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, $\mu_1, \mu_2, \cdots, \mu_k$, and unknown variance, σ^2. Let $n_1, n_2, \cdots, n_k$ 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

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

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 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, $\delta_1,\cdots,\delta_{k-1}$, are specified as follows:

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

where

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

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

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

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

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

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

where

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

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

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

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 $\delta_1, \cdots, \delta_{k-1}$ 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.

$$\Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})]$$

$$= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q|\hat{\sigma}\sqrt{2/n} \text{ for } i = 1, \cdots, k-1\right) \text{and} \left(|q|\hat{\sigma}\sqrt{2/n} < \omega \,/\, 2\right)\right]$$

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

$$\geq 1 - \beta$$

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.

Sample Size and Power – Unbalanced Case

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

$$\Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})]$$

$$= \Pr\left[\left(\mu_i - \mu_k \in \hat{\mu}_i - \hat{\mu}_k \pm |q|\hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \text{ for } i = 1, \ldots, k-1\right)\right.$$

$$\left.\text{and}\left(\min_{i<k}\left[|q|\hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_k}}\right]\right) < \omega / 2\right]$$

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

$$\geq 1 - \beta$$

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, $n_1, n_2, \cdots, n_k$, 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

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

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

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

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.

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

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

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

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

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.

Sample Size and Power – Unbalanced Case

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

$$\Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})]$$

$$= \Pr\left[\begin{array}{l}\left(\begin{array}{l}-\min\left(0, \hat{\mu}_i - \max_{j \neq i}(\hat{\mu}_j) - q^i \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right) \leq \\ \mu_i - \max_{j \neq i}(\mu_j) \leq \\ \max\left(0, \hat{\mu}_i - \max_{j \neq i}(\hat{\mu}_j) + q^i \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right) \text{ for } i = 1, \cdots, k\end{array}\right) \\ \text{and}\left(\min_{j \neq i}\left[q^i \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right] < \omega / 2\right)\end{array}\right]$$

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

$$\geq 1 - \beta$$

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, $n_1, n_2, \cdots, n_k$, 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

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

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

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

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

$$\Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})]$$

$$= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q^*|\hat{\sigma}\sqrt{2/n} \text{ for all } i \neq j\right) \text{and} \left(|q^*|\hat{\sigma}\sqrt{2/n} < \omega / 2\right)\right]$$

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

$$\geq 1 - \beta$$

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

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

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

$$\Pr[(\text{simultaneous coverage}) \text{ and } (\text{narrow})]$$

$$= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q^*|\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} \text{ for all } i \neq j\right)\right.$$

$$\left.\text{and}\left(\min_{i\neq j}\left[|q^*|\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right]\right) < \omega\ /\ 2\right]$$

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

$$\geq 1-\beta$$

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, $n_1, n_2, \cdots, n_k$, the power, $1-\beta$, the significance level, α, and the within-group standard deviation, σ.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters 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.

Design Tab

Solve For .... **Power**
Type of Multiple Comparison .... **All Pairs - Tukey Kramer**
Alpha... **0.05**
N (Sample Size)... **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**

Output

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

Numeric Results

Numeric Results

Solve For: Power
Multiple Comparison Type: Tukey-Kramer (Pairwise)

	Sample Size						
Power	Total N	Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.0113	8	2	4	15.85	5.3	2.9906	0.05
0.0666	16	4	4	15.85	5.3	2.9906	0.05
0.3171	24	6	4	15.85	5.3	2.9906	0.05
0.7371	32	8	4	15.85	5.3	2.9906	0.05
0.9301	40	10	4	15.85	5.3	2.9906	0.05
0.9497	48	12	4	15.85	5.3	2.9906	0.05
0.9500	56	14	4	15.85	5.3	2.9906	0.05

Power Using a 1 - α simultaneous confidence-interval multiple-comparison method, power is the probability that the confidence intervals cover the true group-mean differences and are sufficiently narrow.
N The total sample size of all groups combined.
n The average group sample size. n = N / k.
k The number of groups in the design.
D The Minimum Detectable Difference. This is the smallest difference to be detected between any two group means.
S The within-group standard deviation.
D / S The ratio of Minimum Detectable Difference to the standard deviation.
Alpha The overall (family-wise) Type-I error rate.

Summary Statements

A one-way design with 4 groups will be used to test each group mean against every other group mean (pair-wise). The comparisons will be made using Tukey-Kramer pairwise tests with an overall (family-wise) Type I error rate (α) of 0.05. The common standard deviation within each group is assumed to be 5.3. To detect a minimum difference of 15.85 with group sample sizes of 2, 2, 2, and 2 (totaling 8 subjects), the power is 0.0113.

.
.
.

References

Hsu, Jason. 1996. Multiple Comparisons: Theory and Methods. Chapman & Hall. London.

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.

Total Sample Size (N)

The total sample size of the study.

Average Group Sample Size (n)

The average of the group sample sizes.

Number of Groups (k)

The number of groups in the one-way design.

Minimum Detectable Difference (D)

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.

Effect Size (D / 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.

Alpha

The overall (family-wise) Type-I error rate.

Detailed Results Report

Tukey-Kramer Test Details

Group	n	Percent n of Total N	Alpha	Power	Minimum Detectable Difference	Standard Deviation
1	2	25	0.05	0.0113	15.85	5.3
2	2	25				
3	2	25				
4	2	25				
Total	8	100				

(More Reports Follow)

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.

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 overall (family-wise) Type-I error rate.

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.

Dropout-Inflated Sample Size Report

Dropout-Inflated Sample Size

Average Group Sample Size n	Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
2	1 - 4	20%	2	3	1
	Total		8	12	4
4	1 - 4	20%	4	5	1
	Total		16	20	4
6	1 - 4	20%	6	8	2
	Total		24	32	8
8	1 - 4	20%	8	10	2
	Total		32	40	8
10	1 - 4	20%	10	13	3
	Total		40	52	12
12	1 - 4	20%	12	15	3
	Total		48	60	12
14	1 - 4	20%	14	18	4
	Total		56	72	16

n	The average group sample size.
Group	Lists the group numbers.
Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
Ni	The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
Ni'	The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula Ni' = Ni / (1 - DR), with Ni' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
Di	The expected number of dropouts in each group. Di = Ni' - Ni.

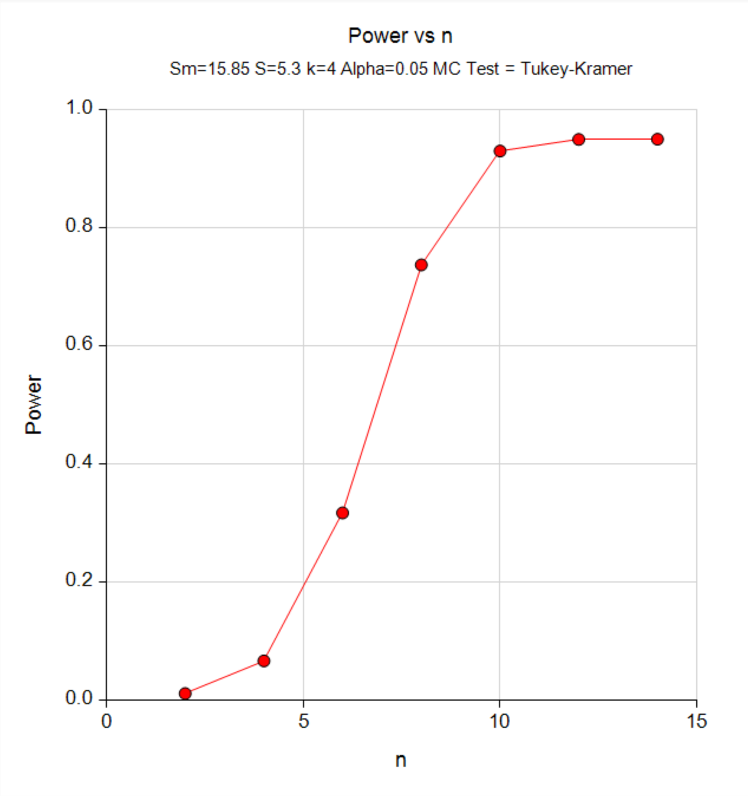
Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 3, 3, 3, and 3 subjects should be enrolled to obtain final group sample sizes of 2, 2, 2, and 2 subjects.

This report shows the sample sizes adjusted for dropout. In this example, dropout is assumed to be 20%. You can change the dropout rate on the Reports tab.

Plots Section

Plots



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.

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 six 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 and F-Test

Model Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Reject Equal Means? (α = 0.05)	Power (α = 0.05)
Between	2	69812.34	34906.17	2.8529	0.08912	No	0.47479
Within (Error)	15	183527.7	12235.18				
Adjusted Total	17	253340					
Total	18						

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=15 MSE=12235.18 Critical Value=2.4393

Treatment Group	Count	Mean	Lower 95.0% Simult.C.I.	Difference With Control	Upper 95.0% Simult.C.I.	Test Result
T1	6	642.6667	-35.94758	119.8333	275.6143	
T2	6	664.5	-14.11425	141.6667	297.4476	

The significance level (Prob Level) was 0.08912—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 or 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 23% higher than the control mean, and the mean for *T2* was about 27% higher than the control mean. He decided to calculate the power of the experiment.

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 $\sqrt{12235.18} = 110.61275$.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab
Solve For .... **Power**
Type of Multiple Comparison .... **With Control - Dunnett**
Alpha.. **0.05**
N (Sample Size).. **6**
k (Number of Groups) **3**
Group Sample Size Pattern .... **Equal**
Minimum Detectable Difference.................... **142**
S (Standard Deviation of Subjects)............... **110.61275**

Output

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

Numeric Results

Solve For: Power
Multiple Comparison Type: Dunnett (With Control)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.0003	18	6	3	142	110.61	1.2838	0.05

The power is only 0.0003. Hence, there was little chance of detecting a difference of 142 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 Power equal to 0.90 and Solve for to Sample Size results in the following report:

Numeric Results

Solve For: Sample Size
Multiple Comparison Type: Dunnett (With Control)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.905	93	31	3	142	110.61	1.2838	0.05

We see that instead of 6 per group, 31 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 Solve For to *Min Detectable Difference* results in the following report:

Numeric Results

Solve For: Min Detectable Difference
Multiple Comparison Type: Dunnett (With Control)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.9	18	6	3	402.13	110.61	3.6355	0.05

We see that a study of this size with these parameters could only detect a difference of 402.13. This explains why the results were not significant.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3a** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Type of Multiple Comparison .... **With Control - Dunnett**
Alpha.. **0.05**
N (Sample Size).. **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

Solve For: Power
Multiple Comparison Type: Dunnett (With Control)

	Sample Size						
Power	**Total N**	**Group n**	**Number of Groups k**	**Minimum Detectable Difference D**	**Standard Deviation of Subjects S**	**Effect Size D / S**	**Alpha**
0.2726	28	9.33	3	2	1	2	0.05

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

Design Tab

N (Sample Size)..**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.

Example 4 – Validation using Hsu (1996)

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4 (a, b, or c)** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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

Numeric Results

Solve For: Sample Size
Multiple Comparison Type: Tukey-Kramer (Pairwise)

	Sample Size						
Power	Total N	Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.9397	80	10	8	10000	3000	3.3333	0.05

PASS also found *n* = 10.

Numeric Results for Hsu's Test

Numeric Results

Solve For: Sample Size
Multiple Comparison Type: Hsu (With Best)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.9087	48	6	8	10000	3000	3.3333	0.05

PASS also found *n* = 6.

Numeric Results for Dunnett's Test

Numeric Results

Solve For: Sample Size
Multiple Comparison Type: Dunnett (With Control)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.9434	64	8	8	10000	3000	3.3333	0.05

PASS also found *n* = 8.

Example 5 – Validation using Pan and Kupper (1999)

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 5 (a or b)** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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

Numeric Results

Solve For: Sample Size
Multiple Comparison Type: Dunnett (With Control)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.9146	212	53	4	0.5	0.5	1	0.05

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

Numeric Results

Solve For: Sample Size
Multiple Comparison Type: Tukey-Kramer (Pairwise)

Power	Sample Size Total N	Sample Size Group n	Number of Groups k	Minimum Detectable Difference D	Standard Deviation of Subjects S	Effect Size D / S	Alpha
0.9057	248	62	4	0.5	0.5	1	0.05

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