Cmpe 343 Assignment2

Alp Tuna

December 2021

1 Question 1

1.1 a)

From the data given in the benan.npy file, I calculated the mean as 54.9993. I attached the code for that one into my submission.
$\mu = 54.9993$
$n = 32$
$\sigma = 0.1$
$z_{\alpha/2} = 1.96$
Since we know the variance of the data, we can use central limit theorem. Putting the values into equation, we get the following confidence interval:
$54.9993 - 1.96 \cdot \frac{0.1}{\sqrt{32}} < \mu < 54.9993 + 1.96 \cdot \frac{0.1}{\sqrt{32}}$
$54.9647 < \mu < 55.0339 \quad (95\% \text{ confidence interval})$

1.2 b)

$Null\ hypothesis = H_0 = \mu = 55$
$Alternative\ hypothesis = H_0 = \mu \neq 55$

Since the mean value of the data which is 54.9993 is inside the confidence interval we found in part A, we fail to reject null hypothesis.

1.3 c)

$Z = \frac{55-54.993}{\frac{0.1}{\sqrt{32}}}$
$Z = 0.04$
$P(Z > 0.04) = 0.48 \quad (Probability\ that\ mean\ is\ greater\ than\ 55)$
$P(|Z| > 0.04) = 0.96 \quad (Probability\ that\ mean\ is\ further\ than\ 55\ than\ 55\ by\ at\ least\ 0.0007)$
We can use both of these values as our p-value depending on our hypothesis, since we are dealing with two-tailed scenario we should choose the below one. Hence, our p-value is 0.96. Since, the p-value is very high, we can safely say that we fail to reject the null hypothesis.

1.4 d)

In part a) we found 95% confidence interval.
In part b) We used the traditional method of fixed α method. Since we found the confidence interval already in part a) this was quite easy.
Part c) was different than the traditional method. We calculated the p-value without fixing the level of significance(i.e α) and derived a conclusion from that value. Hence, the difference in that method is we are more flexible, don't fix the alpha value and evaluate the result accordingly.

1.5 e)

In this case, we can use t-distribution. Degree of freedom is equal to 32-1=31 By looking at t-table we can see that $t_{0.025} = 2.042$. for degree of freedom of 30. I will just use that values for 30 and 31 degree of freedom are approximately equal and we don't have the value for 31 in the table. Also, I calculated the S from the given data file(I shared the python code) and it is equal to 0.1025. Now putting the values into equation, we get:
$54.9993 - 2.042 \cdot \frac{0.1025}{\sqrt{32}} < \mu < 54.9993 + 2.042 \cdot \frac{0.1025}{\sqrt{32}}$
$54.9623 < \mu < 55.0363 \qquad$ (95% confidence interval)

In general, when we know the variance prior to experiment we use central limit theorem and normal probability table. When we don't use pre-known variance and use the one we calculated from the data we use t-distribution and t-table. We we use the t-distrubiton we have less information prior to experiment and correspondingly we have less precision. Also, corresponding t-values are higher than normal values which confirms that. So, it is expected to get a wider interval in part e.

2 Question 2

Since we have a uniform distribution we can calculate probilities easily. Total length of the interval is β and the probability of getting a number from a interval of length L is $\frac{L}{\beta}$ For type-1 error we assume that $\beta = 3$ and calculate the probability that our number is between either 0-0.1 or 2.9-3 Now, using the formula we get:
$\frac{0.1}{3} + \frac{0.1}{3} = 1/15 \qquad$ (Probability of type-1 error)
To calculate the type-2 error we assume $\beta = 3.5$ and calculate the probability of getting a number from 0.1-2.9 By using the formula, we get:
$\frac{2.9-0.1}{3.5} = \frac{2.8}{3.5} = 0.8 \qquad$ (Probability of type-2 error)

3 Question 3

The code for question 3 is given in *q3_1.py* file. I explained the logic of the code in the comments.
If you want to see the output of my uniform distribution, you can uncomment 64th or 65th line. Also you can change the seed values in 28th and 29th line.
If you want to see the histogram of standard normal produced by my uniform distributions, uncomment either 66th or 67th line.
If you want to see the histogram of standard normal produced by the prebuilt normal distributions, uncomment 68th or 69th line.

4 Question 4

For part A, I randomly selected 100000 numbers from uniform distribution between 0 and pi and calculated the weighted average of cosinus of those numbers. Code is relatively simple. You can check the result by running the code. Answer is approximately 0.
Code is in *q4_1.py* file.

For part B, I randomly selectede 100000 numbers from standard normal distrubiton and calculated the weight of the numbers between 0-1. This gives an approximate value of the probability. You can check the result by running the code. Answer is approximately 0.34.
Code is in *q4_2.py* file.

For part C, I calculated the probabilities assumed by chebyshev's inequality and wrote them in the first 4 line of the code as a comment Then I sampled 100000 numbers from standard normal distribution and calculated their weights and probability given by the population. You can check the result by running the code. Answers are approximately: 0.045, 0, 0
Code is in *q4_3.py* file.