Department of AI & DS
CSE and CS&IT

COURSE NAME: PROBABILITY, STATISTICS AND QUEUING THEORY

COURSE CODE: 23MT2005

Topic

Test of significance for Single Mean-t and Z test

Session - 16

AIM OF THE SESSION

To familiarize students with the basic concept of test of significance for single mean using t and Z

INSTRUCTIONAL OBJECTIVES

This Session is designed to:

1. Demonstrate the null and alternative hypothesis for t-test
2. Describe the procedure of t-test for single mean
3. List out the test statistic for t and Z
4. Describe the procedure of Z-test for single mean

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Define Null and alternative hypothesis of test of significance for single mean
2. Describe the procedure for t-test
3. Summarize the importance of t and Z tests in making decision about the sample

Test for single Mean (For small samples)-t-test for Single mean

n<30

Suppose we want to test:

i) If a random sample x_i (i=1,2,…,n) of size n has been drawn from a normal population with a specified mean, say μ_0, or
ii) If the sample mean differs significantly from the hypothetical value μ_0 of the population mean

Under the null hypothesis, H_0:

i) Th sample has been drawn from the population with mean μ_0 or
ii) There is no significant difference between the sample mean $\bar{x}$ and the population mean μ_0,

The test statistic,

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}},$$

Test for single Mean (For small samples)-t-test for Single mean

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

follows student's t-distribution with (n-1) degrees of freedom

We now compute the calculated value of t with the tabulated value at certain level of significance.

If calculated $|t|$>tabulated t, null hypothesis is rejected and if calculated $|t|$<tabulated t, H_0 , may be accepted at the level of significance adopted.

Note:

1. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$

2. We know the sample variance, $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$

$S^2/n = s^2/n-1 \Longrightarrow ns^2 = (n-1)S^2$

Hence for the numerical problems, the test statistic becomes $t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} \sim t_{n-1}$

ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

i) 95% confidence interval for the population mean μ are:

$$\bar{x} \pm t_{0.05} S/\sqrt{n}$$

ii) 99% confidence interval for the population mean μ are:

For Small samples (n<=30)

$$\bar{x} \pm t_{0.01} S/\sqrt{n}$$

iii) 95% confidence interval for the population mean μ are:

$$\bar{x} \pm 1.96(\sigma/\sqrt{n})$$

iv) 99% confidence interval for the population mean μ are:

For Large samples (n>30)

$$\bar{x} \pm 2.58(\sigma/\sqrt{n})$$

TEST OF SIGNIFICANCE FOR SINGLE MEAN (For large sample):

Suppose we wish to test the hypothesis $H_0: \mu=\mu_0$

$H_1: \mu \neq \mu_0$ two tailed test

where μ_0 is a specified constant. We have a random sample $X_1, X_2, \ldots, X_n$ from a normal population. Suppose $\overline{X}$ has a normal distribution (i.e. the sampling distribution of $\overline{X}$ is normal) with mean μ_0 and standard deviation ${}^{\sigma}/_{\sqrt{n}}$ if the null hypothesis is true, we could construct a critical region based on the computed value of the sample mean $\overline{X}$. The test procedure for $H_0: \mu=\mu_0$ uses the test statistic **for large sample test concerning mean**

$$Z = \frac{\overline{X} - \mu_0}{{}^{\sigma}/_{\sqrt{n}}}$$

Under the null hypothesis H_0, that the sample has been drawn from a population with mean 'μ' and variance 'σ^2'. i.e., there is no significant difference between the sample mean and population mean, the test statistic for large samples is

$$Z = \frac{\overline{X}-\mu_0}{{}^{\sigma}/_{\sqrt{n}}} \sim N(0,1)$$

Choose the appropriate level of significance and give the conclusion

ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

Z table

Level of significance	Two tailed test	One tailed test	
		Right tailed test	Left tailed test
0.10 (90% confidence)	1.645	1.28	-1.28
0.05(95%)	1.96	1.645	-1.645
0.01 (99%)	2.58	2.33	-2.33

ACTIVITIES/ CASE STUDIES/ IMPORTANT FACTS RELATED TO THE SESSION

P-Value approach: The P –value is the probability of obtaining a value for the test statistics that is as extreme as or more extreme than the value actually observed. Probability is calculated under the null hypothesis.

If the alternative hypothesis is right sided i.e., $H_1: \mu > \mu_0$ then only values greater than the observed value are more extreme.

If the alternative hypothesis is left sided i.e., $H_1: \mu < \mu_0$ then only values less than the observed value are more extreme.

For two sided alternatives, values in both tails need to be considered.

Z cal compare with z table value, Z cal≤z table we accept null hypothesis, Z cal >z table we reject H_0.

P cal compare with P table, PCal<Ptable we reject null hypothesis, P cal ≥Ptable we accept H_0

EXAMPLE-1

A machinist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a standard deviation of 0.040 inch. Compute the test statistic you would use to test whether the work is meeting the specifications. Also state how you would proceed further.

Solution: Here we are given,

μ =0.700 inch, $\bar{x}$=0.742 inche, s=0.040 inch and n=10

Null Hypothesis: H_0: μ=0.700 inche, i.e., the product is confirming to specifications

Alternative hypothesis: H_1: $\mu \neq 0.700$

Test statistic: Under H_0, the test statistic is

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \sim t_{n-1}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{0.742 - 0.700}{0.040/\sqrt{9}} = 3.15$$

EXAMPLE-1

t-table 2.262

t-cal > t-table value so we reject null hypothesis

The work is not meeting the specifications

How to proceed further: Here the test statistic't' follows student's t-distribution with 10-1=9 degrees of freedom. We will now compare this calculated value with the tabulated value for t for 9 degrees of freedom and at a certain level of significance, say 5%.

i) If calculated 't' =3.15> t-table value, we say that the value of t is significant. This implies that $\bar{x}$ differs significantly from μ and H_0 is rejected at this level of significance and we conclude that the product is not meeting the specifications.

ii) If calculated t<t-table value, we say that the value of t is not significant. There is no significant difference between $\bar{x}$ and μ. We may take the product conforming to specifications.

EXAMPLE-2

A random sample of 10 boys had the following I. Q's: 70, 120, 110, 101, 88,83, 95,98,107,100. Do these data support the assumption of a population mean I. Q. of 100? Find a reasonable range in which most of the mean I. Q. values of samples of 10 boys lie.

Sol**ution**:

Null Hypothesis H_0: The data are consistent with the assumption of a mean I. Q. of 100 in the population,

i.e., H_0: $\mu=100$.

Alternative Hypothesis: H_1: $\mu\neq100$.

Test statistic:

Under H_0, the test statistic is : $t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

Where $\bar{x}$ and S^2 are to be computed from the sample values of I. Q.'s.

EXAMPLE-2

Calculations for Sample Mean and Standard deviation:

x	$(x-\bar{x})$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
Total=972		1833.60

Here n=10, $\bar{x}$=972/10=97.2 and S^2=1833.60/9=203.73

$$|t| = \frac{2.8}{14.27/\sqrt{10}} = 0.62$$

t-table value at 5% LOS for 9 degrees of freedom for two-tailed test is 2.262.

Conclusion: Since calculated t is less than tabulated t H_0,may be accepted at 5% level of significance. Hence we conclude that the data are consistent with the assumption of mean I.Q. of 100 in the population.

The 95% confidence limits within which the mean I. Q. values of samples of 10 boys will be are given by:

$\bar{x} \pm t_{0.05} S/\sqrt{n}$=97.2±2.262*4.514=107.41 and 86.99. Hence the required 95% confidence interval is [86.99, 107.41]

EXAMPLE-3

A sample of 900 members has a mean 3.4 cms and standard deviation 2.61 cms . Is the sample from a large population of mean 3.25 cms and standard deviation 2.61 cms?

If the population is normal and its mean is unknown, find the 95% and 98% fudicial limits of true mean.

Given n=900 (n>30 so we treat this as large sample test)

$$\overline{\mathrm{X}} = 3.4, \qquad S = 2.61$$

Population mean=μ=3.25, $\sigma = 2.61$

Solution: :

Step: 1 H_0: There is no difference between the population mean and sample mean

H_0: $\mu = 3.25$

Step: 2 H_1: There is a difference between the population mean and sample mean.

H_1: $\mu \neq 3.25$ (Two-tailed)

Step: 3: Choose the level of significance α=5% or 1%

Null Hypothesis: The sample has been drawn from the population with mean μ=3.25 cms and S. d. σ=2.61 cms

Alternative Hypothesis: H_1: $\mu \neq 3.25$ (Two-tailed)

EXAMPLE-3

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

Here, we are given that $\overline{X}$=3.4 cms, n=900cms, μ=3.25 cms and S. d. σ=2.61 cms

$$Z = \frac{3.40 - 3.25}{2.61/\sqrt{900}} = 1.73$$

At 5% for two tailed test the fixed value of z is 1.96

Since $|Z| < 1.96$, we conclude that the data don't provide us any evidence against the null hypothesis which may, therefore, be accepted at 5% level of significance

95% fiducial limits for the population mean μ are:

$\bar{x}\pm1.96(\sigma/\sqrt{n})=3.40\pm1.96(2.61/\sqrt{900})=3.40\pm0.1705$, i.e., 3.5705 and 3.2295

98% fiducial limits for the population mean μ are:

$\bar{x}\pm2.33(\sigma/\sqrt{n})=3.40\pm2.33(2.61/\sqrt{900})=3.40\pm0.2027$, i.e., 3.6027 and 3.1973

SUMMARY

In this session, the concept of test of significance for single mean have described

1. Define Null and Alternative hypothesis of significance for single mean
2. Discuss in detail about the level of t-test for single mean.

SELF-ASSESSMENT QUESTIONS

Student's t-test is applicable in case of

A) Small samples
B) for samples of size between 5 and 30
C) large samples
D) none of the above

Degrees of freedom is related to

A. number of observations in a set
B: hypothesis under test
C: number of independent observations in a set
D: none of the above

TERMINAL QUESTIONS

1. Describe the difference between t and Z test for single mean.
2. Summarize the test of significance for single mean.
3. The mean weekly sales of soap bars in departmental store were 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a standard deviation of 17.2 Was the advertising campaign successful?

4. The specifications for a certain kind of ribbon call for a mean breaking strength of 180 pounds. If five pieces of the ribbon have a mean breaking strength of 169.5 pounds with a standard deviation of 5.7 pounds, test the null hypothesis μ=180 pounds against the alternative hypothesis μ<180 pounds at the 0.01 level of significance. Assume that the population distribution is normal.

TERMINAL QUESTIONS

5. Nine measurements were made on a key performance indicator.

123, 106, 114, 128, 113, 109, 120, 102, 111

i) Conduct a test of hypothesis with the intent of showing that the mean key performance indicator is different from 107. Take α=0.05 and assume a normal population.

ii) Based on your conclusion in part (i), what error could you have made? Explain in the context of the problem.

6. The average weekly earnings for female social workers is \$670. Do men in the same positions have average weekly earnings that are higher than those for women? A random sample of n=40 male social workers showed $\bar{x} = \$725$ and s=\$102. Test the appropriate hypothesis using α=0.01.

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Reference Books:

1. William Feller, An Introduction to Probability Theory and Its Applications: Volime 1, Third Edition, 1968 by John Wiley & Sons,Inc.
2. Alex Tsun, Probability & Statistics with Applications to Computing (Available at: http://www.alextsun.com/files/Prob_Stat_for_CS_Book.pdf)

3. Richard A Johnson, Miller& Freund's Probability and statistics for Engineers, PHI, New Delhi, 11th Edition (2011).

Sites and Web links:

1. https://www.khanacademy.org/math/statistics-probability/significance-tests-one-sample/more-significance-testing-videos/v/small-sample-hypothesis-test

THANK YOU

Team – PSQT EVEN SEMESTER 2024-25