Assignment – 10

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. 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 cornstarch had an effect or not.

<u>Solution:</u>

no_of_sample = 36

sample_mean = 108

population_mean = 100

population_sigma = 15

Step-1: State the hypothesis. The population mean is 100.

H_0: μ=100 => Null hypothesis

H_1: ≠100 => Alternate hypothesis

Step-2: Set up the significance level. It is not given in the problem so let's assume it as 5% (0.05).

This 5% is called Significance Level also known as alpha level (symbolized as α).

It means that if random chance probability is less than 5% then we can conclude that there is difference in behavior of two different population.

(1- Significance level) is also known as Confidence Level

i.e. we can say that I am 95% confident that it is not driven by randomness.

Step-3: Calculate Z score

$$Z = \frac{X - \mu}{(\frac{\sigma}{\sqrt{n}})}$$

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

$$Z = 3.2$$

By looking at z-table and p-value associated with **3.20** is **0.9993**

The probability of having value less than 108 is 0.9993 and more than or equals to 108 is $(1 - 0.9993)$ $=0.0007$.

Step-4: Since the probability of having mean glucose level more than or equals to 108 is **0.0007** which is less than **0.05**

So we will reject the Null hypothesis i.e. there is raw cornstarch effect.

---

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 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 the second state than in the first state?

<u>Solution:</u>

P1 = the proportion of Republican voters in the first state

P2 = the proportion of Republican voters in the second state

p1 = the proportion of Republican voters in the sample from the first state

p2 = the proportion of Republican voters in the sample from the second state

The number of voters sampled from the first state (n1) = 100, and the number of voters sampled from the second state (n2) = 100.

Step-1: Make sure the samples from each population are big enough to model differences with a normal distribution. Because

$n1P1 = 100 * 0.52 = 52$

$n1(1 - P1) = 100 * 0.48 = 48$

$n2P2 = 100 * 0.47 = 47$

$n2(1 - P2) = 100 * 0.53 = 53$ are each greater than 10, the sample size is large enough.

Step-2: Find the mean of the difference in sample proportions:

$$E(p1 - p2) = P1 - P2 = 0.52 - 0.47 = 0.05$$

Step-3: Find the standard deviation of the difference.

$$\sigma_d = \sqrt{\left(\frac{P1(1-P1)}{n1}\right) + (\frac{P2(1-P2)}{n2})}$$

$$\sigma_d = \sqrt{\left(\frac{0.52(0.48)}{100}\right) + (\frac{0.47(0.53)}{100})}$$

$$\sigma_d = \sqrt{0.002496 + 0.002491}$$

$$\sigma_d = \sqrt{0.002496 + 0.002491}$$

$$\sigma_d = \sqrt{0.004987}$$

$$\sigma_d = \sqrt{0.004987}$$

$$\sigma_d = 0.070618$$

Step-4: This problem requires us to find the probability that p1 is less than p2. This is equivalent to finding the probability that p1 - p2 is less than zero. To find this probability, we need to transform the random variable (p1 - p2) into a z-score. That transformation appears below.

$$Z_{p1-p2} = \frac{(X - \mu_{p1-p2})}{\sigma_d}$$

$$Z_{p1-p2} = \frac{(0 - 0.05)}{0.070618}$$

$$Z_{p1-p2} = \frac{(0 - 0.05)}{0.070618}$$

$$Z_{p1-p2} = -0.7080$$

We find that the probability of a z-score being **-0.7080** 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:

Step-1: Write your X-value into the z-score equation. For this sample question the X-value is your SAT score, 1100.

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{1100 - \mu}{\sigma}$$

Step-2: Write the mean, μ = 1026, into the z-score equation.

$$Z = \frac{1100 - 1026}{\sigma}$$

Step-3: Write the standard deviation, σ=209 into the z-score equation.

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

Step-4: Calculating the z-score.

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

$$Z = 0.3540$$

This means thatz-score was **0.3540** standard deviations above the mean.

Step-5: Looking up the z-value in the z-table to see whatpercentage of test-takers scored below you.

P(X<1100) =P (z<0.3540) =**0.6368 or 63.68%**

Task 2:

1. Is gender independent of education level? A random sample 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:

H0: Gender and education independent

H1: Gender and education dependent

	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

Chi-Square Test Statistic

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

Where O represents the observed frequency. E is the expected frequency under the null hypothesis and computed by:

$$E = \frac{row\ total * column\ total}{sample\ size}$$

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

Therefore, we have expected frequencies as follows:

	High School	Bachelors	Masters	Ph.d.	Total
Female	50.886	49.868	50.377	49.868	201
Male	49.113	48.131	48.622	48.131	194
Total	100	98	99	98	395

$$\chi^2 = \frac{(60-50.886)^2}{50.886} + \frac{(54-49.868)^2}{49.868} + \frac{(46-50.377)^2}{50.377} + \frac{(41-49.868)^2}{49.868} + \frac{(40-49.113)^2}{49.113} + \frac{(44-48.1315)^2}{48.131} + \frac{(53-48.622)^2}{48.622} + \frac{(57-48.131)^2}{48.131}$$

$\chi^2 = 8.00608$

$Degree\ of\ Freedom = (No. Of\ Columns - 1) * (No. Of\ Rows - 1) = (4 - 1) * (2 - 1) = 3$

$\alpha = 0.05$

The critical value of χ^2 with 3 degree of freedom is **7.815**. Since **8.006 > 7.815**, therefore we reject the null hypothesis and conclude that the education level depends on gender at a 5% level of significance.

---

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:

Total data sets in all groups = $\boldsymbol{N} = 15$

Total of all data sets within each group = $\boldsymbol{n} = 5$

The number of groups = $\boldsymbol{k} = 3$ (So, if there are three groups of measurements, then $\boldsymbol{k} = 3$)

Sample mean ($\bar{X}$) for the Group1: = 48.2

Sample mean ($\bar{X}$) for the Group2: = 35.4

Sample mean ($\bar{X}$) for the Group3: = 69.8

Calculating the group variances:

	Group1	D	D^2	**Group2**	D	D^2	**Group3**	D	D^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		122.56	35.4		103.04	69.8		146.56

Sum of squared deviations from the mean (SS) for the groups:

$$Var_1({S^2}_1) = \frac{612.8}{5-1} = 153.2$$

$$Var_2({S^2}_2) = \frac{515.2}{5-1} = 128.8$$

$$Var_3({S^2}_3) = \frac{732.8}{5-1} = 183.2$$

The degrees of freedom (df) are calculated using the formula below:

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

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

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

Calculating the remaining 'within' terms for the ANOVA table:

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

$$SS_{within} = MS_{within} * df_{within} = MS_{within} * (N-k) = (155.07) * (15-3) = 1860.84$$

Calculating the variance of the sample means:

	$Group\ Mean$	$Group\ D$	$Group\ D^2$
	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		201.52

$$Grand\ Mean\ (\bar{X}_{grand}) = \frac{48.2 + 35.4 + 69.8}{3} = 51.13$$

$$Sum\ of\ squares\ (SS_{means}) = 8.58 + 247.33 + 348.57 = 604.58$$

$$Var_{means}({S^2}_{means}) = \frac{604.58}{3-1} = 302.29$$

$$MS_{between} = {S^2}_{means} * n = (302.29) * (5) = 1511.45$$

Calculating the remaining group terms of the ANOVA table:

$$SS_{group} = MS_{between} * df_{between} = MS_{between} * (k-1) = (1511.45) * (3-1) = 3022.9$$

Test statistic and critical value:

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

$$\boldsymbol{F_{critical}}(\mathbf{2, 12}) = \mathbf{3.89}$$

Since the calculated (absolute value) of F is greater than the tabulated value, we reject the null hypothesis and conclude that at least two of the means are significantly different from each other.

Decision: Reject the 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:

For 10, 20, 30, 40, 50:

Total Inputs= $\boldsymbol{n} = 5$

Calculating Mean:

$$\bar{X} = \frac{1}{\mathrm{n}} * \sum_{i=1}^{n} \mathrm{x_i}$$

$$\bar{X} = \frac{\mathrm{X1+X2+X3+X4+X5}}{\mathrm{n}} = \frac{10+20+30+40+50}{5} = \frac{150}{5} = 30$$

Calculating Variance:

$$S^2 = \frac{\sum(X - \bar{X})^2}{N - 1}$$

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

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

Calculate Variance of second set:

For 5, 10,15,20,25:

Total Inputs= $\boldsymbol{N} = 5$

Calculating Mean:

$$\bar{X} = \frac{\mathrm{X1} + \mathrm{X2} + \mathrm{X3} + \mathrm{X4} + \mathrm{X5}}{\mathrm{n}} = \frac{5 + 10 + 15 + 20 + 25}{5} = \frac{75}{5} = 15$$

Calculating Variance:

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

$$S^2{}_2 = 62.5$$

To calculate F Test:

$$\boldsymbol{F} = \frac{S^2{}_1}{S^2{}_2} = \frac{250}{62.5} = \mathbf{4}$$

The F Test value is **4**.