

Sample Size Determination

Obsan R. (BSc, MPH/RH)

Nov, 2022

- **An essential part of planning any study is to decide how many people need to be studied**

Sample Size

- ***Sample Size:*** The number of study subjects selected to represent a given study population.
- Important to make inferences based on the findings from the sample.
- Should be sufficient to represent the characteristics of interest of the study population.

- In estimating a certain characteristic of a population, sample size calculations are important to ensure that **estimates** are obtained with required precision or confidence
- The accuracy of the envisaged results determine the size of the sample

Example

- A prevalence of 10% from a sample size of 20
 - would have a 95% CI of 1% to 31%,
 - which is not very precise or informative.
- But, a prevalence of 10% from a sample of size 400
 - would have a 95% CI of 7% to 13%,
 - which may be considered sufficiently accurate.

- In studies concerned with **detecting an effect** (e.g. a difference between two groups), sample size calculations are important to ensure the detection of whether association exists or not.
- Large sample size can result in statistical significance when Δ is very small (not of practical, clinical or public health importance).

- Small sample size can result in a non statistically significant finding even when Δ is large (of practical, clinical or public health importance).
- If the sample is too small, then even if large differences are observed, it will be impossible to show that these are due to anything more than sampling variation.

Sample size determination depends on the:

- Objective of the study
- Design of the study
 - Descriptive/Analytic
- Accuracy of the measurements to be made
- Degree of precision required for generalization
- Plan for statistical analysis
- Degree of confidence with which to conclude

- **Common questions:**

- ***“How many subjects should I study?”***

- **Too small sample = Waste of time and resources**

- = Results have no practical use**

- **Too large sample = Waste of resources**

- = Data quality compromised**

When deciding on sample size:



Sample size $\uparrow$ = $\uparrow$ Precision $\neq$ Cost

- **The feasible sample size is also determined by the availability of resources:**

- time**
- manpower**
- transport**
- available facility, and**
- money**

1. Sample Size: Single Sample

- The aim is to have a large enough sample with which to estimate a population mean or proportion within a narrow interval with high reliability.
- Concerned with the precision of the estimate (“narrowness of the CI”).



estimate $\pm$ d units

Sample size for single sample includes:

- A. Sample size for estimating a single population mean
- B. Sample size to estimate a single population proportion

B. Sample size to estimate a single population proportion

- **Aim:** Estimate p
- **Want:** Estimate $\pm d$ units where $d = Z \cdot SE$

(95% CI of width= $2d$)

Steps:

1. Specify d (or $w = 2d$)
2. Use estimated p (use $p=0.5$ if no information)

3. Solve for n

$$d = z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

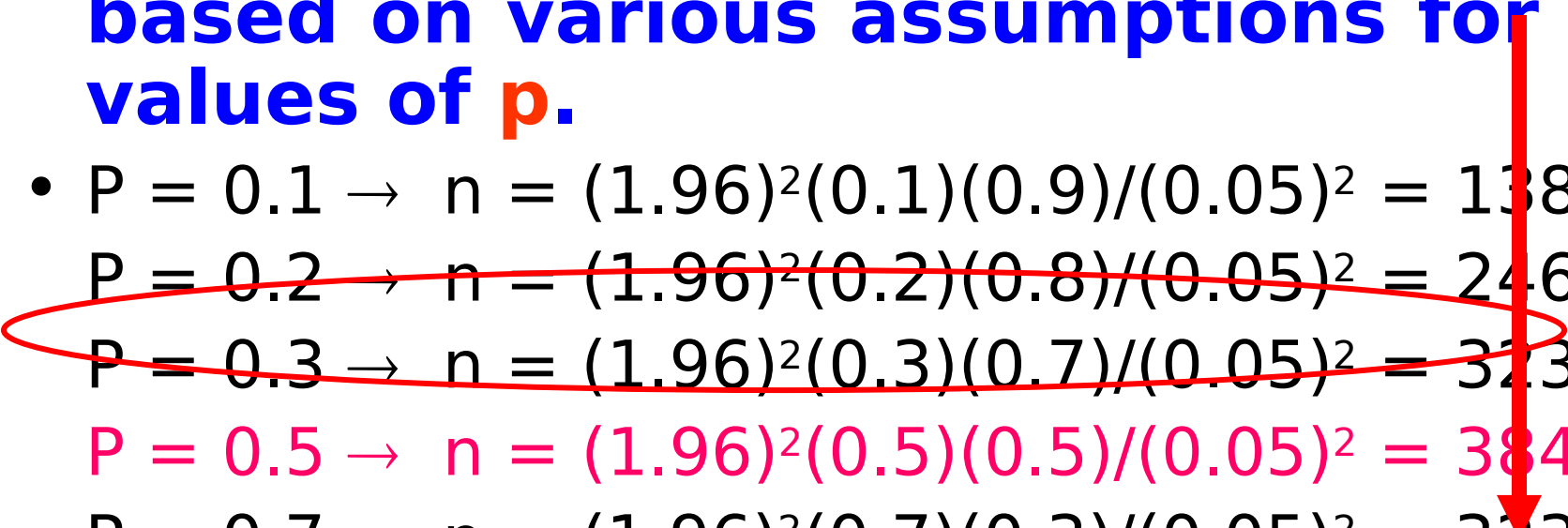
$$d^2 = (z_{\alpha/2})^2 \cdot \frac{pq}{n}$$

$$n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2}$$

- 1. Suppose that you are interested to know the proportion of infants who breastfed >18 months of age in a rural area. Suppose that in a similar area, the proportion (p) of breastfed infants was found to be 0.20. What sample size is required to estimate the true proportion within $\pm 3\%$ points with 95% confidence? Let $p=0.20$, $d=0.03$, $\alpha=5\%$**

$$n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2} = \frac{(1.96)^2 (0.2)(0.8)}{(0.03)^2} = 683$$

- Suppose there is no prior information about the proportion (**p**) who breastfeed
- Assume **p=q=0.5** (most conservative)
- **T**
i $n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2} = \frac{(1.96)^2 (0.5)(0.5)}{(0.03)^2} = 1,068$ **e**

- An estimate of **p** is not always available.
 - However, the formula may also be used for sample size calculation based on various assumptions for the values of **p**.
 - $P = 0.1 \rightarrow n = (1.96)^2(0.1)(0.9)/(0.05)^2 = 138$
 - $P = 0.2 \rightarrow n = (1.96)^2(0.2)(0.8)/(0.05)^2 = 246$
 - $P = 0.3 \rightarrow n = (1.96)^2(0.3)(0.7)/(0.05)^2 = 323$
 - $P = 0.5 \rightarrow n = (1.96)^2(0.5)(0.5)/(0.05)^2 = 384$
 - $P = 0.7 \rightarrow n = (1.96)^2(0.7)(0.3)/(0.05)^2 = 323$
 - $P = 0.8 \rightarrow n = (1.96)^2(0.8)(0.2)/(0.05)^2 = 246$
- 

- For a fixed absolute precision (d), the required sample size increases as P increases from 0 to 0.5, and then decreases in the same way as the prevalence approaches 1.

Summary

- Sample size calculations depend on a number of assumptions:
 - the hypothesized difference of interest, Δ
 - the probability of Type I error (α)
 - the probability of Type II error (β)
 - the variance
- Choice of sample size depends on a balance of reasonable assumptions, time, effort, and expense
- Sample sizes provide a minimum estimate of the desired sample sizes for the study