Parameter-Estimation and Hypothesis-Testing

By
Dr. Anjula Mehto
Assistant Professor
Thapar Institute of Engineering and Technology,
Patiala, Punjab
Email-anjula.mehto@thapar.edu

Parameter Estimation

- There are many different distributions for random variables and all of those distributions have parameters.
- What if we don't know the values of the parameters and we can't estimate them from our own expert knowledge?
- What if instead of knowing the random variables, we have a lot of examples of data generated with the same underlying distribution?

Parameter Estimation

- Parameter Estimation is **a branch of statistics that involves using sample data to estimate the parameters of a distribution**.

Parameters of Different Distributions

- The parameters are the numbers that yield the actual distribution.

Distribution	**Parameters**
Bernoulli(p)	$\theta = p$
Poisson(λ)	$\theta = \lambda$
Uniform (a,b)	$\theta = [a,b]$
Normal (μ, σ^2)	$\theta = [\mu, \sigma$^2]

Parameter Estimations Techniques

- Method of Moments (MoM)
- Maximum Likelihood Estimation (MLE)
- Minimum Mean Square Error (MMSE) Estimation

Method of Moments

- **Estimating the Moments about population mean (μ).**
- Involves equating the sample moments of a statistical distribution to the corresponding population moments.
- The sample moments are computed from the observed data, while the population moments are expressed in terms of the parameters of the distribution.

Method of Moments

- The basic idea of MoM is to estimate the parameters of the distribution by solving a set of equations that equate the sample moments to the population moments.

Method of Moments

- Suppose that we take a random sample from a rectangular distribution, i.e., a uniform distribution over [a, b]. The 2 parameters are 'a' and 'b' here.
- We have to estimate these 2 parameters given a sample of size 'n'.

Method of Moments

- Let $(\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n})$be a random sample of size '*n*' *taken from the variable 'X'.*
- $E(X) = (a + b)/2 = (\boldsymbol{x\_1}, + \boldsymbol{x\_2} + \cdots + \boldsymbol{x\_n})/n = m$ *(say);*
- This gives us **a + b = 2m………….(i)**
- **Var(X)=E(X^2)-(E(X))^2**
- E(X^2)= Var(X) + (E(X))^2
- = (b-a)^2/12 + ((a + b)/2)^2
- = (a^2 + b^2 + ab)/3
- = (x1^2+ x2^2 +… + xn^2)/n = p (say)

Method of Moments

- **This gives us (a^2 + b^2 + ab) = 3p…………(ii)**
- Solving (i) and (ii) for a and b gives the estimates using method of moments.
- a = m -sqrt(3.p -3.m.m)
- b = 2.m –a = m + sqrt(3.p – 3.m.m)

Method of Moments

- Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size 'n' *taken from a Normal Population with parameters: mean* θ_1 *and variance* θ_2.
- Estimate these 2 parameters given a sample of size n ?

Method of Moments

- For the normal distribution, the population mean and variance are given by:
- $\mu = \theta_1$
- $\sigma^2 = \theta_2$
- The sample mean and variance can be calculated from the sample:
- $\bar{x} = (x_1 + x_2 + \ldots + x_n) / n$
- $s^2 = ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2) / (n - 1)$

Method of Moments

- Setting the sample mean equal to the population mean, we get:
- $\bar{x} = \theta 1$
- Solving for $\theta 1$, we get:
- $\theta 1 = \bar{x}$
- Setting the sample variance equal to the population variance, we get:
- s^2 = $\theta 2$

Method of Moments

- Solving for $\theta 2$, we get:
- $\theta 2 = s^2$
- Therefore, the method of moments estimators for $\theta 1$ and $\theta 2$ are:
- $\theta 1 = \bar{x} = (x_1 + x_2 + \ldots + x_n)/n$ (sample mean)
- $\theta 2 = S^2 = ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2) / (n - 1)$ (sample variance)

Maximum Likelihood Estimation

- The maximum likelihood estimation (MLE) estimates the population parameters by finding the values that maximize the likelihood function.
- For a discrete distribution, maximize the PMF of the data.
- For a continuous distribution, maximize the PDF of the data.

Maximum Likelihood Estimation

- Let us take a sample of size 'n' *from an exponential population having the parameter* 'λ'*. So, 'X' is exponentially distributed with parameter* 'λ'.

Maximum Likelihood Estimation

- Solution-
- A sample of size '*n*' *is taken:* $(X\_1, X\_2, \ldots, X\_n)$
- Exponential density is:

$$f(x) = \lambda . e^\wedge(-\lambda . x), \quad \lambda > 0, x \geq 0$$

- The density of X_i is :

$$\lambda . e(-\lambda . x\_i), \quad i = 1, 2, \ldots . n.$$

- Joint density of $(X\_1, X\_2, \ldots, X\_n)$
- **L(x_1, x_2, …., x_n)=** $\lambda \prod_{i=1}^{n} e^\wedge(-\lambda . x\_i)$

Maximum Likelihood Estimation

- Take natural logarithm.
- ln(L(x1, x2, …., xn)) = ln(λ^n* e^((−$\boldsymbol{\lambda}.\boldsymbol{x}\_\mathbf{1}$) + (−$\boldsymbol{\lambda}.\boldsymbol{x}\_\mathbf{2}$) + ⋯ + (−$\boldsymbol{\lambda}.\boldsymbol{x}\_\boldsymbol{n}$)))
- = n*ln($\boldsymbol{\lambda}$)+ ((−$\boldsymbol{\lambda}.\boldsymbol{x}\_\mathbf{1}$) + (−$\boldsymbol{\lambda}.\boldsymbol{x}\_\mathbf{2}$) +⋯ +(−$\boldsymbol{\lambda}.\boldsymbol{x}\_\boldsymbol{n}$))
- = n*ln($\boldsymbol{\lambda}$)−$\boldsymbol{\lambda}$ ($\boldsymbol{x}\_\mathbf{1}$+ $\boldsymbol{x}\_\mathbf{2}$+..+$\boldsymbol{x}\_\boldsymbol{n}$)
- **Differentiate this equation with respect to $\boldsymbol{\lambda}$, and equate to zero, this will give us the estimator that is obtained using this method of maximum likelihood estimation.**

Maximum Likelihood Estimation

- Suppose

  Z= n*ln(λ)–λ (x_1+ x_2+..+x_n)

  $$\frac{\partial Z}{\partial \lambda} = 0$$

  n/ λ - λ (x_1+ x_2+..+x_n) = 0

  Take (x_1+ x_2+..+x_n) = $\sum_{i=1}^{n} x\_i$

  $\lambda = 1/sample(mean)$

Maximum Likelihood Estimation

Q.1 Assume $x_1, x_2, x_3 \ldots\ldots x_n$ be independent and identically distributed sample from a distribution with probability defined as:

$$P_{(x_i)}(x;\theta)=(1-\theta)^{(x-1)}.\theta,$$

where θ is the parameter of the distribution. Find the estimator of the parameter using Maximum Likelihood Estimation.

Maximum Likelihood Estimation

Answer-

$$\theta\_MLE = n / \sum(x\_i)$$

$$\theta\_MLE = 1/mean$$

Maximum Likelihood Estimation

- To find the maximum likelihood estimator of the parameter θ, we need to maximize the likelihood function, which is given by:
- $L(x_1, x_2, \ldots, x_n) = \prod[P_{(X_i)}(x_i;\theta)]$
- where $x_1, x_2, \ldots, x_n$ are the observed values of the random variables $X_1, X_2, \ldots, X_n$.
- Taking the natural logarithm of the likelihood function, we get:
- $\ln L(x_1, x_2, \ldots, x_n) = \sum[\ln P_{(X_i)}]$
- $= \ln (1-\theta)^{\sum(x_i-n)}.\theta^n$
- $=(\sum(x_i - n)) \ln(1 - \theta)]+ n. \ln(\theta)$

Maximum Likelihood Estimation

- To find the maximum of this function, we take the derivative with respect to θ and set it equal to zero:
- $d/d\theta\ [\ln L(x_1, x_2, \ldots, x_n)]$
- $= d/d\theta\ (\sum(x_i - n))\ \ln(1 - \theta) + n.\ \ln(\theta)]$
- $= (\sum(x_i - n))/(1 - \theta).(-1) + n/\theta] = 0$
- Simplifying this expression, we get:
- $(\sum(x_i - n))/(1 - \theta)] = n/\theta$
- Solving for θ, we get the maximum likelihood estimator:
- **$\theta_{MLE} = n / \sum(x_i)$**

Maximum Likelihood (ML) Estimation

Assignment Question-1

- Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n *taken from a Normal Population with parameters: mean* θ_1 *and variance* θ_2*. Find the Maximum Likelihood Estimates of these two parameters.*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Maximum Likelihood Estimation

Assignment Question-2

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $B(m, \theta)$ distribution, where $\theta \in \Theta = (0, 1)$ is unknown and 'm' is a known positive integer. Compute value of θ using the M.L.E.

Minimum Mean Square Error (MMSE) Estimation

- Given some information that is related to an unknown quantity of interest (say random variable X) , the problem is to obtain a good estimate for the unknown (variable X) in terms of the observed data (say random variable Y)
- In general, our estimate **x'** is a function of y:

$$x' = g(y)$$

Minimum Mean Square Error Estimation

- The error ($\tilde{E}$) in our estimate is given by

$$\tilde{E} = X - x' = X - g(y)$$

- In MMSE the objective is to minimize the expected value of residual square, where residual is the difference between the true value and the estimated value.
- The expected residual square is also known as MSE (Mean Square Error):

$$E[(X-x')^2 | Y=y] = E[(X-g(y))^2 | Y=y]$$

Procedure to Solve MMSE

1. Define the estimator
2. Construct the MSE (Mean Square Error)
3. Differentiate the MSE with respect to the parameter and set it to zero.
4. Put the MMSE estimator in step 3 in the MSE expression to find the minimum residual square.

Minimum Mean Square Error Estimation

- For simplicity, let us first consider the case that we would like to estimate X without observing anything.
- What would be our best estimate of X in that case?
- Let `**a**` be our estimate of X. (Step-1)
- Then, the MSE is given by: (Step-2)

$$h(a) = E[(X-a)^2]$$

$$= E[X^2] + a^2 - 2.a.E[X] \quad ..(1)$$

Minimum Mean Square Error Estimation

- This is a quadratic function of `**a**`, and we can find the minimizing value of `**a**` by differentiation and set it to zero (Step-3)

$$h'(a) = -2E[X] + 2a = 0$$

- Therefore, we conclude the minimizing value of `**a**` is:

$$a = E[X]$$

- Put the value of `a` in the Equation (1) (Step-4)

Minimum Mean Square Error Estimation

$$h(a) = E[X^2] + E[X]^2 - 2.E[X].E[X]$$
$$= E[X^2] + E[X]^2 - 2\,E[X]^2$$
$$= E[X^2] - E[X]^2$$
$$= Var(X)$$

- **Interpretation:** The best constant estimator of X is the expected value of X. The minimum MSE using the optimal estimator is the variance of X.

Hypothesis Testing

- Hypothesis testing is a statistical method used to determine whether a hypothesis about a population parameter is supported by the available sample data.
- **A statistical hypothesis is an assumption about a population parameter.**
- **This assumption about the parameter may or may not be true.**

Hypothesis Testing

- An example of hypothesis testing is testing whether the mean weight of apples in a particular fruit garden is 100 grams.

Null Hypothesis

- **Null Hypothesis-**
- It is a statement that there is no significant difference or relationship between two populations or variables.
- **The null hypothesis, denoted by H_0.**
- Example- The null hypothesis is that the mean weight is 100 grams.

Alternative Hypothesis

- It is a statement that there is a significant difference or relationship between the two populations or variables.
- **The alternative hypothesis, denoted by H_1 or H_a.**
- **There are three types of alternative hypotheses. These are: (i) $H1$: weight≠100 , (ii) $H1$:weight>100and (iii) $H1$:weight<100.**
- **We have to consider one of these three hypotheses in a problem.**

Alternative Hypothesis

- Example-The alternative hypothesis is that the mean weight is not equal to 100 grams.

Decision Errors in Hypothesis Testing

- The two hypotheses are complementary to each other.
- We accept (or reject) the null hypothesis; this is equivalent to rejecting (or accepting) the alternative hypothesis.
- In this process, we make two types of decision errors, namely:
  - Type I error and
  - Type II error

Decision Errors in Hypothesis Testing

- ***Type I Error-***
- When we reject a null hypothesis when it is true.
- The probability of committing a Type I error is called the significance level denoted by α.

Decision Errors in Hypothesis Testing

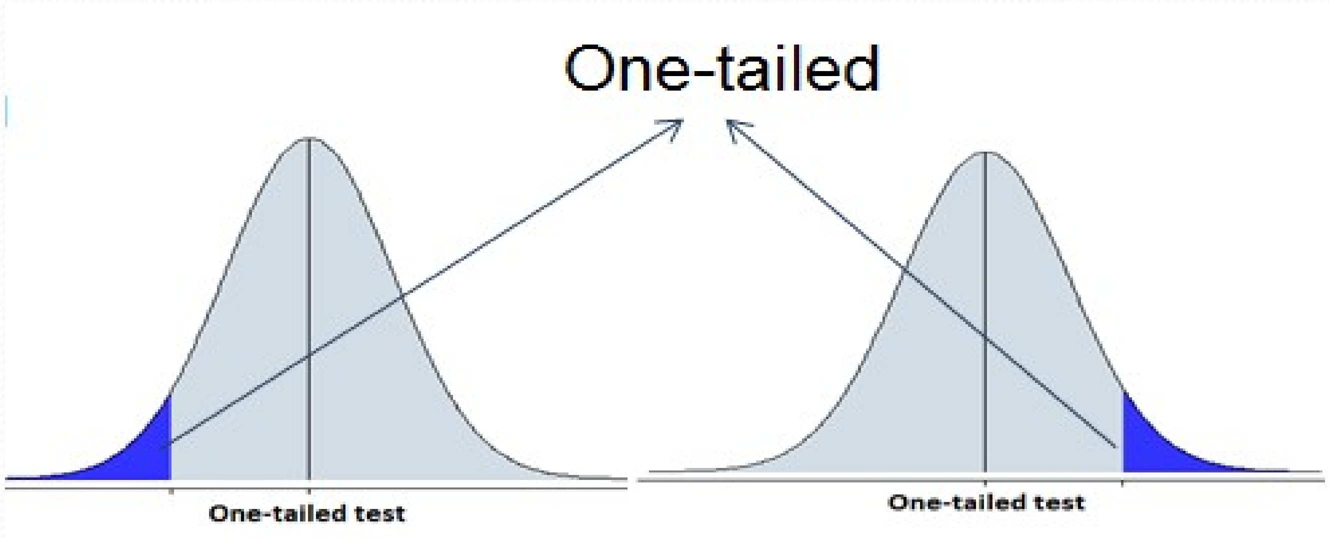
- ***Type II Error-***
- When we accept a null hypothesis when it is false.
- The probability of committing a Type II error is denoted by β.
- The probability of not committing a Type II error is called the Power of the test $(1-\beta)$.

Decision Errors in Hypothesis Testing

	Null hypothesis is TRUE	Null hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct outcome! (True positive)
Fail to reject null hypothesis	Correct outcome! (True negative)	Type II Error (False negative)

Hypothesis Testing

- ***One-Tailed Test***
- A test of hypothesis, in which the region of rejection is only on one side of the distribution used for test statistic, is called a one-tailed test.

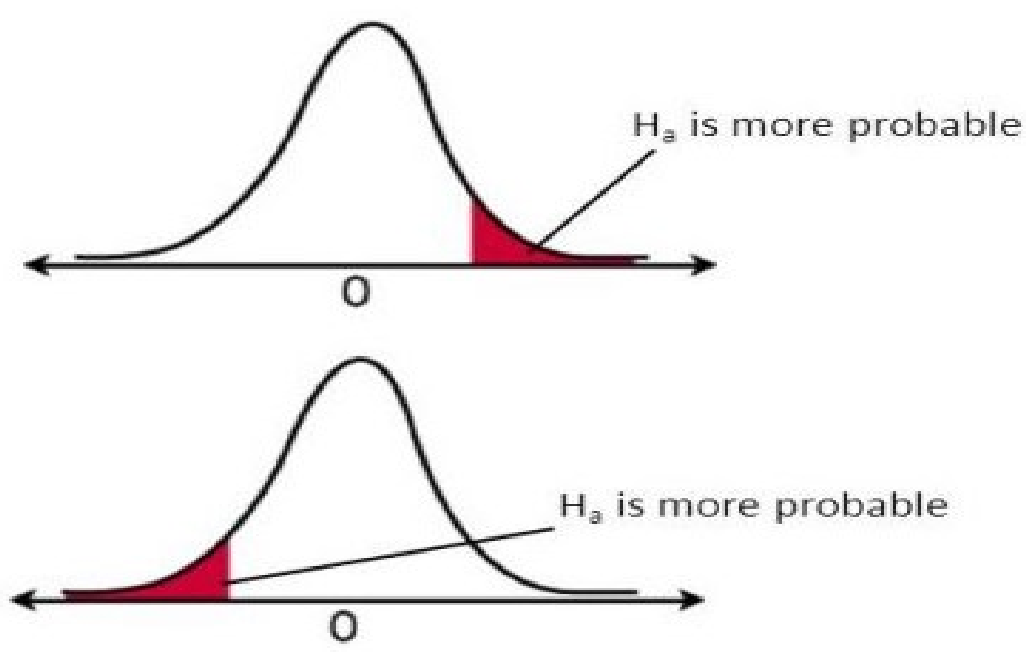


One-Tailed Test

- For example, if we hypothesize that population mean is 12, i.e., null hypothesis is $\mu = 12$.
- We consider that the alternative hypothesis is; mean is less than 12, and then we have to employ a one-tailed test.
- In this situation, alternative hypothesis shall be $\mu < 12$.
- One can see that in some other situation, we may have to consider that $\mu > 12$.

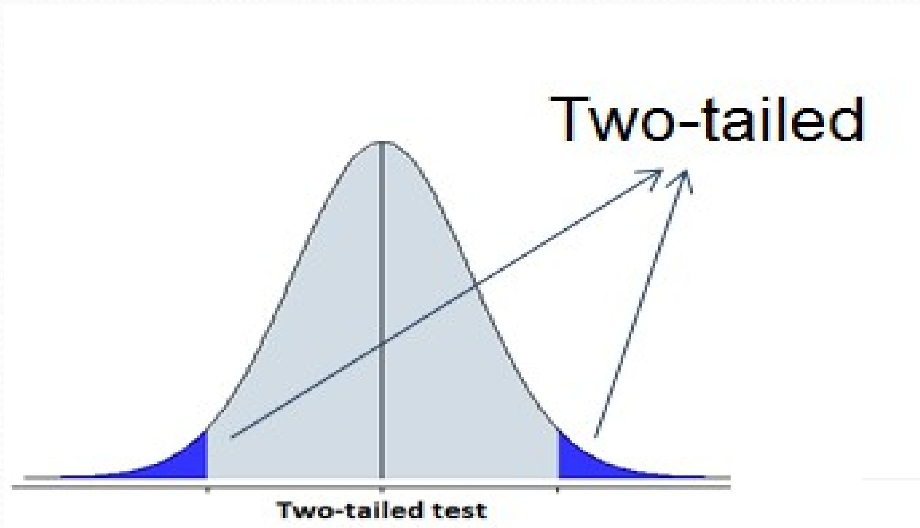
One-Tailed Test

- If the sample being tested falls into the one-sided critical area, the alternative hypothesis will be accepted instead of the null hypothesis.



Hypothesis Testing

- ***Two-Tailed Test***
- A test of hypothesis, in which the region of rejection is on both sides of the distribution used for test statistic, is called a two-tailed test.

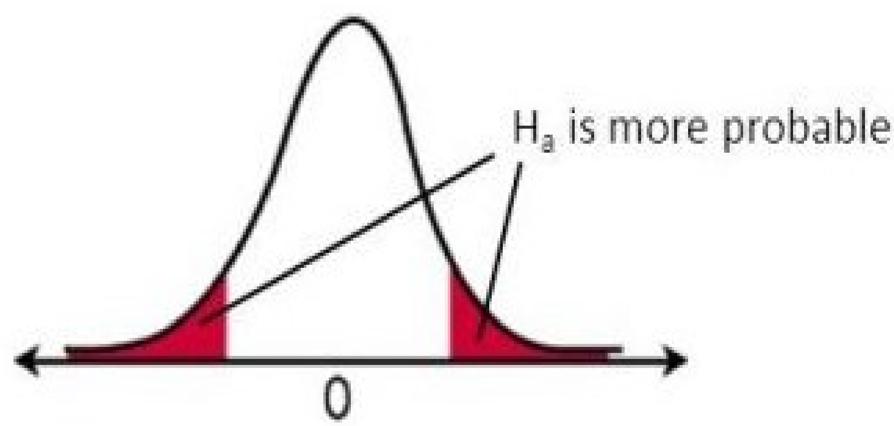


Two-Tailed Test

- For example, if we hypothesize that population mean is 12, i.e., null hypothesis is $\mu = 12$.
- We consider that the alternative hypothesis is; mean is less than 12 or mean is greater than 12, then we have to employ a two-tailed test.
- In such a situation, alternative hypothesis shall be $\mu \neq 12$.

Two-Tailed Test

- If the sample being tested falls into the two-sided critical area, the alternative hypothesis will be accepted instead of the null hypothesis.



Steps in Hypothesis Testing

1. Set up a Hypothesis- State the null and alternative hypotheses.

2. Setup the level of significance- **Significance level is the probability of occurrence of wrong decision.**

3. Choose the appropriate test statistic-
   - Z-Score (sample size is greater than 30)
   - t- Test (sample size is less than 30)

Steps in Hypothesis Testing

4. Computation-

- Z-Test- **Z= x-np/sqrt(npq)**

x-observed success, n-sample size, Z-calculated critical value, p-probability of success, q-probability of failure

$$Z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

$\bar{x}$ is the sample mean, μ is the population mean, σ is the population standard deviation and n is the sample size.

Steps in Hypothesis Testing

4. Computation-
   - t-Test-

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

s- standard deviation of the sample, μ - mean of the population, x̄-sample mean

Steps in Hypothesis Testing

5. Making Decision-

$Z_{cal} > Z_{tab}$, hence (H_o) is rejected

$t_{cal} > t_{tab}$, hence (H_o) is rejected

Hypothesis Testing

- A coin was tossed 484 times and Head turned up 265 times. Test the hypothesis that the coin is unbiased. The critical Z-values are given.

p	Significance Level	Confidence Level	Two-*Tailed Test*	One-*Tailed Test*
0.1	10%	90%	1.65	1.28
0.05	5%	95%	1.96	1.64
0.01	1%	99%	2.58	2.33
0.001	0.1 %	99.9%	3.29	3.10

Hypothesis Testing

Solution-

- Step-1 :
  - H_0 – Coin is unbiased
  - H_a – Coin is biased
- Step-2: Setup a significance level- If not given then consider 5%.
- Step-3: Setting a Test statistic- Z-Score (sample size is greater than 30)
- **Z= x-np/sqrt(npq)**

Hypothesis Testing

Solution-

- Step-4: Calculation

  $n=484$, $x=265$, $p=0.5$, $q=0.5$

  Z= x-np/sqrt(npq)

Put the values:

Z= 265-484.(0.5)/sqrt(484.(0.5).(0.5)

Z_cal =2.09

Z_tab = 1.96

Hypothesis Testing

Solution-

- Step-5: Make a decision
- Check -1: $Z\_cal > Z\_tab$, hence (H_0) is rejected.

or

- Check-2: The Z_cal (2.09) does not lie in the range of Z_tab (i.e. -1.96 to 1.96)

Hence the Null Hypothesis (H_0) is rejected.

Hypothesis Testing

- In 256 throws of a six faced dice, odd points appeared 122 times. Would you say that the dice is fair at 5% level of significance?

Hypothesis Testing

- Solution-

$$Z\_cal = -0.75 < Z\_tab$$

Hence H_o is accepted.

Hypothesis Testing

- A professor claims that all the students in the first-year class possess above-average IQs. Randomly, a test was conducted on thirty students, resulting in a mean IQ of 117. The population mean was 100, and the standard deviation was 27.

Hypothesis Testing

- **Solution- two tailed test**
- **Z_Cal** = 3.44
- Z_Tab =1.96
- Z_cal > Z_tab, hence (H_o) is rejected.

Hypothesis Testing

- The manufacturer of a certain make of Thermometer claims that his thermometers have a mean life of 20 months. A random sample of 7 such thermometers gave the following values. Life of thermometers in months: 19, 21, 25, 16, 17, 14, 21. Can you regard the producer's claim to be valid at 1% level of significance? Given that t_tab= 3.707.

Hypothesis Testing

- Solution-
- μ = 20, n=7, calculate $\bar{x}$ which is the mean of the sample using the data given in the question.
- Then put all the values in the below equation:

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

- t_cal=-0.7 and t_tab= 3.707
- H_0 is accepted (t_cal< t_tab)