**Final Project**

**Data Description and Research Hypothesis on**

**Real-time Data**

**Name- Janmejoy Kar**

**EUID- 11527178**

**Hepatitis C Prediction**

In our midterm project, we framed 4 questions from a real-time dataset to apply our modules understanding.

Questions based on the dataset:

1. What is the probability of ALB being more than 30mg/l? (Module 1)
2. The mean of ALP is more than the mean of protein. (α=.075) (Module 2)
3. The mean of CHOL is more than the mean of ALT. (α=.05) (Module 2)
4. Find that our test statistics are true for the hypothesis of CHOL and ALP. (Module 3)
5. Run the logistics regression analysis for the dataset. And get the Confidence interval.

**Answer 1: -** From the python package attached to the file, we have concluded that there are a total of 24 blood donors who have Albumin content lower than 30mg/l out of 615 donors, which brings down our probability z = .

Z = (30 – 41.62)/ 14.9 = -0.78

P(z>30) = 1 – p(z<30) = 1 - 0.2177 = 0.7823

Here, our probability value indicates that there is a high percentage of blood donors who have high albumin levels which can be harmful to the body as well as the liver.

**Answer 2: -** From the module, we have learned to calculate the difference between two population means.

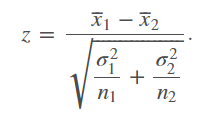
H0 = Mean of **Alkaline Phosphatase** is more than the Mean of **Protein**

Hα = Mean of **Protein** is more than the Mean of **Alkaline Phosphatase.**

H0 = µ1 ≥ µ2

Hα = µ1 < µ2

Here,



X1 = 68.28 1 = 26.03 n1 = 597

X2 = 72.04 2 = 5.4 n2 = 615

Z = = (-3.76)/ (1.087) = -3.459

As our critical value (α) was 0.075 which has a value of 1.440, which means our Null hypothesis has been rejected as our test statistics value is more than the critical value,

Z < -1.440

From this, we can note that our data has more healthy blood donors than a patient who has hepatitis. And donors have a healthy liver. If ALP were more, donors would lose weight rapidly because ALP is responsible for the breakdown of protein in our bodies.

**Answer 3: -** From the module, we have learned to calculate the difference between two population means.

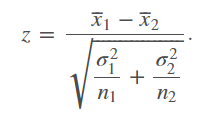
H0 = Mean of **Cholesterol** is more than the Mean of **Alanine Transaminase**

Hα = Mean of **Alanine Transaminase** is more than the Mean of **Cholesterol.**

H0 = µ1 ≥ µ2

Hα = µ1 < µ2

Here,



X1 = 5.37 1 = 1.13 n1 = 605

X2 = 28.45 2 = 25.47 n2 = 614

Z = = (-23.08)/ (1.029) = -22.43

As our critical value (α) was 0.075 which has a value of 1.645, which means our Null hypothesis has been rejected as our test statistics value is more than the critical value,

Z < -1.645

From this, we can note that our data has more healthy blood donors than a patient who has hepatitis. If Cholesterol would have more then there is a high risk of liver and heart failure in near future for the donors.

**Answer 4: -** From the above answers of the test statistics, we can conclude that our hypothesis for the CHOL and ALP has been rejected as the value is outside the boundary of the Not reject region. So, this indicates that our data consist of more healthy donors than that who are suffering from serious diseases.

**Answer 5: -** Refer to the R package for the CI and prediction of the data.