Chapter 527

Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Introduction

This procedure calculates power and sample size for inequality tests which use the difference in the means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to another. It is assumed that there is a *washout* period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a *carryover* effect.

A 2x2 cross-over design refers to two periods (and treatments) and two *sequences* (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present sample size results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Balaam's Design

Balaam's design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

Sequence	Period 1	Period 2
1	A	A
2	B	B
3	A	B
4	B	A

Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

Sequence	Period 1	Period 2	Period 3
1	A	B	B
2	B	A	A

Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

Sequence	Period 1	Period 2	Period 3	Period 4
1	A	B	B	A
2	B	A	A	B

Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

Sequence	Period 1	Period 2	Period 3	Period 4
1	A	A	B	B
2	B	B	A	A
1	A	B	B	A
2	B	A	A	B

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 may be easier to obtain 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. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs 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 provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

A two-sided hypothesis test comparing the means is given by

$$\text{H0: } \mu_1 - \mu_2 = 0 \text{ versus H1: } \mu_1 - \mu_2 \neq 0$$

Rejecting H0 implies that the means are different.

Test Statistic

The presence of a non-zero difference between the means can be tested using an F-test from an appropriate repeated measures model that may include terms for a sequence effect and/or a carry-over effect. The within mean square error from the ANOVA provides an estimate of the within-subject variance, σ_w^2. If prior studies are not available, you can use estimates of the between-subject variance and the within-subject correlation to calculate the within-subject variance using the formula

$$\sigma_w^2 = \sigma_b^2 (1 - \rho)$$

Power Calculation

The power of a tw0-sided test is given by Chen, Chow, and Li (1997). The basic formula is

$$\text{Power} = T_V \left[\left(\frac{|\mu_1 - \mu_2|}{\sigma_w \sqrt{b/n}} \right) - t_{V, 1-\alpha/2} \right]$$

where T represents the cumulative t distribution, V and b depend on the design (shown below), σ_w is defined above, and n is the average number of subjects per sequence. Thus the overall sample size N is given by n times the number of sequences. For a one-sided test, replace α/2 with α in the above formula.

The constants V and b depend on the design as follows:

Design Type	Parameters (V,b)
Balaam's Design	$V = 4n - 3$, $b = 2$.
Two-Sequence Dual Design	$V = 4n - 4$, $b = 3/4$.
Four-Period Design with Two Sequences	$V = 6n - 5$, $b = 11/20$.
Four-Period Design with Four Sequences	$V = 12n - 5$, $b = 1/4$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding *Sample Size*. Select *Sample Size (Equal Per Sequence)* when you want the design to have an equal number of subjects per sequence. Select *Sample Size (Exact)* when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select *Power* when you want to calculate the power of an experiment.

Test

Design Type

Specify which cross-over design that you want to analyze. These designs allow you to compare two means: a treatment (A) and a reference (B). The designs assume that each subject is measured two or more times (or periods) with ample time in between to wash-out the effect of the treatment.

A sequence is an ordering of the how A and B are applied. For example, 'ABA' means that the subjects are measured three times: first treatment A, then B, then A again. Other possible sequences are BAB, AAB, and BBA. Each design includes several sequences.

The design tries to balance the experiment so that each treatment occurs an equal number of times and after each other treatment an equal number of times.

Alternative Hypothesis

Specify whether your test will be two-sided or one-sided.

Two-Sided

Note that usually an inequality test is two-sided. The value of alpha will be divided by 2 and the result will be put evenly into both tails.

One-Sided

The power and sample size will be determined for an upper-, or a lower-, tailed test.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Total Sample Size)

This is the combined sample size of all sequences in the cross-over design. It is assumed that the subjects are divided as evenly as possible among the sequences.

You can enter a single value such as *20* or a series of values such as *10 20 30* or *10 to 100 by 10*.

Effect Size – Mean Difference

|μ1 – μ2|

This is the magnitude of the difference between the two means for which the power is calculated. The value should be in the same scale as the standard deviation, σw or σb. It must be a positive number.

You can enter a single value or a range of values.

Effect Size – Standard Deviation

Specify σ as σw or σb and ρ

This option specifies how you will enter the within standard deviation, σw. You have two choices:

σw

Specify σw directly.

σb and ρ

Specify σb and ρ separately. The value of σw is calculated using the formula: $\sigma w = \sigma b \sqrt{(1 - \rho)}$.

σw (Within Std Dev)

Specify the value(s) of the standard deviation σw. This value can be estimated using $\sigma w = \sqrt{(WMSE)}$ where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design.

If you were to compute the treatment difference for each subject, this would be the standard deviation of those differences. These values must be positive.

A list of values may be entered.

You can press the **σ button** to load the Standard Deviation Estimator window.

σb (Between Std Dev)

This is the between subject standard deviation of the response variable (Y) at a particular time point. It is assumed to be the same for all time points. This represents the variability from subject to subject that occurs when the subjects are treated identically.

As a standard deviation, the number(s) must be greater than zero.

You can enter a list of values separated by blanks or commas, in which case, a separate analysis will be calculated for each value.

You can press the σ button and select 'Covariance Matrix' to obtain help on estimating the standard deviation from an ANOVA table.

ρ (Correlation)

Specify the correlation between two observations made on the same subject. Negative values are not permitted, so it must be between 0 and 1. When no previous information is available, try values between 0.2 and 0.6.

You can press the **σ button** and select 'Covariance Matrix' to obtain help on estimating the standard deviation from an ANOVA table.

You can enter a single value or a list of values.

Example 1 – Finding Power

Researchers want to calculate the power of a two-sided, inequality test using data from a three-period, two-sequence, dual cross-over design. They want to investigate sample sizes between 6 and 66. They want a sample large enough to detect a mean difference of 14 or 16. Similar experiments have had a standard deviation (σw) of about 25. The significance level is 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Difference of Two Means in a Higher-Order Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (Higher-Order) Design**, then clicking on **Test (Inequality)**, and then clicking on **Tests for the Difference of Two Means in a Higher-Order Cross-Over Design**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Design Type...	**3x2 (Three-Period, Two-Sequence Dual: ABB|BAA)**
Alternative Hypothesis	**Two-Sided**
Alpha...	**0.05**
N (Total Sample Size)..............................	**6 to 66 by 10**
|μ1 - μ2|...	**14 16**
Specify σ as σw or σb and ρ....................	**σw (Within Std Dev)**
σw (Within Standard Dev)	**25**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for an Inequality Test of the Mean Difference in a Three-Period, Two-Sequence Dual Design
Treatment Sequences: ABB | BAA
Alternative Hypothesis: Two-Sided

Power	Total Sample Size N	Mean Difference |μ1 - μ2|	Within Standard Deviation σw	Alpha
0.1348	6	14.000	25.0000	0.050
0.1675	6	16.000	25.0000	0.050
0.4139	16	14.000	25.0000	0.050
0.5165	16	16.000	25.0000	0.050
0.6251	26	14.000	25.0000	0.050
0.7419	26	16.000	25.0000	0.050
0.7715	36	14.000	25.0000	0.050
0.8708	36	16.000	25.0000	0.050
0.8658	46	14.000	25.0000	0.050
0.9385	46	16.000	25.0000	0.050
0.9235	56	14.000	25.0000	0.050
0.9718	56	16.000	25.0000	0.050
0.9575	66	14.000	25.0000	0.050
0.9875	66	16.000	25.0000	0.050

References
Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel
   Dekker. New York
Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
Chen, K.W.; Chow, S.C.; and Li, G. 1997. 'A Note on Sample Size Determination for Bioequivalence Studies with
   Higher-Order Crossover Designs.' Journal of Pharmacokinetics and Biopharmaceutics, Volume 25, No. 6, pages
   753-765.

Report Definitions
Power is the probability of correctly rejecting H0 (concluding inequality).
N is the total number of subjects. They are divided evenly among all sequences.
$|\mu1 - \mu2|$ is the mean difference at which the power is computed. $|\mu1 - \mu2|$ = treatment mean - reference mean.
σw is the square root of the within mean square error from the ANOVA table.
Alpha is the probability of falsely rejecting H0 (falsely concluding inequality).

Summary Statements
In an inequality test on data from a three-period, two-sequence dual cross-over design, a
sample size of 6 achieves 13% power at a 5% significance level when the absolute difference
between the two treatment means is 14.000 and the within-subject standard deviation is 25.0000.

This report shows the power for the indicated scenarios.

Plots Section
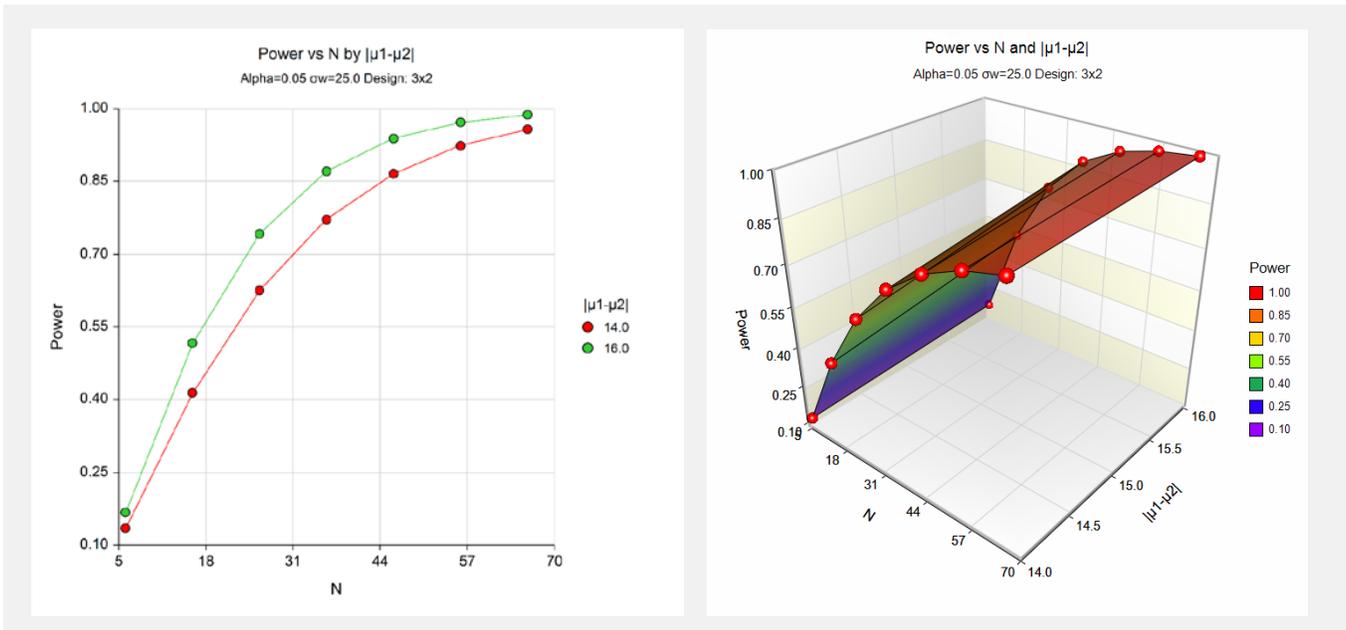
These plots show the power versus the sample size.

Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Difference of Two Means in a Higher-Order Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (Higher-Order) Design**, then clicking on **Test (Inequality)**, and then clicking on **Tests for the Difference of Two Means in a Higher-Order Cross-Over Design**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size (Exact)**
Design Type...	**3x2 (Three-Period, Two-Sequence Dual: ABB|BAA)**
Alternative Hypothesis	**Two-Sided**
Power..	**0.90**
Alpha...	**0.05**
|µ1 - µ2|..	**14 16**
Specify σ as σw or σb and ρ...................	**σw (Within Std Dev)**
σw (Within Standard Dev)	**25**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for an Inequality Test of the Mean Difference in a Three-Period, Two-Sequence Dual Design
Treatment Sequences: ABB | BAA
Alternative Hypothesis: Two-Sided

Power	Total Sample Size N	Mean Difference |µ1 - µ2|	Within Standard Deviation σw	Alpha
0.9039	52	14.000	25.0000	0.050
0.9035	40	16.000	25.0000	0.050

This report shows the sample sizes necessary when the difference is 14 or 16. Note that the N of 52 would be split with 26 devoted to each sequence.

Example 3 – Validation

We will validate this procedure by comparing its results to the *Superiority by a Margin Tests for the Difference of Two Means in a Higher-Order Cross-Over Design* routine which has already been validated. The two routines should achieve the same power when SM is set to zero in that routine.

In the superiority routine, if we set the Design Type to 4x2, Higher Means to Better, Alpha to 0.05, N to 50, SM to 0, Difference to 1.5, and Within Std Dev to 4, the power is computed to be 0.8079.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Difference of Two Means in a Higher-Order Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (Higher-Order) Design**, then clicking on **Test (Inequality)**, and then clicking on **Tests for the Difference of Two Means in a Higher-Order Cross-Over Design**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Design Type...	**4x2 (Four-Period, Two-Sequence: ABBA|BAAB)**
Alternative Hypothesis	**One-Sided**
Alpha...	**0.05**
N	**50**
$|\mu_1 - \mu_2|$..	**1.5**
Specify σ as σw or σb and ρ..................	**σw (Within Std Dev)**
σw (Within Standard Dev)	**4**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for an Inequality Test of the Mean Difference in a Four-Period, Two-Sequence Dual Design
Treatment Sequences: ABBA | BAAB
Alternative Hypothesis: One-Sided

Power	Total Sample Size N	Mean Difference $|\mu_1 - \mu_2|$	Within Standard Deviation σw	Alpha
0.8079	50	1.500	4.0000	0.050

This procedure has also calculated the power to be 0.8079.