**Decision Science**

**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? (10 Marks)**

**Note: Draw the tree diagram for the above problem. Handwritten tree diagram is prohibited.**

**Ans 1.**

**Introduction:**

Bayes' theorem is a statistical concept that allows us to revise our previous beliefs based on new information. It's a useful tool used in many industries, including science, healthcare, engineering, and finance, among others. Bayes' theorem can be used in this paper to solve problems related to periodontal disease and anger.

**Problem statement:**

Gum problems can make you unhappy, comments about the problem. According to studies, 85 percent of people with a bad heart also suffer from periodontal disease, which causes pain in their gums. In short, the disease affects 29 percent of healthy people. How can people with gum disease not be happy? In some communities, the probability of a panic attack is only 10%.

**Solution:**

To solve this problem, we will use Bayes' theorem, which is : P(A|B) = P(B|A) \* P(A) / P(B)

Where:

P (A|B) is the probability of event A, given event B has occurred.

P (B|A) is the probability of event B given event A has occurred.

P (A) is the prior probability of event A

P (B) is the prior possibility of event B

In this case, given that a person has periodontal disease, our goal is to determine what is causing the depression. This can be demonstrated using the following expressions:

P (bad mood | Periodontal disease)

We know that the only way to overcome your anger is 10%. The explanation for the problem also says that the probability of contracting a disease in the body is 0.85 (85%) when in a bad mood, and 0.29 (29%) when in a good mood. This information can be represented using a tree diagram such as:

**Bad mood No bad mood**

**0.1 0.9**

**/ \ / \**

**Periodontal No periodontal Periodontal No periodontal**

**0.85 0.15 0.29 0.71**

The first level of the tree diagram represents the potential for pre-existing fear or lack of fear. In this case, you have a 10% chance of being mentally ill. That is, the probability of being negative is 0.1 and the probability of being negative is 0.9.

The second level of the tree diagram shows that periodontal disease can be caused by fear or negative emotions. The analysis shows that if you have a fever, the probability of contracting a disease in the body is 0.85. By comparison, the probability of periodontal disease in the absence of depression is 0.29.

In other words, the probability of no disease in the body when the mood is bad is 0.15, and when the mood is good, it is 0.71.

Using the tree diagram corresponding to probability, we can determine the probability of illness in the body and mind.

1. We can see from the tree diagram that the likelihood of result 1 is:

P (terrible temper and periodontal disease) = P (bad mood) \* P(Periodontal disease | bad mood)

= 0.1 \* 0.85

= 0.0.5

We additionally recognize that the possibility of having periodontal sickness is:

P (Periodontal sickness) = P (Periodontal disease | bad mood) \* P (bad mood) + P (Periodontal disease | No bad temper) \* P(No bad temper)

= 0.85 \* 0.1 + 0.29 \* 0.9

= 0.346

We can now use Bayes' theorem to calculate the probability that heart failure is caused by periodontal disease:

P (bad mood | Periodontal disorder) = P(Periodontal disease | bad mood) \* P(bad mood) / P(Periodontal disease)

= 0.85 \* 0.1 / 0.346

= 0.2457

Consequently, the chance of having a bad temper given periodontal disease is 0.2457 or about 24.57%.

As a result, we found that according to Bayes' theorem and tree diagram, the probability of stress in the body due to illness is about 24.57%. This finding suggests that there is a link between periodontal disease and negative emotions, with people with negative emotions developing periodontal disease more frequently than those without. The results of this analysis highlight the importance of maintaining health and seeking treatment for oral diseases, general health and mental health.

**Q 2: Using MS-EXCEL 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.**

**(10 Marks)**

|  |  |
| --- | --- |
| **No of followers** | **No of post per day** |
| **439** | **2** |
| **340** | **1** |
| **315** | **4** |
| **444** | **5** |
| **377** | **2** |
| **456** | **5** |
| **495** | **2** |
| **304** | **2** |
| **401** | **5** |
| **305** | **5** |
| **338** | **4** |
| **348** | **2** |
| **402** | **1** |
| **395** | **5** |

**Ans 2.**

To develop a regression model in MS Excel, data must first be entered into a spreadsheet. In this we know how many instagram followers per day and how many posts per day. The number of followers will be the dependent variable and the number of daily posts will be the independent variable.

**Here's how to create a regression version in Excel:**

Step 1: Enter the data in two columns, one for the variable (number of followers) and the other for the individual variable (number of posts per day).

Step 2: Click the Data Analysis button on the Statistics tab.

Step 3: After selecting "Regression" from the list of analysis tools, click the corresponding button.

Step 4: Enter the Input Field (number of cells containing data for individual variable) and Output (number of cells containing truth for difference variable) in the Regression dialog.

Step 5: Select the desired options for recovery verification.

In this example, we can select "Labels" as input and display the different labels, and show "Residuals" to calculate the remainder (the difference between the estimate and the actual).

Step 6: Click the appropriate button to perform an inverse measurement.

When the regression analysis is complete Excel will create a new worksheet containing the regression results. The output includes the regression equation, R-squared coefficient of determination, standard error, and t- and p-values ​​for the coefficients.

**We have the new sheet, as given below.**







The table below shows the results of Excel analysis of a simple horizontal line. The table provides information about the regression coefficients, the significance of the model, and the strength of the model. The number of followers (y) and the number of daily posts (x) are the variables created in this study.

**Interpretation of Excel Tables:**

**Regression statistics:**

**Multiple R:**

The correlation coefficient between these two variables shows that there is a weak relationship between the number of followers and the number of daily posts. The correlation between the variables is negative as indicated by the value 0.0466..

**R square:**

This is a probability that 0.2% of people can be best explained by daily announcements. A low score indicates that more information should be included in the model.

**Adjusted R rectangular:**

Given the number of variables in the example, the R-squared value will change. Since mileage is not good in this case, the model does not fully account for the data.

**Popular error:**

The standard deviation of the estimated residuals is available here. It expresses the difference between actual and estimated values.

**ANOVA:**

Information about the importance of the dependent variable is given in the ANOVA table. The F statistic determines whether the regression model is large. If the F statistic is too small, this version will not be useful in that case.

**Coefficients:**

Information about the regression coefficients is available in the Coefficients table. When the mean variance is zero, the value of the variance is represented by the intersection of 377.21. We can estimate the one day after coefficient as 1.74, which means 1 on average.

74 followers for each additional post daily. Although the p value for this coefficient is small, its deviation from zero is not significant.

In this question, the regression equation is: y = 377.21+ 1.74 x

**Conclusion:**

Regression analysis results show that the diversity of daily posts can be a strong indicator of reach of Instagram followers. Low R-squared, adjusted R-squared value, and low F-statistic all indicate that this version did not fit the data well.

This could be the result of many factors that the model needs to consider, such as time or the best ad. Further research and attention to other aspects are recommended to develop better and more accurate models.

**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? (5 Marks)**

**Note: You are not supposed to use EXCEL or any other software to write this answer.**

**Ans 3a.**

To solve the given problem, we need to determine how often 1000 bulbs need to be replaced at the new factory setting. We also learned that although the average lamp life is 120 days with a standard deviation of 20 days, we cannot leave more than 10% of the bulbs unchanged. Let

X be the life of daylight; under normal distribution, the mean of X is 120 days and its standard deviation is 20 days.

We want to find the value x such that

P (X <= x) = 0.10.

We can find this value using the standard regular distribution table or a calculator.

First, we standardize X using the formula:

Z = (X - μ) / σ

Where μ is the mean and σ is the standard deviation.

Z = (x - 120) / 20

Using the standard regular distribution table, we find that the z-score corresponding to P (Z <= z) = 0.10 is -1.28.

Substituting this value into the formula, we have

-1.28 = (x - 120) / 20

Solving for x, we get

x = -1.28 \* 20 + 120 = 94.4

Therefore, the replacement period should be 94.4 days or less so that the lamp failure does not exceed 10% before replacement.

Bulbs should be changed every 94.4 days to ensure that no more than 10% of the bulbs need to be replaced due to the expiration date. The efficiency of the factory depends on this important measure, because a malfunctioning light can disrupt production or cause safety risks.

It is worth noting that the normal distribution is a useful tool in data analysis and is often used in statistical research. It has a number of important features such as symmetry, clear instructions and distinct patterns. Because of these features, it is useful to read many statistical information from physical measurements to financial information.

The normal distribution is a continuous distribution that is widely used in statistical research because it has many important properties. Symmetrical bell curve, a height of recommendation, positive expression, and standard deviation are some important features of the distribution.

It is worth noting that the normal distribution is a useful tool in data analysis and is often used in statistical research. It has a number of important features such as symmetry, clear instructions and distinct patterns. Because of these features, it is useful to read many statistical information from physical measurements to financial information.

The normal distribution is a continuous distribution that is widely used in statistical research because it has many important properties. Symmetrical bell curve, a height of recommendation, positive expression, and standard deviation are some important features of the distribution.

**Q 3B): calculate the average age of migrants for both the categories of gender and write your interpretation. (5 Marks)**

|  |  |  |
| --- | --- | --- |
| **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** |

**Note: You are not supposed to use EXCEL or any other software to write this answer**

**Ans 3b.**

****

****

To get the average age of immigrants in both sexes, we need to divide the total product of the average of each age group and the number of immigrants in that group by the total number of immigrants in that group.

Among the male population, the total population is 1,455,99803 and the total f\*x (the product of the average of each age group and the population in this group) is 4,613,678,689. Therefore, the following formula can be used to determine the average age of the male floating population:

Average age of male migrants = (Sum of f\*x) / (Total number of migrants)

= 4,613,678,689 / 1, 455, 99803

= 31.687

Hence, the average age of male migrants is approximately 31.69 years.

For females, the total number of migrants is 30 859 1233, and the sum of f\*x is 11,316,123,244. Therefore, the average age of female migrants can be calculated as follows:

Average age of female migrants = (Sum of f\*x) / (Total number of migrants)

= 11,316,123,244 / 30, 859, 1233

= 36.670

Hence, the average age of female migrants is approximately 36.67 years.

Interpretation:

According to the above calculations, the average age of the male swimming population is 31,687, and the female swimming population is around 36,670. Therefore, female immigrants are only slightly older than male immigrants. Different situations may include differences in average age.

**For example,** perhaps because women's migration for family reunification or to register their spouses often happens after men migrate for work-related reasons. Another reason why the legal immigration age is better for women is because women like men.

Knowing the age of immigrants of all genders is important for policy makers because this information helps them formulate policies that address the needs of different age groups. If migrant women are of legal age, policymakers may want to focus on providing services that meet the needs of older people, such as health facilities, security services, and shelters for the elderly. Consider also the low average age of male immigrants.

In this context, the authorities can listen to the openings of jobs such as internship, training and internship that will appeal to young people.

**In conclusion,** The average age for male immigrants is about 31.69, and for female immigrants is about 36.67. Gender disparities in migration motivation and life expectancy are just two of the many causes of the age gap. To promote the effective integration of immigrants into the host society and improve their overall well-being, it is important to know the average age of immigrants by sex.

Based on the availability of this information, the ability to develop packages and policies that address the needs of different age groups, such as providing health facilities for older women and job opportunities for young male immigrants.