**IBEHS 4C03: STATISTICAL METHODS IN BIOMEDICAL ENGINEERING**

**ASSIGNMENT #2**

**Topics:**

Discrete Probability Distributions

Normal Distributions

One sample test for means

P-values and confidence intervals

One-and two-sided alternatives for hypothesis testing

**Total marks: 100**

**Submission Instructions:** Electronic submission to the IBEHS 4C03 Avenue to Learn Assignment 1 folder.

**Due:**

**Assignment Submission Format:**

File Names: StudentLastNameFirstNameAssignment#.doc/pdf/py//ipynb etc.

Both your name and student numbers should appear at the top of the document. If separate documents are submitted, then you should submit fully supported answers to the questions in a single document including the plots you made in python, and refer to coding. Regardless, the python file used to generate the plots and any provided answers should also be submitted with any assignment, along with a pdf or word document conversion of the python file.

**Problem 1: [20 points/100 points] Discrete Probability Distributions**

1. Wastewater samples contain a high level of microplastic particles in 10% of tests. High levels were defined as >0.10g/L. Let X = number of samples high in micro-plastic particles in the next 18 samples tested. Assume samples are independent of each other with regards to particulate concentration.
2. Discuss if the Binomial distribution is a reasonable model to answer questions about sample particulate concentrations and why. Make sure to Identify the random experiment and the random variable and state any assumptions that you make.
3. Specify the parameters of the binomial distribution and discuss the shape of the distribution in this example.
4. What is the probability that all the samples have high levels of particulates?
5. What is the probability that 2 of the samples have high levels of particulates?
6. What is the probability that at least 4 of the samples have high levels of particulates?
7. Plot the PMF of the probability distribution.
8. There are crack-initiating defects along a steel pipe. Your company wants to model the number of defects in a length of pipe knowing that there is a mean of 2.3 defects per meter. Assume that the number of defects in each interval is independent of other intervals.
9. Discuss if the Poisson distribution is a reasonable model to answer questions about defect counts along the pipe and why? Make sure to Identify the random experiment and the random variable and state any assumptions that you make.
10. Specify the parameters of the Poisson distribution and discuss the shape of the distribution in this example.
11. What is the probability that there are 2 defects in 1 meter of pipe?
12. What is the probability that there are 10 defects in 5 meters of pipe?
13. What is the probability that there are at least 1 defect in 2 meters of pipe?
14. Plot the PMF of the probability distribution.

**Problem 2. [20 points/100 points] Normal Distributions**

A shape memory alloy (SMA) wire is used to reconnect torn ligaments to bone through tiny holes drilled through bone. Direct current electrical stimulation through the wires is used to stimulate tendon regeneration and improve outcomes. Wires have a current measurement of 1-microamperes and a variance of 4 microamperes2, and current measurements can be assumed to be independent.

1. Discuss if the Normal distribution is a reasonable model to answer questions about wire current? Make sure to Identify the random experiment and the random variable and state any assumptions that you make.
2. Specify the parameters of the Normal distribution and discuss the shape of the distribution in this example.
3. What is the probability that a randomly selected measurement will exceed 13 microamperes? (i.e. Find P(X>13). You are finding the probability of a certain outcome from a normal population that you have specified.
4. What is the probability that a current measurement is between 9 and 11 microamperes?
5. Determine the value for which the probability that a current measurement is at or below that value is 0.98.
6. Plot the PDF of the normal probability and show the areas under the curve for Problem 2.4 above (i.e.What is the probability that a current measurement is between 9 and 11 microamperes?)

**Problem 3: [40 points/100 points] Inferential Statistics and Univariate Data**

We are interested in the mean burning rate of solid waste at a waste management site through a mass burn process to recover energy from waste combustion. To minimize air emission toxicity, you want to keep the mean burning rate at 50 cm/s with a known standard deviation of 2 cm/s.

1. Assume that burning rate is known to be normally distributed. You take 25 random samples of solid waste and find a mean burning rate. What is the standard error of the mean?
2. At a type I error probability of 0.05, what are the boundary or critical values for the sample mean at which you would make a type I error? Discuss what this means in the context of the example.
3. You find a sample mean of 51.3 cm/s for the 25 samples taken. At a type I error probability of 0.05, follow the steps for statistically testing if the sample mean is different from the specification of 50 cm/s. Make sure to state your conclusion, state the p-value for observing your sample mean and interpret this p-value, and discuss the practical engineering interpretation of your findings.
4. Now assume that you do not know that the burning rate is normally distributed, or the true standard deviation of the population. You take 100 random samples of solid waste and find a mean burning rate of 51.3 cm/s and a sample standard deviation of 2 cm/s. Have the p-value or hypothesis test conclusion changed? Make sure to defend the assumptions of your statistical test of choice.

**Problem 4: [40 points/100 points] Inferential Statistics and Univariate Data**

A novel cortical bone implant material is being developed that is strong but with a Young’s Modulus in transverse testing that is like that of human cortical bone. You want to show that the new material has a elasticity higher than that of 8.2 GPa, and therefore within the specifications of elasticity for your application. Testing of the elasticity of n=20 samples was taken:

**Elasticity of material**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Observation** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| **Sample** | 0.8411 | 0.8580 | 0.8042 | 0.8191 | 0.8532 | 0.8730 | 0.8182 | 0.8483 | 0.8282 | 0.8125 | 0.8276 | 0.8359 | 0.8750 | 0.7983 | 0.8660 |

1. Explore the data through descriptive statistics and visualization and comment if you think the material has met the targeted elasticity. Justify your answer.
2. Show the normal probability plot (the q-q plot) and conclude if the samples are normally distributed.
3. Can you support the claim that the mean elasticity of the new material is greater than of 8.2 GPa, the specification that you are aiming for to mimic the minimal elasticity of cortical bone in transverse measurement? Justify the statistical hypothesis test that you use to answer this question. Show the steps for the hypothesis test and discuss the conclusion.
4. Find the 95%CI of the mean elasticity (two-sided). Provide an interpretation of the intervals.
5. Find the 99%CI of the mean elasticity and provide an interpretation of this interval (two-sided). Why and how is it different than that for the 95%CI.
6. Discuss the relationship between the test of hypothesis and the confidence intervals you have calculated.

**Problem 5: [20 points/100 points]: One-Sided and Two-Sided Tests**

A company manufactures a novel laminated material, and claims it has excellent wear resistance with a mean wear resistance time of 850 hours (testing in a machine measuring wear to failure in hours).

A consumer group thinks that the manufacturer of the material X is overestimating the wear resistance.

An experiment was performed to measure the abrasive wear resistance time of a laminated material. Fifteen pieces of material were tested by exposing each piece to a machine measuring wear resistance time. The time to wear failure (in hours) was observed:

**Depth of wear of material**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Observation** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| **Material** | 768 | 974 | 894 | 859 | 778 | 831 | 867 | 851 | 844 | 893 | 796 | 821 | 671 | 795 | 810 |

1. Can they conclude at the 0.05 level of significance that the time to wear failure of the material is less than the 850 hours that the manufactures claim? Note that the hypothesis here is specifying one direction of difference, so the test will be using a one-tailed or one-sided test. Run the appropriate analysis and discuss your conclusion, supported by a one-side 95%CI.
2. Now run the two-sided test that would allow them to test if the wear failure of the material is different than the 850 hours that the manufactures claim. Run the appropriate analysis and discuss your conclusion, supported by a two-sided 95% CI. What is the difference from the one-sided test?

**END**