1. Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher think that a diet high in raw cornstarch will have a positive effect on blood glucose levels. A sample of 36 patients who have tried the raw cornstarch diet have a mean glucose level of 108. Test the hypothesis that the raw corn startch had an effect or not.

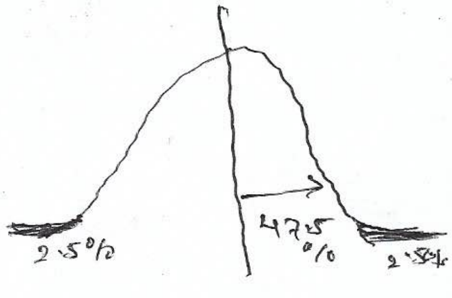
Soln :-

Null Hypothesis Ho is $\mu = 100$

alternate Hypothesis Halt $\neq \mu = 100$

we assumes standard Significance level 0.05 for this proble-m as its not mentioned

$\alpha = 0.05$



As this is two-tailed test

$\alpha = 0.05/2 = 0.025$

As this is two tail test and error area is 2.5% both side.

Remaining area for positive side = 47.5%

Z score for 0.475 is 1.96 as per z-table

$$Z_{stat} \text{ Score as per statistics} = \frac{\bar{x} - \mu_0}{\sigma \sqrt{n}}$$

$$= \frac{108 - 100}{15 \times \sqrt{36}}$$

$$= \frac{\cancel{8}\,4}{15 \times \cancel{6}\,3} = \frac{4}{45} = 0.088$$

If Z_{stat} lies between -1.96 and 1.96 then we can accept the Hypothesis

$Z_{stat} = 0.088$ which lies between -1.96 and 1.96

Hence Hypothesis is accepted means raw corn starch had an effect.

2. In one state, 52% of voters are Republicans and 48% are Democrats. In a second state, 47% of voters are republicans and 53% are Democrats. Suppose a simple random sample of 100 voters are surveyed from each state.
What is the probability that the survey will show a greater percentage of republican voters in second state than in the first state.

Soln,

Mean of the difference in Republican proportion =

$$P_1 - P_2 = 0.52 - 0.47 = 0.05$$

Standard deviation of Republican difference in 2 states =

$$\sigma_d = \sqrt{\left[\frac{P_1(1-P_1)}{n_1}\right] + \left[\frac{P_2(1-P_2)}{n_2}\right]}$$

$$= \sqrt{\left[\frac{0.52 \times 0.48}{100}\right] + \left[\frac{0.47 \times 0.53}{100}\right]}$$

$$= \sqrt{\frac{0.2496}{100} + \frac{0.2491}{100}} = \sqrt{\frac{0.4987}{100}} = \sqrt{0.004987}$$

$$= 0.0706$$

we have to find out the probability that proportion of Republican voter in sample from first state (p_1) is less than proportion of republican voter in sample from 2nd State (p_2)

probability to find out ($p_1 - p_2$) is less than zero

$$Z_{(p_1-p_2)} = \frac{x - \mu_{p_1-p_2}}{\sigma_d}$$

$$= \frac{0 - 0.05}{0.0706} = -0.7082$$

As per Z-Table probability of a Z-Score being -0.7082 or less is 0.24

~~Propa~~ Probability to show greater percentage of republican voters in second State than first state = 0.24.

3. You take the SAT ~~and~~ Score 1100. The mean Score for the SAT is 1026 and the standard deviation is 209. How well did you Score on test compared to the average test taker.

Soln

$$Z = \frac{x-\mu}{\sigma} = \frac{1100 - 1026}{209} = \frac{74}{209}$$

$$= 0.354$$

As per z table value of Z = 0.1368

• Z table has shown Score for the right of the mean.

That's why percentage of test-taker Scored below you = 0.1368+0.50 = 0.6368 ~~=63.68~~

= 63.68 %

TASK - 2

1Q:-

	High-school	Bachelors	Masters	Ph.d	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

Are gender and education level dependent at 5% level of Significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained.

Soln

Hypothesis :-

H_0 :- Gender and education level are independent.

H_a :- Gender and education level are dependent.

The Significance level $= \alpha = 0.05$

We are going to perform Chi-square test for independence.

DF (Degree of Freedom) $= (r-1) \times (c-1)$

$= (2-1) \times (4-1)$

$= 1 \times 3 = 3$

E_{rc} (Expected frequency count for 'r' and 'c') =

$$n_r \times n_c / n$$

$$E_{1,1} = \frac{201 \times \cancel{100}^{20}}{\cancel{395}_{79}} = \frac{402}{7.9} = \frac{4020}{79} = 50.886$$

$$E_{1,2} = \frac{201 \times 98}{395} = \frac{19,698}{395} = 49.868$$

$$E_{1,3} = \frac{201 \times 99}{395} = \frac{19899}{395} = 50.377$$

$$E_{1,4} = \frac{201 \times 98}{395} = \frac{19698}{395} = 49.868$$

$$E_{2,1} = \frac{194 \times 100}{395} = \frac{19400}{395} = 49.113$$

$$E_{2,2} = \frac{194 \times 98}{395} = \frac{19012}{395} = 48.131$$

$$E_{2,3} = \frac{194 \times 99}{395} = \frac{19206}{395} = 48.623$$

$$E_{3,3} = \frac{194 \times 98}{395} = \frac{19012}{395} = 48.131$$

$$X^2 = \sum \left[(O_{rc} - E_{rc})^2 / E_{rc} \right]$$

- O_{rc} :→ observed frequency count for 'r' and 'c'

$= (60-50.886)^2/50.886 + (54-49.868)^2/49.868 +$
$(46-50.377)^2/50.377 + (41-49.868)^2/49.868 +$
$(40-49.113)^2/49.113 + (44-48.131)^2/48.131 +$
$(53-48.623)^2/48.623 + (57-48.131)^2/48.131$

$= \frac{83.064}{50.886} + \frac{17.0734}{49.868} + \frac{19.158}{50.377} + \frac{78.641}{49.868} +$
$\frac{83.047}{49.113} + \frac{17.065}{48.131} + \frac{19.158}{48.623} + \frac{78.659}{48.131}$

$= 1.632 + 0.342 + 0.380 + 1.576 + 1.690 + 0.354 +$
$0.394 + 1.635$

$= 8.003$

The P-value is the probability that a chi-square statistics having 3 degrees of freedom is more extreeme than 8.003

as per chi-square table $= P(X^2 > 8.003) = 0.05$

Since P-value is simillar to the signifiance level (0.05)

So we can accept the null hypothesis (i.e) Gender and education level are independent.

Q2 :-) Using the following data, perform a oneway analysis of variance using $\alpha = 0.05$. write up the results in APA Format.

Group1	Group2	Group3
51	23	56
45	43	76
33	23	74
45	43	87
67	45	56

Soln :-

$$SS_{total} = \left(\sum x_1^2 + \sum x_2^2 + \sum x_3^2\right) - \frac{(\sum x_1 + \sum x_2 + \sum x_3)^2}{n}$$

$$\sum x_1^2 = (51)^2 + (45)^2 + (33)^2 + (45)^2 + (67)^2$$

$$= 2601 + 2025 + 1089 + 2025 + 4489$$

$$= 12229$$

$$\sum x_2^2 = (23)^2 + (43)^2 + (23)^2 + (43)^2 + (45)^2$$

$$= 529 + 1849 + 529 + 1849 + 2025$$

$$= 6781$$

$$\sum x_3^2 = (56)^2 + (76)^2 + (74)^2 + (87)^2 + (56)^2$$

$$= 3136 + 5776 + 5476 + 7569 + 3136$$

$$= 25093$$

$\sum x_1 = 241$ $\quad \sum x_2 = 177$ $\quad \sum x_3 = 349$

$$SS_{total} = (12229 + 6781 + 25093) - \frac{(241 + 177 + 349)^2}{15}$$

$$= 44103 - \frac{(767)^2}{15} = 44103 - \frac{588289}{15}$$

$$= 44103 - 39219.2667 = 4883.73$$

$$SS_{among} = \left[\frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \frac{(\sum x_3)^2}{n_3}\right] - \frac{(\sum x_1 + \sum x_2 + \sum x_3)^2}{N}$$

$$= \cancel{\left[\frac{12229}{5} + \frac{6781}{5} + \frac{25093}{5}\right] - 39219.2667}$$

$$= \cancel{\left[\frac{44103}{5}\right]} -$$

$$= \left[\frac{(241)^2}{5} + \frac{(177)^2}{5} + \frac{(349)^2}{5}\right] - 39219.2667$$

$$= \left[\frac{58081}{5} + \frac{31329}{5} + \frac{121801}{5}\right] - 39219.2667$$

$$= \left[\frac{\overset{42242.5}{\cancel{211211}}}{\cancel{5}}\right] - 39219.2667$$

$$= 42242.5 - 39219.2667$$

$$= 3023.2333$$

$$SS_{within} = SS_{total} - SS_{among}$$

$$= 4883.73 - 3023.23$$

$$= 1860.50$$

$$df_{among} = r - 1 = 3 - 1 = 2$$

$$df_{within} = N - r = 15 - 3 = 12$$

$$MS_{among} = \frac{SS_{among}}{df_{among}} = \frac{3023.23}{2} = 1511.615$$

$$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{1860.50}{12} = 155.04$$

$$F = \frac{MS_{among}}{MS_{within}} = \frac{1511.615}{155.04} = 9.75$$

APA Format Result :-

Source	SS	df	MS	F
Among	3023.2333	2	1511.615	9.75
Within	1860.50	12	155.04	

3Q :-> Calculate F Test for given 10, 20, 30, 40, 50 and 5, 10, 15, 20, 25

For 10, 20, 30, 40, 50

Solⁿ

$$F\ Test = \frac{\text{variance of } (10, 20, 30, 40, 50)}{\text{variance of } (5, 10, 15, 20, 25)}$$

variance of 10, 20, 30, 40, 50 = V_1

variance of 5, 10, 15, 20, 25 = V_2

Mean of 10, 20, 30, 40, 50 = $\bar{x}_1$

Mean of 5, 10, 15, 20, 25 = $\bar{x}_2$

$$\bar{x}_1 = \frac{10+20+30+40+50}{5} = \frac{150}{5} = 30$$

$$\bar{x}_2 = \frac{5+10+15+20+25}{5} = \frac{75}{5} = 15$$

$$v_1 = \frac{(10-30)^2+(20-30)^2+(30-30)^2+(40-30)+(50-30)^2}{5-1}$$

$$= \frac{400+100+0+100+400}{4}$$

$$= \frac{1000}{4} = 250$$

$$v_2 = \frac{(5-15)^2+(10-15)^2+(15-15)^2+(20-15)^2+(25-15)^2}{4}$$

$$= \frac{100+25+0+25+100}{4}$$

$$= \frac{250}{4} = 62.5$$

$$F\ Test = \frac{v_1}{v_2} = \frac{250}{62.5} = 4$$

F Test value = 4