Statistics 1
Chapter 9

Sampling Methods and Sampling distributions

Chapter 9

Sampling Methods and Sampling Distributions

GOALS

When you have completed this chapter, you will be able to:

ONE

Explain why a sample is the only feasible way to learn about a population.

TWO

Explain methods for selecting a sample.

THREE

Define and construct a sampling distribution of the sample means.

FOUR

Explain the central limit theorem.

Chapter 9 *continued*

Sampling Methods and Sampling Distributions

GOALS

When you have completed this chapter, you will be able to:

FIVE

Calculate confidence intervals for means and proportions.

SIX

Determine the sample size for attribute and variable sampling.

Why Sample the Population?

- The physical impossibility of checking all items in the population.
- The cost of studying all the items in a population.
- The sample results are usually adequate.
- Contacting the whole population would often be time-consuming.
- The destructive nature of certain tests.

Probability Sampling

- A probability sample is a sample selected in such a way that each item or person in the population being studied has a known likelihood of being included in the sample.

Methods of Probability Sampling

- Simple Random Sample: A sample formulated so that each item or person in the population has the same chance of being included.
- Systematic Random Sampling: The items or individuals of the population are arranged in some order. A random starting point is selected and then every *k*th member of the population is selected for the sample.

Methods of Probability Sampling

- Stratified Random Sampling: A population is first divided into subgroups, called strata, and a sample is selected from each stratum.
- Cluster Sampling: A population is first divided into subgroups (strata), and a sample of the strata is selected. The sample is then taken from these selected strata.
- A sampling error is the difference between a sample statistic and its corresponding parameter.

Sampling Distribution of the Sample Means

- The sampling distribution of the sample means is a probability distribution consisting of all possible sample means of a given sample size selected from a population, and the probability of occurrence associated with each sample mean.

EXAMPLE 1

- The law firm of Hoya and Associates has five partners. At their weekly partners meeting each reported the number of hours they charged clients for their services last week.

Partner	Hours
Dunn	22
Hardy	26
Kiers	30
Malinowski	26
Tillman	22

- If two partners are selected randomly, how many different samples are possible?

EXAMPLE 1 *continued*

- This is the combination of 5 objects taken 2 at a time. That is, $_5c_2 = (5!)/[(2!)(3!)] = 10$

Partners	Total	Mean
1 , 2	48	24
1 , 3	52	26
1 , 4	48	24
1 , 5	44	22
2 , 3	56	28
2 , 4	52	26
2 , 5	48	24
3 , 4	56	28
3 , 5	52	26
4 , 5	48	24

EXAMPLE 1 *continued*

- Organize the sample means into a sampling distribution.

Sample Mean	Frequency	Relative Frequency probability
22	1	1/10
24	4	4/10
26	3	3/10
28	2	2/10

EXAMPLE 1 *continued*

- Compute the mean of the sample means and compare it with the population mean:
  - The mean of the sample means = [(22)(1) + (24)(4) + (26)(3) + (28)(2)]/10=25.2
  - The population mean = (22+26+30+26+22)/5 = 25.2
  - *Observe that the mean of the sample means is equal to the population mean.*

Central Limit Theorem

- For a population with a mean μ and a variance σ^2, the sampling distribution of the means of all possible samples of size n generated from the population will be approximately normally distributed - with the mean of the sampling distribution equal to μ and the variance equal to σ^2 / n - assuming that the sample size is sufficiently large.

Point Estimates

- A Point estimate is one value (a point) that is used to estimate a population parameter.
- Examples of point estimates are the *sample mean*, the *sample standard deviation*, the *sample variance*, the *sample proportion* etc...
- EXAMPLE 2: The number of defective items produced by a machine was recorded for five randomly selected hours during a 40-hour work week. The observed number of defectives were 12, 4, 7, 14, and 10. So the sample mean is 9.4. Thus a point estimate for the hourly mean number of defectives is 9.4.

Interval Estimates

- An Interval Estimate states the range within which a population parameter probably lies.
- The interval within which a population parameter is expected to occur is called a confidence interval.
- The two confidence intervals that are used extensively are the 95% and the 99%.

Interval Estimates

- A 95% confidence interval means that about 95% of the similarly constructed intervals will contain the parameter being estimated, or 95% of the sample means for a specified sample size will lie within 1.96 standard deviations of the hypothesized population mean.
- For the 99% confidence interval, 99% of the sample means for a specified sample size will lie within 2.58 standard deviations of the hypothesized population mean.

Standard Error of the Sample Means

- The standard error of the sample means is the standard deviation of the sampling distribution of the sample means.
- It is computed by
- $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
- $\sigma_{\bar{x}}$ is the symbol for the standard error of the sample means.
- σ is the standard deviation of the population.
- n is the size of the sample.

Standard Error of the Sample Means

- If σ is not known and $n \geq 30$, the standard deviation of the sample, designated s, is used to approximate the population standard deviation. The formula for the standard error then becomes:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

95% and 99% Confidence Intervals for μ

- The 95% and 99% confidence intervals for μ are constructed as follows when $n \geq 30$
- 95% CI for the population mean is given by

$$\overline{X} \pm 1.96 \frac{s}{\sqrt{n}}$$

- 99% CI for the population mean is given by

$$\overline{X} \pm 2.58 \frac{s}{\sqrt{n}}$$

Constructing General Confidence Intervals for μ

- In general, a confidence interval for the mean is computed by:

$$\overline{X} \pm Z \frac{s}{\sqrt{n}}$$

EXAMPLE 3

- The Dean of the Business School wants to estimate the mean number of hours worked per week by students. A sample of 49 students showed a mean of 24 hours with a standard deviation of 4 hours.
- The point estimate is 24 hours (sample mean).
- What is the 95% confidence interval for the average number of hours worked per week by the students?

EXAMPLE 3 *continued*

- Using the 95% CI for the population mean, we have $24 \pm 1.96(4/7) = 22.88 \, to \, 25.12$
- The endpoints of the confidence interval are the confidence limits. The lower confidence limit is 22.88 and the upper confidence limit is 25.12

Confidence Interval for a Population Proportion

- The confidence interval for a population proportion is estimated by:

$$p \pm z\sigma_{\overline{p}}$$

- where $\sigma_{\overline{p}}$ is the standard error of the proportion:

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$$

EXAMPLE 4

- Matt Williams, a financial planner, is studying the retirement plans of young executives. A sample of 500 executives who own their own home revealed that 175 planned to sell their homes and retire to Arizona. Develop a 98% confidence interval for the proportion of executives that plan to sell and move to Arizona.
- Here, n=500, $\bar{p}$=175/500=.35, and z=2.33
- the 98% CI is

$$.35 \pm 2.33\sqrt{\frac{(.35)(.65)}{500}} \text{ or } .35 \pm .0497$$

Finite-Population Correction Factor

- A population that has a fixed upper bound is said to be finite.
- For a finite population, where the total number of objects is N and the size of the sample is n, the following adjustment is made to the standard errors of the sample means and the proportion:
- Standard error of the sample means:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Finite-Population Correction Factor

› Standard error of the sample proportions:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

- This adjustment is called the *finite-population correction factor.*
- *Note:* If n/N < .05, the finite-population correction factor is ignored.

EXAMPLE 5

- Given the information in EXAMPLE 4, construct a 95% confidence interval for the mean number of hours worked per week by the students if there are only 500 students on campus.
- Since n/N = 49/500 = .098>.05, we have to use the finite population correction factor.

$$24 \pm 1.96\left(\frac{4}{\sqrt{49}}\right)\left(\sqrt{\frac{500-49}{500-1}}\right) = [22.9352,\ 25.0648]$$

Selecting a Sample Size

- There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population. They are:
- The degree of confidence selected.
  - The maximum allowable error.
  - The variation of the population.

Variation in the Population

- *Sample Size for the Mean:* A convenient computational formula for determining n is:

$$n = \left(\frac{Z \bullet S}{E} \right)^2$$

- where : E is the allowable error, Z is the z score associated with the degree of confidence selected, and S is the sample deviation of the pilot survey.

EXAMPLE 6

- A consumer group would like to estimate the mean monthly electric bill for a single family house in July. Based on similar studies the standard deviation is estimated to be \$20.00. A 99% level of confidence is desired, with an accuracy of $\pm$ \$5.00. How large a sample is required?

$$n = [(2.58)(20)/5]^2 = 106.5024 \approx 107$$

Sample Size for Proportions

- The formula for determining the sample size in the case of a proportion is:

$$n = p(1-p)\left(\frac{Z}{E}\right)^2$$

- where p is the estimated proportion, based on past experience or a pilot survey; z is the z value associated with the degree of confidence selected; E is the maximum allowable error the researcher will tolerate.

EXAMPLE 7

- The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = (.30)(.70)(1.96/.03)^2 = 896.3733 \approx 897$$