Sri Lanka Institute of Information Technology

Faculty of Computing

IT2120 - Probability and Statistics

Ms. K. G. M. Lakmali

Year 02 and Semester 01

Lecture 8

STATISTICAL INFERENCE

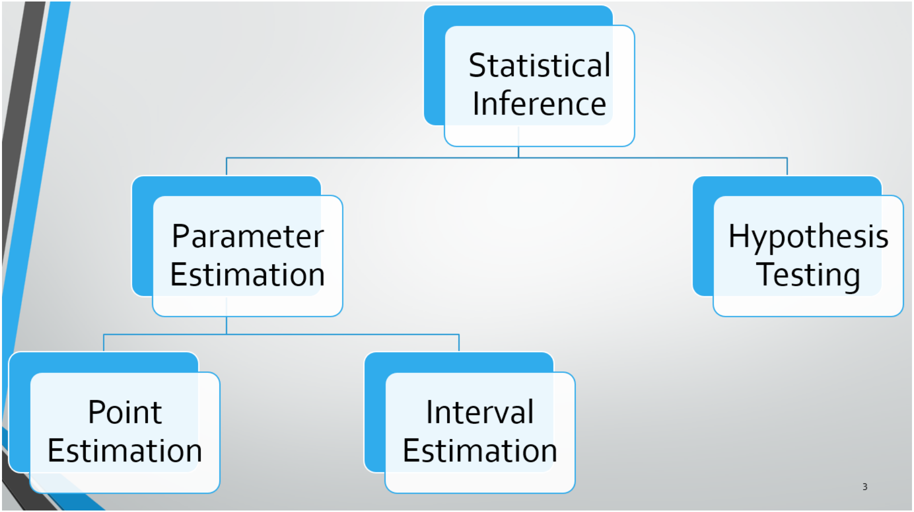
Statistical Inference

- In most researches, we collect data through a sample survey over a census.
- Statistical inference is used when sample survey is conducted over a census.
- Inference: A conclusion reached on the basis of evidence and reasoning.

  - Oxford University Press -

- Statistical Inference: Drawing conclusions about population parameters by using sample statistics.

Statistical Inference



PARAMETER ESTIMATION

Parameter Estimation

- In distribution theory we assumed that distribution parameters are known.
- But practically they should be found or estimated.
- If estimated parameters are wrong, all calculated probabilities will be inaccurate.
- Estimation can be done in two methods.
    - Point estimation
    - Interval estimation

Parameter Estimation

- Point estimation gives a single estimated value for the parameter.
- Interval estimation gives a range of values (interval) as the estimate.
- There are many point and interval estimation methods with their own criteria for use.
- Some interval estimates will be discussed later in this chapter.

HYPOTHESIS TESTING

Hypothesis Testing

- Hypothesis: A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.

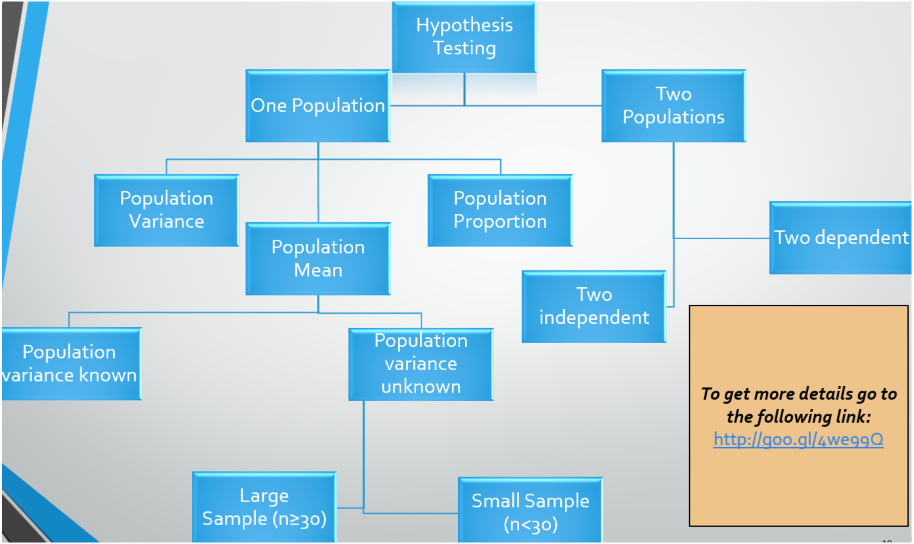
  -Oxford University Press-
- Hypothesis testing is all about checking whether assumptions (research hypothesis) are correct.
- These assumption should be regarding population parameters.

Major Steps under Hypothesis Testing

1. Define the hypothesis (H_0 & H_1)
2. Test statistic and its distribution
3. Define the significance level (α)
4. Define the rejection region.
5. Conduct the test (Calculate test statistic value)
6. Conclusion

- There are various cases under hypothesis testing. The test statistic that you should use depends on the case.
- In this session, we will discuss the hypothesis testing for one population mean.

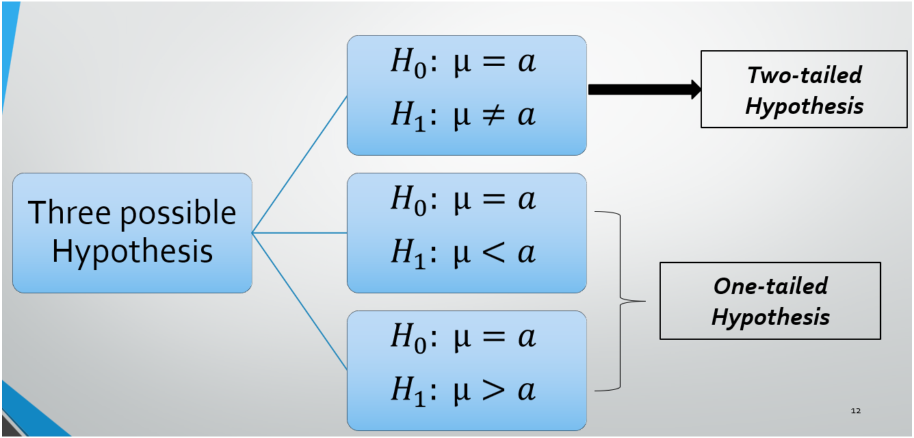
Hypothesis Testing



Defining Hypothesis

- The assumption should be clearly stated in order to test.
- Two statements, null hypothesis (H_0) and an alternative hypothesis (H_1 or H_a) are used for that.
- H_0 and H_1 can be considered as opposites of each other.
- The statement with the equal ($=$) should always come to H_0. Usually if a claim is made, it is selected for H_1.

Defining Hypothesis



Examples

1) In a coin tossing experiment, it should be found whether

a) it's fair coin or not.

b) it's biased in favor of heads.

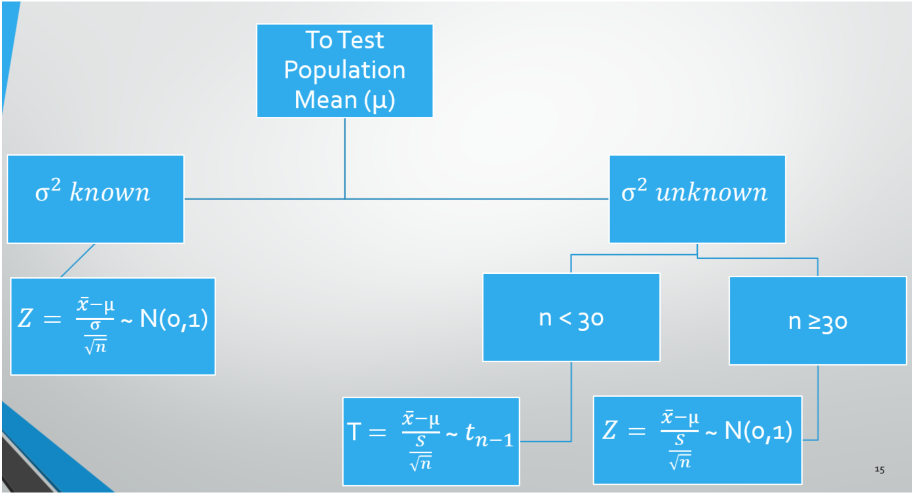
c) it's biased in favor of tails.

2) A company that manufactures cars claims that the gas mileage for its new line of hybrid cars, on the average, is 60 miles per gallon (mpg) with a standard deviation of 4 mpg. It was also found out that the mpg was normally distributed. A random sample of 16 cars yielded a mean of 57 miles per gallon. Is the company's claim about the mean gas mileage per gallon of its cars, correct?

Test Statistic

- **Recap**: A function of observable r.v.s that does not depend on any unknown parameters is called a statistic.
- A test statistic is a quantity associated with the sample.
- The test statistic will depend on the **parameter of interest** as well as the **characteristics of the population**.
- We assume that the assumption (H_0) is correct and find a sampling distribution for the test statistic.

Test Statistic & Distribution



Test Statistic [For μ - When σ^2 known]

- Recap: Let $X_1, \ldots, X_n$ be a random sample of size n from a Normal population with mean μ and variance σ^2. Then,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Then,

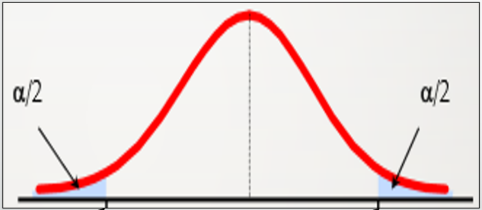
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- If the hypothesis is, $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, then under H_0,

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Rejection Region [For μ]

For a two-tailed hypothesis



$-Z_{\alpha/2}$ or $-t_{(n-1),\alpha/2}$ **(a1)**

$Z_{\alpha/2}$ or $t_{(n-1),\alpha/2}$ **(a2)**

Critical Values

Reject H_0 if $Z_{cal} \geq$ $\boldsymbol{a2}$
OR if $Z_{cal} \leq$ $\boldsymbol{a1}$

17

For a one-tailed hypothesis

$$H_0 : \mu = a$$
$$H_1 : \mu < a$$

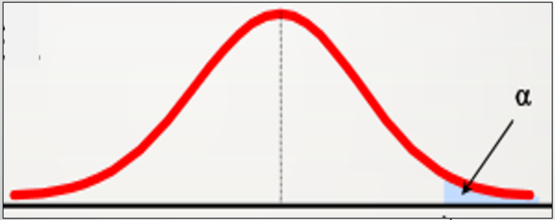
$-Z_\alpha$ *or* $-t_{(n-1),\alpha}$ ***(a1)***

Critical Value

Reject H_0 if $Z_{cal} <$ $\boldsymbol{a1}$

For a one-tailed hypothesis

$$H_0 : \mu = a$$
$$H_1 : \mu > a$$



Critical Value → Z_α or $t_{(n-1),\alpha}$ **(a1)**

Reject H_0 if $Z_{cal} >$ **a1**

Example 02:

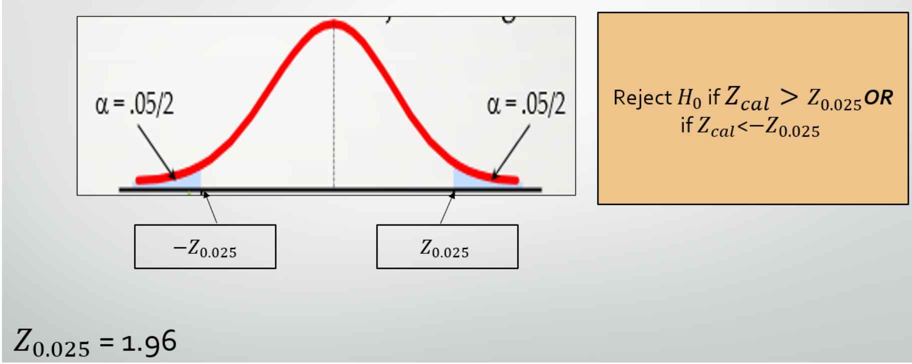
- $H_0 : \mu = 60$
  $H_1 : \mu \neq 60$ $\Bigg\}$ **Two-tailed hypothesis**

- ***Test Statistic***: Under H_0,

$$Z = \frac{\bar{x} - 60}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

- *Consider 5% level of significance.*

Rejection Region:



- ***Test:***

$\bar{x} = 57$, $\sigma = 4$ & $n = 16$

Then,

$$Z_{Cal} = \frac{\bar{x} - 60}{\frac{\sigma}{\sqrt{n}}}$$

$$Z_{Cal} = \frac{57 - 60}{\frac{4}{\sqrt{16}}}$$

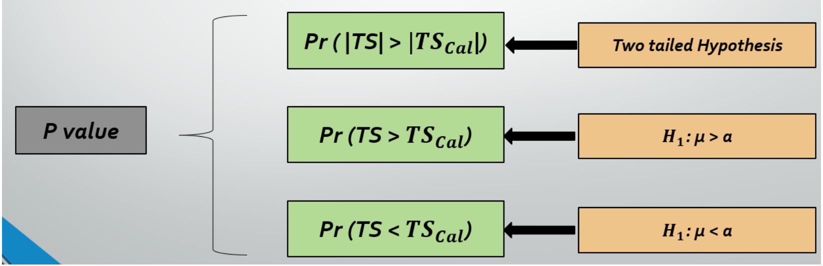
$$Z_{Cal} = -3$$

- ***Conclusion:***

Since $Z_{Cal} = -3 < -1.96 = Z_{0.025}$, we reject H_0 at 5% level of significance. Therefore, we can conclude that company's claim about the mean gas mileage per gallon of its cars is incorrect.

P value Approach

- This is an alternative way of get the decision in hypothesis testing.
- **P value**: The probability of obtaining a test statistic which is more extreme than observed test statistic value given when H_0 is true.



For any test,

$$\textit{If p value} \leq \textit{significance level } (\alpha) \longrightarrow \textit{Reject } H_0$$

$$\textit{If p value} > \textit{significance level } (\alpha) \longrightarrow \textit{Do not Reject } H_0$$

- P value is a measure of the strength of evidence in the data against H_0
- This is the smallest value of α for which H_0 can be rejected and actual risk of committing type I error.
- P value also know as observed significance level.

Errors in Hypothesis Testing

Statistical Decision	True State of the Null Hypothesis	
	H_0 is True	H_0 is False
Reject H_0	Type I Error	Correct
Do not Reject H_0	Correct	Type II Error

$$Pr\ (\text{Type I Error}) = Pr\ (\text{Reject } H_0 | H_0 \text{ true}) = \alpha$$

$$Pr\ (\text{Type II Error}) = Pr\ (\text{Do not Reject } H_0 | H_0 \text{ false}) = \beta$$


THANKS!
Any questions?