CHAPTER 5
(Part 1)

Point Estimation
&
Confidence Interval Estimator

Preview

- In Chapters 1, there are 2 main branches of statistics which are *Descriptive* and *Inferential*.
- In Chapter 4, we used *descriptive statistics* when we summarized data using tools such as graphs and statistics such as the mean and standard deviation.
- In *inferential statistics*, the two major activities of inferential statistics are (1) to use sample data to estimate values of a population parameters, and (2) to test hypotheses or claims made about population parameters.

What Are Point and Interval Estimates of Population Parameters?

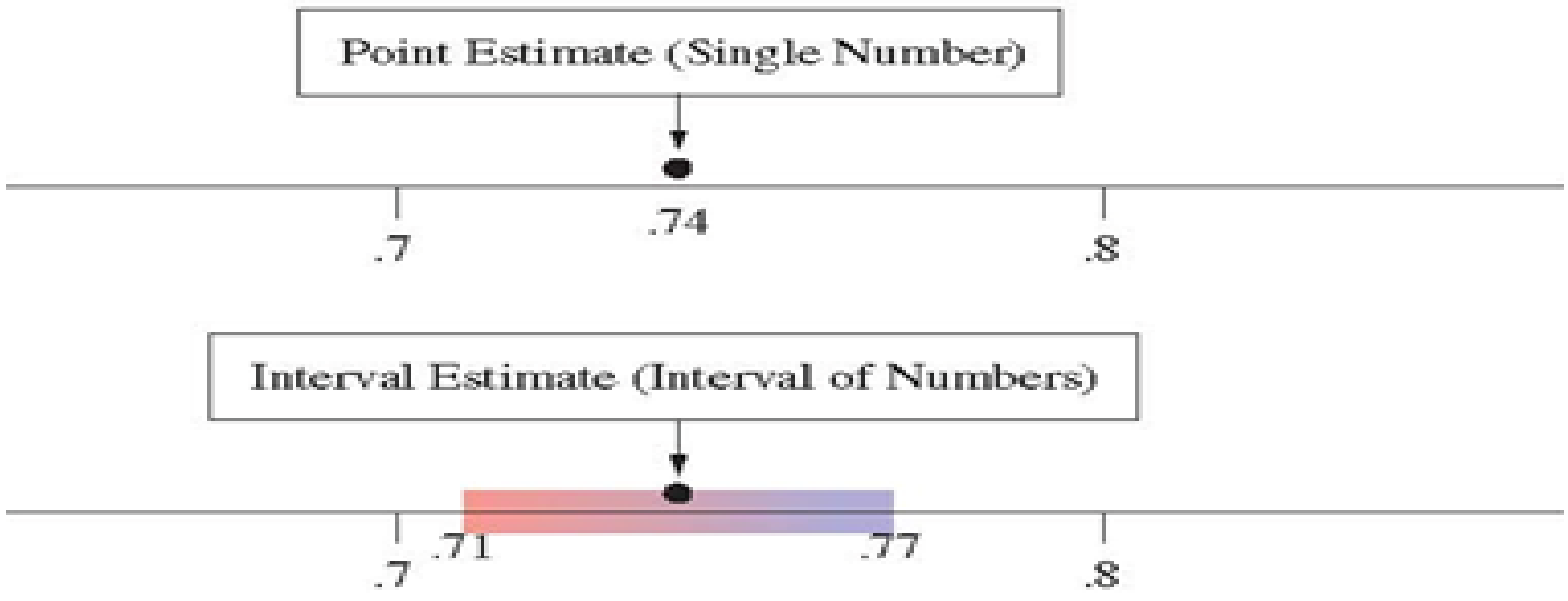
Point Estimate

- A ***point estimate*** is ***a single number*** that is our "best guess" for the parameter.

Interval Estimate

- An ***interval estimate*** is an ***interval of numbers*** within which the parameter value is believed to fall.

Point Estimate vs Interval Estimate



- A point estimate predicts a parameter by a single number.
- An interval estimate is an interval of numbers that are believable values for the parameter.

Why is a point estimate alone not sufficiently informative?

- A ***point estimate*** doesn't tell us how close the estimate is likely to be to the parameter.

- An ***interval estimate*** is more useful.
  - It incorporates a margin of error which helps us to gauge the accuracy of the point estimate.

Point Estimation

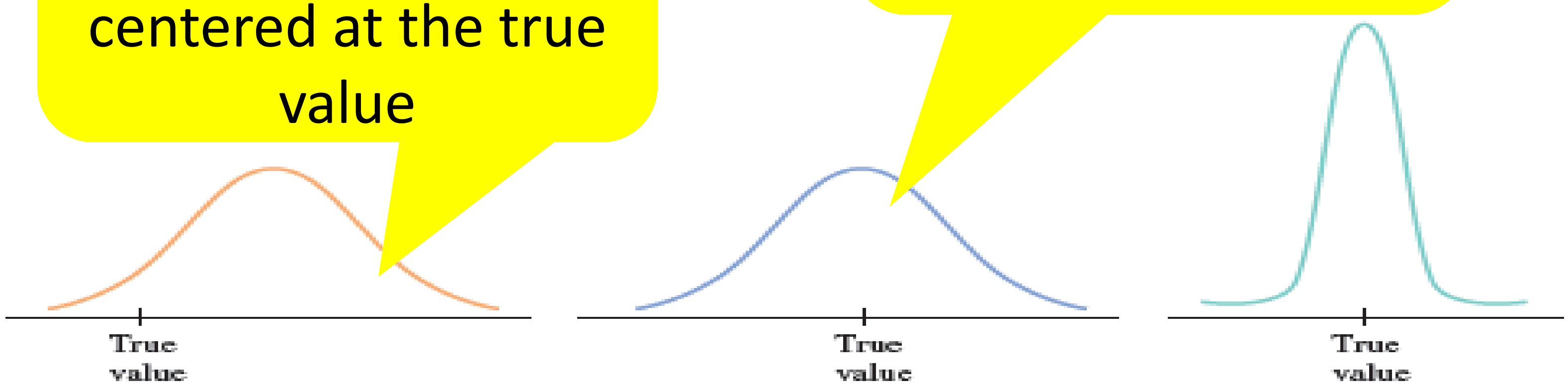
- How do we make a best guess for a population parameter?
- Use an appropriate sample statistic:
  - For the *population mean*, use the *sample mean*
  - For the *population proportion*, use the *sample proportion*
- *Point estimates* are the most common form of inference reported by the mass media.

Properties of Point Estimators

- Property 1: A good estimator has a sampling distribution that is centered at the parameter
  - An estimator with this property is *unbiased*
    - The sample mean is an unbiased estimator of the population mean
    - The sample proportion is an unbiased estimator of the population proportion
- Property 2: A good estimator has a *small standard error* compared to other estimators
    - This means it tends to fall closer than other estimates to the parameter

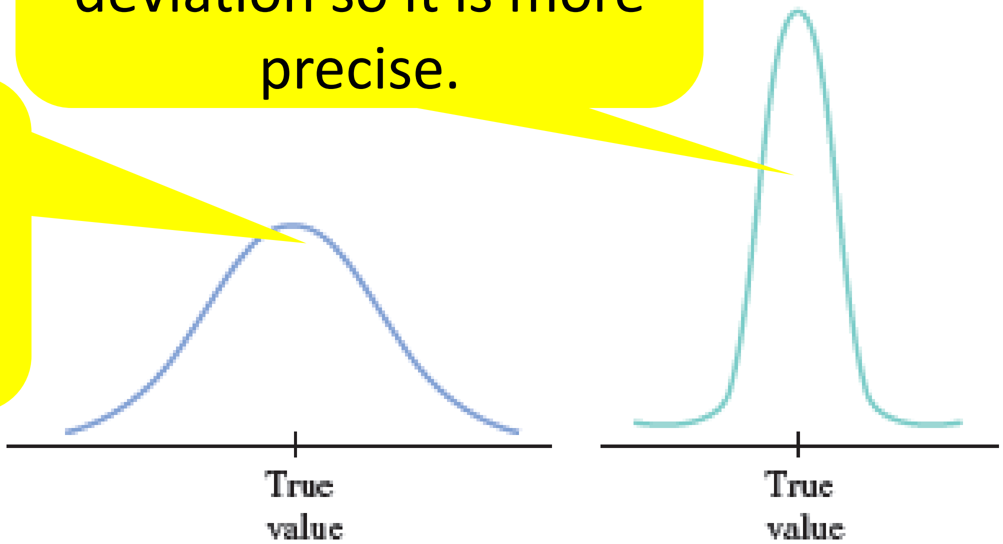
Biased, since the distribution is NOT centered at the true value

Unbiased, since the distribution is centered at the true value



Unbiased, but has a smaller standard deviation so it is more precise.

Unbiased, but has a larger standard deviation so it is not as precise.



Interval Estimation

- Constructing an interval that contains the parameter (we hope!)

- Inference about a parameter should provide not only a point estimate but should also indicate its likely precision

Confidence Interval (CI)

- A *confidence interval (CI)* is an interval containing the most believable values for a parameter.

- The probability that this method produces an interval that contains the parameter is called the *confidence level*
  - This is a number chosen to be close to 1, most commonly 0.95

- What is the logic behind constructing a confidence interval?
  - To construct a confidence interval for a population proportion, start with the *sampling distribution of a sample proportion.*

UTM UNIVERSITI TEKNOLOGI MALAYSIA

The Sampling Distribution of the Sample Proportion

- Gives the possible values for the sample proportion and their probabilities.
- Is approximately a normal distribution for large random samples.
- Has a mean equal to the population proportion.
- Has a standard deviation called the standard error.

Margin of Error

- The ***margin of error*** measures how accurate the point estimate is likely to be in estimating a parameter.

- The distance of 1.96 standard errors in the margin of error for a 95% confidence interval.

Confidence Interval Estimator for Proportion

The General Properties for Sampling Distributions of *p*

Let $\hat{p}$ be the proportion of successes (S) in a random sample of size n from a population whose proportion of S's is p. Denote the mean value of $\hat{p}$ by $\mu_{\hat{p}}$ and the std. deviation by $\sigma_{\hat{p}}$. The following rules hold.

Rule 1: $\mu_{\hat{p}} = p$

Rule 2: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

This rule is exact if the population is infinite, and is approximately correct if the population is finite and no more than 10% of the population is included in the sample.

Rule 3: When n is large and p is not too near 0 or 1, the sampling distribution of $\hat{p}$ is approximately normal.

The Large-Sample Confidence Interval for p

The general formula for a confidence interval for a population proportion:

$$\hat{p} \pm (z \text{ critical value})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

This is also called the margin of error/bound on the error estimation

Three Commonly Used Confidence Levels

- A 90% confidence interval for a population proportion $\hat{p}$

$$\hat{p} \pm 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- A 95% confidence interval for a population proportion $\hat{p}$

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- A 99% confidence interval for a population proportion $\hat{p}$

$$\hat{p} \pm 2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Four Commonly Used Confidence Levels

Confidence Level

$1-\alpha$	α	$\alpha / 2$	$z_{\alpha / 2}$
.90	.10	.05	$z_{.05} = 1.645$
.95	.05	.025	$z_{.025} = 1.96$
.98	.02	.01	$z_{.01} = 2.33$
.99	.01	.005	$z_{.005} = 2.575$

cut & keep handy!

Example #1

In 2000, the GSS asked: “Are you willing to pay much higher prices in order to protect the environment?”

- n = 1154 respondents, 518 were willing to do so.
- Find and interpret a 95% confidence interval for the population proportion of adult Americans willing to do so at the time of the survey.

Example#1 - Solution

- Take a sample of size *n*, calculate $\hat{p}$ (pronounced *p-hat*) as the *sample proportion* of people who have that characteristic.
- Calculate the standard error (SE).
- Calculate confidence interval (i.e., 95%).

Example#1 - Solution

$$\hat{p} = \frac{518}{1154} = 0.45$$

$$SE = \sqrt{\frac{(0.45)(0.55)}{1154}} = 0.015$$

$$\hat{p} \pm 1.96(SE) = 1.96(0.015)$$

$$= 0.45 \pm 0.03 = (0.42, 0.48)$$

Confident interval of *p*-hat

Example # 2

A recent study investigated the effects of a "Buckle Up Your Toddlers" campaign to get parents to use the grocery cart seat belts. Investigators observed a representative sample of parents at grocery stores in a large city, and found 192 out of 594 parents buckling up their toddlers.

a. Compute a point estimate of the true proportion of all parents who buckle up their toddlers.

b. Construct and interpret a 95% confidence interval for the true proportion of all parents who buckle up their toddlers.

Example#2 - Solution

a) point estimate: $\hat{p}$ = 192 / 594 = 0.3232

b) 95% CI: $= 95\%\ \text{CI:}\ 0.3232 \pm 1.960\sqrt{\dfrac{0.3232(0.6768)}{594}} = (0.2856,\ 0.3608).$

I am 95% confident that the true proportion of parents who buckle up their toddlers is between 0.286 and 0.361.

Exercise #1

A sample of 350 students from a college (with a total student population of 25,000) were asked is either both their parents had been to college. In this case, the given sample proportion was 0.789. What is the 90% CI?

Confidence Interval Estimator for Mean

Confidence Interval for mean when σ is known

- **The general formula for a confidence interval (CI) for a population mean when:**
  - x is the sample mean from a random sample,
  - the sample size n is large ($n \geq 30$), and
  - σ, the population standard deviation, is known.

$$\bar{x} \pm (z \text{ critical value})\left(\frac{\sigma}{\sqrt{n}}\right)$$

Example#1

A computer company samples demand during lead time over 25 time periods:

235	374	309	499	253
421	361	514	462	369
394	439	348	344	330
261	374	302	466	535
386	316	296	332	334

Its is known that the standard deviation of demand over lead time is 75 computers. We want to estimate the ***mean*** demand over lead time with 95% confidence level in order to set inventory levels.

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Example#1 - Solution

- "We want to estimate the ***mean*** demand over lead time with 95% confidence level in order to set inventory levels…"
- Thus, the parameter to be estimated is the population mean: μ
- And so our confidence interval estimator will be:

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

Example#1 - Solution

- In order to use our confidence interval estimator, we need the following pieces of data:

$\bar{x}$	370.16	Calculated from the data…
$z_{\alpha/2}$	1.96	$1-\alpha = .95, \ \therefore \alpha/2 = .025$ $so \ \ z_{\alpha/2} = z_{.025} = 1.96$
σ	75	Given
n	25	Given

therefore: $\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 370.16 \pm z_{.025}\dfrac{75}{\sqrt{25}} = 370.16 \pm 1.96\dfrac{75}{\sqrt{25}} = 370.16 \pm 29.40$

The **lower** and **upper** confidence limits are 340.76 and 399.56.

Confidence Interval for μ when σ is unknown

- The confidence interval just developed has an obvious drawback:
  - To compute the interval endpoints, σ must be known.
- If σ is unknown, we must use the sample data to estimate σ.
- If we use the sample standard deviation as our estimate, the result is a different standardized variable denoted by *t*:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

One-sample t confidence interval for μ

- $\bar{x}$ is the sample mean from a simple random sample.
- The population distribution is normal, or the sample size n is large (generally, $n \geq 30$), and
- σ, the population standard deviation, is unknown.

One-sample t confidence interval for μ is,

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

where the t critical value is based on degree of freedom, df = n-1.

Example #2

The article "Chimps Aren't Charitable" (*Newsday*, November 2, 2005) summarized the results of a research study published in the journal *Nature*. In this study, chimpanzees learned to use an apparatus that dispersed food when either of two ropes was pulled. When one of the ropes was pulled, only the chimp controlling the apparatus received food. When the other rope was pulled, food was dispensed both to the chimp controlling the apparatus and also a chimp in the adjoining cage. The accompanying data represent the number of times out of 36 trials that each of seven chimps chose the option that would provide food to both chimps (charitable response).

23 22 21 24 19 20 20

Compute a 99% confidence interval for the mean number of charitable responses for the population of all chimps.

Example#2 - Solution

Data: **23 22 21 24 19 20 20**

- $\bar{x}$ = 21.29 and s = 1.80
- df = 7 – 1 = 6 ; confidence level = 99% (α=0.01)

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

$$21.29 \pm 3.71\left(\frac{1.80}{\sqrt{7}}\right) = (18.77, 23.81)$$

We are 99% confident that the mean number of charitable responses for the population of all chimps is between 18.77 and 23.81.

Exercise # 2

A new species of sea crab has been discovered, and an experiment conducted to determine whether or not the animal can regulate its temperature. If the animal can maintain a body temperature different from the surroundings, this would be considered evidence of regulating capability. Ten of these sea crabs were exposed to ambient temperatures of 24 degrees Celsius. Their body temperatures were measured with the results below:

24.33, 24.61, 24.67, 24.64, 24.42, 24.97, 25.23, 24.73, 24.90, 24.44

For purposes of this example, assume that it is reasonable to regard these 10 crabs as a random sample from the population of all crabs of this species.

a) Calculate a point estimate of the population mean, μ

b) Construct and interpret a 99% confidence interval for μ.

Exercise # 3

Logging activity in forests is thought to affect the behavior of black bears (*Ursus americanus*). An important measure of animal behavior is the home range, the area used by animals in their lives. In a study of black bears in a logged Canadian forest, the spring and early summer home range (in) of 12 radio-collared female black bears was measured with the following results:

39.9, 23.5, 42.1, 29.4, 34.4, 40.9, 27.9, 22.3, 13.0, 20.1, 13.3, 8.6

a) Construct and interpret a 95% confidence interval for the mean home range of female black bears in this logged forest.

b) The typical home range of females in forests with no logging is 20 . Based on the confidence interval from part (a), do you think that the mean home range size of females in this logged forest could be the same as the mean home range size in non-logged forests? Explain, using appropriate statistical terminology.