# Question 1

## PART A

There are three ways to assign probabilities to events:

1. classical approach: If an experiment has n simple outcomes, this method will assign a probability of 1/n to each outcome. In other words, each outcome is assumed to have an equal probability of occurrence. This method is also called the axiomatic approach.
2. relative-frequency approach: Probabilities are assigned based on experimentation or historical data.
3. subjective approach: In the subjective approach, we define probability as the degree of belief that we hold in the occurrence of an event. Thus, judgment is used as the basis for assigning probabilities.

## PART B

* 1. What percentage of the members felt that both job security and salary increment were important?

P(the job security was the important issue) = 74% = 0.74

P(salary increment was an important issue) = 65% = 0.65

P(job security was an important issue | salary increment was an important issue) = 60% = 0.60

P(both job security and salary increment were important)

= P (job security was an important issue | salary increment was an important issue) \*P(salary increment was an important issue)

= 0.60\*0.65

= 0.39

**P (both job security and salary increment were important) = 39%**

* 1. What percentage of the members felt that at least one of these two issues was important?

P (at least one of these two issues was important)

= P (the job security was the important issue) + P (salary increment was an important issue) - P(job security was an important issue | salary increment was an important issue)

= 0.74+0.65-0.39

= 1.00

**P(at least one was important)= 100%**

# Question 2

## PART A

P (X > 12)

= 1- P(X ≤ 12)

= 1- P( Z ≤ (12-10)/3)

= 1- P(Z≤0.6667)

= 0.2546

## PART B

If 25 students consume more than 275 meals, then one student must consume atleast 11 meals. (=275/25)

Therefore

P(X > 11)

= 1- P ( X ≤11)

= 1- P( Z ≤ (11-10)/3)

= 0.3707

# Question 3

ANSWER:

To test the effectiveness of the equipment, let us assume that Mu1 be the average person hours lost before installation. Let Mu2 me the average person hours lost after installation. Let **MuD be the difference between the person hours lost before and after the installation.**

We need to test the person hours lost before the installation is more than after the installation at 10% level of significance.

## PART A

Null hypothesis: Ho: MuD = 0

Alternative hypothesis: H1: MuD > 0 ( this means that there were more person hours lost before installation than after installation)

## PART B

Test statistic for MuD

t = (sample mean – MuD)/ standard error

This will follow a student t distribution with n-1 degree of freedom.

We have chosen the t statistic since population standard deviation is unknown and the test conducted will be a one tailed test since the alternative hypothesis is that there are more person hours lost after the installation.

## PART C

Test statistic is t = (-1.2 – 0)/(5/sqrt (50)) = -1.2 / 0.7071 = -1.697

With Degree of freedom = 49 and for 10% level of significance and one tailed test we have the p-value is .04802 and the critical t value is 1.299

## PART D

The decision criteria are if p value is greater than the chosen level of significance, we do not reject the null hypothesis.

## PART E

Since the p value = 0.048 is less than the chosen level of significance at 0.10 therefore we reject our null hypothesis. Also, since the calculated t-statistic lies in the critical region, so there is enough evidence to reject H₀ at 10% level of significance. Therefore, we can conclude that the person hours lost before the installation was more than after the installation

# Question 4

|  |  |  |
| --- | --- | --- |
| **A** | **B** | **C** |
| 6 | 5 | 6 |
| 5 | 5 | 7 |
| 4 | 4 | 6 |
| 5 | 4 | 5 |
| 6 | 5 | 6 |
| 4 | 4 | 6 |
| 5 | 5 | 6 |
| 4 | 6 | 6 |
| 6 | 5 | 4 |
| 5 | 6 | 5 |

## PART A

null and alternative hypothesis for single factor ANOVA to test for any significant difference in the perception among three groups.

H0: There is no difference in the perception among three groups or: μA = μB = μC

H1: The means are not all equal.

## PART B

State the decision rule at 5% significance level.

The above hypotheses will be tested using an F-ratio for a One-Way ANOVA. Based on the information provided, the significance level is alpha = 0.05 (α=0.05), and the degrees of freedom are df1 = 2 df2 = 2 therefore, the rejection region for this F-test is when F statistic is greater than the critical F value of 3.354

## PART C

F test statistic is calculated as

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Group A** | **Group B** | **Group C** |
|  | 6 | 5 | 6 |
|  | 5 | 5 | 7 |
|  | 4 | 4 | 6 |
|  | 5 | 4 | 5 |
|  | 6 | 5 | 6 |
|  | 4 | 4 | 6 |
|  | 5 | 5 | 6 |
|  | 4 | 6 | 6 |
|  | 6 | 5 | 4 |
|  | 5 | 6 | 5 |
|  |  |  |  |
| sum | 50 | 49 | 57 |
| average | 5 | 4.9 | 5.7 |
| sum of squared value | 256 | 245 | 331 |
| standard deviation | 0.816 | 0.738 | 0.823 |
| SS | 6 | 4.9 | 6.1 |
| n | 10 | 10 | 10 |

The total sample size is N = 30. Therefore, the total degrees of freedom are:

*dftotal* ​= 30−1=29

Also, the between-groups degrees of freedom are

*dfbetween*​=3−1=2,

and the within-groups degrees of freedom are:

*dfwithin*​ = *dftotal*​−*dfbetween*​ = 29−2 = 27

First, we need to compute the total sum of values and the grand mean. The following is obtained.

∑​*Xij*​=50+49+57=156

Also, the sum of squared values is

∑​*Xij*2​=256+245+331=832

Based on the above calculations, the total sum of squares is computed as follows

*SStotal*​= 832−((156)2/30)​=20.8

The within sum of squares is computed as

*SSwithin*​=∑*SSwithingroups*​=6+4.9+6.1=17

Next, we calculate the mean sum of squares:

*MSbetween*​= *SSbetween/ dfbetween*​ ​​=3.8/2​=1.9

*MSwithin*​= ​*SSwithin/ dfwithin* ​​=17/27 ​=0.63

F-statistic is computed as follows:

***F*= *MSbetween/ MSwithin*​ ​​=1.9/ 0.63​=3.018**

## PART D

Based on the calculated test statistics decide whether any significant difference in the mean price of gasoline for three bands.

Answer: Since it is observed that F = 3.018 is less than the critical F value of 3.354 therefore we do not reject the null hypothesis at 5% level of significance. Using the P-value approach: The p-value is p = 0.0656 and since p = 0.0656 is greater than the chosen level of significance at 5% therefore we do not reject the null hypothesis. It is concluded that the null hypothesis Ho is not rejected. Thus, there is no significance difference in he mean price of gasoline for three brands.

# Question 5

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Interest  rate | 3.5 | 3.0 | 2.8 | 3.6 | 2.75 | 3.4 | 3.12 | 2.86 | 3.02 | 2.6 | 3.3 |
| Housing  starts | 100 | 120 | 150 | 130 | 170 | 135 | 130 | 185 | 127 | 190 | 96 |

## PART 1 Derive the regression equation.

**Let X denote interest rate and Y denote Housing start**

|  |  |  |  |
| --- | --- | --- | --- |
| *X* - Mx | *Y* - My | (*X* - Mx)2 | (*X* - Mx)(*Y* - My) |
| 0.4136 | -39.3636 | 0.1711 | -16.2822 |
| -0.0864 | -19.3636 | 0.0075 | 1.6723 |
| -0.2864 | 10.6364 | 0.082 | -3.0459 |
| 0.5136 | -9.3636 | 0.2638 | -4.8095 |
| -0.3364 | 30.6364 | 0.1131 | -10.305 |
| 0.3136 | -4.3636 | 0.0984 | -1.3686 |
| 0.0336 | -9.3636 | 0.0011 | -0.315 |
| -0.2264 | 45.6364 | 0.0512 | -10.3304 |
| -0.0664 | -12.3636 | 0.0044 | 0.8205 |
| -0.4864 | 50.6364 | 0.2365 | -24.6277 |
| 0.2136 | -43.3636 | 0.0456 | -9.264 |
|  |  | SSx = 1.0749 | SP= -77.855 |

Sum of X (interest rate)= 33.95

Sum of Y (housing starts) = 1533

Mean interest rate = 3.0864

Mean housing starts = 139.3636

Sum of squares (SSX) for interest rates = 1.0749

Sum of products (SP) = -77.8555

Regression Equation = ŷ = bX + a

b = SP/SSX = -77.86/1.07 = -72.43

a = MY - bMX = 139.36 - (-72.43\*3.09) = 362.920

ŷ = -72.43X + 362.920

Thus **Housing starts = -72.43 \* interest rates + 362.920**

## PART 2 Estimate the no of housing starts if mortgage interest rate is 2.5%

When X = 2.5 then Housing starts = -72.43 \* 2.5 + 362.920 = 181.845

## PART 3 Calculate and interpret the correlation between interest rate and no of housing starts.

∑X = 33.95

Mean X = 3.086

∑(X - Mx)2 = SSx = 1.075

∑Y = 1533

Mean Y= 139.364

∑(Y - My)2 = SSy = 9850.545

N = 11

∑(X - Mx)(Y - My) = -77.855

Correlation coefficient Calculation

r = ∑((X - My)(Y - Mx)) / √((SSx)(SSy))

r = -77.855 / √((1.075)(9850.545)) = -0.7566

The correlation coefficient implies that there is a negative relationship between the interest rate and the housing starts. As interest rate increases, the number of housing starts decreases and vice versa. The correlation coefficient indicates a strong negative correlation between the interest rate and the housing start as well since the value of the correlation coefficient is close to 75.66%

# Question 6

## PART A

Complete the missing entries from **A to H** in this output

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |
| *Regression Statistics* | | | | | |
| Multiple R | 0.442 |  |  |  |  |
| R Square | **0.195364** |  |  |  |  |
| Adjusted R Square | 0.08 |  |  |  |  |
| Standard Error | 2.587 |  |  |  |  |
| Observations | 25 |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  |  |  |  |  | *Significance F* |
| *Df* | *SS* | *MS* | *F* |
| Regression | **3** | 34.1036 | **11.368** | **1.698** | 0.1979 |
| Residual | 21 | **140.5595** | 6.6933 |  |  |
| Total | **27** | 174.6631 |  |  |  |
|  |  |  |  |  |  |
|  |  | *Standard Error* |  | *P-* |  |
| *Coefficients* | *t Stat* | *value* | *Lower 95%* |
| Intercept | 12.31 | 4.7 | 2.62 | 0.02 | 2.54 |
| Direct | 0.57 | 1.72 | **0.331** | 0.74 | -3.01 |
| Newspaper | 3.32 | 1.54 | 2.16 | 0.04 | 0.12 |
| Television | **0.725** | 1.96 | 0.37 | 0.71 | -3.34 |

## PART B

Assess the independent variables significance at 5% level

For each independent variable we will need to assess their individual significance at 5% level of significance using t test.

For **direct variable** we find that assuming true the null hypothesis that the regression coefficient for direct is 0 the p value is found to be 0.74 which is greater than the chosen level of significance at5%. Therefore, we do not reject the null hypothesis. Thus, direct is not significant at 5% level.

For **NEWSPAPER variable** we find that assuming true the null hypothesis that the regression coefficient is 0 the p value is found to be 0.04 which is lower than the chosen level of significance at 5%. Therefore, we reject the null hypothesis. Thus, NEWSPAPER is significant at 5% level.

For **TELEVISION variable** we find that assuming true the null hypothesis that the regression coefficient is 0 the p value is found to be 0.71 which is greater than the chosen level of significance at 5%. Therefore, we do not reject the null hypothesis. Thus, TELEVISION is not significant at 5% level.

For **INTERECEPT** we find that assuming true the null hypothesis that the regression coefficient is 0 the p value is found to be 0.02 which is lower than the chosen level of significance at 5%. Therefore, we reject the null hypothesis. Thus, INTERCEPT is significant at 5% level.

## PART C

Does the model is significant at 5% level?

We can conduct the F test for overall significance of the regression model. Assuming the null hypothesis to be true that the regression coefficients are all simultaneously 0 the calculated p value for regression model is given to be 0.1979 which is greater than the chosen level of significance at 0.05. Therefore, we do not reject the null hypothesis. Thus, the regression model is not significant at 5% level of significance.