Sample size in Descriptive and Cross-sectional Studies

Sample size formula for estimating Population Mean

$$n = \frac{z^2 \sigma^2}{d^2}$$

If the population standard deviation is **known**

If the population standard deviation is **unknown**

$$n = \frac{z^2 s^2}{d^2}$$

Sample size formula for estimating Population Mean

n=Sample size

S=Standard Deviation

d=Maximum acceptable error in estimation of mean

Z=Confidence level

$$n = \frac{z^2 s^2}{d^2}$$

Some popular values for Confidence level

Confidence Level	Error	Z	Z^2
0.90	0.10	1.64	2.69
0.95	0.05	1.96	3.84
0.99	0.01	2.57	6.60

Example

- In estimating the mean of fasting blood sugar (FBS) level of elderlies in a region, we want to be 95% confident that the difference between estimated and actual mean of FBS to be not more than 3 mg/dL.
- Suppose that in a similar study, the standard deviation of FBS has been 20 mg/dL.
- Please calculate the sample size.

Example

- In estimating the mean of FBS level of elderlies in a region, we want to be 95% confident that the difference between estimated and actual mean of FBS to be not more than 3 mg/dL.
- Suppose that in a similar study, the standard deviation of FBS has been 20 mg/dL.
- Please calculate the sample size.

$$n = \frac{z^2 s^2}{d^2} \quad n = \frac{1.96^2 (20)^2}{3^2} = \frac{3.84 \times 400}{9} = 171$$

Exercise

A researcher likes to estimate the average of Calorie intake of school children. In a similar study The standard deviation of Calorie intake has been reported 100.

Please calculate required sample size for the following situations:

a) 90% confidence and error of 20 Calories
b) 95% confidence and error of 30 Calories
c) 99% confidence and error of 20 Calories

Sample size formula for estimating Population Proportion

$$n = \frac{z^2 p(1-p)}{d^2}$$

Sample size formula for estimating Population Proportion

n=Sample size

p=initial estimate for population proportion

d=maximum acceptable error in estimating population proportion

Z=Confidence level

$$n = \frac{z^2 p(1-p)}{d^2}$$

Some popular values for Confidence level

Confidence Level	Error	Z	Z^2
0.90	0.10	1.64	2.69
0.95	0.05	1.96	3.84
0.99	0.01	2.57	6.60

Example

- We like to estimate the prevalence of metabolic syndrome (Mets) in a region. Previous studies has shown a prevalence of 20%.
- How many sample we need such that by 95% confidence, our error in estimation does not exceed 2%.

Example

- We like to estimate the prevalence of metabolic syndrome (Mets) in a region. Previous studies has shown a prevalence of 20%.
- How many sample we need such that by 95% confidence, our error in estimation does not exceed 2%.

$$n = \frac{1.96^2 \times 0.2(1-0.2)}{0.02^2} = \frac{3.84 \times 0.16}{0.0004} = 1536$$

Example: effect of doubling error rate

- We like to estimate the prevalence of metabolic syndrome (Mets) in a region. Previous studies has shown a prevalence of 20%.
- How many sample we need such that by 95% confidence, our error in estimation does not exceed 4%.

$$n = \frac{1.96^2 \times 0.2(1-0.2)}{0.04^2} = \frac{3.84 \times 0.16}{0.0016} = 384 = \frac{1536}{4}$$

9/9/2024 Sample Size ١٣

Exercises

9/9/2024 Sample Size ١٤

Exercise 1:

- Consider the data shown below. How many subjects would be needed to ensure that a 95% confidence interval estimate of BMI had a margin of error not exceeding 2 units?
- 25 27 31 33 26 28 38 41 24 32 35 40

Exercise 2:

- For multiple sclerosis patients, we wish to estimate the mean age at which the disease was first diagnosed. We want a 90% confidence interval that is 8 years wide.
- If the population variance is estimated to be 85 from previous research, how large of a sample should be taken?

Exercise 3:

- An investigator wants to estimate caffeine consumption in high school students. How many students would be required to ensure that a 95% confidence interval estimate for the mean caffeine intake (measured in mg) is within 15 units of the true mean? Assume that the standard deviation in caffeine intake is 68 mg.

Exercise 4:

- A survey is being planned to estimate, with a 98% confidence interval, the mean amount of time per week that seniors spend watching television. The population standard deviation is known to be 5.2 hours. What sample size will be required to obtain an estimate that is within 0.5 hours of the population mean?

Exercise 5:

- Which of the following statements are true?
  - (a) As the confidence level decreases (from 95% to 90%, say), the width of the interval increases.
  - (b) As the confidence level increases, the margin of error increases.
  - (c) As the confidence level increases, the standard error of the sample mean decreases.
  - (d) The margin of error decreases as the standard deviation decrees.

Exercise 6:

- Weights of women in an age group are normally distributed with a standard deviation of 10 kg. A researcher wishes to estimate the mean weight of all women in this age group. Find how large a sample must be drawn in order to be 90 percent confident that the sample mean will not differ from the population mean by more than 2 Kg.

Exercise 7:

- If you measure the heights of 25 year-old males, how big sample is needed so that the 95% margin of error is 1 cm. Assume that the standard deviation is 5 cm.