ESTIMATION OF SAMPLE SIZE

UNIT-V

"If you haven't adequate sample size then you don't know what you are going to do an experiment"

Learning Objectives

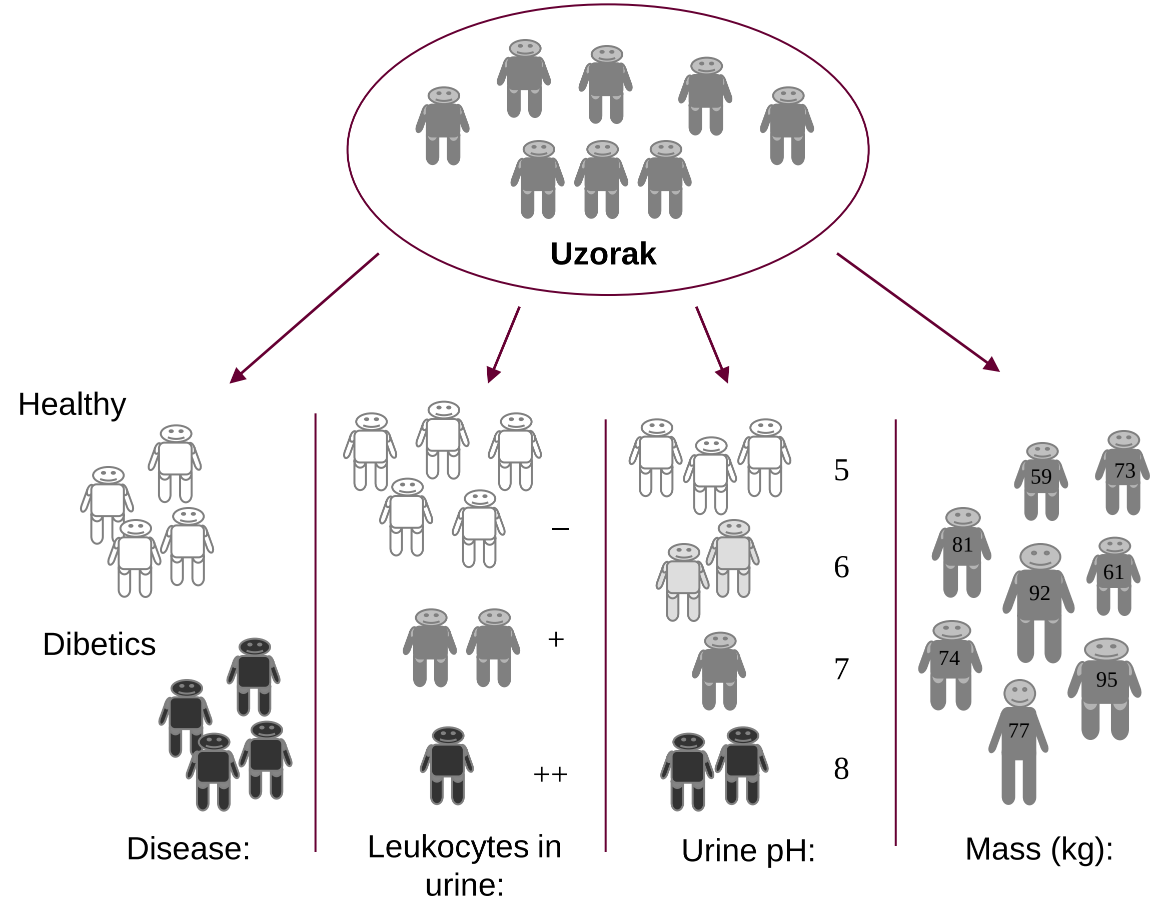
After completing this UNIT, you should be able to

- Importance of sample size
- Estimate the adequate sample size by using standard deviation(s)
- Estimate the adequate sample size by using proportion(s)

**** Applications***

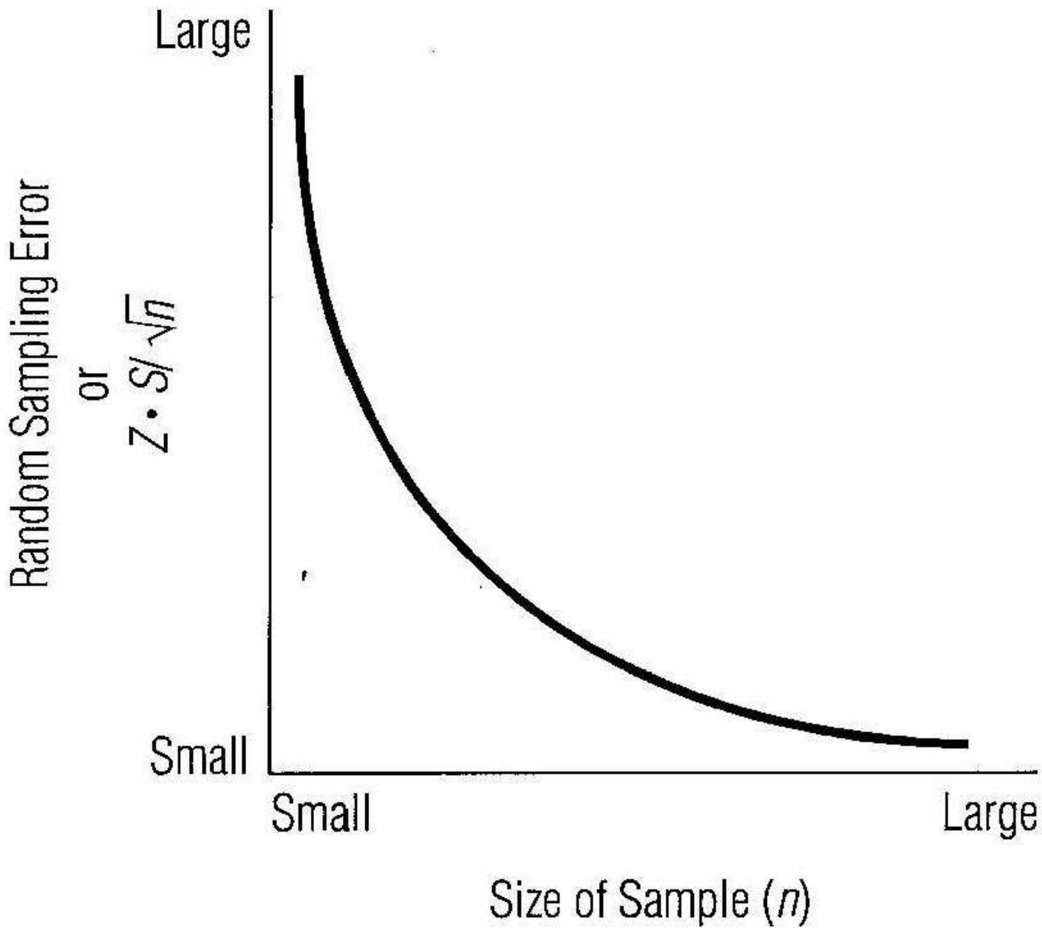

How to
Calculate
Sample Size

Scales of measurements



Nominal scale provides categories (male, female)	**Ordinal scale** Hierarchy (stages, “crosses”)	**Interval scale** difference in the number of units	**Ratio scale** absolute parameter values

Relationship between sample size and error



Am I going to reach my objective?

- I have 4 months to finish my research project, of which only one week is for data collection
- I think I can get data on 50 subjects in a week
- Is 50 a sufficient number of subjects to test my hypothesis with the significance level I want?

Why to calculate sample size?

- To show that under certain conditions, the hypothesis test has a good chance of showing a desired difference (if it exists)
- To show to that the study has a reasonable chance to obtain a conclusive result
- To show that the necessary resources (human, monetary, time) will be minimized and well utilized

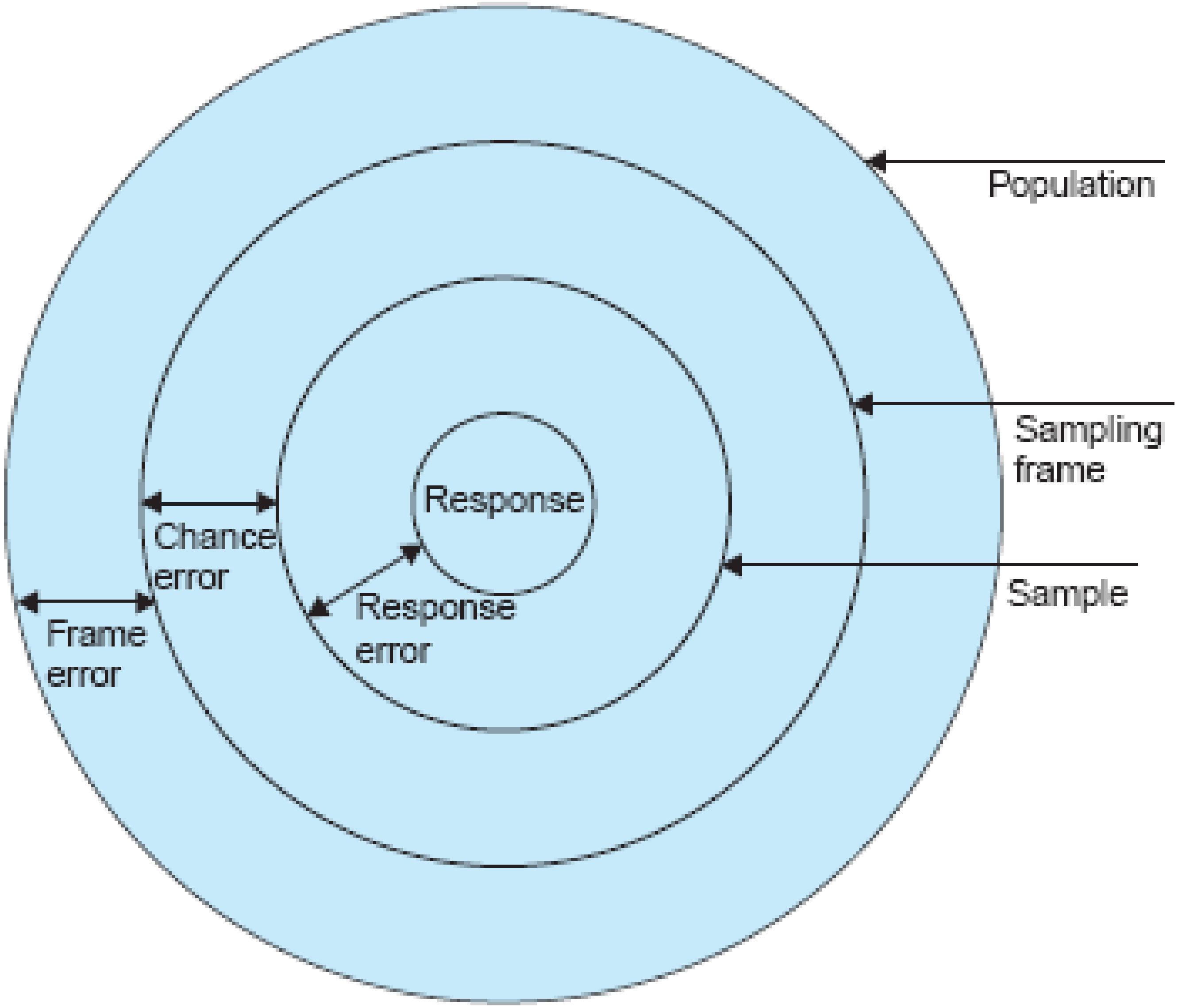
Sample Size

Too Big:

•Requires too much resources

Too Small:

•Won't do the job



Sampling error = Frame error
+ chance error + response error.
(If we add measurement error or the non-sampling error to sampling error, we get total error)

What do I need to know to calculate sample size?

- Most Important: sample size calculation is an educated guess
- It is more appropriate for studies involving hypothesis testing
- There is no magic involved; only statistical and mathematical logic and some algebra
- Researchers need to know something about what they are measuring and how it varies in the population of interest

Sample Size Calculations

- Formulate a PRIMARY question or hypothesis to test (or determine what you are estimating). Write down H_0 and H_1.
- Determine the endpoint. Choose an outcome measure. How do we "measure" or "quantify" the responses?

Where do we get this knowledge?

- Previous published studies
- Pilot studies
- If information is lacking, there is no good way to calculate the sample size!

Errors in sample

- **Systematic error (or bias)**

  Inaccurate response (information bias)

  Selection bias

- **Sampling error (random error)**

Type 1 error

- The probability of finding a difference when compared our sample with population, and in reality there is no difference

- Known as the α (or "type 1 error")

- Usually set at 5% (or 0.05)

Type 2 error

- The probability of not finding a difference that actually exists between two groups (or between sample and population).

- Known as the β (or "type 2 error")

- Power is (1- β) and is usually 80%

Diagnosis and statistical reasoning

	Disease status	
	Present	Absent
Test result		
+ve	True +ve (sensitivity)	False +ve
-ve	False –ve	True -ve (Specificity)

	Significance Difference is	
	Present (Ho *not* true)	Absent (Ho is true)
Test result		
Reject Ho	No error $1-\beta$	Type I err. α
Accept Ho	Type II err. β	No error $1-\alpha$

α : significance level

$1-\beta$: power

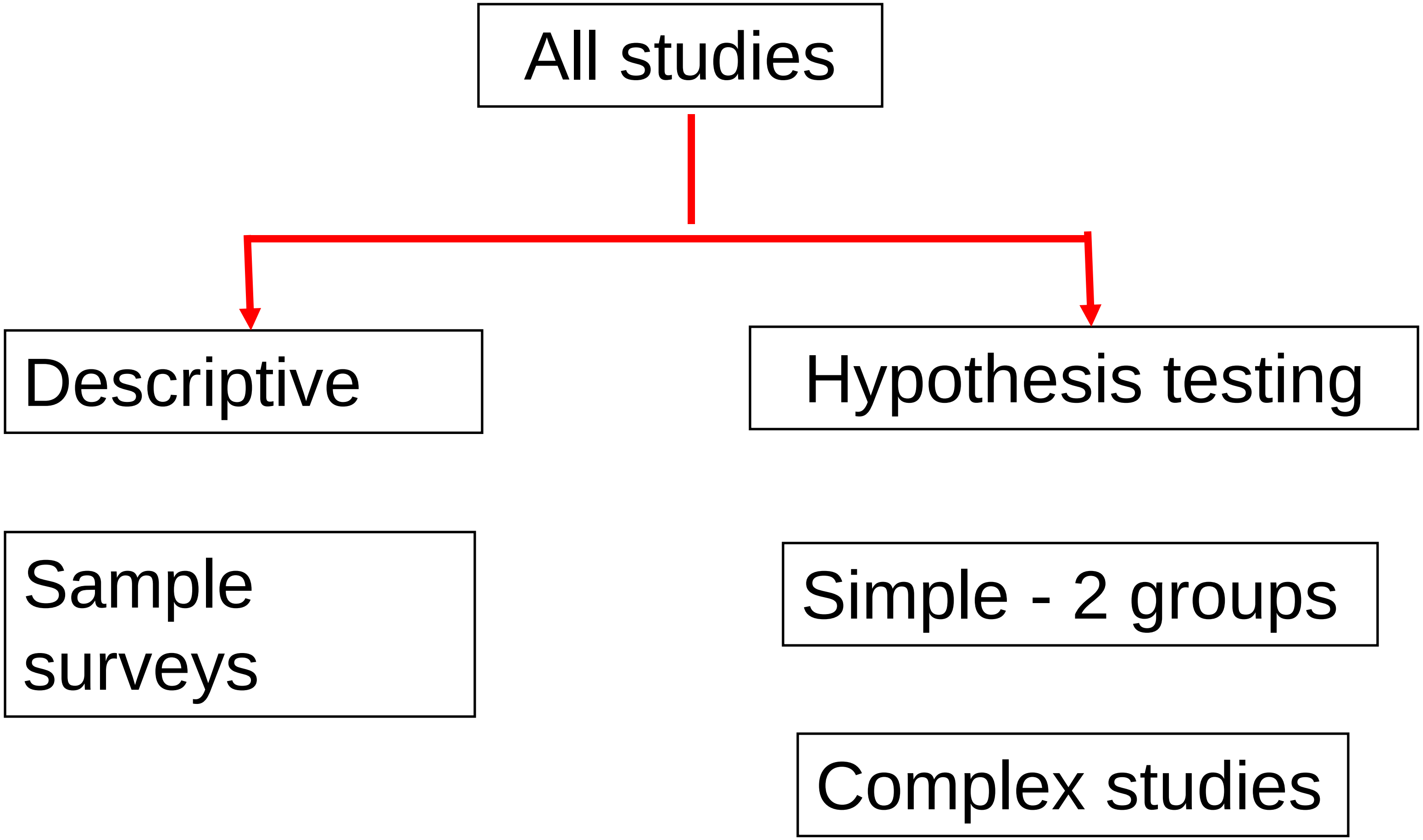
Estimation of Sample Size by Three ways:

By using

(1) Formulae (manual calculations)

(2) Sample size tables

(3) Software's


All studies
Descriptive
Hypothesis testing
Sample surveys
Simple - 2 groups
Complex studies

Sample Size for Adequate Precision

- In a descriptive study,

Summary statistics (mean, proportion)

- Reliability (or) precision

By giving "confidence interval"

Wider the C.I – sample statistic is not reliable and it may not give an accurate estimate of the true value of the population parameter

Sample size formulae: One sample

For single mean : $n = Z^2_{\alpha} S^2 / d^2$

where S=sd (s)

For a single proportion : $n = Z^2_{\alpha} P(1-P)/d^2$

Where , $Z\alpha$ =1.96 for 95% confidence level

$Z\alpha$ = 2.58 for 99% confidence level

Sample Size Formula

$$n = \left(\frac{zs}{d} \right)^2$$

where:

n = sample size

z = confidence interval in standard error units

s = standard error of the mean

d= acceptable magnitude of error

Application-I

- How close to the true mean
- Confidence around the sample mean
- Type I error.

- $n = (Z_{\alpha/2})^2 \sigma^2 / d^2$

  σ: standard deviation

  d: the accuracy of estimate (how close to the true mean).

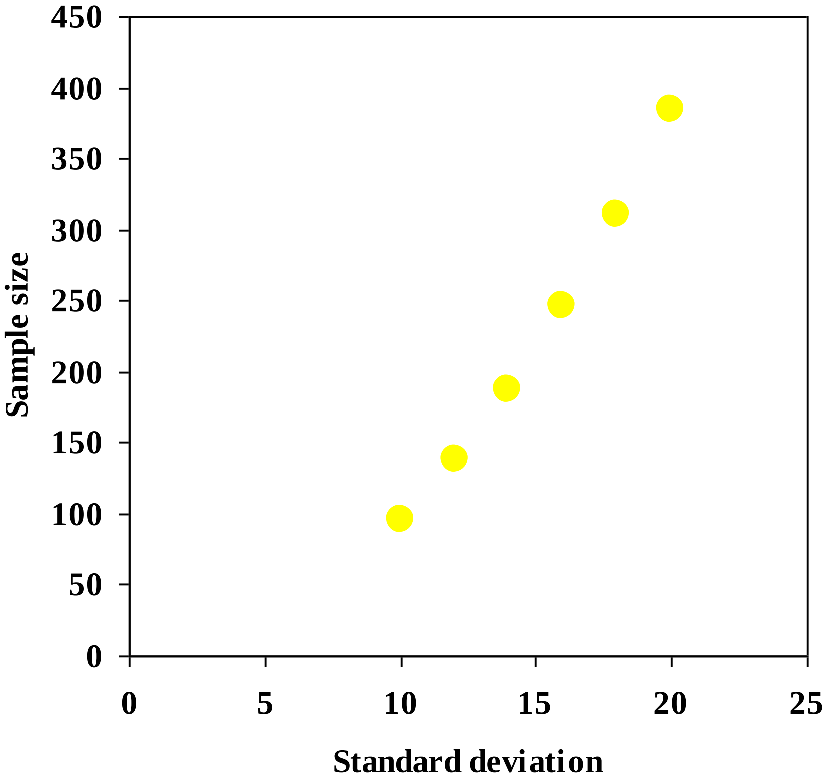
  $Z_{\alpha/2}$: A Normal deviate reflects the type I error.

- **Example:** we want to estimate the average weight in a population, and we want the error of estimation to be less than 2 kg of the true mean, with a probability of 95% (e.g., error rate of 5%).

- $n = (1.96)^2 \sigma^2 / 2^2$

Solution

Std Dev (σ)	Sample size
10	96
12	138
14	188
16	246
18	311
20	384



Application-II

Suppose a survey researcher, studying expenditures on lipstick, wishes to have a 95% confident level (Z) and a range of error (d) of less than $2.00. The estimate of the standard deviation is $29.00.

Solution

$$n = \left(\frac{zs}{d}\right)^2 = \left[\frac{(1.96)(29.00)}{2.00}\right]^2$$

$$= \left[\frac{56.84}{2.00}\right]^2 = (28.42)^2 = 808$$

Application-III

Suppose, in the same example as the one before, the range of error (d) is acceptable at $4.00. By how much is sample size is reduced?

Solution

$$n = \left(\frac{zs}{d}\right)^2 = \left[\frac{(1.96)(29.00)}{4.00}\right]^2$$

$$= \left[\frac{56.84}{4.00}\right]^2 = (14.21)^2 = 202$$

Calculating Sample Size @ 99% CIs

$$n = \left[\frac{(2.57)(29)}{2}\right]^2$$

$$= \left[\frac{74.53}{2}\right]^2$$

$$= [37.265]^2$$

$$= 1389$$

$$n = \left[\frac{(2.57)(29)}{4}\right]^2$$

$$= \left[\frac{74.53}{4}\right]^2$$

$$= [18.6325]^2$$

$$= 347$$

Application-IV

A study is to be performed to determine a certain parameter in a community. From a previous study a SD of 46 was obtained. If a sample error of up to 4 is to be accepted. How many subjects should be included in this study at 99% level of confidence?

Sample Size for a Proportion

$$n = \frac{Z^2 pq}{E^2}$$

Where:

n = number of items in samples

Z^2 = square of confidence interval in standard error units

p = estimated proportion of success

q = (1-p) or estimated the proportion of failures

E^2 = square of maximum allowance for error between true proportion and sample proportion, or zs_p squared.

Application-I

- How close to the true proportion
- Confidence around the sample proportion.
- Type I error.

- $n = (Z_{\alpha/2})^2 \, p(1-p) \, / \, d^2$

  p: proportion to be estimated.

  d: the accuracy of estimate (how close to the true proportion).

  $Z_{\alpha/2}$: A Normal deviate reflects the type I error.

- **Example:** The proportion of preference for male child is around 80%. We want to estimate the preference p in a community within 5% with 95% confidence interval.

- $N = (1.96)^2 \, (0.8)(0.2) \, / \, 0.05^2$

  $= 246$ married women.

Application-II

It was desired to estimate proportion of anemic children in a certain preparatory school. In a similar study at another school a proportion of 30 % was detected.

Compute the minimal sample size required at a confidence limit of 95% and accepting a difference of up to 4% of the true population.

Solution

$n = (Z_{a/2})^2\, p(1-p) \;/\; d^2$

p: proportion to be estimated = 30% (0.30)

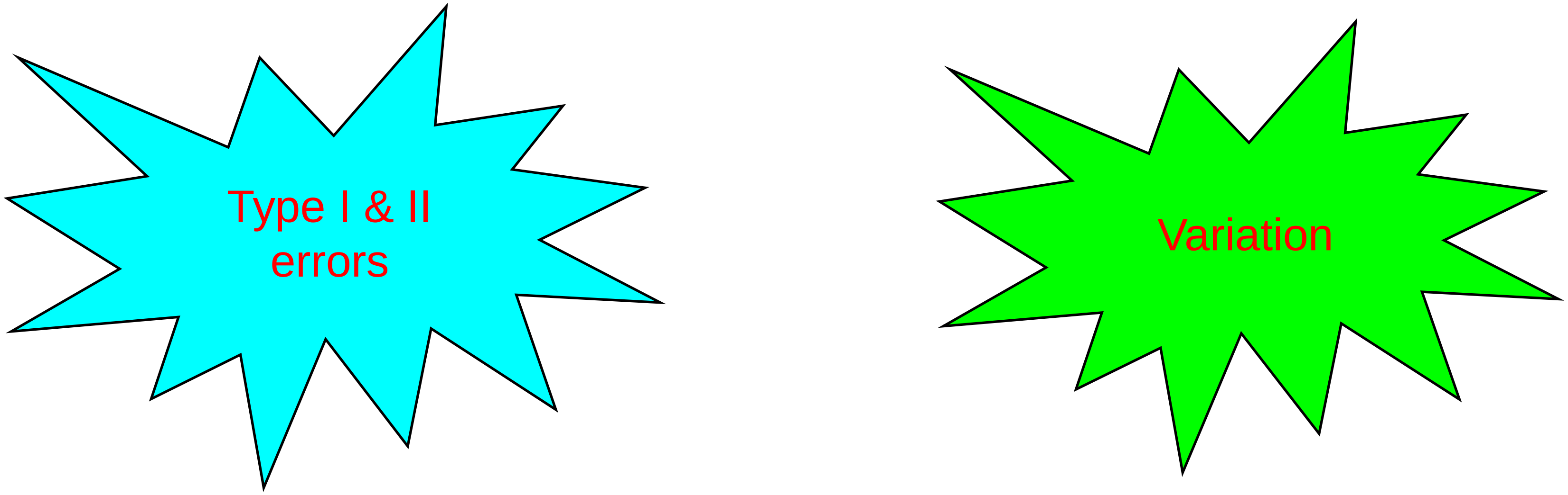
d: the accuracy of estimate (how close to the true proportion) = 4% (0.04)

$Z_{a/2}$: A Normal deviate reflects the type I error

For 95% the critical value =1.96

$$n = \frac{1.96^2 \text{ x } 0.3(1-0.3)}{(0.04)^2} = 504.21 \sim 505$$

Sample Size for two-samples



Sample size formulae: Two sample(s)

For two means :

$$N = n_1 + n_2 = \frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(d = \mu_1 - \mu_2)^2}$$

For two proportions :

$$N = n_1 + n_2 = \frac{4(z_{1-\alpha/2} + z_{1-\beta})^2 \left[\left(\frac{P_1 + P_2}{2} \right) \left(1 - \frac{P_1 + P_2}{2} \right) \right]}{\left(d = P_1 - P_2 \right)^2}$$

$Z\alpha = 1.96$ for 95% confidence level

$Z\alpha = 2.58$ for 99% confidence level ;

$Z_\beta = 0.842$ for 80% power

$Z_\beta = 1.282$ for 90% power

Two Sample test

For the hypothesis: H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$

For a two tailed t-test, the formula is:

$$N = n_1 + n_2 = \frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(d = \mu_1 - \mu_2)^2}$$

Application-I

How large a sample would be needed for comparing two approaches to cholesterol lowering using $\alpha=0.05$, to detect a difference of $d = 20$ mg/dl or more with Power $= 1- \beta = 0.90$ and $\sigma=30$ mg/dl.

Solution

When σ = 30 mg/dl, β = 0.10, α = 0.05; $z_{1-\alpha/2}$ = 1.96

Power = 1- β ; $z_{1-\beta}$ = 1.282 , d = 20mg/dl

$$N = n_1 + n_2 = \frac{4(30)^2(1.96+1.282)^2}{(20)^2}$$

$$= \frac{4 \times 900 \times (3.242)^2}{400} = \frac{37{,}838.03}{400}$$

$$N = 94.6$$

Hence about 50 for each group (rounded value)

Application-II

A study to see the effect of iron fortification for a specific period on haemoglobin level among pregnant women of a community has to be carried. Under a preliminary study or as reported in the literature, it was observed that the SD of Hb in pregnant women from similar community with a intervention was 3.4 g/dl. If α=0.05 and β=0.2, how many pregnant women from each group should be studied if one wants to detect a difference of 2.0 g/dl in Hb in the two groups?

Application-III

Two machines are used for filling plastic bottles with a net volume of 16.0 ounces. The fill volume can be assumed normal, with standard deviation is 0.02 ounces. A member of the quality engineering staff suspects that both machines fill to the same mean net volume, whether or not this volume is 16.0 ounces. what sample size should be used to assure that $\beta=0.1$ if the true difference in means is 0.04? Assume that $\alpha=0.05$.

Two-sample proportions

$H_0 : P_1 = P_2$ vs. $H_1 : P_1 \neq P_2$

$$N = n_1 + n_2 = \frac{4(z_{1-\alpha/2} + z_{1-\beta})^2 \left[\left(\frac{P_1 + P_2}{2} \right) \left(1 - \frac{P_1 + P_2}{2} \right) \right]}{\left(d = P_1 - P_2 \right)^2}$$

Application-I

Children attending public schools have expected dental problem among them as 75%. In another group of children attending govt. schools, 50% are expected to have this problem. How many children should be included from each group to determine whether this difference is significant at $\alpha=0.05$ if we wish to have $\beta=0.1$ chance of detecting the difference if it is real?

Solution

Given, P_1=0.7, P_2=0.5, β = 0.10, α = 0.05; $z_{1-\alpha/2}$ = 1.96

Power = 1- β ; $z_{1-\beta}$ = 1.282 ,

$$(P_1+P_2)/2 = (0.7+0.5)/2 = 0.6$$

$$N = (n_1 + n_2) = \frac{4\left(1.96+1.282\right)^2[(0.6)(1-0.6)]}{(0.2)^2}$$

$$= \frac{4(3.242)^2[(0.6)(0.4)]}{(0.2)^2} = \frac{10.09}{0.04} = 252.25$$

$$N = 252.25$$

Consider using N = 260, or 130 per group

Application-II

The efficacy of ‘treatment A ‘ is expected to be 70%, and for ‘treatment B’ to be 60%. A study is planned to show the difference at the significance level of 1% and power of 80%. Find the adequate sample size is required each group.

“If you haven't observed it you don't know what you are talking about.”

THANK YOU…