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

**ASSIGNMENT #2**

**Topics:**

Normal Distributions

One sample test for means

P-values and confidence intervals

Independent t-tests with and without equal variances

Paired t-tests

Nonparametric tests of location

Chi-Squared Test

**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] 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 mean current measurement of 10-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.

**Problem 2: [20 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 an elasticity different than that of 8.2 GPa, and therefore could be within the specifications of elasticity for your application. Testing of the elasticity of n=15 samples was taken:

**Elasticity of material**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Observation** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| **Sample** | 8.411 | 8.580 | 8.042 | 8.191 | 8.532 | 8.730 | 8.182 | 8.483 | 8.282 | 8.125 | 8.276 | 8.359 | 8.750 | 7.983 | 8.660 |

1. Can you support the claim that the mean elasticity of the new material is different than 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.
2. Find the 95%CI of the mean elasticity (two-sided). Provide an interpretation of the interval.
3. 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.
4. Discuss the relationship between the test of hypothesis you performed and the confidence interval.

**Problem 3: [30 points/100 points] Independent T-tests**

Tumor removal completeness during surgery depends on the surgeon’s ability to differentiate tumor from normal tissue using subjective criteria that are not easily quantifiable. One study describes the use of activatable cell-penetrating peptides (ACPPs) that fluorescently labels tumor cells to aid in the precision of tumor identification. The table below shows the fluorescence after two hours for tumor and normal tissue for two separate groups of 12 mice, with higher fluorescence allowing better tumor identification. You want to test if the fluorescence after two hours is different between the tumor and normal tissue between the two, independent groups.

**Fluorescence After 2 Hours**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Observation** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| **Tumor** | 3037 | 3948 | 3967 | 3164 | 3292 | 2696 | 3975 | 5284 | 2775 | 4083 | 4279 | 3935 |
| **Normal Tissue** | 2805 | 1228 | 2124 | 2272 | 2118 | 1786 | 3847 | 2912 | 3460 | 3029 | 3134 | 4097 |

1. Create a boxplot to graphically show what you are testing and describe the data.
2. What assumptions do you have to make about the data to use a two-sample independent t-test? Discuss the assumptions in general and in the context of this question, and then validate the assumptions. Show the qq-plot of the samples along with any discussion.
3. Run the hypothesis test. Specify the steps of the hypothesis test and show the work for the test (by hand or annotated code). Also run the appropriate test in Python. What do you conclude?
4. Construct a 95% confidence interval for the difference between the means for tumor and normal tissue fluorescence. Discuss this finding and how this adds to your understanding of the analysis.
5. What is the difference in means that you are testing? Find the power of the test for detecting this difference in means? Find the power to detect the difference in means of 800. Feel free to use the normal distribution for the calculations and show your calculations.
6. What sample size would be necessary to detect an actual difference in means of 800 with a power of at least 0.9? Feel free to use the normal distribution for the calculation and show your calculations.

**Problem 4: [15 points/100 points] Parametric and Non-parametric Tests of Location**

A group of biomedical researchers has developed an “image-guided surgery” technology for neurosurgery. Image-guided surgery involves aligning preoperative images with the physical space of the operating theater and visualizing them to help surgeons perform operations more safely and more efficiently. The researchers would like to test if their technology results in shorter recovery times, and to this end, they have gathered data from 21 patients, with recovery time being the subsequent hospital stay in days. Eight of the patients underwent surgery using the new technology (Group 1), and the other 14 had standard conventional surgery (Group 2).

**Recovery time (days) by treatment group.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subject** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| **Image-guided surgery** | 10.7 | 19.1 | 10.2 | 10.9 | 15.8 | 12.3 | 10.2 | 17.1 |  |  |  |  |  |  |
| **Conventional surgery** | 16.4 | 17.2 | 16.8 | 17.1 | 16.9 | 17.0 | 16.2 | 15.0 | 17.9 | 20.2 | 17.2 | 17.3 | 16.0 | 19.1 |

Use this study and associated data to answer the following questions:

1. What type of statistical test do we have here and why? (i.e. one-sample, two-sample, paired/unpaired?)
2. Before you report your conclusions, you become concerned about the distribution of recovery times and the appropriate use of the t-test. Perform an appropriate non-parametric test of location to test if recovery time was significantly different for patients having surgery with the new technology. Specify the steps of the hypothesis test and show the work for the test (by hand or annotated code).
3. Run the appropriate test in Python. What do you conclude?

**Problem 5: [15 points/100 points]: Paired and Unpaired Tests of Location**

A hardness testing machine presses a rod with a pointed tip into a metal specimen with a known force.

By measuring the depth of the depression caused by the tip, the hardness of the specimen is

determined. Two different tips are available for this machine, and you want to test if the two tips produce different mean hardness measurements (using the Mohs scale of mineral hardness). You take 9 samples and will use each sample to take a harness measure with each of the two tips.

**Metal specimen hardness testing by tip.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Metal Specimen** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| **Tip 1 hardness** | 1.19 | 1.15 | 1.32 | 1.34 | 1.20 | 1.40 | 1.37 | 1.54 | 1.56 |
| **Tip 2 hardness** | 1.06 | 0.99 | 1.06 | 1.06 | 1.07 | 1.18 | 1.04 | 1.09 | 1.05 |

1. What type of statistical test do we have here and why? (i.e. one-sample, two-sample, paired/unpaired?)
2. Run the appropriate hypothesis test to see if the two tips produce mean difference hardness measurements. Check assumptions, specify the steps of the hypothesis test, and show the work for the test (by hand or annotated code).
3. Run the appropriate test in Python. What do you conclude?

**Problem 6: [BONUS 5 points] Contingency Tables for Chi-Square Test**

A gene expression profile analysis is done where you want to find genes that are differentially expressed between tumor and normal cells from an individual. Out of 20,000 genes profiled, 1000 were differentially expressed. In the p53 pathway, there are 100 genes, 10 of which were among those genes that were differentially expressed.

1. Test the hypothesis (with a Chi-Square Test of independence and α = 0.05) that the p53 pathway is involved with gene expression in tumors. Calculate this test by hand and show your work. Also show this analysis from Python.

**Contingency table for 5.3: Comparing gene expression between 2 gene pathways**

|  |  |  |  |
| --- | --- | --- | --- |
| **Different gene pathways** | **Number of differentially expressed genes** | **Number of genes that were not differently expressed** | **Total number of al genes** |
| **p53 pathway** | 10 | 90 | 100 |
| **Other genes** | 990 | 18,910 | 19,900 |

**END**