統計學

Fall 2025

政治大學統計系余清祥

2025年10月28日

第八章：點估計與區間估計

http://csyue.nccu.edu.tw

Chapter Contents

8.1 Population Mean: σ Known

8.2 Population Mean: σ Unknown

8.3 Determining the Sample Size

8.4 Population Proportion

8.5 Big Data and Confidence Intervals

Summary

CENGAGE

Introduction

A point estimator is a sample statistic used to estimate a population parameter.

- The sample mean $\bar{x}$ is a point estimator of the population mean μ
- The sample proportion $\bar{p}$ is a point estimator of the population proportion p

An **interval estimate** is computed by adding and subtracting a value, called the **margin of error**, to the point estimate. The general form of an interval estimate is:

Point estimate ± Margin of error

The general form of an interval estimate of a population mean is:

$\bar{x}$ ± Margin of error

And the general form of an interval estimate of a population proportion is:

$\bar{p}$ ± Margin of error

The sampling distributions of $\bar{x}$ and $\bar{p}$ play key roles in computing these interval estimates.

8.1 Population Mean: σ Known

To estimate a population mean, we compute the margin of error using either:

- The population standard deviation, σ
- The sample standard deviation, s

σ is seldom known, but a good estimate can often be obtained based on historical data or other information. We refer to such cases as the **σ known** case.

Let us construct an interval estimate of μ, the mean amount spent per shopping trip for the population of Lloyd's customers, in a case for which it is reasonable to treat σ as known, given:

$n = 100$ customers, a simple random sample

$x =$ the amount spent per shopping trip

$s = \$20$, based on historical data

The population follows a normal distribution

Calculations reveal that the sample mean, a point estimate of μ, is: $\bar{x}$ = $82.

CENGAGE

8.1 Sampling Distribution for the Sample Mean

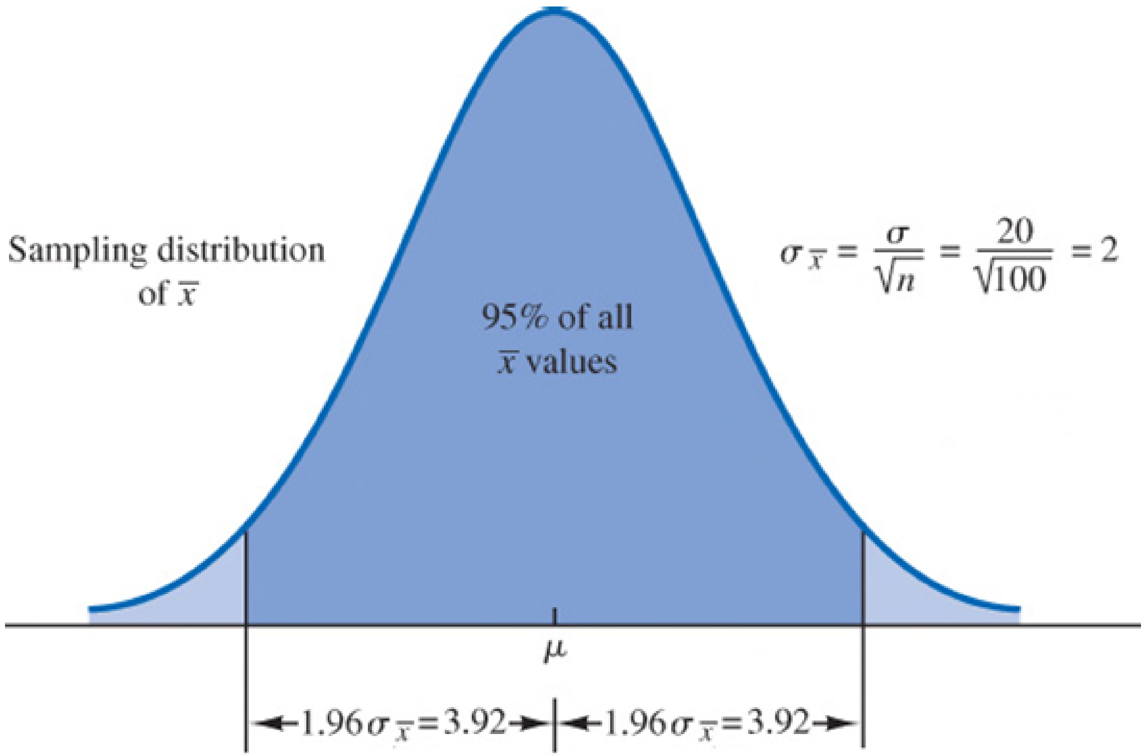
We can conclude that the sampling distribution of $\bar{x}$ follows a normal distribution with a standard error of

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$20}{\sqrt{100}} = 2$$

This sampling distribution, shown to the right, depicts how values of $\bar{x}$ are distributed around the population mean, μ.

It follows that 95% of the values of the normally distributed $\bar{x}$ random variable falls within $\pm 1.96\sigma_{\bar{x}}$ standard deviations of the mean, or within

$$\pm 1.96\sigma_{\bar{x}} = \pm 1.96(2) = \pm 3.92$$



8.1 The Confidence Level

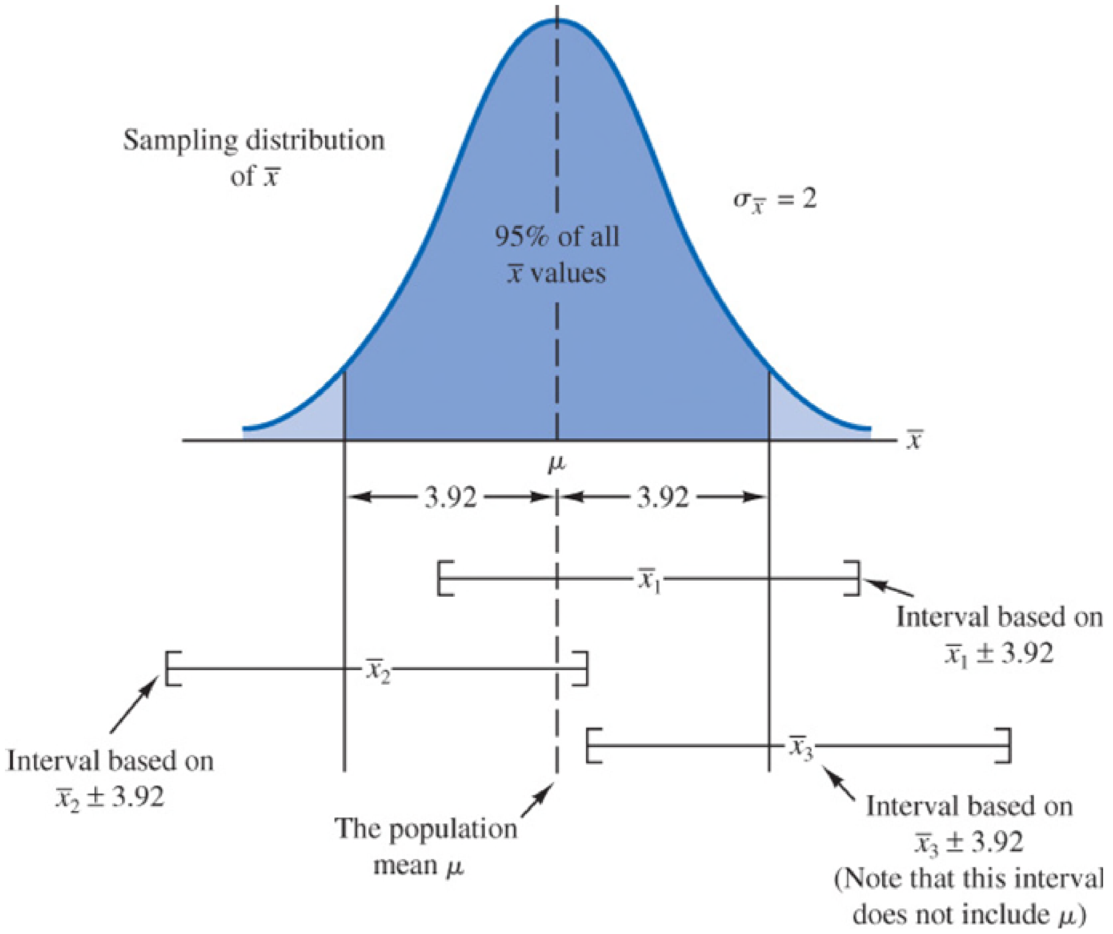
Let us set the margin of error equal to 3.92 and compute the interval estimate of μ using $\bar{x} \pm 3.92$.

The figure shows what could happen if we took three separate samples of 100 Lloyd's customers and built an interval estimate for each of them.

The first two intervals, built around the sample means $\bar{x}_1$ and $\bar{x}_2$, include the population mean, μ, but the third one, $\bar{x}_3$, does not.

It turns out that 95% of all intervals formed by subtracting 3.92 from $\bar{x}$ and adding 3.92 to $\bar{x}$ will include the population mean, μ.

We call an interval built around $\bar{x}$ a **confidence interval** established at the 95% **confidence level**.



8.1 Margin of Error and the Interval Estimate

An interval estimate of the population mean μ, for the σ known case, can be written as

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Where, $1 - \alpha$ is the confidence coefficient, and $z_{\alpha/2}$ the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution.

To construct a 95% confidence interval for the Lloyd's example, we observe that, $1 - \alpha = 0.95$.

Thus, we have

$$\alpha/2 = (1 - 0.95)/2 = 0.05/2 = 0.025, \text{ and } z_{\alpha/2} = 1.96.$$

With, $\bar{x} = 82$, $\sigma = 20$, and $n = 100$, we obtain

$$82 \pm 1.96 \frac{20}{\sqrt{100}} = 82 \pm 3.92$$

8.1 Commonly Used Confidence Levels

In the calculation of the previous slide, we concluded that we are 95% confident the mean amount spent per shopping trip at Lloyd's is between

$$82 - 3.92 = \$78.08 \text{ and } 82 + 3.92 = \$85.92$$

Other confidence levels such as 90% and 99% may also be considered. The table shows the $z_{\alpha/2}$ for 90%, 95%, and 99% confidence levels.

For example, at a 90% confidence level, $z_{\alpha/2} = 1.645$, and the confidence interval becomes

$$82 \pm 1.645 \frac{20}{\sqrt{100}} = 82 \pm 3.29$$

Or between

$$82 - 3.29 = \$78.71 \text{ and } 82 + 3.29 = \$85.29$$

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
90%	0.10	0.05	1.645
95%	0.05	0.025	1.960
99%	0.01	0.005	2.576

CENGAGE

8.1 Practical Advice

Adequate Sample Size

- The sampling distribution of $\bar{x}$ is normal if the underlying population distribution is normal; otherwise, a sample size of $n \geq 30$ is required.
- If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended.
- If the population is not normally distributed but is roughly symmetric, a sample size as small as 15 will suffice.
- If the population is believed to be at least approximately normal, a sample size of less than 15 can be used.
- For small sample sizes, a histogram of the sample data should be built to learn about the shape of the population distribution.

8.2 Population Mean: σ Unknown

The *t* Distribution

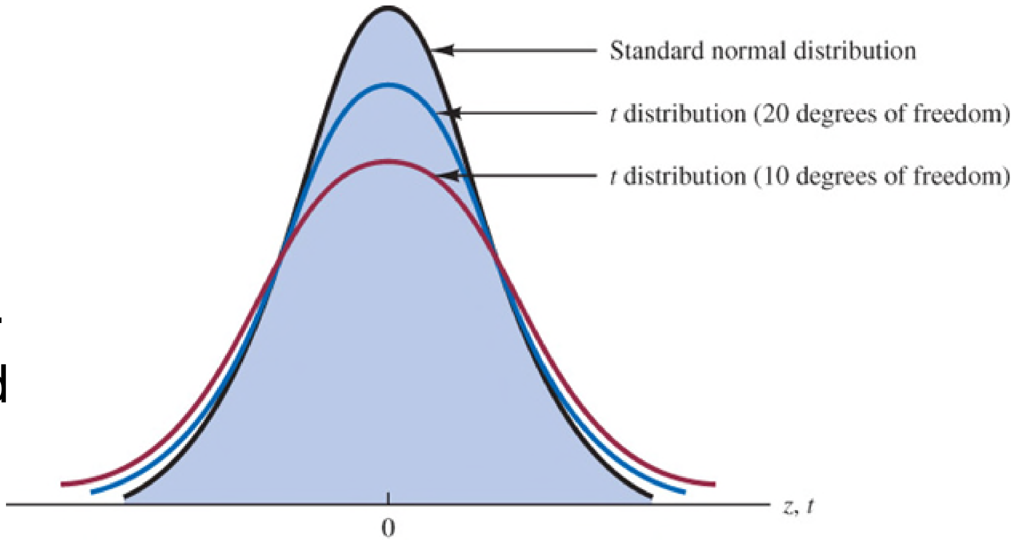
If an estimate of the population standard deviation, σ, cannot be developed prior to sampling, we use the sample standard deviation, *s*, instead.

This is referred to as the σ **unknown** case, with the interval estimate of μ based on the ***t* distribution**, a family of similar probability distributions depending on a parameter called the **degrees of freedom**.

A *t* distribution with more degrees of freedom has less dispersion.

As the degrees of freedom increase, the difference between the *t* distribution and the standard normal probability distribution becomes smaller and smaller.

For more than 100 degrees of freedom, the standard normal *z* value provides a good approximation to the *t* value (*see notes for additional details.)



8.2 Interval Estimate of μ: σ Unknown

An interval estimate of the population mean μ, for the σ unknown case, can be written as

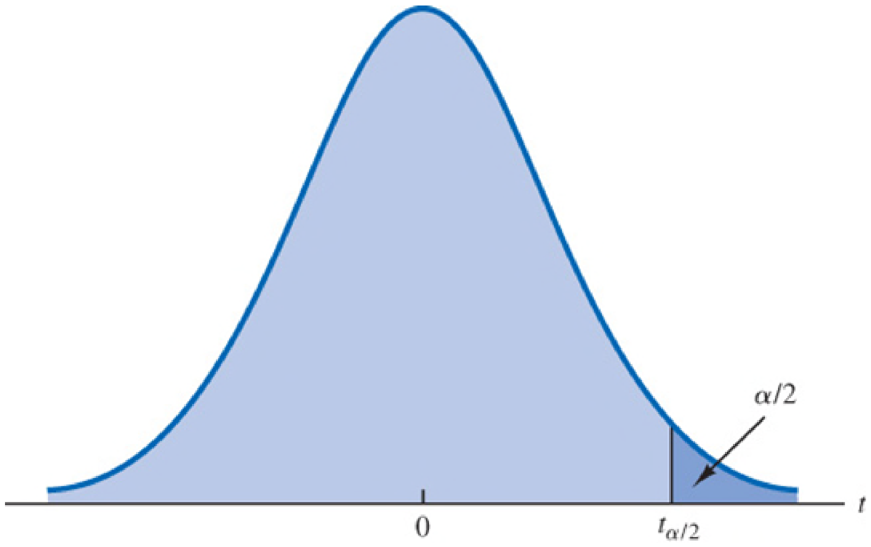
$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Where *s* is the sample standard deviation, $1 - \alpha$ is the confidence coefficient, and $t_{\alpha/2}$ the *t* value providing an area of α/2 in the upper tail of a *t* distribution with $n - 1$ degrees of freedom (*see the notes for details.)

As an application, consider the estimate of the mean credit card debt, σ unknown case, for the population of U.S. households (DATAfile: *NewBalance.)*

The data consists of the household credit card debt, x, for a simple random sample of $n = 70$ households.

Sample calculations reveal a sample mean $\bar{x} = \$9{,}312$, and a sample standard deviation, $s = \$4{,}007$.



8.2 Estimate of the Population Mean Credit Card Balance

With 95% confidence and $n - 1 = 69$ degrees of freedom, Table 2 in Appendix B can be used to obtain $t_{0.025} = 1.995$. We want the *t* value in the row with 69 degrees of freedom, and the column corresponding to 0.025 in the upper tail.

Area in upper tail	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
***t* Value ($df = 69$)**	.847	1.294	1.667	1.995	2.382	2.649

Alternatively, the Excel equation =T.INV(1−0.025,69) will also generate the value 1.995.

We can now compute an interval estimate of the population mean credit card balance as:

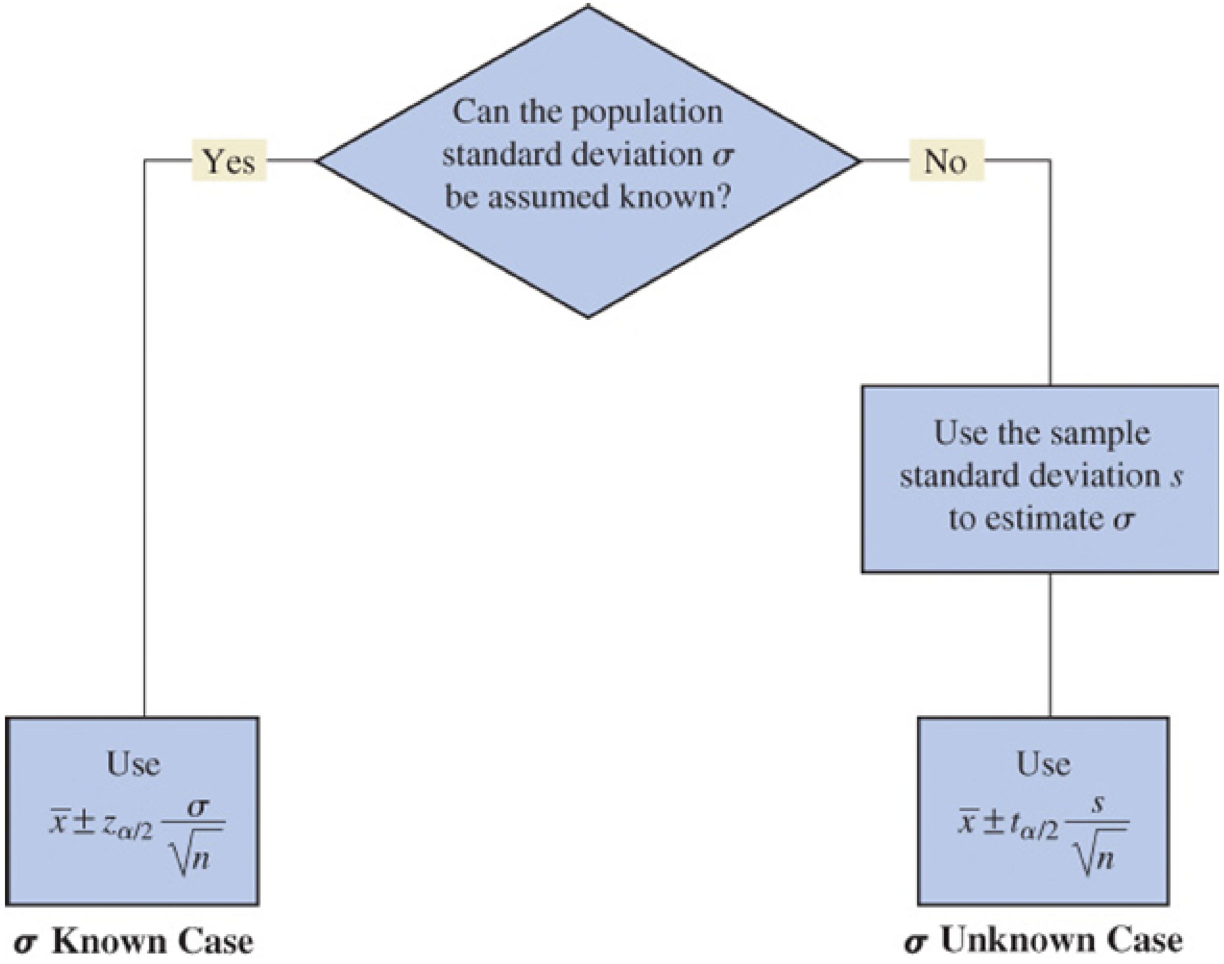
$$9312 \pm 1.995 \frac{4007}{\sqrt{70}} \quad \text{or} \quad 9312 \pm 955$$

Where $9,312 is the point estimate of the population mean and $955 the margin of error.

Thus, we are 95% confident that the mean credit card balance for the population of all households is between $9312 - 955 = \$8{,}357$ and $9312 + 955 = \$10{,}267$.

CENGAGE

8.2 Summary of Interval Estimation Procedures for μ



8.3 Determining the Sample Size for the Estimate of μ

Given a desired the margin of error E, selected prior to sampling such that

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

The sample size necessary for the desired margin of error can be determined solving for n

$$\boldsymbol{n} = \frac{\left(\boldsymbol{z}_{\alpha/2}\right)^2 \boldsymbol{\sigma}^2}{\boldsymbol{E}^2}$$

A preliminary or planning value for the unknown population standard deviation σ, may be determined using:

1. the estimate of the population standard deviation σ, computed in a previous study.
2. a preliminary sample and the sample standard deviation s, selected from a pilot study.
3. judgment or a “best guess” for the value of the population standard deviation, σ.

CENGAGE

8.3 Application of Sample Size Determination

A previous study that investigated the cost of renting midsize automobiles found a mean cost of approximately \$55 per day, with a standard deviation of \$9.65.

Objective: conduct a new study to estimate the mean cost of renting midsize automobiles with a margin of error of $E = \$2$, a 95% confidence level, and a planning value for the population standard deviation, $\sigma = 9.65$.

Thus, with a $(1 - \alpha) = 0.95$ confidence coefficient, the z value is $z_{0.025} = 1.96$.

Plugging the values into the necessary sample size equation, we obtain

$$n = \frac{\left(z_{\alpha/2}\right)^2 \sigma^2}{E^2} = \frac{(1.96)^2(9.65)^2}{(2)^2} = 89.43$$

In cases where the computed n is not an integer, we round up to the next integer value.

Hence, the recommended sample size is 90 midsize automobile rentals.

CENGAGE

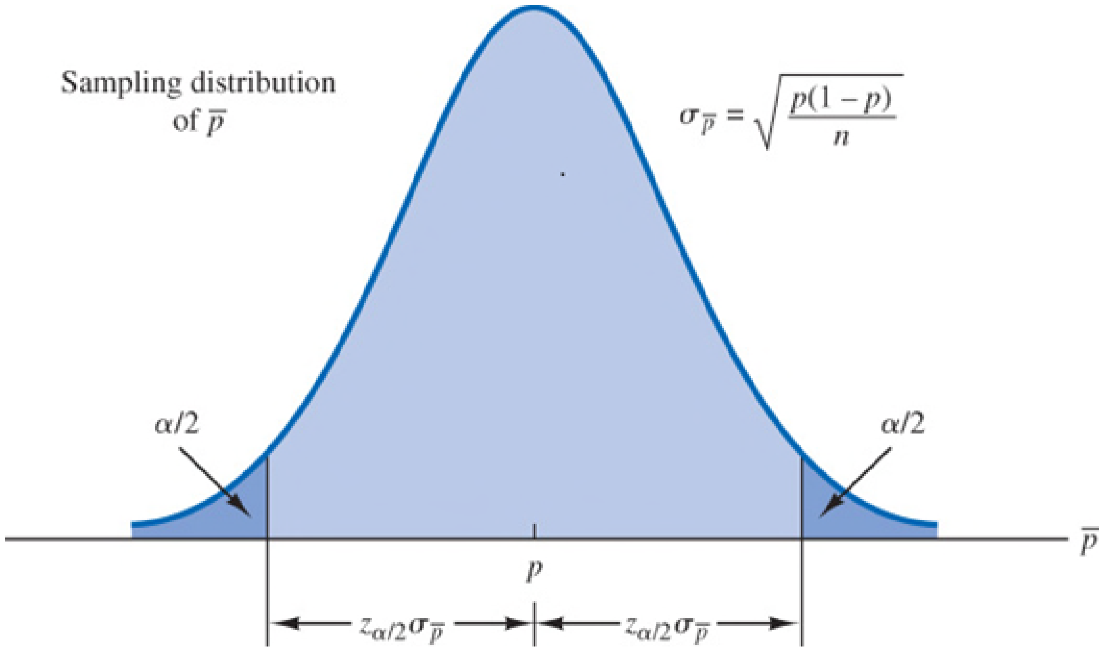
8.4 Interval Estimate of a Population Proportion

The general form of an interval estimate of a population proportion is

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Where,

$1-\alpha$ is the confidence coefficient, $z_{\alpha/2}$ the z value providing an area of α/2 in the upper tail of the standard normal probability distribution, and $\bar{p}$ the sample proportion.



The sampling distribution of $\bar{p}$ plays a key role in computing the margin of error for the interval estimate, and it can be approximated by a normal distribution whenever $n\bar{p} \geq 5$ and $n(1-\bar{p}) \geq 5$.

8.4 Application of Population Proportion Estimate

A national survey of 900 randomly selected patients was conducted to assess the satisfaction with the wait times patients experienced in the medical system.

The survey found that 396 patients were satisfied with the wait times they experienced.

Thus, the point estimate of the proportion of the population of patients who are satisfied with the wait times they experienced is

$$\bar{p} = \frac{396}{900} = 0.44$$

Using the expression for the interval estimate of the population proportion at a 95% confidence level, we obtain

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.44 \pm 1.96\sqrt{\frac{0.44(1-0.44)}{900}} = 0.44 \pm 0.0324$$

Thus, we are 95% confident that between $0.44 - 0.0324 = 0.4076 = 40.76\%$ and $0.44 + 0.0324 = 0.4724 = 47.24\%$ of all patients are satisfied with the wait times they experienced.

CENGAGE

8.4 Determining the Sample Size for the Estimate of *p*

Given a desired the margin of error E, selected prior to sampling such that

$$E = z_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n}}$$

The sample size necessary for the desired margin of error can be determined solving for n

$$\boldsymbol{n = \frac{\left(z_{\alpha/2}\right)^2 p^*(1-p^*)}{E^2}}$$

The planning value p^* replaces the unknown sample proportion, $\bar{p}$, and can be chosen using

1. the sample proportion $\bar{p}$ from a previous sample of the same or similar units.
2. the sample proportion $\bar{p}$ from a preliminary study as the planning value, p^*.
3. a “best guess” for the value of p^*.
4. a planning value of $p^* = 0.50$, if none of the preceding alternatives applies (*see notes.)

8.4 Application of Sample Size Determination

As an application of sample size determination, consider again the national survey conducted to assess the satisfaction with the wait times patients experienced in the medical system.

In its preliminary version, we used a random sample of $n = 900$ patients to estimate at a 95% confidence level the proportion of patients' satisfaction with a 0.0324 margin of error.

If the requirement on the margin of error is 0.025 and the confidence level still 95%, we have

$z_{\alpha/2} = 1.96$

$E = 0.025$

$p^* = \bar{p} = 0.44$ from the preliminary study of $n = 900$ patients

The required sample size to estimate the proportion of patients' wait times is

$$n = \frac{\left(z_{\alpha/2}\right)^2 p^*(1 - p^*)}{E^2} = \frac{(1.96)^2 0.44(1 - 0.44)}{(0.025)^2} = 1{,}514.5 \approx 1{,}515 \text{ patients}$$

Thus, we need to survey an additional $1{,}515 - 900 = 615$ patients.

CENGAGE

8.5 Big Data and Confidence Intervals

A review of the equations for the calculations of confidence intervals for the population mean μ and population proportion *p* reveals that the sampling error decreases as the sample size increases.

This is due to the presence of $\sqrt{n}$ at the denominator of the standard error in the equations of the confidence intervals.

As the sample size becomes arbitrarily large, the margin of error becomes smaller and smaller and the resulting confidence intervals become so extremely narrow to have no longer any **practical significance**.

However, no interval estimate, no matter how precise, will accurately reflect the parameter being estimated unless the sample is relatively free of nonsampling error.

Therefore, when using interval estimation, it is always important to carefully consider whether a random sample of the population of interest has been taken.

Summary

- In this chapter we presented methods for developing interval estimates of a population mean and a population proportion in the form of point estimate ± margin of error.
- We presented interval estimates for a population mean for two cases.
  - σ known: historical data or other information is used to develop an estimate of σ prior to taking a sample, and a confidence interval is built using the standard normal distribution
  - σ unknown: sample data are used to estimate μ and σ, and a confidence interval is built using a t distribution with $n-1$ degrees of freedom.
  - Practical advice about the sample size necessary to obtain good approximations was also included. In most cases, a sample size of at least 30 suffices.
- The general form of the interval estimate for a population proportion based on the standard normal distribution was also given.
- Finally, we discussed the ramifications of extremely large samples on the precision of confidence interval estimates of the mean and proportion.