# STATISTICS FOR DATA SCIENCE

## HYPOTHESIS and INFERENCE

**Dr. Deepa Nair**
Department of Science and Humanities

# STATISTICS FOR DATA SCIENCE

## UNIT-4 HYPOTHESIS and INFERENCE

## Session-5

## Drawing Conclusions from the Results of Hypothesis Tests

**Dr. Deepa Nair**

Department of Science and Humanities

## Choose $H_0$ to Answer the Right Question:

- **When performing a hypothesis test, it is important to choose $H_0$ and $H_1$ appropriately so that the result of the test can be useful in forming a conclusion.**

## Choose $H_0$ to Answer the Right Question:

**Example:**

- Specifications for a water pipe call for a mean breaking strength $\mu$ of more than $2000\ lb$ per linear foot.
- Engineers will perform a hypothesis test to decide whether or not to use a certain kind of pipe.
- They will select a random sample of $1\ ft$ sections of pipe, measure their breaking strengths, and perform a hypothesis test.

## Choose $H_0$ to Answer the Right Question:

**Example:**

- The pipe will not be used unless the engineers can conclude that $\mu > 2000$.
- Assume they test $H_0\!: \mu \le 2000$ *versus* $H_1\!: \mu > 2000$.
- Will the engineers decide to use the pipe if $H_0$ is rejected? What if $H_0$ is not rejected?

## Choose $H_0$ to Answer the Right Question:

**Solution:**

- **If $H_0$ is rejected, the engineers will conclude that $\mu > 2000$, and they will use the pipe.**

- **If $H_0$ is not rejected, the engineers will conclude that $\mu$ *might* be less than or equal to $2000$, and they will not use the pipe.**

## Choose $H_0$ to Answer the Right Question:

**Example:**

- Assume the engineers test $H_0: \mu \geq 2000$ $versus$ $H_1: \mu < 2000.$
- Will the engineers decide to use the pipe if $H_0$ is rejected? What if $H_0$ is not rejected?

## Choose $H_0$ to Answer the Right Question:

**Solution:**

- If $H_0$ is rejected, the engineers will conclude that $\mu < 2000$, and they will not use the pipe.

- If $H_0$ is not rejected, the engineers will conclude that $\mu$ *might* be greater than or equal to $2000$, but that it also might not be. So again, they won't use the pipe.

## Statistical Significance Is Not the Same as Practical Significance:

- The $P$-value does not measure practical significance.
- What it does measure is the degree of confidence we can have that the true value is really different from the value specified by the null hypothesis.

## Statistical Significance Is Not the Same as Practical Significance:

- **When the *P*-value is small, then we can be confident that the true value is really different.**
- **This does not necessarily imply that the difference is large enough to be of practical importance.**

## The Relationship Between Hypothesis Tests and Confidence Intervals:

- The values contained within a two-sided level $100(1-\alpha)\%$confidence interval for a population mean $\mu$ are precisely those values for which the $P$-value of a two-tailed hypothesis test will be greater than $\alpha$.

- Example: the 95% confidence interval consists of precisely those values of μ whose P-values are greater than 0.05 in a hypothesis test.

## The Relationship Between Hypothesis Tests and Confidence Intervals:

- **A one-sided level $100(1-\alpha)\%$ confidence interval consists of all the values for which the *P*-value in a one-tailed test would be greater than $\alpha$.**

- The confidence level is equivalent to $(1 - \alpha)$ level. So, if your significance level is 0.05, the corresponding confidence level is 95%.

## Rejection Region approach for Hypothesis Test

### Critical Point & Rejection Region

- A critical point is a value of the test statistic that produces a P-value exactly equal to $\alpha$.
- The region on the side of the critical point that leads to rejection is called the rejection region.
- The critical point itself is also in the rejection region.

**Example:**

- A machine runs on an average of 125 hours/year.
- A random sample of 49 machines has an annual average use of 126.9 hours with standard deviation 8.4 hours.
- Does this suggest to believe that machines are used on the average more than 125 hours annually at 0.05 level of significance.

**Solution:**

$\mu =$ average number of hours a machine runs in an year.

$H_0. \mu \leq 125$ hours/year , $H_1. \mu > 125$

L.O.S.: $\propto = 0.05$

Calculation: $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{126.9-125}{8.4/\sqrt{49}}$ = 1.58

P- value is $.0571 > 0.05$

So we need to accept $H_o$.

We can not believe that machine works more than 125 hours in an year.

**Example:**

- A manufacture of tyres guarantees that the average lifetime of its tyres is more than 28000 miles.

- If 40 tyres of this company tested, yields a mean lifetime of 28463 miles with s.d. of 1348 miles.

- Can the guarantee be accepted at 0.01 L.O.S.?

**Solution:**

Significance level $\alpha = 0.01$

$$H_0: \mu \leq 28000, \qquad H_1: \mu > 28000$$

$$\bar{X} = 28463\, miles, n = 40 \;\; \sigma \to s = 1348\, miles$$

z$= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{28463-28000}{1348/\sqrt{40}}$=2.17

P value: $0.015 > 0.01$

We need to Reject the null hypothesis.

**Example:**

- Can it be concluded that the average lifespan of Indian is more than 70 years if a random sample of 100 Indians has an average lifespan of 71.8years with a s.d of 8.9 years

**Solution:**

Significance level $\alpha = 0.05$

$H_0: \ \mu \leq 70 \ years, \ \ H_1: \ \mu > 70$years

$\bar{X} = 71.8 \ years, \qquad n = 100 \ \ \sigma \rightarrow s = 8.9 \ years$

z$= \ \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{71.8-70}{8.9/\sqrt{100}}$=2.02

P value: $0.015 > 0.01$

We need to Reject the null hypothesis.

**Dr. Deepa Nair**
Department of Science and Humanities

**deepanair@pes.edu**