MRS.CHANDRALEKHA.E

ASSO.PROFESSOR

ICON

SAMPLE SIZE CALCULATION

SAMPLE SIZE CALCULATION

SAMPLE

• This is the sub-population, to be studied in order to draw a inference from a reference population (a population to which the findings of the Study are to be generalized).

• In Census, the sample size is equal to the population size. However, in research, because of time constraints and budget, a representative sample is normally used.

• Larger the sample, more accurate will be the findings from a Study.

Availability of resources sets upper limit of the sample size.

• Required accuracy sets lower limit of sample size.

• Thus, an optimum sample size is an essential component of any research.

SAMPLE SIZE DETERMINATION

 Sample size determination is the mathematical estimation of the number of subjects/units to be included in a study.

 When a representative sample is taken from a population, the finding are generalized to the population.

 Optimum sample size determination is required for the following reasons:

1. To allow appropriate analysis

2. To provide desired level of accuracy

3. To allow validity to the significance test

AT WHAT STAGE CAN SAMPLE SIZE BE ADDRESSED

It can be addressed at two stages:

1. Calculation of the optimum sample size is required during the planning stage, while designing the Study and information on some parameters.

2. At the stage of interpretation of the result.

PROCEDURE FOR CALCULATING SAMPLE SIZE

There are 3 procedures that could be used for calculating sample size:

1. Use of formula

2. Ready made tables

3. Computer softwares

POWER ANALYSIS AND EFFECT SIZE

- Many published nursing studies (and even more unpublished ones) have nonsignificant findings, and many of these could reflect Type II errors. As indicated earlier, researchers set the probability of committing a Type I error (a false positive) as the significance level, alpha (a).
- The probability of a Type II error (a false negative) is beta (B).
- The complement of beta (1 - B) is the probability of detecting a true relationship or group difference and is the power of a statistical test.
- Found that many published nurs ing studies have insufficient power, placing them at risk for Type II errors-although a more recent study has found that, on average, power has im proved in nursing studies, perhaps because greater attention has been paid to this topic
- Nevertheless, even in the more recent analysis, many studies continued to be underpowered.

Power analysis is used to reduce the risk of **Type II errors** and strengthen statistical conclusion validity by **estimating in advance** how big a sample is needed.

- Researchers typically use power analysis at the outset of a study to estimate the sample size needed to avoid a Type II error.
- To estimate needed sample size (N), researchers must specify a, ES, and 1- R Researchers usually establish the risk of a Type 1 error (a) as .05.
- The conventional standard for 1 - B is .80.
- With power equal to .80, there is a 20% risk of committing a Type II error.
- Although this risk may seem high, a stricter criterion requires sample sizes larger than many researchers could afford.
- With a and 1 - B specified, the information needed to solve for N is ES, the estimated population effect size.

- The effect size is the magnitude of the relationship between the research variables.
- When relationships (effects) are strong, they can be detected at significant levels even with small samples. With modest relationships, large sample.
- sizes are needed to avoid Type II errors.
- In using power analysis to estimate sample size needs, the population effect size is not known:
- if it were known, there would be no need for the new study.
- Effect size must be estimated using available evidence and theory.
- In essence, the effect size esti mate in a power analysis represents the researcher's hypothesis about how strong relationships are.
- More often an effect size is calculated based on findings from earlier studies on a similar problem.
- When there are no relevant carlier findings and when theory offers only broad guidance, researchers use conventions hused on expectations of a small, medium, or large effect. Most nursing studies have modest (small-to- medium) effects.
- Procedures for estimating effects and sample size needs vary from one statistical situation to another. We focus mainly on a two-group situation for which we can estimate mean values.

Sample Size Estimates for Testing Differences between Two Means

QUESTION : Suppose we were testing the hypothesis that cran berry juice reduces the urinary pH of diet-controlled patients. We plan to assign some patients randomly to a control condition (no cranberry juice) and others to an experimental condition in which they will be given 300 mL of cranberry juice for 5 days. How large a sample is needed for this study, given a desired a of .05 and power of .80?

To answer this, we must first estimate ES. In a two-group situation in which mean differences are of interest, ES is usually designated as Cohen's d. the formula for which is

d = u1_u2/ SD

That is, the effect size (d) is the difference be tween the two population means, divided by the population standard deviation . These popula tion values are never known but must be estimated. For example, suppose we found an earlier nonex perimental study that compared the urinary pH of people who had or had not ingested cranberry juice.in the previous 24 hours. The earlier and planned studies are different in many respects, but the ear lier study is a reasonable starting point. Suppose the results were as follows:

X(no cranberry juice) = 5.70 X2(cranberry juice) = 5.50

SD = .50

Thus, the estimated value of d would be 40:

d 5.70 - 5.50 /0.50=.40

Sample Size Estimates for Other Bivariate Tests

Power analysis can be undertaken for the other statistical tests described in this chapter. It is rela tively easy to do a power analysis online (we sug gest several relevant websites in the Toolkit with the Resource Manual), Here we discuss only a few basic features for situations in which ANOVA, Pearson's r. or a chi-square situation would be the basis for doing the power analysis, There are alternative approaches to doing a power analysis in an ANOVA context.

- Estimating sample size requirements for test- ing differences in proportions between groups is complex.
- The effect size for crosstabs tables is in fluenced not only by expected differences in pro- th portions (e.g., 60% in one group versus 40% in another, a 20% point difference) but also by the absolute values of the proportions. Effect sizes are larger (and thus sample size needs are smaller) at the extremes than near the midpoint.

- A 20% point difference is easier to detect if the percentages are 10% and 30% than if they are near the middle, such as 60% and 40%. Because of this fact, it is difficult to offer information on values for small, medium, and large effects in this context.
- We can, however, give examples of differences in proportions that conform to the conventions in a 2 x 2 situation:

Small: .05 versus 10, 20 versus .29, 40 versus

50, .60 versus .70, .80 versus.87

Medium: .05 versus .21, .20 versus .43, .40 versus

.65, .60 versus .82, .80 versus .96

Large: .05 versus .34, .20 versus .58. 78, .60 versus .92.80 versus .9640 versus

As an example, if the expected proportion for a control group were 40, the researcher would need about 385, 70, and 24 per group if higher values were expected for the experimental group and the effect was expected to be small, medium, and large, respectively. As in other situations, researchers are encouraged to avoid using the conventions in favor of more precise estimates based on existing evidence. If the conventions cannot be avoided, conservative estimates should be used to minimize the risk of obtaining nonsignificant results.

Effect Size Calculations in Completed Studies

- Power analysis concepts are sometimes used after analyses are completed to calculate estimated pop ulation effects based on actual Ns.
- In this situation, power, alpha, and N are known, and so the task is to solve for ES. Effect sizes provide readers and clinicians with estimates about the magnitude of effects an important issue in EBP .
- Effect size information can be crucial because, with large samples, even tiny effects can be sta tistically significant.

- P values tell you whether results are likely to be real, but effect sizes can suggest whether they are important.
- Effect size estimates are needed in doing meta-analyses and so when these values are pre sented directly in a report, they are helpful to meta analysts.

Power Is Effected By

- Variation in the outcome ($\sigma 2$) $\downarrow \sigma 2 \rightarrow$ power $\uparrow$
- Significance level (α) $\uparrow \alpha \rightarrow$ power $\uparrow$
- Difference (effect) to be detected (δ) $\uparrow \delta \rightarrow$ power $\uparrow$
- One-tailed vs. two-tailed tests
- Power is greater in one-tailed tests than in comparable two-tailed tests

Power Changes

$2n = 32$, 2 sample test, 81% power, $\delta=2$, $\sigma = 2$, $\alpha = 0.05$, 2-sided test

Variance/Standard deviation

σ: 2 → 1 Power: 81% → 99.99% σ: 2 → 3 Power: 81% → 47%

Significance level

(α) α : 0.05 → 0.01 Power: 81% → 69% α : 0.05 → 0.10 Power: 81% → 94%

Power Changes $2n = 32$, 2 sample test, 81% power, $\delta=2$, $\sigma = 2$, $\alpha = 0.05$, 2-sided test

Difference to be detected (δ)

δ : 2 → 1 Power: 81% → 29% δ : 2 → 3 Power: 81% → 99%

Sample size (n) n: 32 → 64 Power: 81% → 98% n: 32 → 28 Power: 81% → 75%

One-tailed vs. two-tailed tests Power: 81% → 88%

Power Formula

Depends on study design Not hard, but can be VERY algebra intensive May want to use a computer program or statistician

How Big a Sample We Need?

- Fundamental research question Should be addressed after determining the primary objective and study design .
- Too Few Patients in a clinical study – May fail to detect a clinically important difference Too Many – Involve extra patients – Therapy may have risks – Cost more
- Fundamentalresearch question How Big? 18, 180 ,1,800, 18,000, 180,000

Sample Size Formula Information

- Variables of interest
- type of data e.g. continuous, categorical
- Desired power
- Desired significance level
- Effect/difference of clinical importance
- Standard deviations of continuous outcome variables
- One or two-sided tests

Sample Size & Study Design

- Randomized controlled trial (RCT)
- Block/stratified-block randomized trial
- Equivalence trial

- Non-randomized intervention study
- Observational study Prevalence study
- Measuring sensitivity and specificity

Sample Size & Data Structure

- Paired data
- Repeated measures
- Groups of equal sizes
- Hierarchical data

Sample Size

- Non-randomized studies looking for differences or associations require larger sample to allow adjustment for confounding factors
- Absolute sample size is of interest surveys sometimes take % of population approach
- Study's primary outcome is the variable you do the sample size calculation for If secondary outcome variables considered important make sure sample size is sufficient
- Increase the 'real' sample size to reflect loss to follow up, expected response rate, lack of compliance, etc. Make the link between the calculation and increase

Steps

Step 1. Define Primary Objective

To see if feeding milk to 5 year old kids enhances growth.

Step 2. Study Design

Extra Milk Diet

5 yr olds 

Normal Milk Diet

Outcome: height (cm) Step

3. Define clinically significant difference one wishes to detect

Difference (Δ) of 0.5 cm

Step 4. Define degree of certainty of finding this difference

beta (β) or type II error : The probability of NOT detecting a significant difference when there really is one.

Risk of a false-negative finding ie Risk of declaring no significant difference in height between the milk diets when a difference really does exist.

Set at ≤ 20%

Power of the Test: Probability of detecting a predefined clinically significant difference. Power = (1- β) = 1 -20% = 80%

Step 5. Define significance level

Alpha (α) or type I error: The probability of detecting a significant difference when the treatments are really equally effective Risk of a false-positive finding Set at 5% : One has a 5% chance or 1 in 20 odds of declaring a significant difference between the milk diets when in fact they are really equal. We are willing to accept that 1 time out of 20 we will produce a false positive finding

For the Milk Study

- Type I error (α) = 0.05
- Type II error (β) = 0.20
- Power = (1-) = 0.80
- Clinically significant diff (Δ) = 0.5cm

- Measure of variation (SD) = 2.0 cm – Exists in literature or “Guesstimate” Formula Beta

N = 2(SD)2 x f(α, β) /Δ2

f(α Alpha 0.05 0.10 0.20 0.50 Δ2 0.10 10.8 8.6 6.2 2.7 = 2(2)2 x 7.9 / 0.52 0.05 13.0 10.5 7.9 3.8 = 252.8 (each group) 0.02 15.8 13.0 10.0 5.4 0.01 17.8 14.9 11.7 6.6

Simple Method

- Nomogram
- Standardized difference

= smallest medically relevant diff /estimated standard deviation

= 0.5/2.0 = 0.25

Assumptions:

1. 2 sample comparison only

2. Same number of subjects per group

3. Variable is a continuous measure that is normally distributed 500

USE OF READYMADE TABLES FOR SAMPLE SIZE CALCULATION

 How large a sample of patients should be followed up if an investigator wishes to estimate the incidence rate of a disease to within 10% of it’s true value with 95% confidence?

 The table show that for e=0.10 & confidence level of 95%, a sample size of 385 would be needed.

 This table can be used to calculate the sample size making the desired changes in the relative precision & confidence level .e.g if the level of confidence is reduce to 90%, then the sample size would be 271.

 Such table that give ready made sample sizes are available for different designs & situation.

USE OF COMPUTER SOFTWARE FOR SAMPLE SIZE CALCULATION

The following softwares can be used for calculating sample size .

- Epi-info (epiinfo.codeplex.com)
- nQuerry (nquery.codeplex.com)
- STATA (www.stata.com)
- SPSS (www.spss.co.in)

CONCLUSIONS

- Sample size determination is one of the most essential components of every research Study.
- The larger the sample size, the higher will be the degree of accuracy, but this is limited by the availability of resources.
- It can be determined using formulae, readymade tables and computer softwares.