**NMIMS Global Access**

**School for Continuing Education (NGA-SCE)**

**Course:** **Decision Science**

**Internal Assignment Applicable for June 2023 Examination**

**Q 1: Bad gums may mean a bad mood. Researchers discovered that 85% of people who have suffered a bad mood had periodontal disease, an inflammation of the gums. Only 29% of healthy people have this disease. Suppose that in a certain community bad moods are quite rare, occurring with only 10% probability. If someone has periodontal disease, what is the probability that he or she will have a bad mood?**

To solve this problem, we can use Bayes' theorem. Let's define the events:

A: Having periodontal disease

B: Having a bad mood

We are given the following probabilities:

P(B) = 0.10 (probability of having a bad mood)

P(A|B) = 0.85 (probability of having the periodontal disease given a bad mood)

P(A') = 0.29 (probability of not having periodontal disease, which is equal to 1 - P(A))

We want to find P(B|A), the probability of having a bad mood given periodontal disease.

Using Bayes' theorem:

**P(B|A) = (P(A|B) \* P(B)) / P(A)**

P(A) can be calculated using the law of total probability:

**P(A) = P(A|B) \* P(B) + P(A|B') \* P(B')**

Calculate P(A):

P(A) = P(A|B) \* P(B) + P(A|B') \* P(B')

= 0.85 \* 0.10 + 0.29 \* (1 - 0.10)

= 0.085 + 0.29 \* 0.9

= 0.085 + 0.261

= 0.346

Now we can substitute the values into Bayes' theorem:

P(B|A) = (P(A|B) \* P(B)) / P(A)

= (0.85 \* 0.10) / 0.346

= 0.085 / 0.346

≈ 0.2459

Therefore, if someone has periodontal disease, the probability that he or she will have a bad mood is approximately 0.2459, or about 24.59%.

**Q2. Using MS-EXCEL to show the Regression model, consider ‘Instagram followers’ as dependent variable and ‘no f post per day’ as an independent variable. Write the interpretation of EXCEL Tables. Write the conclusion on the fitting of your model also.**



To interpret the Excel tables:

1. The regression output will provide coefficients for the intercept and the independent variable (no posts per day). The coefficient for the independent variable represents the estimated change in the number of followers for a one-unit increase in the number of posts per day.

2. The standard error column provides an estimate of the standard deviation of the coefficient. Lower standard errors indicate greater precision in the coefficient estimation.

3. The t-value column represents the ratio of the coefficient estimate to its standard error. It is used to test the hypothesis that the coefficient is equal to zero. Generally, larger t-values indicate greater significance.

4. The p-value column indicates the probability of obtaining the observed t-value or more extreme values if the null hypothesis (coefficient = 0) is true. Lower p-values indicate greater evidence against the null hypothesis.

5. The R-squared value represents the proportion of the total variation in the dependent variable (followers) that is explained by the independent variable (no of post per day). A higher R-squared value indicates a better fit of the model to the data.

From the provided summary output, we can draw the following conclusions about the relationship between the number of posts per day and the number of followers on Instagram:

1. Coefficients: The coefficient for the intercept is 365.0211268, which represents the estimated number of followers when the number of posts per day is zero. The coefficient for the independent variable (2) is 4.063380282, indicating the estimated change in the number of followers for a one-unit increase in the number of posts per day.

2. Standard Error: The standard error for the coefficients represents the estimated standard deviation of the coefficient. In this case, the standard error for the intercept is 40.39703934, and for the independent variable (2) is 11.01037698.

3. t Stat and P-value: The t Stat column provides the t-value for each coefficient, which is used to test the hypothesis that the coefficient is equal to zero. The corresponding P-value column indicates the significance of the coefficient. In this case, both coefficients have P-values greater than 0.05 (e.g., 0.719095592), suggesting that they are not statistically significant at the conventional 5% significance level.

4. Confidence Intervals: The Lower 95% and Upper 95% columns provide the 95% confidence interval for each coefficient. These intervals give a range within which the true population coefficient is likely to fall. For example, the 95% confidence interval for the intercept is 276.1078427 to 453.9344109.

5. R-squared: The R-squared value represents the proportion of the total variation in the dependent variable (number of followers) that is explained by the independent variable (number of posts per day). In this case, the R-squared value is 0.012230202, indicating that only about 1.23% of the variation in the number of followers can be explained by the number of posts per day.

Overall, the regression model does not provide strong evidence of a significant relationship between the number of posts per day and the number of followers on Instagram. The coefficients are not statistically significant, and the low R-squared value suggests that the number of posts per day has limited explanatory power for the number of followers.

**Q 3A): 1000 light bulbs with a mean life of 120 days are installed in a new factory and their length of life is normally distributed with standard deviation of 20 days. If it is decided to replace all the bulbs together, what interval should be allowed between replacements if not more than 10% should expire before replacement?**

To determine the interval that should be allowed between replacements such that not more than 10% of the bulbs expire before replacement, we can use the concept of z-scores and the standard normal distribution.

Given information:

Mean life (μ) = 120 days

Standard deviation (σ) = 20 days

To find the interval, we need to find the z-score corresponding to the 10th percentile (0.10) of the standard normal distribution. The z-score represents the number of standard deviations away from the mean.

Using a standard normal distribution table or a statistical calculator, we can find the z-score that corresponds to a cumulative probability of 0.10. The z-score is approximately -1.28.

Now, we can calculate the interval as follows:

Interval = μ + (z-score \* σ)

Interval = 120 + (-1.28 \* 20)

Interval = 120 - 25.6

Interval ≈ 94.4

Therefore, the interval that should be allowed between replacements is approximately 94.4 days. This means that the bulbs should be replaced before 94.4 days to ensure that not more than 10% of them expire before replacement.

**Q 3B): calculate the average age of migrants for both the categories of gender and write your interpretation.**

Age Group | Male | Female

---------------------------------------

0-4 | 98,34,738 | 91,27,975

5-9 | 1,09,59,506 | 99,58,059

10-14 | 1,24,25,108 | 1,14,51,227

15-19 | 1,26,83,733 | 1,65,18,666

20-24 | 1,31,97,283 | 3,36,58,466

25-29 | 1,30,45,214 | 3,75,22,017

30-34 | 1,21,34,009 | 3,42,86,096

35-39 | 1,20,60,030 | 3,30,54,887

40-44 | 1,09,00,143 | 2,72,61,236

45-49 | 97,04,026 | 2,34,47,716

50-54 | 79,40,152 | 1,78,42,986

55-59 | 61,61,754 | 1,51,92,910

60-64 | 54,01,736 | 1,43,47,372

65-69 | 36,87,082 | 1,01,41,196

70-74 | 26,62,421 | 70,33,728

75-79 | 13,41,572 | 34,93,001

80-85 | 14,61,296 | 42,53,695

To calculate the average age for each gender:

For Males:

Total Male Migrants = Sum of the number of male migrants in each age group

Total Male Migrants = 98,34,738 + 1,09,59,506 + 1,24,25,108 + 1,26,83,733 + 1,31,97,283 + 1,30,45,214 + 1,21,34,009 + 1,20,60,030 + 1,09,00,143 + 97,04,026 + 79,40,152 + 61,61,754 + 54,01,736 + 36,87,082 + 26,62,421 + 13,41,572 + 14,61,296

Total Male Migrants = 16,58,76,367

Average Male Age = Total Male Migrants / Number of Age Groups

Average Male Age = 16,58,76,367 / 17

Average Male Age ≈ 97,58,604

For Females:

Total Female Migrants = Sum of the number of female migrants in each age group

Total Female Migrants = 91,27,975 + 99,58,059 + 1,14,51,227 + 1,65,18,666 + 3,36,58,466 + 3,75,22,017 + 3,42,86,096 + 3,30,54,887 + 2,72,61,236 + 2,34,47,716 + 1,78,42,

986 + 1,51,92,910 + 1,43,47,372 + 1,01,41,196 + 70,33,728 + 34,93,001 + 42,53,695

Total Female Migrants = 28,75,34,012

Average Female Age = Total Female Migrants / Number of Age Groups

Average Female Age = 28,75,34,012 / 17

Average Female Age ≈ 1,69,14,353

Interpretation:

The average age of male migrants is approximately 97,58,604, while the average age of female migrants is approximately 1,69,14,353. This suggests that, on average, female migrants tend to be older than male migrants in the given population. However, it's important to note that the age distribution within each gender category may vary across different age groups.