## Sample Size For Ratios of Means (New Rule 2.13)

### Introduction

Requests for sample size calculations come in an almost limitless variety of formulations. The basic considerations are still present: Type I and II error, treatment effect, and variability. The current formulation came about because the client asked about the power to detect a $k$-fold change in means of treatment groups compared with a control group. This formulation is very similar to the formulation for Rule 2.3 in *Statistical Rules of Thumb* but with enough of a twist to qualify for a new rule. The rule is formulated in terms of coefficient of variation and a required $k$-fold ratio of the means. This was discussed briefly in the ROM for January 2003. This month's discussion extends that discussion to include multiple comparisons and power considerations.

### Rule of Thumb

Given a coefficient of variation, CV, the sample size needed to detect a $k$-fold ratio of means, ($k = \frac{\mu_1}{\mu_2}$), is

$$n = \frac{16(CV)^2}{(\ln(k))^2}.$$

This formulation is for a two-sample comparison with $n$ units per group, a power of 0.80, a Type I error of 0.05, and a two-sided alternative.

### Example

The researcher is interested in a two-fold difference in means with a coefficient of variation about 50%. The required sample size per group is

$$n = \frac{16(0.50)^2}{(\ln(2))^2} = \frac{4}{0.48045} = 8.32 \approx 9$$

experimental units per group.

### Basis of the Rule

As has been shown, for example van Belle (2002), a coefficient of variation in the arithmetic scale is approximately equal to the standard deviation in the logarithmic scale. Since the question is in terms of the ratio of the means this translates to a difference in the means in the logarithmic scale. So equation 2.3 in *Statistical Rules of Thumb* can be applied immediately.

## Discussion and extensions

A question that comes up immediately is whether the lognormal model is the appropriate model for the question. It would appear to be reasonable on two grounds. First, the hypothesis formulation is in terms of ratios of means. This is not an uncommon formulation as has been noted in a previous discussion (ROM January 2003). Second, the data are bounded on the left by zero and in these situations the standard deviation is often proportional to the mean and a logarithmic transformation stabilizes the variance and tends to induce a more symmetric distribution.

It is not uncommon to have the question of sample size formulated in terms of detectable effects given a specified sample size. This can be the case when, for example, it is known how much work can be practically accomplished. This leads to the following elegant formulation. The detectable ratio of means, *k*, in two groups is

$$k = \frac{\mu_1}{\mu_2} = e^{\frac{4}{\sqrt{n}}CV}$$

where,

$k$= $k$-fold ratio of means, $\mu_1$ and $\mu_2$, to be detected,
$n$ = sample size per group,
4 = the multiplier necessary for a Type I error of 0.05, power of 0.80, two-sided alternative, and two-sample test,
$CV$= coefficient of variation.

This formulation can be generalized in three ways. First, the sample size per group, *n*, can actually be considered the harmonic mean of the two sample sizes. In an unequal sampling situation it's the harmonic mean of the two groups that is the effective sample size,

$$\frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} = n$$

where *n* is now the harmonic mean. For example,

$$\frac{1}{20} + \frac{1}{30} = \frac{1}{24} + \frac{1}{24}.$$

Second, in many situations, a multiple comparison is implied because the analyses are exploratory. The Type I error can be adjusted by the usual Bonferroni inequality to maintain the per-experiment error rate. Third, the power can be adjusted as desired. All these can be combined by specifying the appropriate quantities in the following formula

$$k = \frac{\mu_1}{\mu_2} = e^{\frac{\sqrt{2}(z_{1-\alpha/2m} + z_{1-\beta})}{\sqrt{n}} CV}$$

where *m* is the number of comparisons to be made.

To illustrate, the question was asked about the possibility of detecting a 1.5-or 2-fold difference in means with sample sizes of 30 in the control group and 20 in the experimental group. The investigator wanted to look at twenty different substances. Table 1 gives the detectable *k*-fold differences in these situations.

**Table 1**. Detectable *k*-fold ratio of means based on sample sizes of 20 in experimental and 30 in control group (harmonic mean=24) for specified coefficients of variation. The Bonferroni adjusted detectable ratio assumes 20 comparisons (adjusted Type I error of 0.05/20= 0.0025).

| | Detectable *k*-fold ratio | |
|---|---|---|
| Coefficient of Variation | Simple | Bonferroni Adjusted |
| 0.25 | 1.23 | 1.32 |
| 0.50 | 1.50 | 1.75 |
| 0.75 | 1.84 | 2.31 |

The table indicates that except in the extreme case of a very large coefficient of variation and adjustment for multiple comparisons we will be able to detect *k*-fold differences well below our objective of 2.
These calculations are conservative for two reasons. The alternative hypothesis could be considered one-sided in this particular case and the Type I error made one-sided. On the whole, we are reluctant to deal with one-sided alternative hypotheses and prefer to keep the two-sided structure. Second, the analyses would actually start out with analyses of variance and the (global) *F*-tests will provide protection against the multiple comparison problem.

The investigator should be advised that these sample size calculations are only good to “order of magnitude.” Many things can happen that change the scenario under which the calculations were made. But they are a useful help in planning the experimental effort.