

Chapter 575

Multiple Comparisons

Introduction

This module computes sample sizes for multiple comparison procedures. The term “*multiple comparison*” refers to the individual comparison of two means selected from a larger set of means. The module emphasizes one-way analysis of variance designs that use one of three multiple-comparison methods: Tukey’s all pairs (MCA), comparisons with the best (MCB), or Dunnett’s all versus a control (MCC). Because these sample sizes may be substantially different from those required for the usual F test, a separate module is provided to compute them.

There are only a few articles in the statistical literature on the computation of sample sizes for multiple comparison designs. This module is based almost entirely on the book by Hsu (1996). We can give only a brief outline of the subject here. Users who want more details are referred to Hsu’s book.

Although this module is capable of computing sample sizes for unbalanced designs, it emphasizes balanced designs.

Technical Details

The One-Way Analysis of Variance Design

The summarized discussion that follows is based on the common, one-way analysis of variance design. Suppose the responses Y_{ij} in k groups each follow a normal distribution with respective means, $\mu_1, \mu_2, \dots, \mu_k$, and unknown variance, σ^2 . Let n_1, n_2, \dots, n_k denote the number of subjects in each group.

The analysis of these responses is based on the sample means

$$\hat{\mu}_i = \bar{Y}_i = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i}$$

and the pooled sample variance

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{\sum_{i=1}^k (n_i - 1)}$$

The F test is the usual method for analyzing such a design, and tests whether all of the means are equal. However, a significant F test does not indicate which of the groups are different, only that at least one is different. The analyst is left with the problem of determining which group(s) is(are) different and by how much.

The probability statement associated with the F test is simple, straightforward, and easy to interpret. However, when several simultaneous comparisons are made among the group means, the interpretation of individual probability statements becomes much more complex. This is called the problem of *multiplicity*. *Multiplicity* here

Multiple Comparisons

refers to the fact that the probability of making at least one incorrect decision increases as the number of statistical tests increases. The method of *multiple comparisons* has been developed to account for such multiplicity.

Power Calculations for Multiple Comparisons

For technical reasons, the definition of power in the case of multiple comparisons is different from the usual definition. Following Hsu (1996) page 237, power is defined as follows.

Using a $1 - \alpha$ simultaneous confidence interval multiple comparison method, power is the probability that the confidence intervals cover the true parameter values and are sufficiently narrow. Power is still defined to be $1 - \beta$. Note that $1 - \beta < 1 - \alpha$. Here, *narrow* refers to the width of the confidence intervals. The definition says that the confidence intervals should be as narrow as possible while still including the true parameter values. This definition may be restated as the probability that the simultaneous confidence intervals are *correct* and *useful*.

The parameter ω represents the maximum width of any of the individual confidence intervals in the set of simultaneous confidence intervals. Thus, ω is used to specify the narrowness of the confidence intervals.

Multiple Comparisons with a Control (MCC)

A common experimental design compares one or more *treatment* groups with a *control* group. The control group may receive a placebo, the standard treatment, or even an experimental treatment. The distinguishing feature is that the mean response of each of the other groups is to be compared with this control group.

We arbitrarily assume that the last group (group k), is the control group. The $k-1$ parameters of primary interest are

$$\begin{aligned}\delta_1 &= \mu_1 - \mu_k \\ \delta_2 &= \mu_2 - \mu_k \\ &\vdots \\ \delta_i &= \mu_i - \mu_k \\ &\vdots \\ \delta_{k-1} &= \mu_{k-1} - \mu_k\end{aligned}$$

In this situation, Dunnett's method provides simultaneous, one- or two-sided confidence intervals for all of these parameters.

The one-sided confidence intervals, $\delta_1, \dots, \delta_{k-1}$, are specified as follows:

$$\Pr\left(\delta_i > \hat{\mu}_i - \hat{\mu}_k - q_{\alpha, df, \lambda_1, \dots, \lambda_{k-1}} \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \text{ for } i = 1, \dots, k-1\right) = 1 - \alpha$$

where

$$\lambda_i = \sqrt{\frac{n_i}{n_i + n_k}}$$

and q , found by numerical integration, is the solution to

$$\int_0^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \left[\Phi\left(\frac{\lambda_i z + qs}{\sqrt{1 - \lambda_i^2}}\right) \right] d\Phi(z) \gamma(s) ds = 1 - \alpha$$

where $\Phi(z)$ is the standard normal distribution function and $\gamma(z)$ is the density of $\hat{\sigma} / \sigma$.

Multiple Comparisons

The two-sided confidence intervals for $\delta_1, \dots, \delta_{k-1}$ are specified as follows:

$$\Pr\left(\delta_i \in \hat{\mu}_i - \hat{\mu}_k - |q_{\alpha, df, \lambda_1, \dots, \lambda_{k-1}}| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \text{ for } i = 1, \dots, k-1\right) = 1 - \alpha$$

where

$$\lambda_i = \sqrt{\frac{n_i}{n_i + n_k}}$$

and $|q|$, found by numerical integration, is the solution to

$$\int_0^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \left[\Phi\left(\frac{\lambda_i z + |q|s}{\sqrt{1 - \lambda_i^2}}\right) - \Phi\left(\frac{\lambda_i z - |q|s}{\sqrt{1 - \lambda_i^2}}\right) \right] d\Phi(z) \gamma(s) ds = 1 - \alpha$$

where $\Phi(z)$ is the standard normal distribution function and $\gamma(z)$ is the density of $\hat{\sigma} / \sigma$.

Interpretation of Dunnett's Simultaneous Confidence Intervals

There is a specific interpretation given for Dunnett's method. It provides a set of confidence intervals calculated so that, if the normality and equal-variance assumptions are valid, the probability that all of the $k-1$ confidence intervals enclose the true values of $\delta_1, \dots, \delta_{k-1}$ is $1 - \alpha$. The presentation below is for the two-sided case. The one-sided case is the same as for the MCB case.

Sample Size and Power – Balanced Case

Using the modified definition of power, the two-sided case is outlined as follows.

$$\begin{aligned} & \Pr\left[\left(\text{simultaneous coverage}\right) \text{ and } \left(\text{narrow}\right)\right] \\ &= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q| \hat{\sigma} \sqrt{2/n} \text{ for } i = 1, \dots, k-1\right) \text{ and } \left(|q| \hat{\sigma} \sqrt{2/n} < \omega / 2\right)\right] \\ &= k \int_0^u \int_{-\infty}^{\infty} \left[\Phi\left(z + \sqrt{2}|q|s\right) - \Phi\left(z - \sqrt{2}|q|s\right) \right]^{k-1} d\Phi(z) \gamma(s) ds \\ &\geq 1 - \beta \end{aligned}$$

where

$$u = \frac{\omega / 2}{\sigma |q| \sqrt{\frac{2}{n}}}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample size, n , the power, $1 - \beta$, the significance level, α , and the within-group standard deviation, σ .

Multiple Comparisons

Sample Size and Power – Unbalanced Case

Using the modified definition of power, the unbalanced case is outlined as follows.

$$\begin{aligned} & \Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})] \\ &= \Pr \left[\left(\mu_i - \mu_k \in \hat{\mu}_i - \hat{\mu}_k \pm |q| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \text{ for } i = 1, \dots, k-1 \right) \right. \\ & \quad \left. \text{and } \left(\min_{i < k} \left[|q| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \right] < \omega/2 \right) \right] \\ &= k \int_0^u \int_{-\infty}^{\infty} [\Phi(z) - \Phi(z - \sqrt{2} |q| s)]^{k-1} d\Phi(z) \gamma(s) ds \\ &\geq 1 - \beta \end{aligned}$$

where

$$u = \frac{\omega/2}{\min_{i < k} \left[\sigma |q| \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \right]}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample sizes, n_1, n_2, \dots, n_k , the power, $1 - \beta$, the significance level, α , and the within-group standard deviation, σ .

Multiple Comparisons with the Best (MCB)

The method of multiple comparisons with the best (champion) is used in situations in which the best group (we will assume the best is the largest, but it could just as well be the smallest) is desired. Because of sampling variation, the group with the largest sample mean may not actually be the group with the largest population mean. The following methodology has been developed to analyze data in this situation.

Perhaps the most obvious way to define the parameters in this situation is as follows

$$\max_{j=1, \dots, k} \mu_j - \mu_i, \text{ for } i = 1, \dots, k$$

Obviously, the group for which all of these values are positive will correspond to the group with the largest mean.

Another way of looking at this, which has some advantages, is to use the parameters

$$\theta_i = \mu_i - \max_{i \neq j} \mu_j, \text{ for } i = 1, \dots, k$$

since, if $\theta_i > 0$, group i is the best.

Hsu (1996) recommends using constrained MCB inference in which the intervals are constrained to include zero. Hsu recommends this because inferences about which group is best are sharper. For example, a confidence interval for θ_i whose lower limit is 0 indicates that group i is the best. Similarly, a confidence interval for θ_i whose upper limit is 0 indicates that group i is not the best.

Multiple Comparisons

Hsu (1996) shows that $100(1 - \alpha)\%$ simultaneous confidence intervals for θ_i are given by

$$- \min \left[0, \left(\hat{\theta}_i - q_{\alpha, df, \lambda_1, \dots, \lambda_{k-1}}^i \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \right) \right], \max \left[0, \left(\hat{\theta}_i + q_{\alpha, df, \lambda_1, \dots, \lambda_{k-1}}^i \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \right) \right], i = 1, \dots, k$$

where q^i is found using Dunnett's one-sided procedure discussed above assuming that group i is the control group.

Sample Size and Power – Balanced Case

Using the modified definition of power, the balanced case is outlined as follows

$$\begin{aligned} & \Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})] \\ &= \Pr \left[\begin{array}{l} \left(- \min \left(0, \hat{\mu}_i - \max_{j \neq i}(\hat{\mu}_j) - q \hat{\sigma} \sqrt{2/n} \right) \leq \right. \\ \left. \mu_i - \max_{j \neq i}(\mu_j) \leq \right. \\ \left. \max \left(0, \hat{\mu}_i - \max_{j \neq i}(\hat{\mu}_j) + q \hat{\sigma} \sqrt{2/n} \right) \text{ for } i = 1, \dots, k \right) \\ \text{and } (q \hat{\sigma} \sqrt{2/n} < \omega/2) \end{array} \right] \\ &= k \int_0^u \int_{-\infty}^{\infty} [\Phi(z + \sqrt{2}qs)]^{k-1} d\Phi(z)\gamma(s)ds \\ &\geq 1 - \beta \end{aligned}$$

where

$$u = \frac{\omega/2}{\sigma q \sqrt{\frac{2}{n}}}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample size, n , the power, $1 - \beta$, the significance level, α , and the within-group standard deviation, σ .

Multiple Comparisons

Sample Size and Power – Unbalanced Case

Using the modified definition of power, the unbalanced case is outlined as follows

$$\begin{aligned}
 & \Pr\left[\left(\text{simultaneous coverage}\right) \text{ and } \left(\text{narrow}\right)\right] \\
 &= \Pr\left[\left[\begin{array}{l} -\min\left(0, \hat{\mu}_i - \max_{j \neq i}(\hat{\mu}_j) - q^i \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right) \leq \\ \mu_i - \max_{j \neq i}(\mu_j) \leq \\ \max\left(0, \hat{\mu}_i - \max_{j \neq i}(\hat{\mu}_j) + q^i \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right) \text{ for } i = 1, \dots, k \end{array}\right] \right. \\
 & \quad \left. \text{and } \left(\min_{j \neq i} \left[q^i \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right] < \omega / 2 \right) \right] \\
 &= k \int_0^u \int_{-\infty}^{\infty} \left[\Phi(z + \sqrt{2}|q|s) \right]^{k-1} d\Phi(z) \gamma(s) ds \\
 &\geq 1 - \beta
 \end{aligned}$$

where

$$u = \max_{i \neq j} \left(\frac{\omega / 2}{\sigma |q^i| \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \right)$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample sizes, n_1, n_2, \dots, n_k , the power, $1 - \beta$, the significance level, α , and the within-group standard deviation, σ .

All-Pairwise Comparisons (MCA)

In this case you are interested in all possible pairwise comparisons of the group means. There are $k(k-1)/2$ such comparisons. A popular method in this case is that developed by Tukey.

Balanced Case

The Tukey method provides simultaneous, two-sided confidence intervals. They are specified as follows

$$\Pr\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q^*| \hat{\sigma} \sqrt{\frac{2}{n}} \text{ for } i \neq j\right) = 1 - \alpha$$

When all the sample sizes are equal, q^* may be found by numerical integration as the solution to the equation

$$\int_0^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \left[\Phi(z) - \Phi(z - \sqrt{2}|q^*|s) \right] d\Phi(z) \gamma(s) ds = 1 - \alpha$$

Multiple Comparisons

where $\Phi(z)$ is the standard normal distribution function and $\gamma(z)$ is the density of $\hat{\sigma} / \sigma$. Note that $q' = \sqrt{2}|q^*|$ is the critical value of the Studentized range distribution.

Sample Size and Power

Using the modified definition of power, the balanced case is outlined as follows

$$\begin{aligned} & \Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})] \\ &= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q^*| \hat{\sigma} \sqrt{2/n} \text{ for all } i \neq j\right) \text{ and } \left(|q^*| \hat{\sigma} \sqrt{2/n} < \omega/2\right)\right] \\ &= k \int_0^u \int_{-\infty}^{\infty} \left[\Phi(z) - \Phi\left(z - \sqrt{2}|q^*|s\right)\right]^{k-1} d\Phi(z)\gamma(s)ds \\ &\geq 1 - \beta \end{aligned}$$

where

$$u = \frac{\omega/2}{\sigma |q^*| \sqrt{\frac{2}{n}}}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample size, n , the power, $1 - \beta$, the significance level, α , and within-group standard deviation, σ .

Unbalanced Case

The simultaneous, two-sided confidence intervals are specified as follows

$$\Pr\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_k \pm |q^e| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \text{ for } i \neq j\right) = 1 - \alpha$$

Unfortunately, $|q^e|$ cannot be calculated as a double integral as in previous cases. Instead, the Tukey-Kramer approximate solution is used. Their proposal is to use $|q^*|$ in place of $|q^e|$.

Sample Size and Power

Using the modified definition of power, the unbalanced case is outlined as follows

$$\begin{aligned} & \Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})] \\ &= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q^*| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \text{ for all } i \neq j\right) \right. \\ & \quad \left. \text{and } \left(\min_{i \neq j} \left[|q^*| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right]\right) < \omega/2\right] \\ &= k \int_0^u \int_{-\infty}^{\infty} \left[\Phi(z) - \Phi\left(z - \sqrt{2}|q^*|s\right)\right]^{k-1} d\Phi(z)\gamma(s)ds \\ &\geq 1 - \beta \end{aligned}$$

Multiple Comparisons

where

$$u = \frac{\omega / 2}{\min_{i \neq j} \left[\sigma |q^*| \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right]}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample sizes, n_1, n_2, \dots, n_k , the power, $1 - \beta$, the significance level, α , and the within-group standard deviation, σ .

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 most of the parameters and options that you will be concerned with.

Solve For

Solve For

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

Test

Type of Multiple Comparison

Specify the type of multiple comparison test to be analyzed. The tests available are

- **All Pairs - Tukey Kramer**
All possible paired comparisons.
- **With Best - Hsu**
Constrained comparisons with the best.
- **With Control - Dunnett**
Two-sided versus a control group.

Multiple Comparisons

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. In the case of multiple comparisons, power is the probability that the confidence intervals will cover the true parameter values and be sufficiently narrow to be useful. This is a modified definition of power; since beta equals 1-power, it has a modified definition as well.

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. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

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 / Groups – Sample Size Multiplier

n (Sample Size Multiplier)

This is the base, per-group sample size. One or more values separated by blanks or commas may be entered. A separate analysis is performed for each value listed here.

The group sample sizes are determined by multiplying this number by each of the Group Sample Size Pattern numbers. If the Group Sample Size Pattern numbers are represented by $m_1, m_2, m_3, \dots, m_k$ and this value is represented by n , the group sample sizes $N_1, N_2, N_3, \dots, N_k$ are calculated as follows:

$$N_1 = [n(m_1)]$$

$$N_2 = [n(m_2)]$$

$$N_3 = [n(m_3)]$$

etc.

where the operator, $[X]$ means the next integer after X , e.g. $[3.1] = 4$.

For example, suppose there are three groups and the Group Sample Size Pattern is set to $1, 2, 3$. If n is 5, the resulting sample sizes will be 5, 10, and 15. If n is 50, the resulting group sample sizes will be 50, 100, and 150. If n is set to $2, 4, 6, 8, 10$, five sets of group sample sizes will be generated and an analysis run for each. These sets are:

2	4	6
4	8	12
6	12	18
8	16	24
10	20	30

Multiple Comparisons

As a second example, suppose there are three groups and the Group Sample Size Pattern is $0.2, 0.3, 0.5$. When the fractional Pattern values sum to one, n can be interpreted as the total sample size of all groups and the Pattern values as the proportion of the total in each group.

If n is 10, the three group sample sizes would be 2, 3, and 5.

If n is 20, the three group sample sizes would be 4, 6, and 10.

If n is 12, the three group sample sizes would be

$(0.2)12 = 2.4$ which is rounded up to the next whole integer, 3.

$(0.3)12 = 3.6$ which is rounded up to the next whole integer, 4.

$(0.5)12 = 6$.

Note that in this case, $3+4+6$ does not equal n (which is 12). This can happen because of rounding.

Sample Size / Groups – Groups

k (Number of Groups)

This is the number of group means being compared. It must be greater than or equal to three.

You can enter a list of values, in which case a separate analysis will be calculated for each value. Commas or blanks may separate the numbers. A TO-BY list may also be used.

Note that the number of items used in the Hypothesized Means box and the Group Sample Size Pattern box is controlled by this number.

Examples:

3,4,5

4 5 6

3 to 11 by 2

Group Sample Size Pattern

A set of positive, numeric values (one for each group) is entered here. The sample size of group i is found by multiplying the i^{th} number from this list times the value of n and rounding up to the next whole number. The number of values must match the number of groups, k . When too few numbers are entered, 1's are added. When too many numbers are entered, the extras are ignored.

Note that in the case of Dunnett's test, the last group corresponds to the control group. Thus, if you want to study the implications of having a group with a different sample size for the control group, you should specify that sample size in the last position.

- **Equal**

If all sample sizes are to be equal, enter "Equal" here and the desired sample size in n . A set of k 1's will be used. This will result in $N1 = N2 = N3 = n$. That is, all sample sizes are equal to n .

Multiple Comparisons

Effect Size

Minimum Detectable Difference

Specify one or more values of the minimum detectable difference. This is the smallest difference between any two group means that is to be detectable by the experiment. Note that this is a positive amount. This value is set by the researcher to represent the smallest difference between two means that will be of practical significance to other researchers. Note that in the case of Dunnett's test, this is the maximum difference between each treatment and the control.

Examples:

3

3 4 5

10 to 20 by 2

S (Standard Deviation of Subjects)

This option refers to σ , the standard deviation within a group. It represents the variability from subject to subject that occurs when the subjects are treated identically. It is assumed to be the same for all groups. This value is approximated in an analysis of variance table by the square root of the mean square error.

Since they are positive square roots, the numbers must be strictly greater than zero. You can press the *SD* button to obtain further help on estimating the standard deviation.

Note that if you are using this procedure to test a factor (such as an interaction) from a more complex design, the value of standard deviation is estimated by the square root of the mean square of the term that is used as the denominator in the *F* test.

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

Examples of valid entries:

1,4,7,10

1 4 7 10

1 to 10 by 3

Options Tab

The Options tab contains a search precision option.

Precision

Search Precision

Specify the precision to be used in the routines that search for the value of the minimum detectable difference. Note that the closer this value is to zero, the longer a search will take.

Multiple Comparisons

Example 1 – Calculating Power

An experiment is being designed to compare the means of four groups using the Tukey-Kramer pairwise multiple comparison test with a significance level of 0.05. Previous studies indicate that the standard deviation is 5.3. The typical mean response level is 63.4. The researcher believes that a 25% increase in the mean will be of interest to others. Since $0.25(63.4) = 15.85$, this is the number that will be used as the minimum detectable difference.

To better understand the relationship between power and sample size, the researcher wants to compute the power for several group sample sizes between 2 and 14. The sample sizes will be equal across all groups.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Comparisons** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Comparisons**. 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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Type of Multiple Comparison	All Pairs - Tukey Kramer
Alpha	0.05
n (Sample Size Multiplier)	2 to 14 by 2
k (Number of Groups)	4
Group Sample Size Pattern	Equal
Minimum Detectable Difference	15.85
S (Standard Deviation of Subjects)	5.3

Annotated Output

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

Numeric Results Report

Numeric Results for Multiple Comparison Test: Tukey-Kramer (Pairwise)

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.0113	2.00	4	8	0.0500	0.9887	15.85	5.30	2.9906
0.0666	4.00	4	16	0.0500	0.9334	15.85	5.30	2.9906
0.3171	6.00	4	24	0.0500	0.6829	15.85	5.30	2.9906
0.7371	8.00	4	32	0.0500	0.2629	15.85	5.30	2.9906
0.9301	10.00	4	40	0.0500	0.0699	15.85	5.30	2.9906
0.9497	12.00	4	48	0.0500	0.0503	15.85	5.30	2.9906
0.9500	14.00	4	56	0.0500	0.0500	15.85	5.30	2.9906

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.

n is the average group sample size.

k is the number of groups.

Total N is the total sample size of all groups.

Alpha is the probability of rejecting a true null hypothesis. It should be small.

Beta is the probability of accepting a false null hypothesis. It should be small.

The Minimum Detectable Difference between any two group means.

S is the within group standard deviation.

The Diff / S is the ratio of Min. Detect. Diff. to standard deviation.

Multiple Comparisons

Summary Statements

In a single factor ANOVA study, sample sizes of 2, 2, 2, and 2 are obtained from the 4 groups whose means are to be compared. The total sample of 8 subjects achieves 1% power to detect a difference of at least 15.85 using the Tukey-Kramer (Pairwise) multiple comparison test at a 0.0500 significance level. The common standard deviation within a group is assumed to be 5.30.

This report shows the numeric results of this power study. Following are the definitions of the columns of the report.

Power

This is the probability that the confidence intervals will cover the true parameter values and be sufficiently narrow to be useful.

Average n

The average of the group sample sizes.

k

The number of groups.

Total N

The total sample size of the study.

Alpha

The probability of rejecting a true null hypothesis. This is often called the significance level.

Beta

Equal to 1 - power. Note the definition of power above.

Minimum Detectable Difference

This is the value of the minimum detectable difference. This is the minimum difference between two means that is thought to be of practical importance. Note that in the case of Dunnett's test, this is the minimum difference between a treatment mean and the control mean that is of practical importance.

Standard Deviation (S)

This is the within-group standard deviation. It was set in the Data window.

Diff / S

This is an index of relative difference between the means standardized by dividing by the standard deviation. This value can be used to make comparisons among studies.

Detailed Results Report

Tukey-Kramer Test Details						
Group	n	Percent n of Total N	Alpha	Power	Minimum Detectable Difference	Standard Deviation
1	2	25.00	0.0500	0.0113	15.85	5.30
2	2	25.00				
3	2	25.00				
4	2	25.00				
Total	8	100.00				

This report shows the details of each row of the previous report.

Group

The group identification number is shown on each line. The second to the last line represents the last group. When Dunnett's test has been selected, this line represents the control group.

Multiple Comparisons

The last line, labeled *Total*, gives the total for all the groups.

n

This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

Percent n of Total N

This is the percentage of the total sample that is allocated to each group.

Alpha

The probability of rejecting a true null hypothesis. This is often called the significance level.

Power

This is the probability that the confidence intervals will cover the true parameter values and be sufficiently narrow to be useful.

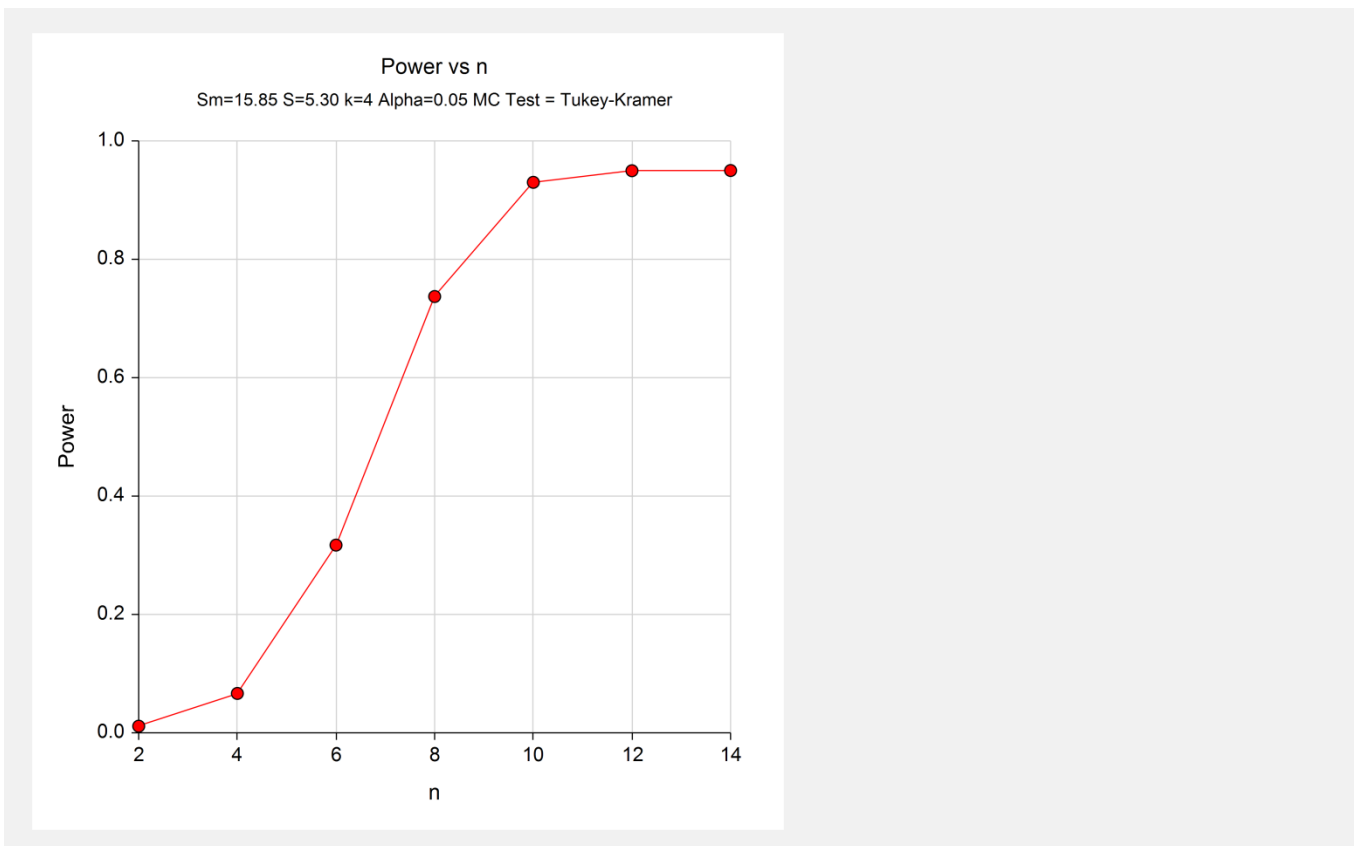
Minimum Detectable Difference

This is the value of the minimum detectable difference. This is the minimum difference between two means that is thought to be of practical importance. Note that in the case of Dunnett's test, this is the minimum difference between a treatment mean and the control mean that is of practical importance.

Standard Deviation (S)

This is the within-group standard deviation. It was set in the Data window.

Plots Section



This plot gives a visual presentation to the results in the Numeric Report. We can see the impact on the power of increasing the sample size.

When you create one of these plots, it is important to use trial and error to find an appropriate range for the horizontal variable so that you have results with both low and high power.

Multiple Comparisons

Example 2 – Power after Dunnett’s Test

This example covers the situation in which you are calculating the power of Dunnett’s test on data that have already been collected and analyzed.

An experiment included a control group and two treatment groups. Each group had seven individuals. A single response was measured for each individual and recorded in the following table.

Control	T1	T2
554	774	786
447	465	536
356	759	653
452	646	685
674	547	658
654	665	669

When analyzed using the one-way analysis of variance procedure in NCSS, the following results were obtained.

Analysis of Variance Table					
Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
A (...)	2	75629.8	37814.9	3.28	0.061
S(A)	18	207743.4	11541.3		
Total (Adjusted)	20	283373.3			
Total	21				

Dunnett's Simultaneous Confidence Intervals for Treatment vs. Control						
Response: Control,T1,T2						
Term A:						
Control Group: Control						
Alpha=0.050 Error Term=S(A) DF=18 MSE=11541.3 Critical Value=2.3987						
Treatment Group	Count	Mean	Lower 95.0% Simult.C.I.	Difference With Control	Upper 95.0% Simult.C.I.	Test Result
T1	7	660.43	-5.17	132.57	270.31	
T2	7	649.14	-16.46	121.29	259.03	

The significance level (Prob Level) was only 0.061—not enough for statistical significance. Since the lower confidence limits are negative and the upper confidence limits are positive, Dunnett’s two-sided test did not find a significant difference between either treatment and the control group.

The researcher had hoped to show that the treatment groups had higher response levels than the control group. He could see that the group means followed this pattern since the mean for *T1* was about 25% higher than the control mean and the mean for *T2* was about 23% higher than the control mean. He decided to calculate the power of the experiment.

Setup

The data entry for this problem is simple. The only entry that is not straight forward is finding an appropriate value for the standard deviation. Since the standard deviation is estimated by the square root of the mean square error, it is calculated as $\sqrt{11541.3} = 107.4304$.

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Comparisons** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Comparisons**. 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**.

Multiple Comparisons

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Type of Multiple Comparison	With Control - Dunnett
Alpha	0.05
n (Sample Size Multiplier)	7
k (Number of Groups)	3
Group Sample Size Pattern	Equal
Minimum Detectable Difference	133
S (Standard Deviation of Subjects)	107.4304

Output

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

Numeric Results

	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
Power	7.00	3	21	0.0500	0.9998	133.00	107.43	1.2380

The power is only 0.0002. Hence, there was little chance of detecting a difference of 133 between a treatment and a control group.

It was of interest to the researcher to determine how large of a sample was needed if the power was to be 0.90. Setting Beta equal to 0.90 and 'Find/Solve For' to n resulted in the following report.

	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
Power	33.00	3	99	0.0500	0.0958	133.00	107.43	1.2380

We see that instead of 7 per group, 33 per group were needed.

It was also of interest to the research to determine how large of a difference between the means could have been detected. Setting 'Find' to Min Detectable Difference resulted in the following report.

	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
Power	7.00	3	21	0.0500	0.1000	348.81	107.43	3.2468

We see that a study of this size with these parameters could only detect a difference of 348.8. This explains why the results were not significant.

Multiple Comparisons

Example 3 – Using Unequal Sample Sizes

It is usually advisable to design experiments with equal sample sizes in each group. In some cases, however, it may be necessary to allocate subjects unequally across the groups. This may occur when the group variances are unequal, the costs per subject are different, or the dropout rates are different. This module can be used to study the power of unbalanced experiments.

In this example which will use Dunnett's test, the minimum detectable difference is 2.0, the standard deviation is 1.0, alpha is 0.05, and k is 3. The sample sizes are 7, 7, and 14.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Comparisons** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Comparisons**. You may then make the appropriate entries as listed below, or open **Example 3a** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Type of Multiple Comparison	With Control - Dunnett
Alpha	0.05
n (Sample Size Multiplier)	1
k (Number of Groups)	3
Group Sample Size Pattern	7 7 14
Minimum Detectable Difference	2
S (Standard Deviation of Subjects)	1

Output

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

Numeric Results

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.2726	9.33	3	28	0.0500	0.7274	2.00	1.00	2.0000

Alternatively, this problem could have been set up as follows (**Example3b** template):

<u>Option</u>	<u>Value</u>
Data Tab	
n (Sample Size Multiplier)	7
Group Sample Size Pattern	1 1 2

The advantage of this method is that you can try several values of n while keeping the same allocation ratios.

Multiple Comparisons

Example 4 – Validation using Hsu

Hsu (1996) page 241 presents an example of determining the sample size in an experiment with 8 groups. The minimum detectable difference is 10,000 psi. The standard deviation is 3,000 psi. Alpha is 0.05 and beta is 0.10. He finds a sample size of 10 per group for the Tukey-Kramer test, a sample size of 6 for Hsu's test, and a sample size of 8 for Dunnett's test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Comparisons** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Comparisons**. You may then make the appropriate entries as listed below, or open **Example4**, **Example4b**, or **Example4c** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Type of Multiple Comparison	All Pairs - Tukey Kramer
Power	0.90
Alpha	0.05
k (Number of Groups)	8
Group Sample Size Pattern	Equal
Minimum Detectable Difference	10000
S (Standard Deviation of Subjects)	3000

Output

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

Numeric Results for Tukey-Kramer Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9397	10.00	8	80	0.0500	0.0603	10000.00	3000.00	3.3333

PASS also found $n = 10$.

Numeric Results for Hsu's Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9087	6.00	8	48	0.0500	0.0913	10000.00	3000.00	3.3333

PASS also found $n = 6$.

Numeric Results for Dunnett's Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9434	8.00	8	64	0.0500	0.0566	10000.00	3000.00	3.3333

PASS found $n = 8$.

Multiple Comparisons

Example 5 – Validation using Pan and Kupper

Pan and Kupper (1999, page 1481) present examples of determining the sample size using alternative methods. It is interesting to compare the method of Hsu (1996) with theirs. Although the results are not exactly the same, they are very close.

In the example of Pan and Kupper, the minimum detectable difference is 0.50. The standard deviation is 0.50. Alpha is 0.05, and beta is 0.10. They find a sample size of 51 per group for Dunnett's test and a sample size of 60 for the Tukey-Kramer test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Comparisons** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Comparisons**. You may then make the appropriate entries as listed below, or open **Example5a** or **Example 5b** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Type of Multiple Comparison	With Control - Dunnett
Power	0.90
Alpha	0.05
k (Number of Groups)	4
Group Sample Size Pattern	Equal
Minimum Detectable Difference	0.50
S (Standard Deviation of Subjects)	0.50

Output

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

Numeric Results for Dunnett's Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9146	53.00	4	212	0.0500	0.0854	0.50	0.50	1.0000

PASS found $n = 53$. This is very close to the 51 that Pan and Kupper found using a slightly different method.

Numeric Results for Tukey-Kramer Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9057	62.00	4	248	0.0500	0.0943	0.50	0.50	1.0000

PASS also found $n = 62$. This is very close to the 60 that Pan and Kupper found using a slightly different method.