8/9/24

UNIT - 5

Small Sample Tests

In this chapter we discuss significance test for small samples in which the sample size $n \leq 30$. Following are the small sample tests

(1) Students t-distribution / t-test

(2) F-distribution test

(3) χ^2 - ~~distribution~~ (Chi-square) test.

(1) T-distribution / t-test:

If $x_1, x_2, x_3, \ldots x_n$ performs a sample of size 'n' having mean $\bar{x}$, std. deviation 'S' which is taken from a population with mean μ & std. deviation σ, then t-distribution or t-test is defined as

$$\boxed{t = \frac{|\bar{x} - \mu|}{\frac{S}{\sqrt{n-1}}}}$$

with $\nu = n-1$ degree of freedom

Note

(1) If the sample mean $\bar{x}$, sample std. deviation 'S' is not given we can find

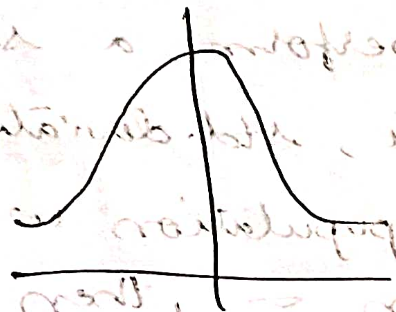
$$\boxed{S = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}}$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

* Degree of freedom: It is a number which indicates how many of the values of a variable may be freely choosen/selected.
It is represented by 'ν'

* Properties of t-distribution:-

i) The shape of the t-distribution is in bell shape which is similar to a normal distribution.



$t=0$

ii) T-distribution curve is symmetric about the line $t=0$

iii) The mean, median & mode of t-distribution are same.

* Method - I

Students t-test for single mean:-

If $\bar{x}$ is the mean of the sample having std. deviation 's' & size 'n' which is taken from a population with mean μ.

→ We apply this method to check whether the population has a specified mean ie., $\mu = \mu_0$ (or) not.

→ We use the test statistic

$$t = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}} \quad \text{or} \quad \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n-1}}}$$

→ Degree of freedom in this method is $\gamma = n - 1$.

Q1) The average breaking strength of steel rod is specified to be 18.5 thousand pounds. To test the sample of 14 rods were tested the mean & std. deviation are obtained as 17.85 thousand pounds & ~~19.55~~ 1.955 thousand pounds. is the result of experiment significant.

A) Given that $n = 14$ (small sample).

$\bar{x} = 17.85$

$s = 1.955$

$\mu = \cancel{17.85}\ 18.5$

i) Null hypothesis: Let the test is significant

ii) Alternate hypothesis: Let the test is not significant (two tail test)

(iii) Level of significance ~~statistics~~ : $\alpha = 5\%$ (Assumed)

(iv) Test statistic:

$$t = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n-1}}}$$

$$= \frac{|17.85 - 18.5|}{\frac{1.955}{\sqrt{13}}} = 1.199$$

(v) Conclusion:

The calculated value of 't' (1.199) is less than the tabulated value (2.160) for two tail test at 5% level

∴ Accept null hypothesis.

9/02/24

Q2) A random sample of 10 boys ~~have~~ had the following I.Q's.

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

(i) Do this data support the assumption of a population mean I.Q. of 100?

(ii) Construct the confidence intervals at 5% level

A) Given that sample size $n = 10$ (small sample).

WKT sample mean, $\bar{x} = \frac{70+120+110+101+88+83+95+98+107+100}{10}$

$\bar{x} = 97.2$

sample std. deviation, etc

$$s = \sqrt{\frac{(70-97.2)^2+(120-97.2)^2+(110-97.2)^2+(101-97.2)^2+(88-97.2)^2+(83-97.2)^2+(95-97.2)^2+(98-97.2)^2+(107-97.2)^2+(100-97.2)^2}{9}}$$

$$s = \sqrt{\frac{(-27.2)^2+(22.8)^2+(12.8)^2+(3.8)^2+(-9.2)^2+(-14.2)^2+(-2.5)^2+(0.8)^2+(9.8)^2+(2.8)^2}{9}}$$

$\Rightarrow$ $s = 14.29$

$\Rightarrow$ $\vartheta = n-1 = 9$

(i) Null hypothesis(H_0): Let the population mean IQ, $\mu = 100$.

(ii) Alternate hypothesis (H_1): Let the population mean IQ, $\mu \neq 100$.

(iii) Level of significance (α): $\alpha = 5\%$

$\alpha = 0.05$

(iv) Test statistic:

$$t = \frac{|\bar{x}-\mu|}{\frac{s}{\sqrt{n-1}}} = \frac{|97.2-100|}{\frac{14.29}{\sqrt{9}}} = 0.589$$

(v) Conclusion: The calculated value of 't' (0.589) is less than to the tabulated value (2.262) for two tail test at

5% level with 9 degree of freedom.

∴ Accept null hypothesis i.e., the population mean IQ = 100.

(ii) WKT confidance intervals are

$$\left(\bar{x} \pm t_\alpha \cdot \frac{s}{\sqrt{n}}\right)$$

$$97.2 \pm (2.262)\left(\frac{14.27}{\sqrt{10}}\right) = 107.4$$

Method-II : Test of significance for difference of two small sample means:

If n_1, n_2 are two small sample sizes having means $\bar{x}_1, \bar{x}_2$ resp. with std. deviations s_1, s_2 which are taken from two different populations having means μ_1, μ_2

→ we apply this method to check whether the two samples drawn from same population.

→ we apply the test statistic

$$\boxed{t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$

where the common variance,

$$\boxed{s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

→ Here degree of freedom, $\nu = n_1 + n_2 - 2$

Q) Samples of two types of electric light bulbs were tested ~~according to the time~~ for length of life & following data was obtained.

Type-I	Type-II
$n_1 = 8$	$n_2 = 7$
mean = 1234 hrs	mean = 1.036 hrs
S.D = 36 hrs	S.D = 40 hrs.

Is the difference in means significant to say that type I is superior to type-II regarding length of life.

A). Given that $n_1 = 8$, $n_2 = 7$

$\bar{x}_1 = 1234$, $\bar{x}_2 = 1036$

$S_1 = 36$, $S_2 = 40$

(i) Null hypothesis (H_0): Let there is no difference ~~in~~ b/w type-I & type-II regarding length of life

(ii) Alternate hypothesis (H_1): Let type I is superior to type-II regarding length of life. (one tail test) (right tail test)

(iii) Level of significance: $\alpha = 5\%$ (Assumed)

(iv) Test statistics:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{|1234 - 1036|}{\sqrt{S^2\left(\frac{1}{8}+\frac{1}{7}\right)}}$$

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{(8)(36)^2 + (7)(40)^2}{8+7-2}$$

$$S^2 = \cancel{166}\ 1659.077$$

$$t = \frac{|1234 - 1036|}{\sqrt{(1659.077)^2\left(\frac{1}{8}+\frac{1}{7}\right)}} = 9.392$$

(v) Conclusion: The calculated value of t (9.392) is more than the tabulated value (1.771) for one tail test at 5% level with '13' degree of freedom.

∴ Reject the null hypothesis i.e., the type-I is superior to type-II towards length of life.

---

2) Two horses 'A' & 'B' were tested according to the time (in sec) to run a particular track with the following results.

Horse - A	28	30	32	33	33	29	34
Horse - B	29	30	30	24	27	29	-

$n_1 = 7$
$n_2 = 6$

Test whether the two horses have the same running capacity at 1% level

A) Given that $n_1 = 7$, $n_2 = 6$.

$$\bar{x}_1 = \frac{28+30+32+33+33+29+34}{7}$$

$$\bar{x}_1 = 31.286$$

$$\bar{x}_2 = \frac{29+30+30+24+27+29}{6}$$

$$\bar{x}_2 = 28.177.$$

$$S_1 = \sqrt{\frac{(28-31.28)^2+(30-31.28)^2+(32-31.28)^2+(33-31.28)^2+(33-31.28)^2+(29-31.28)^2+(34-31.28)^2}{6}}$$

$$S_1 = 2.28$$

$$S_2 = \sqrt{\frac{(29-28.17)^2+(30-28.17)^2+(30-28.17)^2+(24-28.17)^2+(27-28.17)^2+(29-28.17)^2}{5}}$$

$$= \cancel{2.257}\ 2.31$$

(i) Null hypothesis (H_0): Let ~~there~~ is the two horses have the same running capacity

(ii) Alternate hypothesis (H_1): Let the two horses do not have the same running capacity

(iii) Level of significance: $\alpha = 1\%$.

(iv): test statistis:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2-2} = \frac{7(2.28)^2 + 6(2.31)^2}{7+6-2}$$

$$= \cancel{6.269}\quad 6.266$$

$$t = \frac{|31.286 - 28.177|}{\sqrt{(6.266)^2\left(\frac{1}{7}+\frac{1}{6}\right)}}$$

$= \cancel{1.603}\ 2.24.$

(v) Conclusion: The calculated value of t (2.24) is ~~grea~~ less than the tabulated value (3.106) for two tail test at 1% level with '11' degree of freedom

$\therefore$ Accept the null hypothesis i.e.. the two horses have the same running capacity.

---

* F- distribution or F-test:

We apply F-distribution to check the significant difference b/w two variances.

→ It is defined as

$$\boxed{F = \frac{s_1^2}{s_2^2}}$$ where $s_1^2 > s_2^2$ with (ν_1, ν_2) degree of freedom

(or)

$$\boxed{F = \frac{s_2^2}{s_1^2}}$$ where $s_2^2 > s_1^2$ with (ν_2, ν_1) degree of freedom

* properties of F-distribution:

① F-distribution is free from population parameters & It depends upon degree of freedom.

② F-distribution curve lies entirely in the first quadrant since it is always positive more than 1.

③ The mode of F-distribution is less than unity.

Q1) ~~If~~ ~~8'~~ In a sample of 8 observations from a normal population the sum of the squares of the deviations from the sample mean is 84.4. In another sample of 10 observations, it was 102.6. Test at 5% level whether the two samples have the same variance.

A) Given that $n_1 = 8$, $n_2 = 10$

$$\Sigma (x_i - \bar{x}_1)^2 = 84.4$$

$$\Sigma (x_i - \bar{x}_2)^2 = 102.6$$

wkt $$s_1^2 = \frac{\Sigma(x_i - \bar{x}_1)^2}{n_1 - 1} = \frac{84.4}{7} = 12.06$$

$$s_2^2 = \frac{\Sigma(x_i - \bar{x}_2)^2}{n_2 - 2} = \frac{102.6}{9} = 11.4$$

(i) Null hypothesis (H0): Let both the variances are same.

(ii) Alternate hypothesis (H_1): Let ~~the~~ both the variances are not same

(iii) Level of significance: $\alpha = 5\%$ (Assumed)

(iv) Test statistic:

$$F = \frac{S_1^2}{S_2^2} = \frac{12.06}{11.4} = 1.06$$

(v) Conclusion: The calculated value of F (1.06) is less than the tabulated value (3.29) for two tail test at 5% level with (7, 9) degree of freedom.

∴ Accept null hypothesis i.e., both the variances are same.

13/2/24

(1) The nicotin content in milligrams in 2 samples of tobacco were found to be as follows

sample-A	24	27	26	21	25	–
sample-B	27	30	28	31	22	36

Check whether there is any significant difference b/w ~~in~~ the 2 sample variances.

Sol: Given that

sample ~~mean~~ size $n_1 = 5$

" " $n_2 = 6$

I sample mean, $\bar{x}_1 = \frac{24+27+26+21+25}{5}$

$\bar{x}_1 = 24.6$

II sample mean, $\bar{x}_2 = \frac{27+30+28+31+22+36}{6}$

$\bar{x}_2 = 29.$

wkt I sample variance, $s_1^2 = \frac{\Sigma(x_i-\bar{x}_1)^2}{n_1-1}$

$$s_1^2 = \frac{(24-24.6)^2+(27-24.6)^2+(26-24.6)^2+(21-24.6)^2+(25-24.6)^2}{4}$$

$s_1^2 = 5.3$

$s_2^2 = \frac{\Sigma(x_i-\bar{x}_2)^2}{n-1}$

$$s_2^2 = \frac{(27-29)^2+(30-29)^2+(28-29)^2+(31-29)^2+(22-29)^2+(36-29)^2}{5}$$

$s_2^2 = 21.6.$

(i) Null hypothesis (H0): Let there is no difference b/w two sample variances ($s_1^2 = s_2^2$)

(ii) Alternate hypothesis (H_1): Let there is a difference b/w 2 variances ($s_1^2 \neq s_2^2$) (two tail test).

(iii) Level of significance: $\alpha = 5\%$ (Assumed)

(iv) Test statistic

$$F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3} = 4.075 \quad (\because S_2^2 > S_1^2)$$

(v) Conclusion: The calculated value of F (4.075) is less than the ~~cal~~ tabulated value (6.26) at ~~with~~ 5% level for two tail test.

$\therefore$ Accept null hypothesis.

i.e., both the sample variances are same

---

* χ^2-distribution.

We apply χ^2-distribution to check the difference b/w two frequencies (observed & expected freq)

We use the test statistic

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

* The degree of freedom is $\nu = n - 1$

Properties of χ^2-distribution.

1. χ^2-curve is not symmetric
2. χ^2 curve lies entirely in the 1st quadrant & varies from 0 to ∞
3. χ^2 value depends only on degree of freedom

① The no. of automobile accidents per week in a certain community are as follows 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accidents conditions were same during this 10 week period?

Sol: Given that

sample size, $n = 10$

Degree of freedom, $\nu = n-1 = 10-1 = 9$

Observed freq = 12, 8, 20, 2, 14, 10, 15, 6, 9, 4

$$\text{Expected freq}, E = \frac{12+8+20+2+14+10+15+6+9+4}{10}$$

$$\Rightarrow \quad E = 10$$

(i) Null hypothesis (H_0): Let the accident conditions were same during 10 weeks period.

(ii) Alternate hypotheses (H_1): Let the accident conditions were not same during 10 weeks period (two tail test)

(iii) Level of significance: $\alpha = 5\%$.

(iv) Test statistic:

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i}$$

$\longrightarrow$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	4	0.4
8	10	-2	4	0.4
20	10	10	100	10
2	10	-8	64	6.4
14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	-4	16	1.6
9	10	-1	1	0.1
4	10	-6	36	3.6
				26.6

(v) Conclusion : The calculated value of χ^2 (26.6) is more than the tabulated value (16.916) at 5% level for two tail test.

$\therefore$ Reject null hypothesis

i.e., accident conditions were not same during 10weeks period.

(3) A survey of 240 families with 4 children each revealed the following distribution

Male births	4	3	2	1	0
observed freq	10	55	105	58	12

Check whether the freq. of male & female births are equal

A) Given that
sample size $n = 5$
Observed freq. $= 10, 55, 105, 58, 12$

Expected freq. of observed freq are obtained by Binomial freq distribution ie.. $N(p+q)^n$
where $N = 240, n = 5$
Let 'p' be the probability of getting a male birth
'q' " " " female ".

$\Rightarrow p = \frac{1}{2}, q = \frac{1}{2}$

$\Rightarrow 240\left[\frac{1}{2} + \frac{1}{2}\right]^4$

$\Rightarrow 240\left[{}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 + {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 + {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + {}^4C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + {}^4C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4\right]$

$\Rightarrow \frac{240}{16}\left[1+4+6+4+1\right]$

$\Rightarrow 15(1+4+6+4+1)$

$\therefore$ The expected freq. ~~are~~ of male births are
15, 60, 90, 60, 15

(i) Null hypothesis (H_0): let both male & female births frequencies be same.

(ii) Alternate hypothesis (H_1): Let the male & female birth frequencies are not same

(iii) Level of significance: $\alpha = 5\%$ (Assumed)

(iv) Test statistic:

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	15	-5	25	1.67
55	60	-5	25	0.41
105	90	15	225	2.5
58	60	-2	4	0.06
12	15	-3	9	0.6
				5.24

$$\chi^2 = 5.24$$

(v) Conclusion: The calculated value of χ^2 (5.24) is less than the tabulated value (9.488) for two tail test at 5% level

∴ Accept null hypothesis.

ie., the ~~act~~ frequencies of male & female birth are same.

χ^2 test for independent of attributes :-

Literally, an attribute means a quality or characteristic. Ex: drinking, smoking, honestly etc

If the observed freq are in matrix form

	I	II	
I	a	b	a+b
II	c	d	c+d
	a+c	b+d	(Grand Total)

Then the expected freq of each observed freq is obtained as

$$E(a) = \frac{(a+b)(a+c)}{G.T}$$

$$E(b) = \frac{(a+b)(b+a)}{G.T}$$

$$E(c) = \frac{(c+d)(a+c)}{G.T}$$

$$E(d) = \frac{(c+a)(b+d)}{G.T}$$

We take degree of freedom = (no. of rows - 1) × (no. of col - 1)

Q1) From the following data find whether there is any significant liking in the habit of taking soft drinks. Among the categories of employees.

	clerks	Teachers	officers	
pepsi	10	25	65	~~100~~
Thumps up	15	3	65	~~110~~
Fanta	50	60	30	~~140~~
	~~75~~	~~115~~	~~160~~	$\boxed{350}$

Sol: i) Null hypothesis (H_0): Let there is no difference in the habit of taking soft drinks

ii) Alternate hypothesis (H_1): Let there is a significant ~~test~~ difference in the habit of taking soft drinks.

(iii) Level of significance: $\alpha = 0.05\%$ (Assumed)

(iv) Test statistic:

$$\frac{\sum (O_i - E_i)^2}{E_i}$$

The expected freq are obtained by

	Clerks	Teachers	Officers	
pepsi	10	25	65	100
Thumps up	15	3	65	110
Fanta	50	60	30	140
	75	115	160	$\boxed{350}$

$$\Rightarrow E(10) = \frac{100 \times 75}{350} = 21.42$$

$$\Rightarrow E(25) = \frac{100 \times 115}{350} = 32.85$$

$$\Rightarrow E(65) = \frac{100 \times 160}{350} = 45.71$$

$$\Rightarrow E(15) = \frac{110 \times 75}{350} = 23.57$$

$$\Rightarrow E(30) = \frac{110 \times 115}{350} = 36.14$$

$$\Rightarrow E(65) = \frac{110 \times 160}{350} = 50.28$$

$$\Rightarrow E(50) = \frac{140 \times 75}{350} = 30$$

$$\Rightarrow E(60) = \frac{140 \times 115}{350} = 46$$

$$\Rightarrow E(30) = \frac{140 \times 160}{350} = 64$$

$\longrightarrow$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	21.42	-11.42	129.96	6.073
25	32.85	-7.85	62.41	1.897
65	45.17	19.3	372.49	8.151
15	23.57	-8.6	73.96	3.134
30	36.14	-6.1	37.21	1.031
65	50.28	14.7	216.09	4.3
50	30	20	400	13.333
60	46	14	196	4.261
30	64	-34	1156	18.062
				60.2425

Conclusion: The calculated value of χ^2 (60.2425) is more than tabulated value (9.488) for two tail test with ~~80~~ 4 degree of freedom

$\therefore$ Reject null hypothesis

ie., There is no significant difference in the habit of taking soft ~~dri~~ drinks