

Hypothesis Testing

Problem Statement: Aravind Productions(AP) is a newly formed movie production house based out of Mumbai, India. AP was interested in understanding the production cost required for producing a Bollywood movie. The industry believes that the production house will require at least INR 500 million(50 crore) on average. It is assumed that the Bollywood movie production cost follows a normal distribution. Production costs of 40 Bollywood movies in millions of rupees are given in the Table. Conduct an appropriate hypothesis test at $\alpha=0.05$ to check whether the belief about average production cost is correct.

Table: Production cost of Bollywood movies(million INR)

601	627	330	364	562	353	583	254	528	470
125	60	101	110	60	252	281	227	484	402
408	601	593	729	402	530	708	599	439	762
292	636	444	286	636	667	252	335	457	632

Solution:

Industry belief: The production house will require at least INR 500 million(50 crore) on average.

Assumption: Bollywood movie production cost follows a normal distribution.

The null hypothesis (H_0): ($\mu = 500$ million)

The alternative hypothesis (H_1): ($\mu > 500$ million)

Here, the standard deviation of the sample is not given hence T-test is used for testing the hypothesis.

Calculate the mean and standard deviation of the sample:

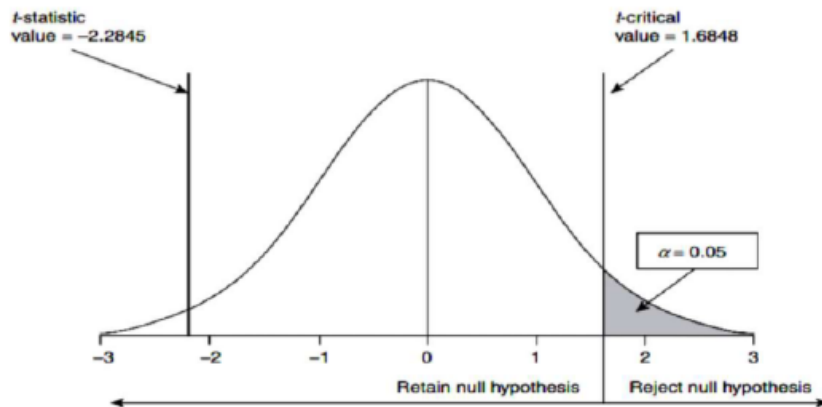
$$\begin{aligned}n &= 40 \\ \bar{X} &= 429.55 \\ S &= 195.0337\end{aligned}$$

$$\begin{aligned}t\text{-statistic} &= \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{429.55 - 500}{195.0337/\sqrt{40}} \\ &= -2.2845\end{aligned}$$

Calculate the p-value:

Degree of freedom($n-1$) = 39

t-critical = 1.6848 [computed using Excel by inserting the formula, T.INV(0.05, 39)]



t-statistic < t-critical

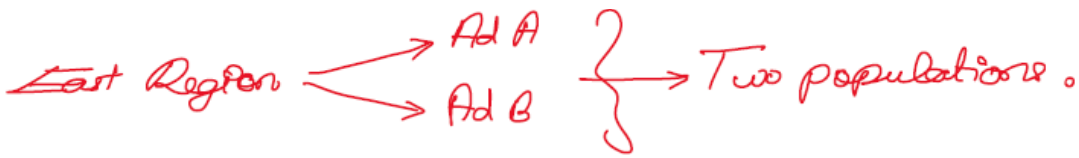
Fail to reject the null hypothesis.

Problem Statement: You are planning the campaign of the Chief Minister of a major state going to election in a month's time. You want to figure out the effectiveness of a 10 sec YouTube ad in increasing the popularity of the CM's party in a particular region. If a person clicks on the ad, then that ad is effective in drawing attention. Assume that YouTube can deliver ads to a particular location and that "Skip Ad" is disabled as the video is only 10 sec. Currently, Ad A is running in both regions. You propose to use Ad B. So you isolate 2 hours for running a test (Serve Ad A for 1 hour and Serve Ad B for 1 hour), and obtain the following data:

Region	Ad	Shown to	Clicked by
East	A	10,000	100
East	B	5,000	75
West	A	9,000	100
West	B	10,000	115

1. Estimate, statistically at 1% significance, if Ad B is preferred in the East region.

Solution:



Click through rate \rightarrow Fraction

A/B testing

Two modalities :-

Control group : Ad A

Treatment group : Ad B

Steps:

1. Formulating Hypothesis
2. Define test statistic
3. Calculate the p-value
4. Compare against a threshold value

Step 1: Formulating the hypothesis

Null hypothesis (H_0): There is no difference between the control and treatment group.

$$P_a = P_b = P$$

Alternative hypothesis (H_1): There is a real effect to the treatment.

$$P_b > P_a$$

Step 2: Set the test statistic

$$Z_{obs} = \frac{\hat{p}_B - \hat{p}_A}{SE_0(\hat{p}_B - \hat{p}_A)}$$

$$\hat{p}_A = \frac{100}{10000} = 0.01, \hat{p}_B = \frac{75}{5000} = 0.015, \hat{p} = \frac{100+75}{15000} = 0.0117$$

SE_0 is calculated using the pooled probability \hat{p} as

$$SE_0 = \sqrt{0.0117 \times 0.9883 \times \left(\frac{1}{10000} + \frac{1}{5000} \right)} = 0.0019$$

$$Z_{obs} = \frac{0.015 - 0.01}{0.0019}$$

$$Z_{obs} = 2.63$$

Step 3: Calculate the p-value

$$\text{P-Value} = P(Z \geq Z_{obs}) = P(Z \geq 2.6316) \approx 1 - 0.9958 = 0.0042$$

Step 4: Compare against a threshold

P-Value is less than the threshold of 0.01. Hence there evidence against the null hypothesis. We reject the null hypothesis and conclude that there is a statistically significant preference for Ad B in East Region.

2. Repeat the same for west region and combined state.