Question 5:- T - Test:-

A company manufactures likes batteries with an average life span of 2 year or more year. An Engineer believes this value to be less using 10 samples, he measures the average life span to be 1.8 years. with a standard deviation of 0.15.

(a) State the Null and Alternate hypothesis?

(b) At a 99% C.I., is there enough evidence to discard the Null Hypothesis Ho?

Ans: $\mu = 2$, $n = 10$, $\bar{x} = 1.8$, $S = 0.15$, $C.I = 99\%$

$S.V = 0.01$

Step 1:- Null Hypothesis (Ho): $\mu \geq 2$

Alternate Hypothesis (H1): $\mu < 2$

Step 2:- Confidence Interval = 99%.

Significance value = 0.01.

As this is a one tail test, and $n < 30$ so we have to perform t test.

$\boxed{\text{Note}}$ If $n < 30$ and sample standard deviation is given we have to perform t - test.

To perform t-test we require Degree of Freedom:-

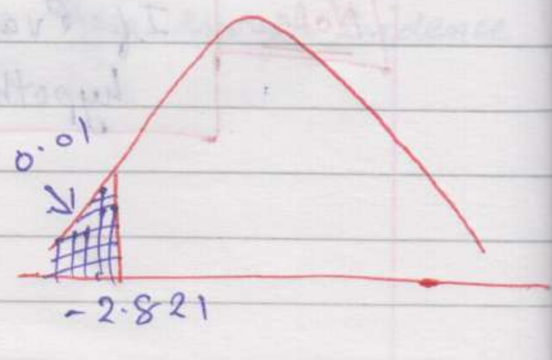
Degree of Freedom $= n - 1 = 10 - 1 = 9$

Now according to t - table.

Step 3:- Decision Boundaries:-

$\alpha = 0.01$

$t_{0.01} = -2.821$



Step 4:- Calculate the t-score.

$$t\ score = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{1.8 - 2}{0.15/\sqrt{10}} = -4.216$$

$$t = -4.216.$$

Conclusion:- $-4.216 < -2.821$. {Reject the Null Hypothesis}

The average lifespan of the battery is less then 2 Years.

★ $\boxed{\text{Z test with Proportions:-}}$

⇒ Question:- A tech company believes that the percentage of residents in town xyz that owns a cell phone is 70%. A marketing manager believes that this value to be different. He conducts a survey of 200 individuals and find that 130 responded. Yes. owning a cell phone?

(a) State Null and Alternate Hypothesis?

(b) At a 95% Confidence interval, is there enough evidence to reject the null hypothesis?

Ans. Null Hypothesis: $P_0 = 0.70$ } Step 1

Alternate Hypothesis: $P_0 \neq 0.70$

$$q_0 = 1 - P_0 = 0.30$$

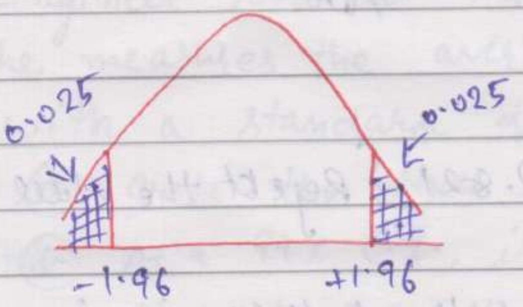
Here

$n = 200 \quad x = 130$

$$\therefore \text{proportion } \hat{P} = \frac{130}{200} = 0.65$$

Step 2:- $C.I = 0.95 \quad \alpha = 0.05$

Step 3:- Decision Boundaries.



This is a two tail test so decision boundaries are -1.96 and $+1.96$.

Also we have to perform Z test here because $n \geq 30$

Step 4:- Z test with proportion.

$$\boxed{Z\,test = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}} = \frac{0.65 - 0.70}{\sqrt{\frac{(0.70)(0.30)}{200}}} \approx -1.54$$

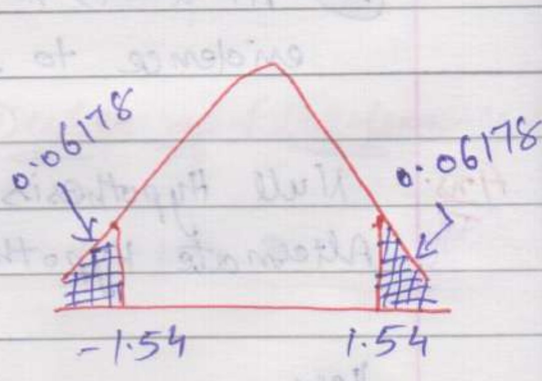
Conclusion:-

$Z\,test = -1.54$

$-1.54 > -1.96$ {We Fail to reject the Null Hypothesis}

$\therefore$ 70% of residents owns a cell phone.

P value:-



P value :- $0.06178 + 0.06178$
$= 0.12356$

Here Pvalue > Significance value.
then we fail to reject the Null Hypothesis.

Question :- 7 :- A car company believes that the percentage of residents in city ABC that owns a vehicle is 60% or less. A sales manager disagrees with this. He conducts a hypothesis testing surveying 250 residents and find that 170 responded Yes. to owning a vehicle.

(a) State the null & Alternate hypothesis.

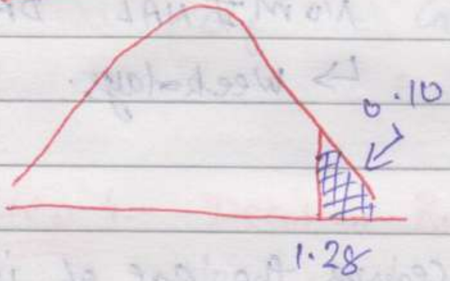
(b) At 10% significance level, is there enough evidence to support the idea that vehicle ownership in city ABC is 60% or less?

Ans. In this we will Perform Right tail test as

$$\left.\begin{array}{l}\text{Null Hypothesis } p_0 \le .60 \\ \text{Alternate Hypothesis } p_1 > 60 \\ \qquad q_0 = 0.40\end{array}\right\} \text{Step 1}$$

Step 2:- $S.V = 0.10 = \alpha$ $\qquad \hat{P} = \frac{170}{250} = 0.68$

Step 3:- Decision Boundaries.



$Z_{0.10} = 1.28$
According to Z table.

Step 4:- Z test with proportion.

$$Z\,test = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.68 - 0.60}{\sqrt{\frac{0.6 \times 0.4}{250}}} = \frac{0.08}{0.03098}$$

$Z\,test = 2.58$

Conclusion:-

$Z\ score = 2.58$

$2.58 > 1.28$ {Reject the Null Hypothesis}

$\therefore$ vehicle ownership in the city ABC is more then 60%.

P value:-

Chi Square test:-

=> Chi Square test claims about population proportions.

It is a non parametric test that is performed on categorical data.

Categorical data include.

ORDINAL DATA → Ranks.

NOMINAL DATA → Weekdays.

Question:- In the 2000 U.S. census the age of individuals in a small town found to be the following.

<18	18-35	>35
20%	30%	50%

In 2010, ages of $n = 500$ individuals were sampled. Below are the results.

<18	18-35	>35
121	286	91

Using $\alpha = 0.05$, would you conclude the population distribution of ages has changed in the last 10 years?

Ans:

	<18	18 - 35	>35
Expected	20%	30%	50%

n = 500

	<18	18 - 35	>35
Observed	121	288	91
Expected	100	150	250

Step 1:-

Null Hypothesis Ho: The data meets the expected distribution.

Alternate Hypothesis H1: The data does not meet the expected distribution.

Step 2:- $\alpha = 0.05$ Confidence Interval = 95%

Step 3:- Degree of Freedom {categories}

df = Categories - 1

= 3 - 1 = 2.

Step 4:- Decision Boundaries:-

Now we have to see Chi Square Table.

We have Degree of Freedom (df) = 2

$\alpha = 0.05$

If Chi Square test $(\chi^2) > 5.991$ {Reject the Null Hypothesis}

[5.991 comes from chi Square table]

Step 5:- Chi Square test statistics.

$$\boxed{\chi^2 = \sum \frac{(fo - fe)^2}{fe}}$$

Where fo = observed
fe = Expected.

$$= \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-250)^2}{250}$$

$$\chi^2 = \frac{441}{100} + \frac{19044}{150} + \frac{(-25281)}{250}$$

$$\boxed{\chi^2 = 30.246}.$$

Conclusion:-

Chi Square test $\chi^2 = 30.246$

As $\chi^2 > 5.991$ {Reject the Null Hypothesis}

Therefore, we can conclude the population distribution of ages has changed in the last 10 years.

★ ~~ANOVA TEST:- [F Test]~~

NOTE:-

T-test:- It is essentially, testing the significance of the difference of the mean values when the sample size is small (i.e less then 30) & when population standard deviation is not available.

Assumption:- For t test:-

- Population distribution is normal, and.

- Samples are random & independent.
- Sample size is small.
- Population standard deviation is not known.

One Sample T-Test

To compare sample mean with that of the population mean.

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

Two Sample T-Test:

To compare means of two different samples.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = Sample mean of 1st group.
$\bar{x}_2$ = Sample mean of 2nd group.
S_1 = Sample1 Standard deviation.
S_2 = Sample2 Standard deviation.
n = sample size.

Z-Test:-

It is used to determine whether the means are different when the population variance is known & the sample size is large (i.e greater than 30)

It is also a parametric test so population distribution is normal.

Condition when to used which test.

Sample Size : Small
Popupalation variance : Known } T-Test.

Sample Size : Large
Population variance : Known } Z-Test.

Sample Size : Large } Z - test.
Population variance : Known

Sample size : Small } Z - test.
Population variance : Known

F - Test:-

- It is also Parametric test, So population distribution is normal.
- It is a test for the null hypothesis that two normal populations have the same variance.
- F-statistic is simply a ratio of two variances.

Formula:-

$$F = \frac{s_1^2}{s_2^2} \quad \text{where } s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

S^2 = Sample variance.

Uses of F-Test:-

↳ Test the overall significance for a regression model.
↳ To compare the fits of different models and.
↳ To test the equality of means.

ANOVA Test:- Analysis of Variance

-> It is parametric test of hypothesis testing.
-> It is an extention of T-test & Z-test.
-> It is used to test the significance of the differences of the mean values among more

than two sample groups.

→ In simple words, If you have more then two sample groups, and in that case if you want to compare mean value, then we use ANOVA Test.

→ It uses F-test to statistically test the equality of means & the relative variance between them.

Types of ANOVA Test:-

1. One Way ANOVA : One Independent Variable.
2. Two Way ANOVA : Two Independent Variable.

Formula:-

$$F\text{-statistic} = \frac{\text{variance between the sample mean.}}{\text{variance within the samples.}}$$

★ Question:- ONE WAY ANOVA TEST:-

To assess the significance of possible variation in performance in a certain test between the convent schools of a city, a common test was given to a number of students taken at random from the fifth class of the 3 schools concerned the results given below.

A	B	C
9	13	14
11	12	13
13	10	17
9	15	7
8	5	9

Solution:-

A	B	C
9	13	14
11	12	13
13	10	17
9	15	7
8	5	9
50	55	60.

$\bar{x}_A = \frac{50}{5} = 10$

$\bar{x}_B = \frac{55}{5} = 11$

$\bar{x}_C = \frac{60}{5} = 12$

$$\bar{\bar{X}} = \frac{\bar{x}_A + \bar{x}_B + \bar{x}_c}{3} = \frac{10+11+12}{3} = \frac{33}{3} = 11$$

TABLE - I

Source of variance.	Sum of Square	Degree of Freedom.	Mean Square	F
Between the Sample	SSC $= 10$	$V_1 = C-1$ $= 3-1 = 2$	$MSC = SSC/V_1$	$\frac{MSC}{MSE}$
Within the Sample	SSE	$V_2 = n - C$	$MSE = SSE/V_2$	

★ Calculate SSC :- (Variance between the Sample)

$(\bar{x}_A - \bar{\bar{x}})$	$(x_A - \bar{\bar{x}})^2$	$(\bar{x}_B - \bar{\bar{x}})$	$(x_B - \bar{\bar{x}})^2$	$(\bar{x}_c - \bar{\bar{x}})$	$(x_c - \bar{\bar{x}})^2$
$(10-11) = -1$	1	$(11-11) = 0$	0	$(12-11) = 1$	1
$(10-11) = -1$	1	$(11-11) = 0$	0	$(12-11) = 1$	1
$(10-11) = -1$	1	$(11-11) = 0$	0	$(12-11) = 1$	1
$(10-11) = -1$	1	$(11-11) = 0$	0	$(12-11) = 1$	1
$(10-11) = -1$	1	$(11-11) = 0$	0	$(12-11) = 1$	1
$\Sigma(\bar{x} - \bar{\bar{x}})$	5		0		5

$$SSC = \Sigma(x_A - \bar{\bar{x}})^2 + \Sigma(x_B - \bar{\bar{x}})^2 + \Sigma(x_c - \bar{\bar{x}})^2$$

$$= 5 + 0 + 5 = \boxed{10}$$

* Calculate SSE :- (Variance within the sample)

$(A-\bar{x}_A)$	$(A-\bar{x}_A)^2$	$(B-x_B)$	$(B-x_B)^2$	$(C-x_C)$	$(C-\bar{x}_C)^2$
$(9-10)=-1$	1	$(13-11)=2$	4	$(14-12)=2$	4
$(11-10)=1$	1	$(12-11)=1$	1	$(13-12)=1$	1
$(13-10)=3$	9	$(10-11)=-1$	1	$(17-12)=5$	25
$(9-10)=-1$	1	$(15-11)=4$	16	$(7-12)=-5$	25
$(8-10)=-2$	4	$(5-11)=-6$	36	$(9-12)=-3$	9
$\Sigma(A-\bar{x}_A)$	16		58		64

$$SSE = \Sigma(A-\bar{x}_A)^2 + \Sigma(B-x_B)^2 + \Sigma(C-x_C)^2$$

$$SSE = 16+58+64 = \boxed{138}$$

Now the value of Table 1 will become.

$SSC = 10$ [Variance between the sample]

$SSE = 138$ [Variance within the sample]

$V_1 = C-1$ {C = No of Columns}

$= 3-1 = 2$

$V_2 = n-C$

$= 15-3$ {n = Columns × rows

$= 12$ $= 3\times5$}

$$MSC = SSC/V_1 = \frac{10}{2} = 5$$

$$MSE = \frac{SSE}{V_2} = \frac{138}{12} = 11.5$$

$$F = \frac{MSC}{MSE} = \frac{5}{11.5} = 0.435$$

Calculated F value = 0.435

Now According to F distribution table for ANOVA And Degree of Freedom are $V_1 = 2$ and $V_2 = 12$

If we see F table., the value where df 2 and df = 12 intersect is 3.89