Module 5

Hypothesis Testing II

- Small sample tests
- Student’s t-test
- F-test
- chi-square test
- goodness of fit
- independence of attributes
- Design of Experiments
- Analysis of variance
- one and two way classifications - CRD- RBD- LSD.

Small Sample Tests

In this study, we discuss test of significance for small samples. The important tests for small samples are

- t-test
- F-test

Note:

Large Sample: Size of the sample is greater than or equal to 30.

Small Sample: Size of the sample is less than 30.

T-test

When the population standard deviation is not known and the size of the sample is less than thirty, we use t-test.

t-distribution is also known as "students t-distribution".

Let $x_1, \ldots, x_n$ be the members of a random sample drawn from a normal population with mean μ and variance σ^2. We define the test statistic as

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$$ (Note: we are using n-1)

where sample standard deviation $= s = \sqrt{\sum \frac{(x_i-\bar{x})^2}{n}}$

Or one can write $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$, where $S = \sqrt{\sum \frac{(x_i-\bar{x})^2}{n-1}}$

(students t)

A random variable *X* is said to follow *t*−distribution, if its probability density function is given by

$$f(t) = \frac{k}{\left(1+\frac{t^2}{v}\right)^{{v+1}/{2}}}, \quad -\infty < t < \infty$$

where *v* is known as the degrees of freedom and *k* is constant.

The constant value *k* is chosen in such a way that $\int_{-\infty}^{\infty} f(t)dt = 1$

After simplification, we get $k = \dfrac{1}{\sqrt{v}\,\beta\left(\frac{1}{2},\frac{v}{2}\right)}$.

Note on Degree of Freedom

The number of degrees of freedom can be interpreted as the number of useful bits of information generated by a sample of given size for estimating a population parameter. Suppose we wish to find the mean of a sample with observations $x_1, x_2, \dots, x_n$. We have to use all the n values taken by the variable with full freedom for computing x . Hence x is said to have n degrees of freedom.

Assumption of t-distribution:

1) The population from which the sample is drawn is normal.
2) The sample is random and size $n \leq 30$.
3) The population S.D. σ is not known.

Properties of t-distribution:

1) The probability curve of t-distribution is symmetrical.
2) The tails of the curve are asymptotic to x-axis.
3) When $n \to \infty$, t- distribution tends to normal distribution.
4) The form of the t-dist. varies with the degrees of freedom.

Application of t- distribution: The t- distribution is used

1) To test significance of the mean of sample.
2) To test the difference between two means or to compare two samples.

Test of significance of the difference between single sample mean and population mean

To test the significance of a mean of a small sample the test statistic is $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n-1}}$,

where standard deviation = s = $\sqrt{\sum \frac{(x_i - \overline{x})^2}{n}}$.

The degrees of freedom is $v = n - 1$.

Then, find out tabular value $t_\nu(\alpha)$ based on the level of significance and degree of freedom. To find $t_\nu(\alpha)$ value follow t-distribution table. Compare calculated value with tabular value and then conclude the result.

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										

Computation of t value by using deviation method:

Consider the deviation $d_i = x_i - A$. Then, $t = \frac{\overline{x}-\mu}{s/\sqrt{n-1}}$,

Where $s = \sqrt{\frac{1}{n}\sum d_i^2 - \left(\frac{1}{n}\sum d_i\right)^2}$.

Conclusions:

If the calculated value of |t| is less than the table value $t_\nu(\alpha)$, then we accept null hypothesis for the degrees of freedom and level of significance.

If the calculated value of |t| is greater than the table value $t_\nu(\alpha)$, then we reject null hypothesis for the degrees of freedom and level of significance.

Interval estimate of population mean:

To find confidence interval or critical region for single mean of a small sample we need critical points. The critical points can be calculated by the formula:

$$\overline{x} \pm t_\nu(\alpha)\frac{s}{\sqrt{n-1}}.$$

Problem 1: A random sample of size 7 from a normal population gave a mean of 977.51 and a standard deviation of 4.42. Find 95% confidence interval for the population mean.

Solution: Here we have n = 7 (Small Sample), $\bar{x}$ = 977.51, degree of freedom= d.f. = v = $n - 1$ = 6 and s = 4.42.

Level of significance= 0.05

Hence, 95% confidence interval for the population mean μ is

$$\overline{x} - t_{0.05}(6)\frac{s}{\sqrt{n-1}} < \mu < \overline{x} + t_{0.05}(6)\frac{s}{\sqrt{n-1}}.$$

From t−table, the tabular value of $t_{0.05}(6)$ = 2.447 (Check t-table). After substituting all the values in this formula we will get

$$973.094 < \mu < 981.9256.$$

The required confidence interval is (973.094, 981.9256).

Problem 2: A random blood sample for test of fasting sugar for 10 boys gave the following data in mg/dl:

70, 120, 110, 101, 88, 83, 95, 107, 100, 98.

Does this support the assumption of population mean of 100 mg/dl? Find a reasonable range in which most of the mean fasting sugar test of the 10 boys lie.

Solution: We have n = 10. $\bar{x} = \frac{\sum x_i}{n} = \frac{972}{10} = 97.2$

Null hypothesis H0 : $\mu = 100$

Alternative hypothesis H1 :$\mu \neq 100$

Degree of freedom= v = n-1 = 9.

Level of significance= 5%

Table value of t for 9 d.f. at 5% L.S = 2.262.

Test statistic is

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n-1}} = \frac{-2.8}{4.5137} = -0.6203.$$

$\therefore |t| = 0.6203 < 2.262$; that is, calculated value of t at 5% L.S for 9 d.f is less than the table value.

Hence we accept null hypothesis.

The 95% confidence limit for μ are $\overline{x} \pm t_{\alpha}(v)\frac{s}{\sqrt{n-1}} = [87, 107.40].$

Test of significance of the difference between means of two small samples drawn from the same normal population

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means for two small samples drawn from a normal population. The test statistics is

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The degrees of freedom is $v = n_1 + n_2 - 2$.

Test of significance of the difference means of two small samples drawn from the normal populations having different mean values

Let $\bar{x}$ and $\bar{y}$ be the sample means for two small samples drawn from the normal population with means μ_1 and μ_2. The test statistics is

$$t = \frac{(\overline{x} - \overline{y}) - (\mu_1 - \mu_2)}{\sqrt{S\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $S = \sqrt{\frac{1}{n_1+n_2-2}\left(\sum(x_i - \overline{x})^2 + \sum(y_i - \overline{y})^2\right)}$

and the degrees of freedom is $v = n_1 + n_2 - 2$.

Critical points: $(\overline{x} - \overline{y}) \pm t_v(\alpha) S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$

Two types of batteries are tested for their length of life (in hours). The following data is the summary descriptive statistics

Type	Number of batteries	Average life (in hours)	Sample standard deviation
A	14	94	16
B	13	86	20

Is there any significant difference between the average life of the two batteries at 5% level of significance?

Solution:

: **Hypotheses**

Null Hypothesis $H_0 : \mu_X = \mu_Y$

i.e., there is no significant difference in average life of two types of batteries A and B.

Alternative Hypothesis $H_0 : \mu_X \neq \mu_Y$

i.e., there is significant difference in average life of two types of batteries A and B. It is a

Under null hypotheses H_0 :

$$t_0 = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

where s is the pooled standard deviation given by,

$$\mathrm{s_p} = \sqrt{\frac{ms_X^2 + ns_Y^2}{m + n - 2}}$$

$$= \sqrt{\frac{(14)(16)^2 + (13)(20)^2}{14 + 13 - 2}} = \sqrt{351.36} = 18.74$$

The value of T is calculated for the given information as

$$t_0 = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{94 - 86}{18.74\sqrt{\frac{1}{14} + \frac{1}{13}}} = \frac{8}{7.22} = 1.11$$

Critical value

Since H_1 is two-sided alternative hypothesis, the critical value at $\alpha = 0.05$ is $t_e = t_{m+n-2, \frac{\alpha}{2}} = t_{25, 0.025} = 2.060$.

Decision

Since it is a two-tailed test, elements of critical region are defined by the rejection rule $|t_0| < t_e = t_{m+n-2, \frac{\alpha}{2}} = t_{25, 0.025} = 2.060$. For the given sample information $|t_0| = 1.15 < t_e = 2.060$. It indicates that given sample contains insufficient evidence to reject H_0. Hence, there is no significant difference between the average life of the two types of batteries.

Problem 3: The nicotine content in milligrams of the samples of tobacco were found as follows

Sample A: 24, 27, 26, 21, 25.

Sample B: 27, 30, 28, 31, 22, 36.

Can it be said that the two samples come from normal populations with the same mean?