Power Analysis, Sample Size Calculation

Drs. Cheng, and Simon

One Way ANOVA

- Single factor design with non-repeated measurements
- LOM for DV=ration/interval
- Research hypothesis:

  Different treatment would result significantly differences in response outcomes (DV).

Single factor design LOM=ratio/interval

Statistical Hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

H_A: at least one equality does not hold

Errors in Making Decision

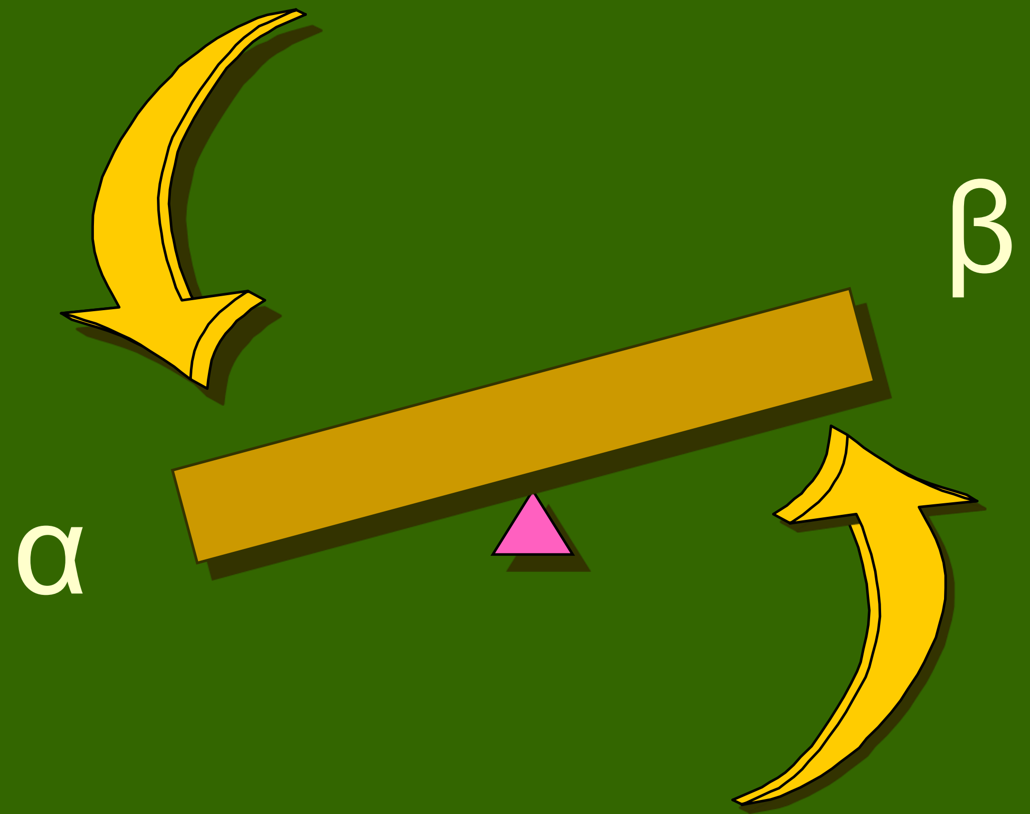
- Type I Error
  - Reject True Null Hypothesis
  - Has Serious Consequences
  - Probability of Type I Error Is α (Alpha)
    - Called Level of Significance
- Type II Error
  - Do Not Reject False Null Hypothesis
  - Probability of Type II Error Is β (Beta)

Decision Results

H0: Innocent

Jury Trial			H_0 Test		
	Actual Situation			Actual Situation	
Verdict	Innocent	Guilty	Decision	H_0 True	H_0 False
Innocent	Correct	Error	Do Not Reject H_0	$1 - \alpha$	Type II Error (β)
Guilty	Error	Correct	Reject H_0	Type I Error (α)	Power ($1 - \beta$)

α & β Have an Inverse Relationship



Factors Affecting β

- True Value of Population Parameter
  - Increases when difference (Effect size) decreases. The *effect size* is the impact made by the independent variable.
- Significance Level, α
  - Increases when α decreases
- Sample Size, *n*
  - Increases when *n* decreases

Outline for Power Analysis

- Prior Power Analysis

  Conducting at the planning stage of the research and is typically used to determine an appropriate sample size to achieve adequate power.

- Post Hoc Power analysis

  Conducting after the study is done and uses the obtained samples to determine the power of statistical analysis.

  -ANOVA (F-test)

  -t-test

Prior Estimation of Power- Sample Size determination

- It is important to plan the sample size so that needed protection against both Type I and Type II errors can be obtained, or so that the estimates of interest have sufficient precision to be useful.

Effect Size

- Effect size needs to be determine before we can calculate the sample size (n)

effect size

0.2   small

0.5   medium

>.8   large

Example

A company owning a large fleet of trucks whishes to determine whether or not four different brands of snow tires have the same mean trend life. From previous studies similar to this, we know the effect size is about 0.3. How many tires per brand are necessary to have power=0.7 when α=0.05??

Example

- Due to the complex of the calculation of power, there are many power tables available to answer such kind of problem.
- Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. 1988. ISBN 0-8058-0283-5.

TABLE C.3

Power of F Test at $\alpha = .05$, $u = 3$

Group Size n	f (effect size)											
	.05	.10	.15	.20	.25	.30	.35	.40	.50	.60	.70	.80
4	05	06	07	08	10	12	15	18	27	38	50	62
5	05	06	07	09	12	15	19	24	36	50	64	76
6	05	06	08	10	13	18	23	29	44	60	75	86
7	05	06	08	11	16	21	27	35	52	69	83	92
8	05	07	09	12	17	24	31	40	59	77	89	96
9	05	07	09	14	19	27	36	46	66	82	93	98
10	05	07	10	15	21	30	40	51	71	87	96	99
11	06	07	11	16	24	33	44	55	76	91	97	99
12	06	08	11	17	26	36	48	60	81	93	98	*
13	06	08	12	19	28	39	52	64	84	95	99	
14	06	08	13	20	30	42	56	68	87	97	99	
15	06	08	13	21	32	45	59	71	90	98	*	
16	06	09	14	23	34	48	62	75	92	98		
17	06	09	15	24	37	51	65	78	94	99		
18	06	09	16	26	39	53	68	80	95	99		
19	06	09	16	27	41	56	71	83	96	99		
20	06	10	17	28	43	59	73	85	97			
22	06	10	18	31	47	63	78	88	98			
24	06	11	20	34	51	68	82	91	99			
26	06	11	22	37	54	72	85	94	99			
28	07	12	23	39	58	75	88	95				
30	07	13	25	42	61	79	90	96				
32	07	13	26	45	65	81	92	97				
34	07	14	28	47	68	84	94	98				
36	07	14	29	50	70	86	95	99				
38	07	15	31	52	73	88	96	99				
40	07	16	32	54	76	90	97	99				
44	08	17	35	59	80	93	98					
48	08	18	39	63	84	95	99					
52	08	20	42	67	87	96	99					
56	08	21	45	71	89	97						
60	09	22	47	74	91	98						
64	09	24	50	77	93	99						
68	09	25	53	80	95	99						
72	09	27	56	82	96	99						
76	10	28	58	84	97	*						
80	10	29	61	86	97							
100	11	36	71	93	99							
140	14	49	86	99								
200	19	66	96									

Example

- The formulas for sample size vary from problem to problem.
- The sample size needed for a **comparison of means from two independent groups** is

$$n = \frac{\left(\sigma_1^2 + \sigma_2^2\right) \times \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{D^2}$$

Example

- We use the letter "z" to represent a **standard normal distribution**. Alpha represents the probability of a **Type I error** (usually .05). Beta represents the probability of a **Type II error** (we usually want this to somewhere between .05 and .20). Sigma represents the **standard deviation**, and this formula allows for the possibility of different standard deviations in group 1 and group 2. Don't forget that the formula requires you to square these standard deviations. Finally, D is the **clinically relevant difference**.

Example

$$n = \frac{\left(\sigma_1^2 + \sigma_2^2\right) \times \left(z_{.975} + z_{.80}\right)^2}{D^2}$$

$$= \frac{\left(1.5^2 + 1.5^2\right) \times \left(1.96 + 0.84\right)^2}{1^2}$$

$$= 35.3$$

So in order to achieve **80% power** for detecting a **one unit difference in the Oucher score**, which has a **reported standard deviation of 1.5**, we would need to sample **36 patients in each group**.

Dichotomous Response Variables

- Two Independent Samples
  - e.g. presence or absence of a virus
  - to calculate the needed N for a trial with dichotomous response variables

$$2N = 4\left(Z_{\alpha} + Z_{\beta}\right)^{2}\,\overline{p}\left(1 - \overline{p}\right) / \left(p_{C} - p_{I}\right)^{2}$$

Values of $Z\alpha$ *(critical value that corresponds to a given* α *level) and* $Z\beta$ *(corresponds to the power* $1-\beta$*) are provided in tables*

Formula applies to equal sample sizes in both groups

$$\overline{p} = \left(p_{C} + p_{I}\right)/2$$

Some quick sample size calculations.

Rule of 50: If you have a binary outcome variable (e.g. mortality), then you should plan your study so that **each group has 25 to 50 events** (e.g., deaths)

Some quick sample size calculations.

Rule of 15: For a **multiple linear regression model**, your sample size should be at least 15 times as large as the number of independent variables.

For a **multiple logistic regression model**, the number of events in your sample should be 15 times as large as the number of independent variables.

- Think carefully about **how quickly you can recruit patients** for a clinical trial. A trial that takes 15 years to complete will be probably be useless.
- Try to estimate how many patients will **refuse to participate**, how many will **fail to meet all eligibility requirement**, and how many patients will **drop out prior to the completion of the study**. Adjust your sample size upward to account for these factors.

Remarks

- For sample size calculation, we can estimate the effect size based on prior studies.
- We should generally assume that all levels have equal sample sizes.
- The α level used in the decision rule for determining whether or not the factor level means are equal is often set relatively high (e.g., 0.05 or 0.1 instead of 0.01) so as to increase the power of the test.
- We usually want power > 0.8 for a study.
- When we fail to reject H0 under $\alpha = 0.05$, we should exam if the test has achieved the desired power.

Biostatistics vs. Lab Research

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

Sample Multiple Choice Question

1. Which statement about power is NOT TRUE when estimating the sample size for a dichotomous response variable in an RCT?

(a) The total sample size will need to be lower when fewer events are expected

(b) The total sample size will need to be lower when a large effect is expected

(c) The total sample size will need to be higher if the alpha value is set lower