Hypothesis Testing :-

Estimation

- We have been discussing "models" — linear model

- Ingredients :- Population of interest

  Sample data from the population

  Assume: it comes from a certain model

  - Estimate the parameters

Focus now shifts :-

- interested whether two sub-populations are different or the same?

- are the measured attributes independent of each other?

Examples :-

- Are temperatures today higher than they were 100 years ago?

- Does smoking reduce life expectancy?

- Is treatment A genuinely different from treatment B?

— Test of a hypothesis —

Two steps:

(a) Make a conjecture

(b) Perform some computation to test the conjecture.

Probability Theory

Statistics

calculations to test conjecture

Example :- • Given a coin and we are interested in the probability $-p-$ of showing heads, when tossed.

• Toss the coin 100 times, $X_1, X_2, \ldots, X_{100}$

$$X_i \sim \text{Bernoulli}(p)$$

Find: $\sum_{i=1}^{100} X_i = 67$

– till now we use this to estimate p

Different Question: Is $p = 0.5$ a valid hypothesis given the finding?

observation :- If the coin had an equal chance of Heads and tails then chance of 67 heads in 100 tosses

$$ {}^{100}C_{67} \left(\tfrac{1}{2}\right)^{100} \approx 0.04$$

Z-test :- Suppose $X_1, X_2, \ldots$ are i.i.d Normal(μ, σ^2) when σ is known but μ is unknown.

let $X_1, X_2, \ldots, X_n$ i.i.d sample from the population

- Compute & Find: $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

c - known - conjectured value.

Hypothesis Test :- Is $\underbrace{\mu = c}_{\text{null hypothesis}}$ or $\underbrace{\mu > c}_{\text{alternate hypothesis}}$?

Question: Given $\bar{X}$; If null hypothesis was true how likely is it that we would have a sample mean as large as the observed value $\bar{X}$?

Answer :- $X_1, X_2, \ldots, X_n$ are known.

let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. Normal (c, σ^2)

Compute: $\mathbb{P}(\bar{Y} \geq \bar{X})$

Random variable

Deterministic quantity.

(Test statistic)

$\mathbb{P}(\bar{Y} \geq \bar{X}) \equiv$ describes how likely it is that the test statistic $\bar{X}$ would be at least as far away from c as what was observed.

<u>Define</u> :- $Z = \frac{\sqrt{n}\,(\bar{Y} - c)}{\sigma} \sim \text{Normal}(0,1)$

$$\mathbb{P}(\bar{Y} \geq \bar{X})$$

$$= \mathbb{P}\left(\frac{\sqrt{n}\,(\bar{Y} - c)}{\sigma} \geq \frac{\sqrt{n}\,.(\bar{X} - c)}{\sigma}\right)$$

$$= \mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{x}-c)}{\sigma}\right)$$

$$\boxed{\text{Compute :- } \mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{x}-c)}{\sigma}\right)}$$

"We Reject null hypothesis if the <u>probability is small</u>"

↑ variable at hands of the user.

- $\boxed{\text{Fix level } \alpha \in (0,1)}$

  - If $\mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{x}-c)}{\sigma}\right) < \alpha$

    then we can conclude that the sample average $\bar{X}$ is so far from c that hypothesis $\mu = c$ is not true!

Example 1:- Suppose $X \sim \text{Normal}(\mu, 9)$

A sample $X_1, X_2, \ldots X_{16}$ is drawn from X and we observe $\bar{x} = 10.2$

Null hypothesis :- $\mu = 9.5$

Alternative hypothesis :- $\mu > 9.5$

level of significance :- $\alpha = 0.05$

Answer: $\bar{x} = 10.2$, $c = 9.5$, $\sigma = 3$, $n = 16$

Compute:-

$$P\left(Z \geq \frac{\sqrt{n}(\bar{x} - c)}{\sigma}\right)$$

$$= P\left(Z \geq \frac{\sqrt{16}(10.2 - 9.5)}{3}\right)$$

$$= P\left(Z \geq \frac{4(0.7)}{3}\right)$$

$\boxed{Z \sim N(0,1)}$

Tables or R

$$\approx 0.175$$

observe:- $\alpha = 0.05$

we have

$$P\left(Z \geq \frac{\sqrt{n}(\bar{x} - c)}{\sigma}\right) \geq \alpha$$

Conclusion: At level α we cannot reject the null hypothesis $\square$

Example 2 :-

Suppose $X \sim \text{Normal}(\mu, 9)$

A sample $X_1, X_2, \ldots X_{16}$ is drawn from X and we observe $\bar{X} = 10.2$

Null hypothesis :- $\mu = 8.5$

Alternative hypothesis :- $\mu > 8.5$

level of significance :- $\alpha = 0.05$

Answer :- $\bar{X} = 10.2$, $c = 8.5$, $n = 16$, $\sigma = 3$, $\alpha = 0.05$

Compute :- $\mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{X} - c)}{\sigma} \right)$

$$= \mathbb{P}\left(Z \geq \frac{4(10.2 - 8.5)}{3} \right)$$

$$= \mathbb{P}\left(Z \geq \frac{4(1.7)}{3} \right)$$

As $Z \sim N(0,1)$ from tables

$$\simeq 0.012$$

observe :- $\alpha = 0.05$

we have

$$\mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{x} - c)}{\sigma} \right) \simeq 0.012 < \alpha$$

Conclusion: At level α, as $\mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{x} - c)}{\sigma} \right)$ is less than α, we reject the null hypothesis.