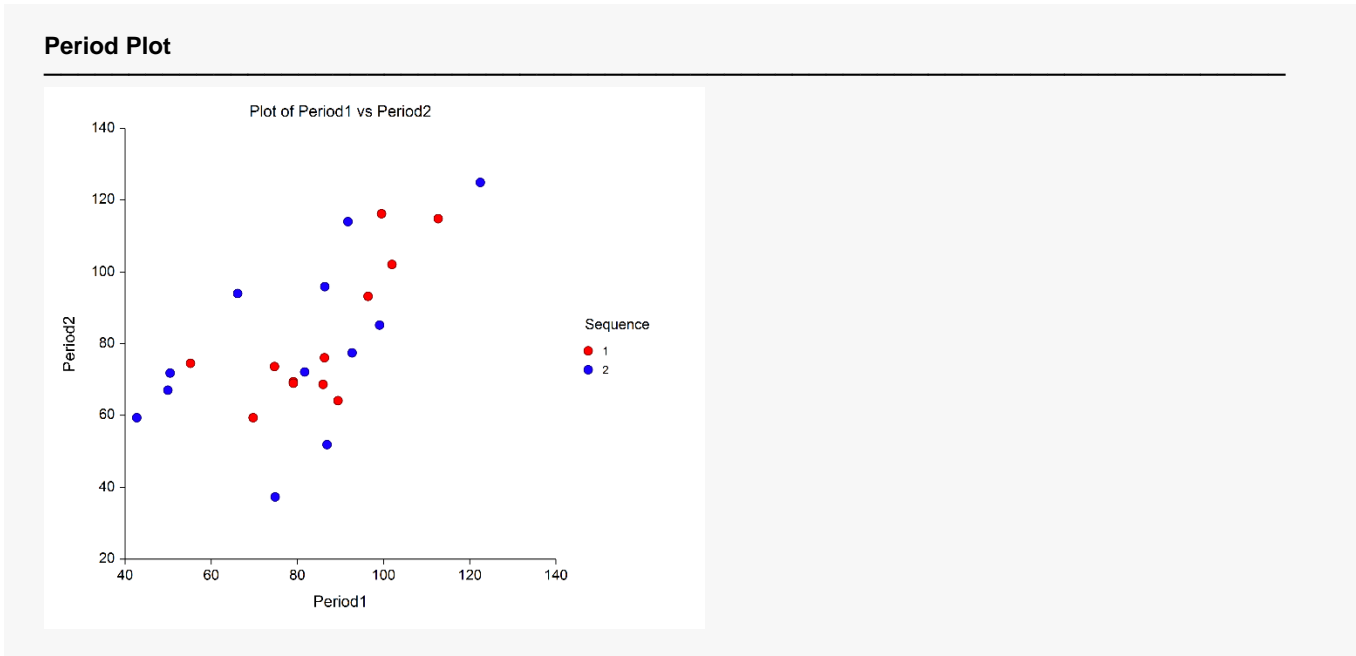
Each point represents the sum and difference of a single subject. Different plotting symbols are used to denote the subject's sequence. A horizontal line has been added at zero to provide an easy reference from which to determine if a difference is positive (favors treatment R) or negative (favors treatment T).

The degree to which the plotting symbols tend to separate along the horizontal axis represents the size of the carryover effect. The degree to which the plotting symbols tend to separate along the vertical axis represents the size of the treatment effect.

Outliers are easily detected on this plot. Outlying subjects should be reviewed for data-entry errors and for special conditions that might have caused their responses to be unusual. Outliers should not be removed from an analysis just because they are different. A compelling reason should be found for their removal and the removal should be well documented.

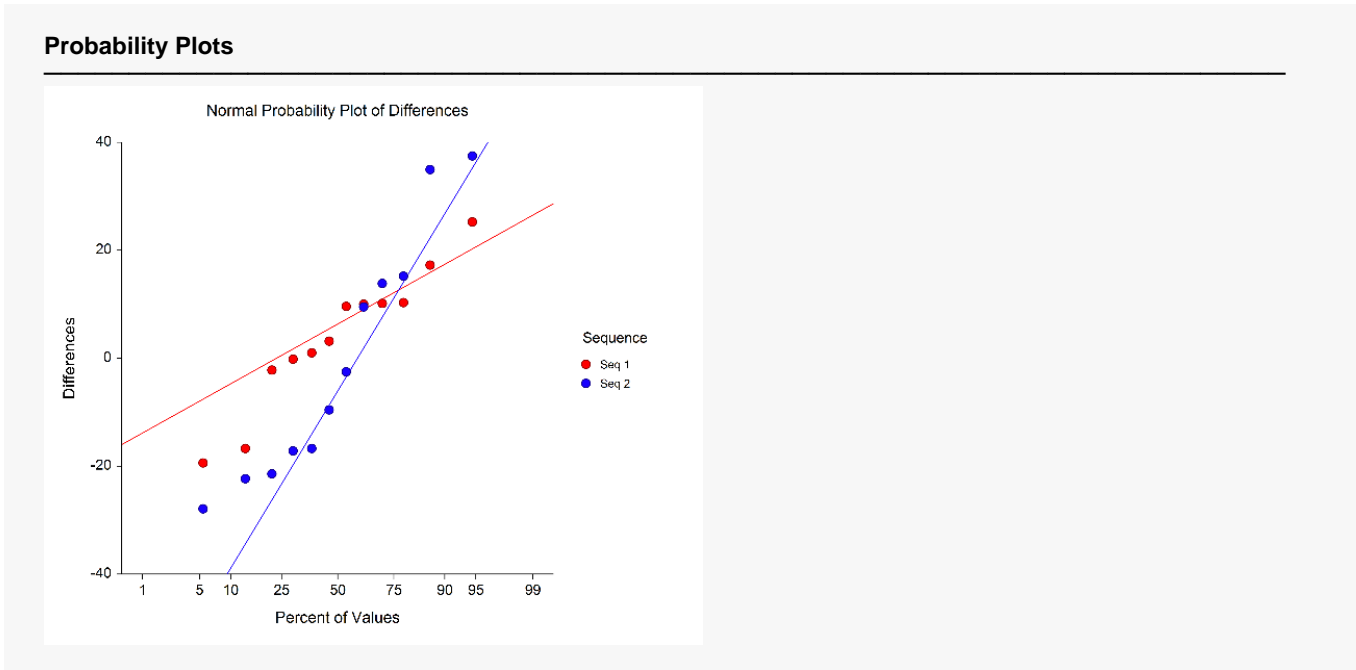


## Period Plot



The Period Plot displays a subject's period 1 response on the horizontal axis and the period 2 response on the vertical axis. The plotting symbol is the sequence number. The plot is used to find outliers and other anomalies.

## Probability Plots



These plots show the differences ( $P1-P2$ ) on the vertical axis and values on the horizontal axis that would be expected if the differences were normally distributed. The first plot shows the differences for sequence 1 and the second plot shows the differences for sequence 2.

## Analysis of 2x2 Cross-Over Designs using T-Tests for Non-Inferiority

If the assumption of normality holds, the points should fall along a straight line. The degree to which the points are off the line represents the degree to which the normality assumption does not hold. Since the normality of these differences is assumed by the  $t$ -test used to test for a difference between the treatments, these plots are useful in assessing whether that assumption is valid.

If the plots show a pronounced pattern of non-normality, you might try taking the square roots or the logs of the responses before beginning the analysis.