1 Hypothesis Testing

When conducting a hypothesis test, there are two adjustable values that effect the sample size, n. They are α and β, and are responsible for controlling Type I and Type II error.

$$P(\text{Type I Error: false positive}) = \alpha$$

$$P(\text{Type II Error: false negative}) = \beta$$

α and β should be as small as possible, which is done by increasing the sample size, n.

1.1 Means

1.1.1 One group

With one group, the hypotheses are quite simple. Given the statistic X being tested (mean, for example)

$$H_0 : X = \mu_0$$

$$H_1 : X = \mu_1$$

Where $X \sim N(\mu_0, \sigma_0^2)$ if H_0 is correct, and $X \sim N(\mu_1, \sigma_1^2)$ if H_1 is correct. Lachin in [4] gives a detailed description of how to manipulate the above probabilities to get the following equation for n.

$$n \geq \left[\frac{z_\alpha \sigma_0 + z_\beta \sigma_1}{\mu_1 - \mu_0} \right]^2 \tag{1}$$

1.1.2 Two independant groups

Two groups, i.e, experimental and control, not necessarily with equal distribution. Let $n_e = Q_e n$ and $n_c = Q_c n$ define the sample sizes of each group, where $Q_e + Q_c = 1$. Note that now there are two statistics: $X_e \sim N(\mu_e, \sigma_e^2)$ and $X_c \sim N(\mu_c, \sigma_c^2)$. Let σ^2 be the pooled estimation of σ_e^2 and σ_c^2.

$$H_0 : \mu_e - \mu_c = 0$$

$$H_1 : |\mu_e - \mu_c| = \delta$$

$$n \geq \frac{\sigma^2 (Q_e^{-1} + Q_c^{-1})(Z_\alpha + Z_\beta)^2}{\delta^2} \tag{2}$$

For practical purposes, note that Q_c is not a real random variable, but defined as $Q_c = 1 - Q_e$. Also, for simplicity, the function takes as input the pooled estimate of σ.

1.1.3 Paired Observations

Instead of comparing two independant observations, two observations at different times from the same subject are taken. Let $\bar{d} = X_e - X_c$, and redefine the variance as $\sigma_d^2 = 2\sigma^2(1 - \rho)$, where ρ is the correlation coefficient.

$$H_0 : \bar{d} = 0$$

$$H_1 : \bar{d} = \delta$$

$$n \geq \frac{\sigma_d^2 (Z_\alpha + Z_\beta)^2}{\delta^2} \tag{3}$$

1.2 Proportions

Proportions are a special case of the means, using the following relationship.

$$X \sim Bin(n, \pi) \sim N\left(\pi, \frac{\pi(1 - \pi)}{n}\right)$$

Using this, the sample size necessary for one and two samples can be calculated.

1.2.1 One sample

Testing the hypotheses $H_0 : X = \pi_0$ against $H_1 : X = \pi_1$, then N is defined as

$$n \geq \left[\frac{Z_\alpha \sqrt{\pi_0(1-\pi_0)} + Z_\beta \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2 \tag{4}$$

1.2.2 Two independant observations

This function is developed the same way that it was for the independent observations of a mean calculation. Define $\bar{\pi} = Q_e \pi_e + Q_c \pi_c$.

$$H_0 : \pi_e - \pi_c = 0$$

$$H_1 : \pi_e - \pi_c = \delta$$

$$n \geq \frac{(Z_\alpha + Z_\beta)^2 4\bar{\pi}(1-\bar{\pi})}{(\pi_e - \pi_c)^2} \tag{5}$$

Note that this equation uses a slight estimation, and that the function `bsamsize` in R does not, so it's results, used in the interactive function, will differ slightly from the above equation.

1.2.3 Paired Observations

Now the two observations are not independent, but are taken at times t and s on the same subject. Consider the following table. Where a, b, c and d could be converted into probabilities π_a, π_b, π_c and π_d by dividing

	Time t		
	+	-	
Time s +	a	b	n_s
-	c	d	
	n_t		N

by the sample size N. Also, define π_t and π_s as the marginal probabilities. Then the hypotheses are defined as

$$H_0 : \pi_s - \pi_t = 0$$

$$H_1 : |\pi_s - \pi_t| = \delta$$

And note that the calculation of $\pi_s - \pi_t = \pi_b - \pi_c$, therefore the problem ends up being defined in terms of the number of discordant pairs. Using this fact, and the standard variance calculations, the sample size can be defined as

$$n \geq \left[\frac{Z_\alpha \sqrt{\pi_b + \pi_c} + Z_\beta \sqrt{\frac{4\pi_b \pi_c}{\pi_b + \pi_c}}}{\pi_b - \pi_c} \right]^2 \tag{6}$$

Though this function does not seem to depend on pi_s and pi_t, they are used to determine the probabilities pi_b and pi_c, so are still taken as input.

1.3 Survival Analysis

Assume k groups being analyzed, each with an exponential distribution with mean μ_j. Let $\rho_j = ln(\mu_j)$ and d_j be the number of failures per group. These equations are taken from [4] and [6], and depend heavily on their calculations.

k	τ(k-1,0.05,0.05)	τ(k-1,0.05,0.10)	τ(k-1,0.05,0.20)
2	12.995	10.507	7.849
3	15.443	12.654	9.635
4	17.170	14.171	10.903
5	18.572	15.405	11.935
6	19.780	16.469	12.828

Table 1: $\tau(k-1, \alpha, \beta)$ value from the non central chi-squared distribution required to achieve power (1-β) in testing H_0 at level α

1.3.1 comparing means

The most basic hypothesis is that all the groups have the same mean, with the alternative being that there are at least two groups with different means.

$$H_0 : \mu_j = \mu, \forall j$$

$$H_1 : \exists k, l \ s.t \ \mu_k \neq \mu_l$$

Using this hypothesis, it can be shown that, under the alternative hypothesis, the data follows a non-central chi-squared distriubtion on k-1 degrees of freedom, with non-central parameter

$$\tau = \sum d_j (\rho_j - \bar{\rho})^2 \tag{7}$$

Where $\bar{\rho}$ is the mean of the ρ_j weighted by the number of failures per group, d_j. The following table is a sample of the necessary τ values need to satisfy α and β. Given a set of ρ_j values, a family of d_j values can be found that will satisfy the Type I and II errors.

Note that, if all the d_j values are taken as the same, then the equation for τ can be solved for explicitly. Note that the equation for τ would not change if the group means, μ_j were scaled, so it turns out that the sample size is dependant on the ratio of ρ values, and not there differences. If the alternative hypothesis is defined by $a_j = \frac{\mu_j}{\mu_1}, \forall j > 1, \mu_1 = 1$, and the number of failures per group are taken to all be the same, d, then the equation can be solved explicitly.

$$d \geq \frac{\tau(k-1, \alpha, \beta)}{\sum (ln(a_j) - \bar{lna})^2} \tag{8}$$

Finally, if the alternative hypothesis is not concerned with the individual means, but with the ratio between the largest and the smallest means, then d can be calculated even easier. Define a_k as the largest difference between two means, $a_k = max\left(\frac{\mu_j}{\mu_k}, \forall j, k\right)$. Then d is defined by

$$d \geq \frac{2 * \tau(k-1, \alpha, \beta)}{(lna_k)^2} \tag{9}$$

1.3.2 An approach for two groups

Because of the complexity of the above method, and the fact that the τ's seem only defined for a select group, it is worth looking at another approach. Consider a two-treatment trial that is going to be run up to time T, with patients entering any time up to time point T. Thinking of the two groups as an experimental and a control group with means μ_e and μ_c respectively, the hypotheses are

$$H_0 : |\mu_e - \mu_c| = 0$$

$$H_A : |\mu_e - \mu_c| \neq 0$$

Without going into the complex details of the problem, define the equation

$$\psi(\mu) = \frac{\mu^3 T}{\mu T - 1 + e^{-\mu T}}$$

	Exposed	Not Exposed	
Disease	a	b	
No Disease	c	d	

And the variable $\bar{\mu} = Q_e\mu_e + Q_c\mu_c$, then the sample size is defined as

$$n \geq \frac{(Z_{\alpha/2}\sqrt{\psi(\bar{\mu})(Q_e^{-1} + Q_c^{-1})} + Z_\beta\sqrt{\psi(\mu_e)Q_e^{-1} + \psi(\mu_c)Q_c^{-1}})^2}{(\mu_e - \mu_c)^2} \tag{10}$$

1.4 Miscellaneous Statistics

1.4.1 Kappa

The paper by Donner and Eliasziw, [3], describes several methods of computing kappa, finding confindence intervals for kappa, and on performing hypotheses tests, on top of the sample size calculations. Their calculations use the chance corrected calculation of κ outlined in the agreement section of the course. The sample size calculation is based on the hypothesis test:

$$H_0 : \kappa = \kappa_0$$

$$H_A : \kappa = \kappa_1$$

And, as in the survival analysis section, the assumption is that, under the alternative hypothesis, the distribution falls under the non-central chi-sqaured distribution with parameter $\tau(1, \alpha, \beta)$. π is the chance of a subject as being rated as positive. Given the above assumptions, the sample size is calculated using the equation:

$$n \geq \tau(1, \alpha, \beta)\left[\frac{\pi^2(1-\pi)^2(\kappa_1 - \kappa_0)^2}{\pi^2 + \pi(1-\pi)\kappa_0} + \frac{2\pi^2(1-\pi)^2(\kappa_1 - \kappa_0)^2}{\pi(1-\pi)(1-\kappa_0)} + \frac{\pi^2(1-\pi)^2(\kappa_1 - \kappa_0)^2}{(1-\pi)^2 + \pi(1-\pi)\kappa_0}\right]^{-1} \tag{11}$$

1.4.2 Odds Ratio

The equations here are taken from [5]. The design of an experiment to calculate odds ratio is similar to that for paired observations. Consider the following table. The hypotheses for this test are

$$H_0 : OR = 1$$

$$H_A : OR \neq 1$$

Given the oabove hypotheses, and confidence and power defined by α and β, then the sample size is definde as

$$n \geq \left(\frac{z_{\alpha/2}^2\sqrt{2p_2(1-p_2)} + z_\beta\sqrt{p_1(1-p_1) + p_2(1-p_2)}}{p_1 - p_2}\right)^2 \tag{12}$$

Note that the $2p_2(1-p_2)$ is being used instead of the more logical $2\bar{p}(1-\bar{p})$. According to Lwanga, [5], this is done because "the study population is likely to be made up of many more controls than cases, and the exposurte rate among the controls os often known with a high degree of precision; under the null hypothesis this is the exposure rate for the cases as well. If the investgator is in doubt about the exposure rate among the controls, however, the formula should be modified", ie, $\bar{p}$ should be used, where $\bar{p} = (p_1 + p_2)/2$.

1.4.3 Relative Risk

The equations here are also taken from [5], and are very similar to those in the odds ratio equations. The hypotheses for this test are

$$H_0 : RR = 1$$

$$H_A : RR \neq 1$$

The relative risk is defined by two probabilities: the probability of contracting a disease given exposure, p_1 and the probability of contracting a disease given no exposure, p_2. The relative risk is defined as p_1/p_2, and the sample size required for an accurate calculation, ie, within the range of ϵ, is

$$n \geq \left(\frac{z^2_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + z_\beta \sqrt{p_1(1-p_1)+p_2(1-p_2)}}{p_1 - p_2} \right)^2 \tag{13}$$

2 Point Estimation

2.1 SRS

2.1.1 Means

Assuming we're calculating the sample mean, we want it to be correct within a certain value, a certain percentage of them time.

$$P(|\bar{y_s} - \mu_y| < \epsilon) = 1 - \alpha$$

Unfortunately, ϵ is often dificult to calculate, so we'll use a *relative precision*.

$$P(|\frac{\bar{y_s} - \mu_y}{\mu_y}| < \epsilon) = 1 - \alpha$$

Now, using the CLT for a simple random sample, and re-arranging [2], page 14, we get

$$n \geq \frac{z^2_{\alpha/2}\sigma^2}{\epsilon^2\mu^2 + \frac{z^2_{\alpha/2}\sigma^2}{N}}$$

Now because the population mean and variance are often difficult to estimate, the **coefficient of variation of the population** is often used, because it tends to be simpler to guess at.

$$CV_U = \frac{\sigma^2}{\mu}$$

Which redefines the equation for n as

$$n \geq \frac{z^2_{\alpha/2}CV^2_U}{\epsilon^2 + \frac{z^2_{\alpha/2}CV^2_U}{N}} \tag{14}$$

In this equation, ϵ is the only unknown, so it's the variable that should be manipulated in the Tk widget.

2.1.2 Proportions

Remember that proportions are the same as means, except that $y_k = 1/0$, so the sample size calculation can be approached the same way, ie, The desired precision of the estimation can be expressed as

$$P(|\bar{y_s} - \mu_y| < \epsilon) = P(|\hat{p} - p| < \epsilon) = 1 - \alpha$$

Where the term $\hat{p}$ is the estimate of the proportion from the sample, and p is the true sample proportion. Note that there is no need to use relative precision in this case as there was in the regular case: the range of p is already restricted to [0,1], so the computation can proceed as follows, producing the following equation for n

$$n \geq \frac{z^2_{\alpha/2}p(1-p)}{\epsilon^2} \tag{15}$$

Note that the above equation is dependent on knowing the true proportion. This is often unknown, but since the range of p is restricted to [0,1], taking a value of $p = 0.5$ ensures that the sample will cover all possible values of p.

	Exposed	Not Exposed	
Disease	a	b	
No Disease	c	d	

2.2 Miscellaneous Statistics

2.2.1 Intraclass correlation

The paper by Bonett, [1], gives an excellent summary of the following equation, which will not be explained in any detail here. Since the intra-class correlation is calculated for an ANOVA situation, two sample sizes are needed: the number of subjects: n, and the number of treatments per subject, k. Also, define ρ_I as the intra-class correlation, then treating the number of treatments, k, as fixed, the number of subjects required to estimate ρ_I within a certain range ϵ is defined as

$$n \geq 8z^{\alpha/2}\left[\frac{(1-\rho_I)^2(1+(k-1)\rho_I)^2}{(k(k-1)\epsilon^2}\right]+1 \tag{16}$$

Like most sample size estimations, this requires an initial guess at the value of ρ_I.

2.2.2 Odds Ratio

The equations here are taken from [5]. The design of an experiment to calculate odds ratio is similar to that for paired observations. Consider the following table.

From a design such as this, the odds ratio is defined as ad/bc. To estimate the odds ratio with relative precision, ϵ, the following equation is used. p_1 is the probability of exposure given the subject has the disease: $a/(a+b)$ and p_2 is the probability of exposure given that the subject does not have the disease: $c/(c+d)$.

$$n \geq \frac{z^2_{\alpha/2}}{ln(1-\epsilon)^2}\left(\frac{1}{p_1(1-p_1)}+\frac{1}{p_2(1-p_2)}\right) \tag{17}$$

Note that either p_1 or p_2 can be extrapolated from the equation for the odds ratio if the odds ratio and the other probability are known.

$$OR = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$$

2.2.3 Relative Risk

The equations here are taken from [5]. This time, the relative risk is being estimated within ϵ of its true value. The equation for n is

$$n \geq \frac{z^2_{\alpha/2}}{ln(1-\epsilon)^2}\left(\frac{(1-p_1)}{p_1}+\frac{(1-p_2)}{p_2}\right) \tag{18}$$

3 Power Calculation

Though hypothesis tests are of ten designed with the goal of obtaining a certain power level, by obtaining a certain sample size, the value of n used is not always the same as what was outlined. The following equations are versions of equation previously defined in the paper, solved for the power variable. Given an alpha value and a sample size n, along with the equation specific random variables, a power rating can be calculated.

3.1 Means

Equation (1)

$$Z_\beta = \frac{\sqrt{n}|\mu_1-\mu_0|-Z_\alpha\sigma_0}{\sigma_1} \tag{19}$$

Equation (2)

$$Z_\beta = \frac{\delta}{\sigma}\sqrt{\frac{n}{Q_e^{-1} + Q_c^{-1}}} - Z_\alpha \tag{20}$$

Equation (3)

$$Z_\beta = \frac{\delta}{\sigma_d}\sqrt{n} - Z_\alpha \tag{21}$$

3.2 Proportions

Equation (4)

$$Z_\beta = \frac{\sqrt{n}|\pi_1 - \pi_0| - Z_\alpha\sqrt{\pi_0(1-\pi_0}}{\sqrt{\pi_1(1-\pi_1)}} \tag{22}$$

Equation (5)

$$Z_\beta = \sqrt{\frac{n}{4\bar{\pi}(1-\bar{\pi})}}|\pi_e - \pi_c| - Z_\alpha \tag{23}$$

Equation (6)

$$Z_\beta = \left(\sqrt{n}|\pi_b - \pi_c| - Z_\alpha\sqrt{\pi_b + \pi_c}\right)\sqrt{\frac{\pi_b + \pi_c}{4\pi_b\pi_c}} \tag{24}$$

3.3 Survival Analysis

Note that, for the following equations, the left side $\tau(k-1, \alpha, \beta)$ set to a known value is solvable for β.

Equation (8)

$$\tau(k-1, \alpha, \beta) = d\sum (ln(a_j) - l\bar{n}a)^2 \tag{25}$$

Equation (9)

$$\tau(k-1, \alpha, \beta) = \frac{d(lna_k)^2}{2} \tag{26}$$

Equation (10)

$$Z_\beta = \frac{\sqrt{n}|\mu_e - \mu_c| - z_{\alpha/2}\sqrt{\psi(\bar{\mu})(Q_e^{-1} + Q_c^{-1})}}{\sqrt{\psi(\mu_e)Q_e^{-1} + \psi(\mu_c)Q_c^{-1}}} \tag{27}$$

3.4 Miscellaneous Statistics

Equation (11). Note that $\tau(1, \alpha, \beta) = (z_{\alpha/2} + z_\beta)^2$.

$$z_\beta \geq \sqrt{n}\sqrt{\frac{\pi^2(1-\pi)^2(\kappa_1-\kappa_0)^2}{\pi^2 + \pi(1-\pi)\kappa_0} + \frac{2\pi^2(1-\pi)^2(\kappa_1-\kappa_0)^2}{\pi(1-\pi)(1-\kappa_0)} + \frac{\pi^2(1-\pi)^2(\kappa_1-\kappa_0)^2}{(1-\pi)^2 + \pi(1-\pi)\kappa_0}} - z_{\alpha/2} \tag{28}$$

Equation (12)

$$z_\beta = \frac{\sqrt{n}|p_1 - p_2| - z_{\alpha/2}\sqrt{2p_2(1-p_2)}}{\sqrt{p1_(1-p_1) + p_2(1-p_2)}} \tag{29}$$

Equation(13)

$$z_\beta = \frac{\sqrt{n}|p_1 - p_2| - z_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})}}{\sqrt{p1_(1-p_1) + p_2(1-p_2)}} \tag{30}$$

References

[1] Bonett, Douglas G. *Sample Size Requirements for Estimating Intraclass Correlations with Desired Precision.* Statistics in Medicine, 2002; **21**: 1331-1335.

[2] Boudreau,C. STAT454/854: Sampling Theory and Practice: Course Notes. University of Waterloo, Dept. of Statistics & Actuarial Science. Winter 2007.

[3] Donner, Alland and Eliasziw, Michael. *A Goodness-of-Fit Approach to Inference Procedures for the Kappa Statistic: Confidence Interval Construction, Significance-Testin and Sample Size Estimation.* Statistics in Medicine, Vol. 11: 1511-1519.

[4] Lachin, John M. *Introduction to Sample Size Determination and Power Analysis for Clinical Trials.* Control Clinical trials **2** 93-113. 1981

[5] Lwanga, S. K., and Lemeshow, S. *Sample Size Determination in Health Studies: A Practical Manual.* World Health Organization; Geneva: 1991. Available online at: http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/.

[6] Makuch, Robert W. and Simon, Richard M. *Sample Size Requirements For Comparing Time-To-Failure Among k Treatment groups.*??.