Task 1:

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

Solution:

The accepted fact is that the population mean is 100

State the Null hypothesis. $\mathbf{H_0}: \boldsymbol{\mu} = \mathbf{100}$

State the Alternate hypothesis $\mathbf{H_1}: \boldsymbol{\mu} \neq \mathbf{100}$

Set up the significance level. It is not given in the problem so assume it as **5% (0.05)**.

Compute the random chance probability using Z score and Z-table.

Population mean $\boldsymbol{\mu} = \mathbf{100}$

Standard Deviation $\boldsymbol{\sigma} = \mathbf{15}$

Sample Mean $\bar{\mathbf{x}} = \mathbf{108}$

Total Sample Data $\mathbf{n} = \mathbf{36}$

$$\mathbf{Z} = \frac{\bar{\mathbf{x}} - \boldsymbol{\mu}}{\boldsymbol{\sigma}/\sqrt{\mathbf{n}}}$$

$$\mathrm{Z} = \frac{(108 - 100)}{15/\sqrt{36}}$$

$\mathbf{Z = 3.20}$

The P-value associated with the Z-value of 3.20 is **0.9993**. This means that the probability of having a value less than 108 is 0.9993, while the probability of having a value equal to (or more than) 108 is (1-0.9993), which is equal to 0.0007.

As the computed value of 0.0007 is less than the significance level of 0.05, the Null hypothesis test to determine the raw corn-starch effect can be rejected.

---

2. In one state, 52% of the voters are Republicans, and 48% are Democrats. In a second state, 47% of the voters are Republicans, and 53% are Democrats. Suppose simple random samples of 100 voters are surveyed from each state.

What is the probability that the survey will show a greater percentage of Republican voters in the second state than in the first state?

Solution:

The proportion of Republican voters in the first state: $\boldsymbol{R_1 = 0.52}$

The proportion of Democrats voters in the first state: $\boldsymbol{D_1 = 0.48}$

The proportion of Republican voters in the second state: $\boldsymbol{R_2 = 0.47}$

The proportion of Democrats voters in the second state: $\boldsymbol{D_2 = 0.53}$

The number of voters sampled from the first state: $\boldsymbol{n_1 = 100}$

The number of voters sampled from the second state: $\boldsymbol{n_2 = 100}$

Mean of the difference in sample proportion of $\boldsymbol{R_1 \text{ and } R_2 = \mu = 0.52 - 0.47 = 0.05}$

Standard deviation of the difference

$$\boldsymbol{\sigma = \sqrt{\frac{R_1 D_1}{n_1} + \frac{R_2 D_2}{n_2}}}$$

$$\sigma = \sqrt{((0.52 \times 0.48)/100 + (0.47 \times 0.53)/100)}$$

$$\boldsymbol{\sigma = 0.0706}$$

Find the probability. This problem requires us to find the probability that $\boldsymbol{R_1 < R_2}$. This is equivalent to finding the probability that $\boldsymbol{R_1 - R_2 < 0}$. To find this probability, we need to transform the random variable ($\boldsymbol{R_1 - R_2}$) into a z-score. That transformation appears below.

$$\boldsymbol{Z = \frac{x - \mu}{\sigma}}$$

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

$$\boldsymbol{Z = -0.7082}$$

The corresponding percentage probability of $\boldsymbol{Z = -0.7082}$ is **0.24**

Therefore, the probability that the survey will show a greater percentage of Republican voters in the second state than in the first state is 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 the test compared to the average test taker?

Solution:

SAT score: $\mathbf{x = 1100}$

Mean SAT score: $\boldsymbol{\mu} = \mathbf{1026}$

Standard Deviation: $\boldsymbol{\sigma} = \mathbf{209}$

To calculate the z-score we use the formula:

$$\mathbf{Z = \frac{x - \mu}{\sigma}}$$

$$\mathrm{Z} = \frac{1100 - 1026}{209}$$

$$\mathbf{Z = 0.354}$$

From the z-table for the z-value **(Z = 0354)** we get **.6368**

Therefore 63.38% of test-takers scored below you

Task 2:

1. Is gender independent of education level? Random samples of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

	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

Question: 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?

Solution:

$\mathbf{H_0}$**:** Gender and education independent

$\mathbf{H_1}$**:** Gender and education dependent

The Observed Frequencies:

	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**

The Expected frequency under the null hypothesis is given

$$\boldsymbol{Expected\ Frequencies} = \frac{(\boldsymbol{Row\ Total} \times \boldsymbol{Column\ Total})}{\boldsymbol{Sample\ Size}}$$

Row totals and column totals are given above sample size = 395

The Expected Frequencies:

	High School	Bachelors	Masters	Ph.d	Total
Female	50.886	49.868	50.377	49.868	**201**
Male	49.114	48.132	48.623	48.132	**194**
Total	**100**	**98**	**99**	**98**	**395**

$$\chi^2 = \sum \frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}$$

$$\chi^2 = \frac{(60-50.886)^2}{50.886} + \frac{(54-49.868)^2}{49.868} + \frac{(46-50.377)^2}{50.377} + \frac{(41-49.868)^2}{49.868} + \frac{(40-49.114)^2}{49.114} + \frac{(44-48.132)^2}{48.132} + \frac{(53-48.623)^2}{48.623} + \frac{(57-48.132)^2}{48.132}$$

$$\chi^2 = 8.006066$$

Formula to calculate Degree of Freedom:

$$\boldsymbol{Degree\ of\ Freedom} = (\boldsymbol{No.of\ Columns} - \boldsymbol{1}) \times (\boldsymbol{No.of\ Rows} - \boldsymbol{1})$$

$$\boldsymbol{Degree\ of\ Freedom} = (4-1) \times (2-1) = \boldsymbol{3}$$

Now, referring Chi Squared table with Degree of freedom = 3 and level of significance = 0.05, the critical value $\boldsymbol{\chi^2}_{\boldsymbol{critical}} = \boldsymbol{7.815}$

It's found that the calculated $\boldsymbol{\chi^2} > \boldsymbol{\chi^2}_{\boldsymbol{critical}}$ $(\boldsymbol{8.006066} > \boldsymbol{7.815})$

Inference: We reject the null hypothesis. Therefore, gender and education are dependent

2. Using the following data, perform a one way analysis of variance using α=.05. Write up the results in APA format.

[Group1: 51, 45, 33, 45, 67]

[Group2: 23, 43, 23, 43, 45]

[Group3: 56, 76, 74, 87, 56]

Solution:

Using one way ANOVA, we conduct Hypothesis Testing.

Total number of data points across all groups: $\boldsymbol{N = 15}$

Total number of data points within an individual group: **n**$= \mathbf{5}$

Total number of levels of factor: $\boldsymbol{a = 3}$

$\boldsymbol{\alpha = 0.05}$

	Group 1	Δ	Δ^2	Group 2	Δ	Δ^2	Group 3	Δ	Δ^2
	51	2.8	7.84	23	-12.4	153.76	56	-13.8	190.44
	45	-3.2	10.24	43	7.6	57.76	76	6.2	38.44
	33	-15.2	231.04	23	-12.4	153.76	74	4.2	17.64
	45	-3.2	10.24	43	7.6	57.76	87	17.2	295.84
	67	18.8	353.44	45	9.6	92.16	56	-13.8	190.44
Sum	**241**		**612.8**	**177**		**515.2**	**349**		**732.8**
Mean	**48.2**			**35.4**			**69.8**		

Mean of Groups:

$\bar{X}_1 = 48.2$

$\bar{X}_2 = 35.4$

$\bar{X}_3 = 69.8$

Sum of Squared Deviation:

$SS_1 = 612.8$

$SS_2 = 515.2$

$SS_3 = 732.8$

Variances for the group:

$S^2{}_1 = \frac{612.8}{4} = \mathbf{153.2}$

$S^2{}_2 = \frac{515.2}{4} = \mathbf{128.8}$

$S^2{}_2 = \frac{732.8}{4} = \mathbf{183.2}$

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

$df_{total} = N - 1 = 15 - 1 = 14$

$df_{between} = a - 1 = 3 - 1 = 2$

$MS_{within} = \frac{153.2+128.8+183.2}{3} = \mathbf{155.07}$

$SS_{within} = MS_{within} \times df_{within} = 155.07 \times 12 = \mathbf{1860.8}$

	Group Mean	Group Δ	Group Δ^2
	48.2	-2.93	8.58
	35.4	-15.73	247.43
	69.8	18.67	348.57
Sum	**153.4**		**604.58**
Mean	**51.13**		

Grand Mean:

$$\bar{X}_{group} = \frac{48.2+35.4+69.8}{3} = 51.13$$

Sum of Squared Deviation:

$$SS_{group} = 8.58 + 247.43 + 348.57 = \mathbf{604.58}$$

Variances for the group:

$S^2{}_{group} = \frac{\mathbf{604.58}}{3-1} = \mathbf{302.29}$

$MS_{between} = 302.29 \times 5 = \mathbf{1511.45}$

$SS_{between} = MS_{between} \times df_{between} = 1511.45 \times 2 = \mathbf{3022.9}$

$F = \frac{MS_{between}}{MS_{within}} = \frac{1511.45}{155.07} = \mathbf{9.75}$

Referring F table for α = 0.05, the corresponding value for $df_{between}$ and df_{within} is **3.8853** which means if the calculated F is greater than $\boldsymbol{F_{critical}(2, 12) = 3.8853}$ then we reject Null Hypothesis.

Inference: Reject Null Hypothesis

3. Calculate F Test for given 10, 20, 30, 40, 50 and 5, 10, 15, 20, 25.

Solution:

Calculate Variance of first set: 10, 20, 30, 40, 50

$\boldsymbol{X_1 = 10,\ X_2 = 20, X_3 = 30, X_4 = 40, X_5 = 50}$

$\boldsymbol{n = 5}$

To calculate the average or the mean of the above data, we use the formula:

$$\boldsymbol{\overline{X} = \frac{\sum_{i=1}^{n}(X_i)}{n}}$$

$$\overline{X} = \frac{10 + 20 + 30 + 40 + 50}{5}$$

$\boldsymbol{\overline{X} = 30}$

To calculate the variance we use the formula:

$$\boldsymbol{S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

$$\boldsymbol{S^2}_1 = \frac{(10-30)^2 + (20-30)^2 + (30-30)^2 + (40-30)^2 + (50-30)^2}{5-1}$$

$\boldsymbol{S^2{}_1 = 250}$

The Variance is **250**

Calculate Variance of first set: 5, 10, 15, 20, 25

$\boldsymbol{X_1 = 5,\ X_2 = 10, X_3 = 15, X_4 = 20, X_5 = 25}$

$\boldsymbol{n = 5}$

$$\overline{X} = \frac{5 + 10 + 15 + 20 + 25}{5}$$

$\boldsymbol{\overline{X} = 15}$

$$\boldsymbol{S^2}_2 = \frac{(5-15)^2 + (10-15)^2 + (15-15)^2 + (20-15)^2 + (25-15)^2}{5-1}$$

$\boldsymbol{S^2{}_2 = 62.50}$

The Variance is **62.50**

To calculate F Test

$$F = \frac{s^2{}_1}{s^2{}_2}$$

Where $\mathbf{S^2{}_1}$ and $\mathbf{S^2{}_2}$ are the variance

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

$$F = 4$$

The F Test value is 4.