STRUTS: Statistical Rules of Thumb

Gerald van Belle

Departments of Environmental Health and Biostatistics

University of Washington

Seattle, WA

98195-4691

Steven P. Millard

Probability, Statistics and Information

Seattle, WA

98115-5117

©

Last date worked on this material 10/5/98

Chapter 2

Sample Size

2.1 The Basic Formula

Introduction

The first question faced by a statistical consultant, and frequently the last, is, "How many subjects (animals, units) do I need?" This usually results in exploring the size of the treatment effects the researcher has in mind and the variability of the observational units. Researchers are usually less interested in questions of Type I error, Type II error, and one-sided versus two-sided alternatives. You will not go far astray if you start with the basic sample size formula for two groups, with a two-sided alternative, normal distribution with variances homogeneous.

Rule of Thumb

The basic formula is:

$$n = \frac{16}{\Delta^2}$$

where,

$$\Delta = \frac{\mu_1 - \mu_2}{\sigma}$$

is the treatment difference to be detected in units of the standard deviation.

Illustration

If the standardized distance is expected to be 0.5 then $16/0.5^2 = 64$ subjects per treatment will be needed. If the study requires only one group then a total of 32 subjects will be needed.

Derivation

For $\alpha = 0.05$, $\beta = 0.20$ the values of $z_{1-\alpha/2} + z_{1-\beta}$ are 1.96 and 0.84 respectively and $2(z_{1-\alpha/2} + z_{1-\beta})^2 = 15.68$ which can be rounded up to 16. So a quick rule of thumb for sample size calculations is:

$$n = \frac{16}{\Delta^2}.$$

Discussion and Extensions

This formula is convenient to memorize. The key is to think in terms of standardized units of Δ. The multiplier can be calculated for other values of Type I and Type II error. In addition, for a given sample size the detectable difference can be calculated.

2.2   Sample Size and Coefficient of Variation

Introduction

Consider the answer to the following question posed in a consulting session,

"What kind of treatment effect are you anticipating?"

"Oh, I'm looking for a 20% change in the mean."

"Mm, and how much variability is there in your observations?"

"About 30%"

How are we going to address this question? It turns out, fortuitously, that the question can be answered.

Rule of Thumb

The sample size formula becomes:

$$n = \frac{8(CV)^2}{(PC)^2}[1 + (1 - PC)^2].$$

where PC is the proportionate change in means $(PC = (\mu_1 - \mu_2)/\mu_1)$ and CV is the coefficient of variation $(CV = \sigma_1/\mu_1 = \sigma_2/\mu_2)$.

Illustration

For the situation described in the consulting session the sample size becomes,

$$n = \frac{8(0.30^2)}{(0.20)^2}[1 + (1 - 0.20)^2].$$

$$= 29.52 \simeq 30$$

and the researcher will need to aim for about 30 subjects per group. If the treatment is to be compared with a standard, that is, only one group is needed then the sample size required will be 15.

Derivation

Since the coefficient of variation is assumed to be constant this implies that the variances of the two populations are not the same and the variance σ_2 in the sample size formula is replaced by the average of the two population variances: $(\sigma_1^2 + \sigma_2^2)/2$. Replacing σ_i by $\mu_i CV$ for $i = 1, 2$ and simplifying the algebra leads to the equation above.

Discussion and Extensions

Sometimes the researcher will not have any idea of the variability inherent in the system. For biological variables a variability on the order of 35% is not uncommon and you will be able to begin the discussion by assuming a sample size formula of:

$$n \simeq \frac{1}{(PC)^2}[1 + (1 - PC)^2].$$

References

For additional discussion see van Belle and Martin (1993).

2.3 Sample Size Confidence Interval Width

Introduction

Frequently the question is asked to calculate a sample size for a fixed confidence interval width. We consider two situations where the confidence in the original scale is w and is $w^* = w/\sigma$ in units of the standard deviation

Rule of Thumb

For w and w^* as defined above the sample size formulae are:

$$n = \frac{16 * \sigma^2}{w^2},$$

and,

$$n = \frac{16}{(w^*)^2}.$$

Illustration

If $\sigma = 4$ and the confidence interval width desired is 2 then the required sample size is 64. In terms of standardized units the value for $w^* = 0.5$ leading to the same answer.

Derivation

The width, w, of a 95% confidence interval is,

$$w = 2 * 1.96 \frac{\sigma}{\sqrt{n}}.$$

Solving for n and rounding up to 16 leads to the result for w, substituting $w = \sigma * w^*$ leads to the result for w^*.

Discussion and Extensions

The sample size formula for the confidence interval width is identical to the formula for sample sizes comparing two groups. Thus you have to memorize only one formula. If you switch back and forth between these two formulations in a consulting session you must point out that you are moving from two sample to one sample situations.

This formulation can also be used for setting up a confidence interval on a difference of two means. You can show that the multiplier changes from 16 to 32. This makes sense because the variance of two independent means is twice the variance of each mean.

2.4   Sample Size and the Poisson Distribution

Introduction

A rather elegant result for sample size calculations can be derived in the case of Poisson variables. It is based on the square root transformation of Poisson random variables.

Rule of Thumb

$$n = \frac{4}{\left(\sqrt{\lambda_1} - \sqrt{\lambda_2}\right)^2}.$$

where λ_1 and λ_2 are the means of the Poisson distribution.

Illustration

Suppose two Poisson distributed populations are to be compared. The hypothesized means are 30 and 36. Then the number of sampling units per group are required to be $4/(\sqrt{30} - \sqrt{36})^2 = 14.6 = 15$ observations per group.

Derivation

Let Y_i be Poisson with mean λ_i for i =1, 2. Then it is known that $\sqrt{Y_i}$ is approximately normal ($\mu = \sqrt{\lambda_i}, \sigma = 0.5$) Using equation (1) the sample size formula for the Poisson case becomes:

Discussion and Extensions

The sample size formula can be rewritten as:

$$= \frac{2}{(\lambda_1 + \lambda_2)/2 - \sqrt{\lambda_1 \lambda_2}}.$$

This is a very interesting result, the denominator is the difference between the arithmetic and the geometric means of the two Poisson distributions! The denominator is always positive since the arithmetic mean is larger than the geometric mean (xxxx inequality). So n is the number of observations per group that are needed to detect a difference in Poisson means as specified.

Now suppose that the means λ_1 and λ_2 are means per unit time (or unit volume) and that the observations are observed for a period of time, T. Then Y_i are Poisson with mean $\lambda_i T$. Hence the sample size required can be shown to be:

$$n = \frac{2}{T[(\lambda_1 + \lambda_2)/2 - \sqrt{\lambda_1 \lambda_2}]}.$$

This formula is worth contemplating. By increasing the observation period T we reduce the sample size proportionately, not as the square root! Suppose we choose T so that the number per sample is 1. To achieve that effect we must choose T to be of length:

$$T = \frac{2}{(\lambda_1 + \lambda_2)/2 - \sqrt{\lambda_1 \lambda_2}}.$$

This, again is reasonable since the sum of independent Poisson variables is Poisson.

2.5 Poisson Distribution With Background Radiation

Introduction

The Poisson distribution is a common model for describing radioactive scenarios. Frequently there is background radiation over and above which a signal is to be detected. It turns out that the higher the background radiation the larger the sample size is needed to detect differences between two groups.

Rule of Thumb

Suppose that the background level of radiation is θ^* and let θ_1 and θ_2 now be the additional radiation over background. Then, X_i is Poisson $((\theta^* + \theta_i))$. The rule-of-thumb sample size formula is:

$$n = \frac{16(\theta^* + (\theta_1 + \theta_2)/2)}{(\theta_1 - \theta_2)^2}.$$

Illustration

Suppose the means of the two populations are 1 and 2 with no background radiation. Then the sampling effort is $n = 24$. Now assume a background level of 1.5. Then the sample sizes per group become 48. Thus the sample size has doubled with a background radiation halfway between the two means.

Derivation

The variance of the response in the two populations is estimated by $\theta^* + (\theta_1 + \theta_2)/2$. Thi formula is based on the normal approximation to the Poisson distribution.

Discussion and Extensions

We did not use the square root transformation. The reason is that the background radiation level is more transparently displayed in the original scale and, second, if the square root transformation is used then an expansion in terms in the $\theta's$ produces exactly the formula above. The denominator does not include the background radiation but the numerator does. Since the sample size is proportional to the numerator, increasing levels of background radiation require larger sample sizes to detect the same difference in radiation levels. When the square root transformation formula is used in the first example the sample size is 23.3, and in the second example, 47.7. These values are virtually identical to 24 and 48. While the formula is based on the normal approximation to the Poisson distribution the effect of background radiation is very clear.

2.6   Sample Size and the Binomial Distribution

Introduction

The binomial distribution provides a model for the occurrence of independent Bernoulli trials. The sample size formula in equation (1) can be used for an approximation to the sample size question involving two independent samples. We use the same labels for variables as in the Poisson case. Let Y_i be independent binomial random variables with probability of success π_i, respectively. Assume that equal samples are required.

Rule of Thumb

$$n = \frac{4}{(\pi_1 - \pi_2)^2}.$$

Illustration

For $\pi_1 = 0.5$ and $\pi_2 = 0.7$ the required sample size per group is $n = 100$.

Derivation

$$\Delta = \frac{\pi_1 - \pi_2}{\sigma},$$

where,

$$\sigma = \sqrt{1/2[\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}.$$

An upper limit on the required sample size is obtained at the maximum values of π_i which occurs at $\pi_i = 1/2$ for $i = 1, 2$. For these values $\sigma = 1/2$ and the sample size formula becomes as above.

Discussion and Extensions

Some care should be taken with this approximation. It is reasonably good for values of n that come out between 10 and 100. For larger (or smaller) resulting sample sizes using this approximation, more exact formulae should be used. For more extreme values use tables of exact values given by Haseman (1978) or use more exact formulae (see Fisher and van Belle, 1993). Note that the tables by Haseman are for one-tailed tests of the hypotheses.

References

Haseman (1978) contains tables for "exact" sample sizes based on the hypergeometric distribution. See also Fisher and van Belle (1993)

2.7 Sample Size and Precision

Introduction

In some cases it may be useful to have unequal sample sizes. For example, in epidemiological studies in may not be possible to get more cases but more controls are available. Suppose n subjects are required per group but only n_1 are available for one of the groups where we assume that $n_1 < n$. We desire to know the number of subject, kn_1 required in the second group in order to obtain the same precision as with n in each group.

Rule of Thumb

The required value of k is,

$$k = \frac{n}{(2n_1 - n)}.$$

Illustration

Suppose that sample size calculations indicate that $n = 16$ cases and controls are needed in a case-control study. However, only 12 cases are available. How many controls will be needed to obtain the same precision? The answer is

$k = 16/8 = 2$ so that 24 controls will be needed to obtain the same precision as with 16 cases and controls.

Derivation

For two independent samples of size n, the variance of the estimate of difference (assuming equal variances) is proportional to,

$$\frac{1}{n} + \frac{1}{n}.$$

Given a sample size $n_1 < n$ available for the first sample and a sample size kn for the second sample, then equating the variances for the two designs,

$$\frac{1}{n} + \frac{1}{n} = \frac{1}{n_1} + \frac{1}{kn_1},$$

and solving for k produces the result.

Discussion and Extensions

This approach can be generalized to situations where the variances are not equal. The derivations are simplest when one variance is fixed and the second variance is considered a multiple of the first variance (analogous to the sample size calculation).

Now consider two designs, one with n observations in each group and the other with n and kn observations in each group.

The relative precision of these two designs is,

$$\frac{SE_k}{SE_1} = \sqrt{\frac{1}{2}\left(1 + \frac{1}{k}\right)},$$

where SE_k and SE_1 are the standard errors of the designs with kn and n subjects in the two groups respectively.

For $k = 1$ we are back to the usual two-sample situation with equal sample size. If we make $k = \infty$ the relative precision is $\sqrt{0.5} = 0.71$. Hence, the best we can do is to decrease the standard error of the difference by 29%. For $k = 4$ we are already at 0.79 so that from the point of view of precision there is no reason to go beyond four or five more subjects in the second group than the first group. This will come close to the maximum possible precision in each group.

2.8   Sample Size and Cost

Introduction

In some two sample situations the cost per observation is not equal and the challenge then is to choose the sample sizes in such a way so as to minimize

cost and maximize precision, or minimize the standard error of the difference (or, equivalently, minimize the variance of the difference). Suppose the cost per observation in the first sample is c_1 and in the second sample is c_2. How should the two sample sizes n_1 and n_2 be chosen?

Rule of Thumb

The ratio of the two sample size is:

$$\frac{n_2}{n_1} = \sqrt{\frac{c_1}{c_2}}.$$

This is known as the square root rule: pick sample sizes inversely proportional to square root of the cost of the observations. If costs are not too different then equal sample sizes are suggested (because the square root of the ratio will be closer to 1).

Illustration

Suppose the cost per observation for the first sample is 160 and the cost per observation for the second sample is 40. Then the rule of thumb states that you should take twice as many observations in the second group as compared to the first. To calculate the specific sample sizes, suppose that on an equal sample basis 16 observations are needed. To get equal precision with n_1 and $2n_1$ we solve the same equation as in the previous section to produce 12 and 24 observations, respectively.

Derivation

The cost, C, of the experiment is:

$$C = c_1 n_1 + c_2 n_2,$$

where n_1 and n_2 the number of observations in the two groups, respectively, and are to be chosen to minimize:

$$\frac{1}{n_1} + \frac{1}{n_2}$$

subject to the total cost being C. This is a linear programming problem with solutions:

$$n_1 = \frac{C}{c_1 + \sqrt{c_1 c_2}},$$

and

$$n_2 = \frac{C}{c_2 + \sqrt{c_1 c_2}}.$$

When ratios are taken the result follows.

Discussion and Extensions

The argument is similar as that in connection with the unequal sample size rule of thumb.

2.9   The Rule of Threes

Introduction

The rule of threes can be used to address the following type of question, "I am told by my physician that I need a serious operation and have been informed that there has not been a fatal outcome in the twenty operations carried out by the physician. Does this information give me an estimate of the potential post operative mortality?" The answer is "yes!"

Rule of Thumb

Given no observed events in n trials, a 95% upper bound on the rate of occurrence is,

$$\frac{3}{n}.$$

Illustration

Given no observed events in 20 trials a 95% upper bound on the rate of occurrence is $3/20 = 0.15$. Hence, with no fatalities in twenty operations the rate could still be as high as 0.15.

Derivation

Formally, we assume Y is Poisson(μ). We use n samples. For the Poisson we have the useful property that the sum of independent Poisson variable is also Poisson. Hence in this case, $Y_1 + Y_2 + ... + Y_n$ is Poisson $(n\mu)$ and the question of at least one Y_i not equal to zero is the probability that the sum, $\sum Y_i$, is greater than zero. We want this probability to be, say, 0.95 so that:

$$P\left(\sum Y_i = 0\right) = e^{-n\mu} = 0.05.$$

Taking logarithms we get:

$$n\mu = -ln(0.05) = 2.996 = 3$$

Solving for μ we get:

$$\mu = \frac{3}{n}.$$

This is one version of the "rule of threes."

Discussion and Extensions

We solved the equation $n\mu = 3$ for μ. We could have solved it for n as well. To illustrate this derivation, consider the following question, "'The concentration of cryptosporidium in a water source is μ per liter. How many liters must I take to make sure that I have at least one organism?" The answer is, "Take $n = 3/\mu$ liters to be 95% certain that there is at least one organism in your sample.

For an interesting discussion see Hanley and Lippman-Hand (1983). For other applications see Fisher and van Belle (1993).

References

Fisher, L. and van Belle G. (1993). *Biostatistics: A Methodology for the Health Sciences.* Wiley and Sons, New York, NY.

Hanley, J.A. and Lippman-Hand, A. (1983) If nothing goes wrong, is everything alright? *Journal of the American Medical Association,* **249:** 1743-1745.

Haseman, J.K. (1978) Exact sample sizes for the use with the Fisher-Irwin test for 2x2 tables. *Biometrics,* **34:** 106-109.

van Belle, G. and Martin D. C. (1993) Sample size as a function of coefficient of variation and ratio of means. *American Statistician,* **47:** 165-167.

2.10   WEB sites

In the next few years there will be an explosion of statistical resources available on WEB sites. Here are some that are already available. All of these programs allow you to calculate sample sizes for various designs.

1. Designing clinical trials
   http://hedwig.mgh.harvard.edu/size.html

2. Martindale's "The Reference Desk: Calculators On-Line"
   This is a superb resource for all kinds of calculators. If you use this URL you will go directly to the statistical calculators. It will be worth your while to browse all the resources that are available.
   http://www-sci.lib.uci.edu/HSG/RefCalculators2.html#STAT

3. Power Calculator
   http://www.stat.ucla.edu/~ jbond/HTMLPOWER/index.html

4. Russ Lenth's power and sample-size page
   http://www.stat.uiowa.edu/~ rlenth/Power/index.html

5. Power analysis for ANOVA designs
   http://www.math.yorku.ca/SCS/Demos/power/