Name: Hoyoung Kim

UVU id: 10829866

1. (40 points) A batch of 𝑛 resistors have an average of 101.5 Ohms. Assuming a population sample variance of 25. We are interested to see whether the population mean is 100 Ohms at a level of significant 0.05 (meaning that our criterion for drawing the conclusion from the p-value is based on 0.05). The hypothesis testing is as follows.

a) 𝐻\_0: 𝜇 = 100 Versus 𝐻\_1: 𝜇 ≠ 100.

For 𝑛 = 100, compute the p-value and draw a conclusion for our test.

t = (101.5 - 100) / (5 / sqrt(100)) = 3

according to the z- table(since sample size is large), the p-value turns out to be 0.9987.

then we find the tail which is 1-0.9987= 0.0013.

since the nature of the question requires 2-tailed test, we multiply the 0.0013 by 2 = 0.0026

since 0.0026 is less then the significance level of 0.05, we can safely reject the hypothesis.

b) 𝐻\_0: 𝜇 ≥ 100 Versus 𝐻\_1: 𝜇 < 100.

For 𝑛 = 100, compute the p-value and draw a conclusion for our test.

t = (101.5 - 100) / (5 / sqrt(100)) = 3

according to the z- table(since sample size is large), the p-value turns out to be 0.9987.

then we find the tail which is 1-0.9987= 0.0013.

since this is a one tailed test, we can just compare this to the significance level 0.05

and since this is less than 0.05 we can conclude that population mean is greater than or equal to 100 ohms.

c) 𝐻\_0: 𝜇 ≥ 100 Versus 𝐻\_1: 𝜇 < 100.

For 𝑛 = 10, compute the p-value and draw a conclusion for our test.

t = (101.5 - 100) / (5 / sqrt(10)) = 0.9486

we also find that the df = 9 since n-1

to get the p- value from t-value we find that this is between 0.25 and 0.1 in the t-table and since they are both greater then 0.05 (0.75 from 1-0.25 and 0.9 from 1-0.10) we can conclude that the population mean in greatern then or ruqal to 100 ohms

2. (30 points) A clinical trial (N # 50 patients) has been performed in which the volume of distribution of a new anti-diabetes drug has been calculated as 9.5 ± 5.0 L. Calculate the 65% and 95% confidence limits of the mean value (assuming that the data originated from a normal distribution).

Mean 9.5L

SD = 5L

n = 50

65%

α = (1 + 0.65)/2 = 0.825 (Divide α by 2 since it's a two-tailed test)

z-value according to table is between 0.93 or 0.94

Confidence Interval = 9.5 ± (0.93 \* (5.0 / sqrt(50))) ≈ (8.842, 10.157) L or 9.5 ± (0.94 \* (5.0 / sqrt(50))) ≈ (8.8353, 10.164) L

95%

In the text book 95% is given as



So per text book we can say

Confidence Interval = 9.5 ± (1.96 \* (5.0 / sqrt(50))) ≈ (8.114, 10.885) L

3. (30 points) We are given the following set of input-output (x,y) data.

𝑥: later Payments 𝑦: Credit Score

0.43 772

0.67 735

0.40 774

0.45 769

0.80 723

Suppose that we want to model the above set of data with a linear model 𝑦 = 𝑎x + 𝑏. Since our model may be perfect and there might be some noise in the measurements 𝑦 , we assume 𝑦 = ax + 𝑏 + 𝑒, where 𝑒 is the error in our modeling.

a) Write down the matrix structure for this problem using the given model and the input-out

data.

[772, 735, 774, 769, 723] = [[1, 0.43], [1, 0.67], [1, 0.40], [1, 0.45], [1, 0.80]] [b, a] + [e\_1, e\_2, e\_3, e\_4]

Y = x aug b + e

b) Write down the structure of the solution for the coefficients 𝑎 and 𝑏. Then, use MATLAB

or Python to solve for the coefficients of the model, i.e., 𝑎 and 𝑏 using least-squares error

method. Include your code and the obtained results.

b = (Xaug^t xaug)^-1 xaug^t y or

b = ([[0.43, 0.67, 0.40, 0.45, 0.8],[1, 1, 1, 1, 1]] [[1, 0.43], [1, 0.67], [1, 0.40], [1, 0.45], [1, 0.80]])^-1\* [[0.43, 0.67, 0.40, 0.45, 0.8],[1, 1, 1, 1, 1]] \* [772, 735, 774, 769, 723]

    x = np.array([0.43, 0.67, 0.40, 0.45, 0.80])

    y = np.array([772, 735, 774, 769, 723])

    an = np.vstack([x, np.ones(len(x))]).T

    AtA = np.dot(an.T, an)

    AtY = np.dot(an.T, y)

    AtA\_inv = np.linalg.inv(AtA)

    coefficients = np.dot(AtA\_inv, AtY)

    print("a:", coefficients[0])

    print("b:", coefficients[1])

a: [-134.81421648]

b: [828.74781906]

c) Fill out the last column of the following table using the model and obtained coefficients.

x: Later Payment 𝑦: Credit Score Estimated Credit Score using ŷ = ax + 𝑏

𝑥\_1 = 0.43 𝑦\_1 = 772 ŷ1 = -134.81 \* 0.43 + 828.75 = 770.78

𝑥\_2 = 0.67 𝑦\_2 = 735 ŷ2 = -134.81 \* 0.67 + 828.75 = 738.42

𝑥\_3 = 0.40 𝑦\_3 = 774 ŷ3 = -134.81 \* 0.40 + 828.75 = 774.82

𝑥\_4 = 0.45 𝑦\_4 = 769 ŷ4 = -134.81 \* 0.45 + 828.75 = 768.0

𝑥\_5 = 0.80 𝑦\_5 = 723 ŷ5 = -134.81 \* 0.80 + 828.75 = 720.90

d) Find the normalized squared-error defined as follows

𝑒 = sqrt(1/5∑ (𝑦\_𝑛 – ŷ\_𝑛)^2) from 𝑛=1 to 5, where 𝑦\_𝑛 is the 𝑛th measurement and ŷ is the output of the

model evaluated at the 𝑛th input data.

This basically boils down to (772 - 770.78)^2 + (735 - 738.42)^2 + (774 - 774.82)^2 + (769 - 768.00)^2 + (723 - 720.90)^2 as n1 to 5 have been calculated in c. so the sum of the e = 19.2672