**Final Project Report**

**Introduction to the Business Decision Process**

**Name: Abhignu Reddy Alety Student ID: 11443224**

**About Dataset:**

The dataset is regarding 11 clinical features for predicting heart disease events.

<https://www.kaggle.com/fedesoriano/heart-failure-prediction>

**Data set Description:**

The dataset is based on 11 clinical features for predicting heart disease events. It has 918 rows and 12 columns

The following are the attributes of the data set which will be using in the project:

Date

Age: age of the patient

Sex: Sex of the patient

Chest pain type: [Typical angina, ATA, NAP]

Resting BP: resting blood pressure

Cholesterol: Serum cholesterol

Fasting BS: Fasting blood sugar

Resting ECG: resting electrocardiogram

Max HR: maximum heart rate achieved

Exercise Angina: exercise-included angina

Old peak: oldpeak= ST [Numeric value measured in depression]

ST\_Slope: the slope of the peak exercise ST segment

Heart disease: output class

**Research questions:**

Using the above dataset to perform the following analysis:

Q1. Devices manufactured by TSI Electronics to detect heart attacks have life spans that have a normal distribution with a standard deviation of 3000 hours and a mean life span of 15,000 hours. If a monitor is selected at random, find the probability that the life span of the monitor will be more than 13,000 hours. Round your answer to four decimal places.

Solution:

Given Mean = 15000

Standard Deviation = 3000

Standard normal random variable,

Graphical user interface, application

Description automatically generated with medium confidence

Z = (15000 – 13000)/3000 = - 0.67

The Probability that the life span of the monitor will be more than 13000 hours (p more than z) is 0.7475.

Q2. The mean of people with heart failures due to Cholesterol is 198.7996 with a standard deviation of 3.610215. (8.3 - 11)

If a sample of 69 people were randomly selected, what is the probability that the sample mean would differ from the true mean by less than 0.5. Round your answer to four decimal places.

Solution:

Since the true value of the population mean is 198.7996, the value of X bar should not fall between 198.2996 and 199.2996 in order for the error of estimation to be less than 40.

Since the distribution of x‾ is approximately normally distributed for large samples, the distribution of x‾ will be approximately normal with

μx‾ = μ = 198.7996

σ= Standard Deviation of the population = 3.610215

n = Sample Size = 69

A picture containing clock, watch

Description automatically generated0.43461

Q3. The Denton County Public Health Services claims that the Heart attack due to Cholesterol has 197 rating. An independent testing firm has been contracted to test the rating for this Cholesterol rating since it is believed that Denton Health Services has an incorrect Cholesterol rating. After testing 918 patients, they found a mean Cholesterol rating of 198.80. Assume the standard deviation is known to be 109.38. A level of significance of 0.05 will be used. Find the value of test statistic. Round your answer to 2 decimal places.

Solution: Given, mu = 197

X bar = 198.80

n= 918

σ = 109.38

Z- Test Statistic:

Graphical user interface, text, application, Word

Description automatically generated

Z = (198.80 – 197) / (109.38)/Sqrt Root (918)) = 0.50.

The Value of Z Test Statistic: 0.50

Q4. A Hospital management polled a sample of 725 Male patients and 192 female patients. Each of the patients’ records were checked whether or not the cholesterol leads to heart disease. As per the insurance medical records, 525 male students and 82 female patients it mentions that the cholesterol leads to heart disease. Give a 95% confidence interval for the difference between the proportions of male and female candidates where the cholesterol leads to heart disease.

Construct the 95% confidence level. Round your answers to three decimal players.

Solution:

In this solution, we are assuming that the male undergraduate students are "Population 1" and the female undergraduate students are "Population 2". However, the alternative representation of the two populations is also valid.

We substitute the values and calculate the sample proportions, rounding the values to three decimal places.

A whiteboard with writing on it

Description automatically generated with low confidence

We can substitute these values into the formula to construct the confidence interval for the difference between two proportions, rounding the values to three decimal places.

Substitute the values in the below equation:

A picture containing chart

Description automatically generated

Lower Limit: 0.221

Upper Limit: 0.373

Q5. The following data give the number of 5 patients spent studying and their corresponding hours hospitalized on their admission.

Hours Hospitalized 10 21 31 41 51 61

Resting BP 160 120 110 140 155 135

Calculate the sum of squared errors (SSE), the estimated variance of errors , the

estimated variance of slope , & construct the 98% confidence interval for the slope.

Round your answers to three decimal places. Use the values b0=77.7894 and b1=3.9211

for the calculations. Round your answer to three decimal places.

From the summation table, we have

Table

Description automatically generated

The calculation of sum of squared errors (SSE) can be tabulated as follows:

∑(y−yˆ)^2 = 69341.648

A picture containing text

Description automatically generated

= 69341.648/(6-2)

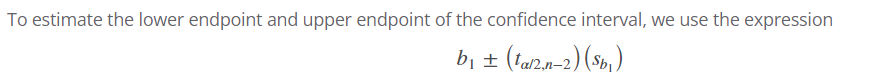
= 17335.412

Graphical user interface, text, application

Description automatically generated with medium confidence

= 17335.41/(46225 – 7704.166667) = 9.626

Construct the 98% confidence interval for the slope. Round your answers to three decimal places.



Df = n-2 = 6-2 = 4

Sb1 = 3.10258

T alpha/2 = 3.747

E = t alpha/2 \* Sb1 = 11.62537

Lower C.I. : -7.704

Upper C.I. : 15.546