Confidence Interval Estimation

One mean and one proportion

Learning Objectives

In this topic, you learn:

- To construct and interpret confidence interval estimates for the mean and the proportion
- How to determine the sample size necessary to develop a confidence interval estimate for the mean or proportion
- How to use confidence interval estimates in auditing

Outline

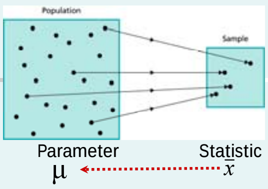
- Confidence Intervals for the Population Mean, μ
  - when Population Standard Deviation σ is Known
  - when Population Standard Deviation σ is Unknown
- Confidence Intervals for the Population Proportion, p
- Determining the Required Sample Size

Introduction

- Estimate
  - The value(s) assigned to a population parameter based on the value of a sample statistic is called an ***estimate***.
- Estimator
  - The sample statistic used to estimate a population parameter is called an ***estimator***.

Source: P.S. Mann, Introductory Statistics, 5th ed. John Wiley, 2005

Estimation Procedure



1. Select a sample.
2. Collect the required information from the members of the sample.
3. Calculate the value of the sample statistic.
4. Assign value(s) to the corresponding population parameter.

Source: P.S. Mann, Introductory Statistics, 5th ed. John Wiley, 2005

Estimation Methods

- **A Point Estimate**
  - The value of a sample statistic that is used to estimate a population parameter is called a ***point estimate***.
- **An Interval Estimate**
  - In ***interval estimation***, an interval is constructed around the point estimate, and it is stated that this interval is likely to contain the corresponding population parameter.

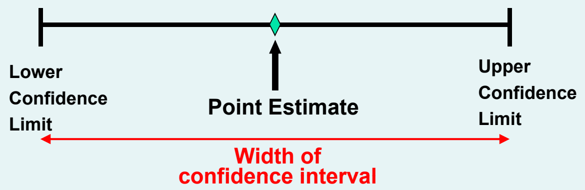
- 區間估計
  - 常態母體
    - 平均數
      - 一個母體
      - 二個母體(平均數差)
    - 變異數
      - 一個母體
      - 二個母體(變異數比)
  - 點二項母體
    - 一個母體比例
    - 二個母體比例差

Topic-1

Point and Confidence Interval Estimates for a Population Mean

Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about the variability of the estimate



Point Estimates

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{X}$
Proportion	p	$\hat{p}$

Confidence Intervals

- How much uncertainty is associated with a point estimate of a population parameter?
- An interval estimate provides more information about a population characteristic than does a point estimate.
- Such interval estimates are called confidence intervals.

Confidence Interval Example

Population has μ = 368 and σ = 15.

Sample #	$\bar{X}$	Lower Limit	Upper Limit	Contain μ?
1	362.30	356.42	368.18	Yes
2	369.50	363.62	375.38	Yes
3	360.00	354.12	365.88	No
4	362.12	356.24	368.00	Yes
5	373.88	368.00	379.76	Yes

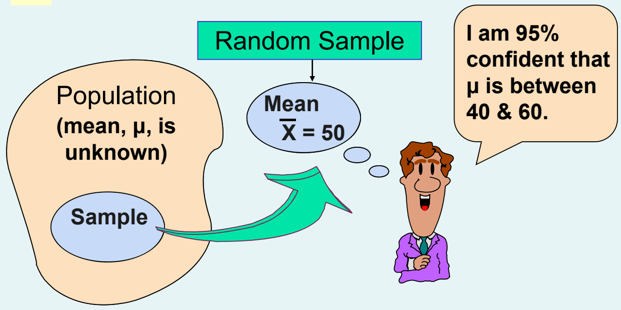
Confidence Interval Example

(continued)

- In practice you only take one sample of size n
- In practice you do not know μ so you do not know if the interval actually contains μ
- However, you do know that 95% of the intervals formed in this manner will contain μ
- Thus, based on the one sample you actually selected, you can be 95% confident your interval will contain μ (this is a 95% confidence interval)

Note: 95% confidence is based on the fact that we used Z = 1.96.

Estimation Process



Confidence Level: Confidence of the interval will contain the unknown population parameter.

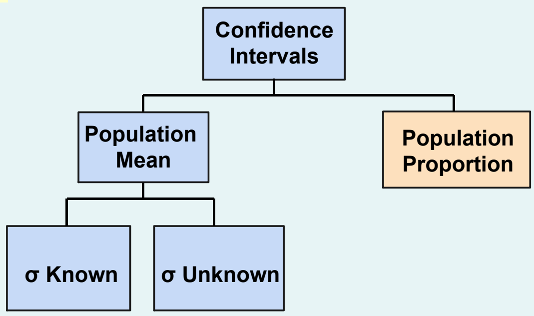
General Formula

- The general formula for <u>all confidence intervals</u> is:

Point Estimate ± (Critical Value)(Standard Error)

Where:
- **Point Estimate** is the sample statistic estimating the population parameter of interest
- **Critical Value** is a table value based on the sampling distribution of the point estimate and the desired confidence level
- **Standard Error** is the standard deviation of the point estimate

Confidence Intervals



Confidence Interval for μ (σ Known)

- Assumptions
  - Population standard deviation σ is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

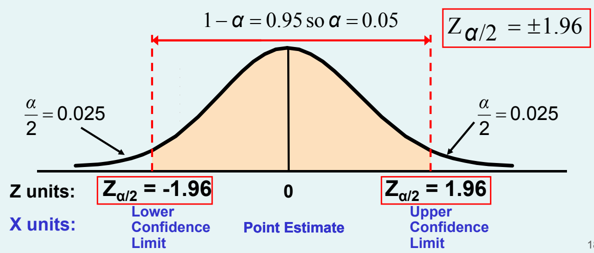
$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $\bar{X}$ is the point estimate
$Z_{\alpha/2}$ is the normal distribution critical value for a probability of $\alpha/2$ in each tail
$\sigma/\sqrt{n}$ is the standard error

Confidence Interval for *μ*

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Consider a 95% confidence interval:

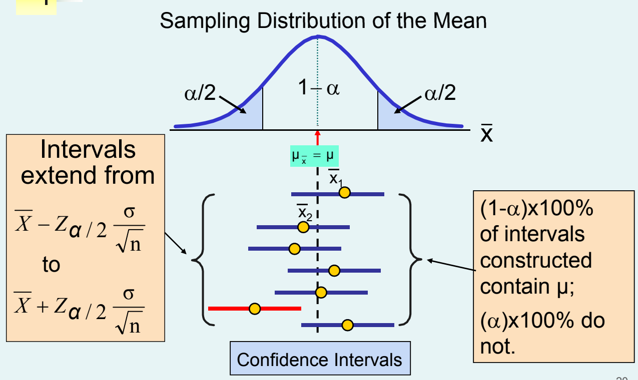


Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, and 99%

Confidence Level	*Confidence Coefficient,* $1-\alpha$	*$Z_{\alpha/2}$ value*
80%	**0.80**	**1.28**
90%	**0.90**	**1.645**
95%	**0.95**	**1.96**
98%	**0.98**	**2.33**
99%	**0.99**	**2.58**
99.8%	**0.998**	**3.08**
99.9%	**0.999**	**3.27**

Intervals and Level of Confidence

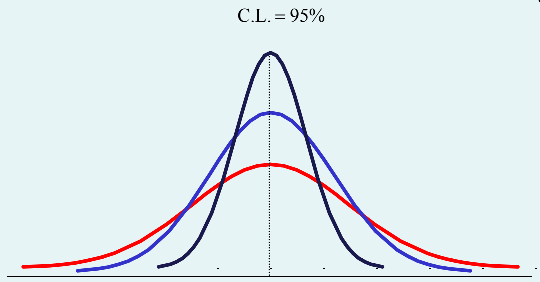


Do You Ever Truly Know σ?

- Probably not!
- In virtually all real world business situations, σ is not known.
- If there is a situation where σ is known, then μ is also known (since to calculate σ you need to know μ.)
- If you truly know μ there would be no need to gather a sample to estimate it.

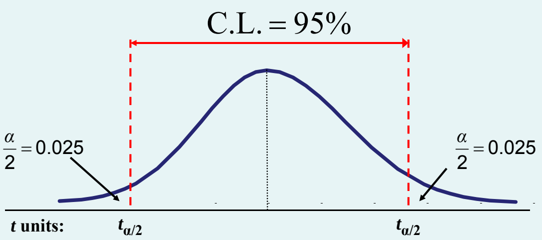
Confidence Interval for μ (σ Unknown)

- Consider a 95% confidence interval: $\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$

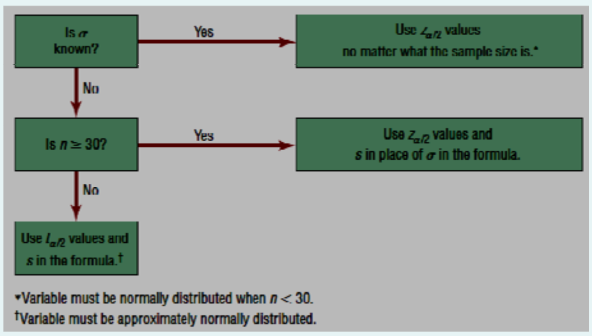


Confidence Interval for μ (σ Unknown)

- Consider a 95% confidence interval: $\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$



When to use the *z* or *t* distribution



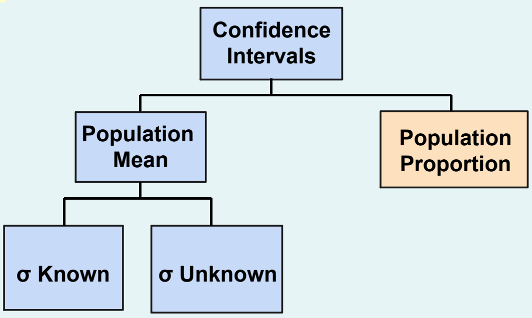
課堂練習

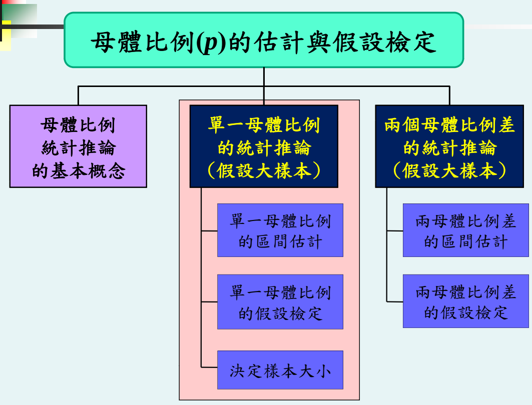
- 分組練習：4~5人一組
- 自訂題目
  - 上網時間、睡眠時間、手機費用、通話時間、體重、身高、慢跑3000m所需時間...(不限於此)
  - 定義問題及確認母體
  - 蒐集資料(樣本數自訂)
  - 計算信賴區間
    - 計算平均數、標準差
    - 需滿足之假設條件有哪些
    - 信賴水準
    - 建立信賴區間

Topic-2

Point and Confidence Interval Estimates for a Population Proportion

Confidence Intervals





Confidence Intervals for Proportions

- p = population proportion

$\hat{p}$ (read p "hat") = sample proportion

For a sample proportion,

$$\hat{p} = \frac{X}{n} \quad \text{and} \quad \hat{q} = \frac{n-X}{n} \quad or \quad \hat{q} = 1 - \hat{p}$$

where

X = number of sample units that possess the characteristics of interest, ***n*** = sample size.

Example: Air Conditioned Households

In a recent survey of 150 households, 54 had central air conditioning. Find $\hat{p}$ and $\hat{q}$, where $\hat{p}$ is the proportion of households that have central air conditioning.

The Binomial Distribution

Mean, Variance, and Standard Deviation for the Binomial Distribution

The mean, variance, and standard deviation of a variable that has the *binomial distribution* can be found by using the following formulas.

Mean $\mu = n \cdot p$ Variance $\sigma^2 = n \cdot p \cdot q$ Standard deviation $\sigma = \sqrt{n \cdot p \cdot q}$

Example: Likelihood of Twins

The *Statistical Bulletin* published by Metropolitan Life Insurance Co. reported that 2% of all American births result in twins. If a random sample of 8000 births is taken, find the mean, variance, and standard deviation of the number of births that would result in twins.

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

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

- We will estimate this with sample data

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Formula for a Specific Confidence Interval for a Proportion

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

when $np \geq 5$ and $nq \geq 5$.

Rounding Rule: Round off to three decimal places.

Example: Male Nurses

A sample of 500 nursing applications included 60 from men. Find the 90% confidence interval of the true proportion of men who applied to the nursing program.

Example: Male Nurses

A sample of 500 nursing applications included 60 from men. Find the 90% confidence interval of the true proportion of men who applied to the nursing program.

Example: Religious Books

A survey of 1721 people found that 15.9% of individuals purchase religious books at a Christian bookstore. Find the 95% confidence interval of the true proportion of people who purchase their religious books at a Christian bookstore.

看臉書算閱讀？ 調查：近半國人不看書

Example: Religious Books

A survey of 1721 people found that 15.9% of individuals purchase religious books at a Christian bookstore. Find the 95% confidence interval of the true proportion of people who purchase their religious books at a Christian bookstore.

Example

- A random sample of 100 people shows that 25 are left-handed.
- Form a 95% confidence interval for the true proportion of left-handers

Example

(continued)

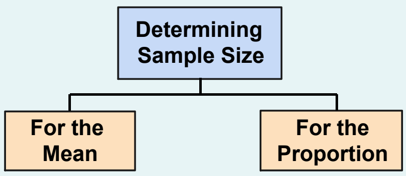
- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

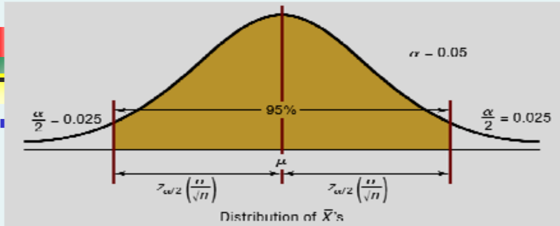
Make sure the sample is big enough → $np = 100*0.25 = 25 > 5$ & $n(1-p) = 100*0.75 = 75 > 5$

Interpretation

- We are 95% confident that the true percentage of left-handers in the population is between
  16.51% and 33.49%.
- Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

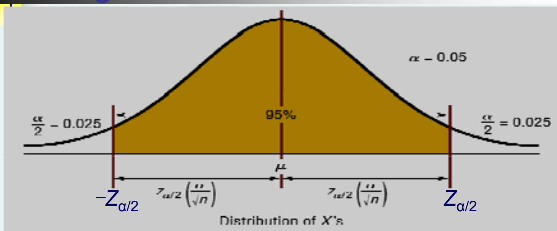
Determining Sample Size





- The margin of error is also called sampling error
  - the amount of imprecision in the estimate of the population parameter
  - the amount added and subtracted to the point estimate to form the confidence interval

Margin of Error for the estimate



Source: Bluman, Chapter 7

Margin of Error for the Estimate of μ (Maximum Error of the Estimate)

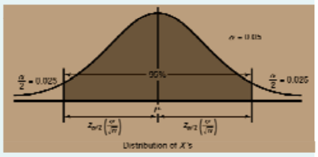
The **margin of error** for the estimate of μ is

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

Determining Sample Size

Determining Sample Size

For the Mean



$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

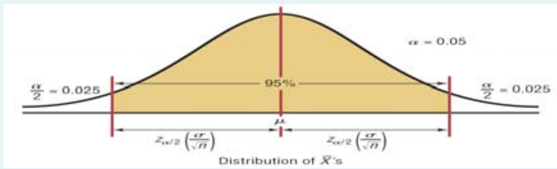
$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$n = \frac{Z_{\alpha/2}^{\ 2}\ \sigma^2}{e^2}$$

Determining Sample Size

(continued)

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence (1 - α), which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The standard deviation, σ



Required Sample Size Example

If σ = 45, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{Z^2\sigma^2}{e^2} = \frac{(1.645)^2(45)^2}{5^2} = 219.19$$

So the required sample size is **n = 220**

(Always round up)

Determining Sample Size

(continued)

Determining Sample Size

For the Proportion

$$e = Z\sqrt{\frac{p(1-p)}{n}}$$ Now solve for n to get $$n = \frac{Z^2\ p(1-p)}{e^2}$$

Determining Sample Size

(continued)

- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence (1 - α), which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The true proportion of events of interest, p
    - p can be estimated with a pilot sample if necessary (or conservatively use 0.5 as an estimate of p)

Required Sample Size Example

How large a sample would be necessary to estimate the true proportion defective in a large population within ±3%, with 95% confidence?
(Assume a pilot sample yields $\hat{p}$ = 0.12)

Required Sample Size Example

(continued)

Solution:

For 95% confidence, use $Z_{\alpha/2} = 1.96$

$e = 0.03$

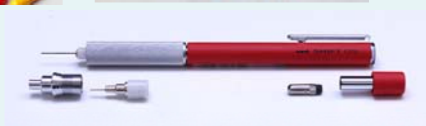
$\hat{p} = 0.12$, so use this to estimate p

Topic-3

Point and Confidence Intervals for Variances and Standard Deviations

Introduction





Confidence Intervals for σ^2 & σ

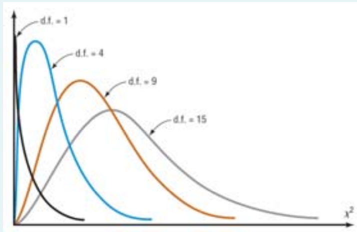
- When products that fit together (such as pipes) are manufactured, it is important to keep the variations of the diameters of the products as small as possible; otherwise, they will not fit together properly and will have to be scrapped.
- In the manufacture of medicines, the variance and standard deviation of the medication in the pills play an important role in making sure patients receive the proper dosage.
- For these reasons, confidence intervals for variances and standard deviations are necessary.

Chi-Square Distributions

- The chi-square distribution must be used to calculate confidence intervals for one population variance and standard deviation.
- The chi-square variable is similar to the *t* variable in that its distribution is a family of curves based on the number of degrees of freedom.
- The symbol for chi-square is χ^2 (Greek letter chi, pronounced "ki").
- A chi-square variable cannot be negative, and the distributions are skewed to the right.

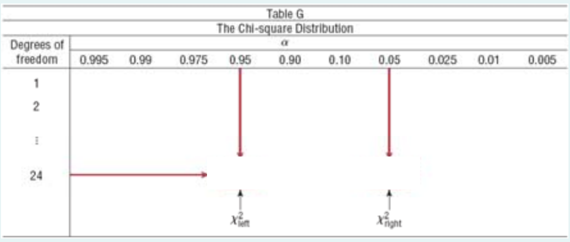
Chi-Square Distributions

- At about 100 degrees of freedom, the chi-square distribution becomes somewhat symmetric.
- The area under each chi-square distribution is equal to 1.00, or 100%.



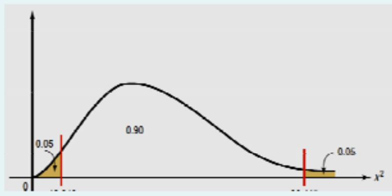
Example 7-13: Using Table G

Use the 0.95 and 0.05 columns and the row corresponding to 24 d.f. in Table G.



Example 7-13: Using Table G

Find the values for χ^2_{right} and χ^2_{left} for a 90% confidence interval when *n* = 25.



Formula for the C.I. for a Variance

- For a Variance

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}, \qquad \text{d.f.} = n-1$$

- For a Standard Deviation

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}, \qquad \text{d.f.} = n-1$$

Confidence Intervals for σ^2 & σ

Rounding Rule

When you are computing a confidence interval for a population variance or standard deviation by using raw data, round off to one more decimal places than the number of decimal places in the original data.

When you are computing a confidence interval for a population variance or standard deviation by using a sample variance or standard deviation, round off to the same number of decimal places as given for the sample variance or standard deviation.

Example 7-14: Nicotine Content

Find the 95% confidence interval for the variance and standard deviation of the nicotine content of cigarettes manufactured if a sample of 20 cigarettes has a standard deviation of 1.6 milligrams.

Solution:

Example 7-15: Cost of Ski Lift Tickets

Find the 90% confidence interval for the variance and standard deviation for the number of named storms per year in the Atlantic basin. A random sample of 10 years has been used. Assume the distribution is approximately normal.

10	5	12	11	13
15	19	18	14	16

Solution: