PBIO 504
Experimental Design and Analysis

Sample Size and Power in Study Design

Textbook Chapter 10

How many study subjects do I need?

http://jama.jamanetwork.com/article.aspx?articleid=1886173&resultClick=1

http://www.youtube.com/watch?v=PbODigCZqL8

Review: Truth and Errors in HypothesisTesting

<table>
<tr><td colspan="2" rowspan="2">Contrast between the test result and the truth in the population</td><td colspan="2">Truth in the population:</td></tr>
<tr><td>Association between risk factor and disease</td><td>No association exists</td></tr>
<tr><td rowspan="2">Result of statistical test in the study sample:</td><td>Reject null hypothesis</td><td>Correct</td><td>type I error</td></tr>
<tr><td>Accept null hypothesis</td><td>type II error</td><td>Correct</td></tr>
</table>

Questions

- **How large of a difference we want to find?**
  –What is the ‘minimal worthwhile/clinically relevant difference’? (this is more of a clinical question not a statistical one)

- **What is the effect size?**
- **Endpoint: Looking for a difference in what?**
- **Do you want to follow the patients up?**
- **Do you have prior estimates of variability?**
- **What precision, significance level, power do you desire?**
- **What is the sampling plan? Random sample?**
- **What is the available resource (accrual and $) ?**
- **What are your study hypotheses after all?**

Why don’t we make alpha and beta very small?

- Ideally, alpha and beta would be set at or near zero, to eliminate any possibility of false inference, but in practice they are set as small as possible and practical.
- As we know, <u>alpha and beta are inversely related</u>: lowering one of will increase the other, and vice-versa.
- This is where the **sample size** comes in, and where we need to carefully consider the consequences of manipulating alpha, beta, and the sample size.

What is the Optimal Size of a Study

- Studies that are too large (in terms of numbers of experimental units or study subjects) may be wasteful of resources
- On the other hand, studies that are too small may fail to detect true associations (type II errors)
- The optimal sample size is derived based on the statistical test appropriate to test the proposed research hypotheses under various parameters

Example of alpha and power in the design of a study of Colon Cancer Risk

- An investigator plans to study the association of dietary intake of beta carotene and colon cancer.
- She sets *alpha* at .05 and *beta* at 0.10 for the statistical analysis (i.e. 5% chance of incorrectly rejecting the null hypothesis).
- Suppose that the study found that there was a 30% reduction in colon cancer incidence when a high level of beta carotene was consumed. With 90% power (*beta= 0.10*), it could be concluded that 90 times out of 100 the investigator would observe a risk of colon cancer reduction of 30% or greater.

Tumor Xenograft Experiment

- An investigator plans an experiment to examine if new anti-cancer agents would be effective in animal models

Group 1: TMZ 42 mg/kg+CPT11 0.61 mg/kg;

Group 3: CPT11 0.61 mg/kg

Preliminary data: next slide

Xenograft Data:

Tumor Volumes (cm^3) (Rh18 cell-line)

Group	No.	Weeks 0	1	2	3	4	5	6	7	8	9	10	11	12
I	1	0.03	0.26	0.24	0.17	0.01	*	*	*	*	*	*	*	*
	2	0.52	0.15	0.14	0.01	*	*	*	*	*	*	*	*	*
	3	1.57	0.34	0.18	0.15	0.01	*	*	*	*	*	*	*	*
	4	0.89	0.38	0.24	0.36	0.12	0.14	0.10	0.14	0.14	0.30	1.42	2.91	4.57
	5	1.28	0.48	0.22	0.25	0.01	*	*	*	0.01	0.20	1.31	2.62	5.73
	6	0.53	0.34	0.16	0.20	0.01	*	*	*	0.01	0.18	1.06	1.94	3.13
	7	1.49	0.12	0.10	0.01	*	–	–	–	–	–	–	–	–
II	1	0.70	0.32	0.14	0.19	0.01	*	*	*	0.01	0.17	0.65	1.08	2.74
	2	0.37	0.09	0.09	0.01	*	*	*	*	*	*	*	*	*
	3	0.30	0.08	0.01	*	*	*	*	*	*	*	*	*	*
	4	0.26	0.15	0.01	*	*	0.01	0.28	0.26	0.25	0.39	0.58	0.67	0.83
	5	0.22	0.11	0.08	0.06	0.01	*	*	*	*	–	–	–	–
	6	1.08	0.23	0.09	0.15	0.01	0.01	0.24	0.24	0.27	0.92	2.39	4.12	6.75
	7	0.87	0.31	0.15	0.22	–	–	–	–	–	–	–	–	–
III	1	1.75	0.95	0.54	0.64	0.24	0.69	1.16	1.80	1.22	3.17	5.96	6.06	8.17
	2	0.16	0.09	0.09	0.14	0.10	0.13	0.21	0.22	0.28	0.42	0.93	–	–
	3	0.83	0.39	0.29	0.43	0.36	0.34	0.61	0.74	0.57	1.46	1.99	2.59	3.33
	4	1.54	0.94	0.30	0.42	0.30	0.41	0.87	1.32	0.96	2.10	4.27	5.76	6.63
	5	0.91	0.24	0.10	0.01	*	*	*	*	*	*	*	*	*
	6	0.78	0.22	0.07	0.01	*	*	*	*	*	*	*	*	*

'–:" the mouse died or the tumor quadrupled, "*": tumor volume $< 0.01cm^3$.

Two-sample t test with unequal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
TVol1	7	.1642857	.0475022	.1256791	.0480519	.2805195
TVol3	6	.275	.1059796	.2595958	.0025709	.5474291
combined	13	.2153846	.0549664	.1981841	.0956232	.335146
diff		-.1107143	.1161384		-.3855281	.1640996

diff = mean(TVol1) - mean(TVol3)   t = -0.9533

Ho: diff = 0   Satterthwaite's degrees of freedom = 6.97618

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.1861	Pr(|T| > |t|) = 0.3723	Pr(T > t) = 0.8139

.

Classic Sample Size Formula

- Comparing means of two groups (two-sided)
- The # of subjects **per group**

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_0)^2} = \frac{2K^2\sigma^2}{\Delta^2}$$

- K is determined by α, β, one-sided or two-sided
- When n is small, the formula is approximate (if n‹10， plus 2)
- **If it is one-sided,** $K = (Z_{\alpha} + Z_{\beta})^2$

Important Elements in Sample Size Estimation

- Alpha
- Power (1-Beta)
- Effect size (e.g. difference in means, or the size of the odds ratio)
- Precision (variability)

Steps in Estimating Sample Size

1. Decide the levels of alpha and power.
2. Decide on the effect size you wish to detect.
3. Obtain data on (or make an educated guess about) the standard deviation in the study sample.
4. Consult software to input the above figures to determine the required sample size.

An Example

- Planning a study and find required n for comparing mean Lp(a) levels between subjects age 30-60 who have CHD and those without CHD
- Hypotheses: ean Lp(a) levels are different for cases and controls
- Assume Lp(a) approx. normally distributed within each group.
- Plan to use 2-sample t-test
- Pilot Data on 42 Normals: mean Lp(a) = 32 mg/dl s
- Assume alpha .05 (2-sided)
- *Beta* .10 (Power = . 90), $z_{0.10} = 1.28$
- Delta 10 mg/dl
- Sigma=18 mg/dl

$$n = \frac{2(1.96 + 1.28)^2 18^2}{10^2}$$

n=68 per group

How does n change when the other parameters change?

	Δ	σ	α	1-β	n
Ref	5	20	.05	.80	100
Δ ↑	10	20	.05	.80	25
σ ↑	5	25	.05	.80	155
α ↓	5	20	.01	.80	160
β ↓	5	20	.05	.90	138
2-sided	5	20	.05	.80	127

Sample Size Example Using Software

- A researcher is studying a drug to reduce cholesterol levels. If the drug is effective, we would expect to see a lowering of the mean cholesterol level when the study is finished.
- Two groups of adults will be studied: those in Group 1 will not be taking any cholesterol-related medications and will get a placebo, and those in Group 2 will take the new drug.
- Prior research (a large, random survey of adults in the U.S. population) reported that the mean total cholesterol level of people who didn’t use any cholesterol lowering drugs was 200 mg/dl, with a standard deviation of 30.
- Assume that the cholesterol readings are normally distributed; therefore the statistical inference test in our study will be the 2-sample (Student’s) t-test.

Sample Size Example (cont.)

- We set *alpha* at 0.01 because we do not want to risk marketing an ineffective new drug.
- We want to design a study with 95% *power* to detect a mean lowering of 20 mg/dl (*effect size*), since we think this will be a clinically meaningful drop in cholesterol.
- How many subjects per group are needed to observe the desired effect or greater?

→ SWOG statool total sample size n=161

https://stattools.crab.org/Calculators/twoArmNormalColored.htm

SWOG
Leading cancer research. Together.

Two Arm Normal

Select Calculation and Test Type

(•) Sample Size	( ) 1 Sided
( ) Power	(•) 2 Sided

Select Hypothesis Test Parameters

Mean: Arm 1	Mean: Arm 2	Standard Deviation	Ratio of sample size Arm2/Arm1	Alpha
200	180	30	1	.01

Power	Total Sample Size
.95	161

Calculate   Help Document

Xenograft two groups experiment revisited

- *alpha* = .05 and power=90%
- The true standard deviation (average of the two groups) set at 0.075
- If we want to show an observed difference of 0.10 cm3/kg or greater, how many mice do we need?

- SWOG Statool with two arm normal: total n=24

Colon Cancer Risk Example Revisited

- An investigator plans a study of the association of dietary intake of beta carotene and colon cancer risk in men who are 50 years of age or older.
- Previous studies suggest that 2% (true proportion) of these men will develop colon cancer in 20 years.
- Choose *alpha* = .05 and *beta* = 0.20, power=0.80
- Suppose that the study intends to show a 30% reduction in colon cancer incidence when a high level of beta carotene was consumed.

SWOG Statool: alpha=0.05, power=0.8, p_1=0.014, p_2=0.02, n_2 to n_1 ratio=1, → Total n=15234 ($n_2 = n_1$)

Leading cancer research. **Together.**

Two Arm Binomial

Select Calculation and Test Type

◉ Sample Size ○ Power	○ 1 Sided ◉ 2 Sided

Select Hypothesis Test Parameters

Null Proportion	Alternative Proportion	Alpha	Sample Size Ratio 2-to-1
0.014	0.02	.05	1

Calculate Power/Sample Size

Power	Sample Size
.80	15234

Help Document

Study Sample Size and Power

Review

Truth and Errors in HypothesisTesting

Contrast between the test result and the truth in the population		**Truth in the population:**	
		Association between risk factor and disease	No association exists
Result of statistical test in the study sample:	Reject null hypothesis	**Correct**	**type I error**
	Accept null hypothesis	**type II error**	**Correct**

Type I and Type II Errors in Statistical Inference

- **Alpha** is the probability of making a type I error
- A **type I error** (false positive result) occurs when the investigator rejects a null hypothesis which is actually true in the population.

- **Beta** is the probability of making the type II error
- A **type II error** (false negative result) occurs when the investigator fails to reject the null hypothesis that is actually false in the population.

- Both of these errors lead to incorrect statistical inference.

Beta in Hypothesis Testing

- Recall that the probability of making a type II error is *beta* (failing to reject the null hypothesis when it is actually false).
- If *beta* is set at 0.10, for example, one is willing to accept a 10% chance of missing a true association.
- By convention, *beta* is often set in the range of 0.10 to 0.20 but it can be set at any level that makes sense for the specific study aim.

Statistical Power in Study Design

- Recall that, in the context of interpreting statistical test results, such as the t-test, we specified a level of alpha when evaluating the hypotheses.
- By convention, alpha is often set at 0.05 (5% level of confidence), but it can be set at any level that makes sense for the specific study aim.

Power

- The function 1- *beta* is called **power**, defined in the textbook as "*the probability of avoiding a type II error*."
- If *beta* is set at 0.10, for example, one is willing to accept a 10% chance of missing a true association in the data; this corresponds to 90% power, i.e. a 90% chance of avoiding a type II error.
- In practice, if the power of a study to detect a meaningful effect difference is less than 80% then we say that the study is underpowered. In other words,

  1-beta<0.80 is considered insufficient