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

**ASSIGNMENT #3**

**Topics:**

Independent t-tests with and without equal variances

Paired t-tests

Nonparametric tests of location

Choice of Sample Size

ROC curves

Chi-square tests and tables

**Total marks: 100**

**Submission Instructions:** Electronic submission to the IBEHS 4C03 Avenue to Learn Assignment 3 folder.

**Due:**

**Assignment Submission Format**:

File Names: StudentLastNameFirstNameAssignment#.doc/pdf/jpynb etc.

Both your name and student numbers should appear at the top of the document. If separate documents are submitted, then you should submit the answers to the questions in a document including the plots you made in python. Regardless, the python file used to generate

the plots and any provided answers should also be submitted with any submission. Submit a single zip file per assignment.

**Problem 1: [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). Analyze the hypothesis test using Python.
4. Construct the 95% confidence interval for the fluorescence after two hours for tumor and the 95% confidence interval for the fluorescence after two hours in normal tissue. Show your work and specify your conclusion based on these findings. Discuss how this conclusion does or does not support your hypothesis test?
5. 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.
6. 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.
7. 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 2: [20 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. Plot boxplots to visualize recovery times and describe the groups. Do you think there will be a statistically significant difference in recovery time?
3. We will assume that the underlying distribution is known to be normal. Perform the appropriate parametric t-test to determine if recovery time was significantly different for patients having surgery with the new technology.
4. 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).
5. Run the appropriate test in Python. What do you conclude?

**Problem 3: [30 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. Plot boxplots to visualize hardness measures and describe the groups. Do you think there will be a statistically significant difference in hardness measures between the tips?
3. 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).
4. Run the appropriate test in Python. What do you conclude?

**Problem 4: [10 points] Contingency Tables as Confusion Matrices and ROC Curves**

The aSAH dataset summarizes several clinical and laboratory variables of 113 patients with an aneurysmal subarachnoid hemorrhage (aSAH). Serum s100 calcium binding protein B (s100B) is a well-studied marker for neurologic injury and has been suggested as a candidate for predicting outcome after ASH. Outcome is reported as a good outcome (coded as a 0 here based on being a 4 or 5 on the Glasgow Outcome Scale) or a poor outcome (coded here as a 1 based on being a 1,2, or 3 on the Glasgow Outcome Scale).

Chart, histogram

Description automatically generatedThe data on the s100b values in patients with good outcomes after ASH and in patients with poor outcomes after ASH is shown graphically to the right. This same data is shown in table form below, as well.

**s100b by Outcome after ASH**.

|  |  |  |
| --- | --- | --- |
| **s100b (microg/L)** | **Good Outcome (0)** | **Poor Outcome (1)** |
| <0.01 | 0 | 0 |
| 0.01 to 0.019 | 0 | 0 |
| 0.02 to 0.029 | 0 | 0 |
| 0.03 to 0.039 | 0 | 1 |
| 0.04 to 0.049 | 5 | 0 |
| 0.05 to 0.059 | 3 | 0 |
| 0.06 to 0.069 | 2 | 0 |
| 0.07 to 0.079 | 6 | 3 |
| 0.08 to 0.089 | 6 | 1 |
| 0.9 to 0.099 | 6 | 2 |
| 0.10 to 0.19 | 30 | 8 |
| 0.20 to 0.29 | 2 | 5 |
| 0.30 to 0.39 | 4 | 4 |
| 0.40 to 0.49 | 6 | 5 |
| 0.50 to 0.59 | 2 | 3 |
| 0.60 to 0.69 | 0 | 0 |
| 0.70 to 0.79 | 0 | 5 |
| 0.80 to 0.89 | 0 | 2 |
| 0.90 to 0.99 | 0 | 1 |
| >1.00 | 0 | 1 |
| **TOTAL** | **72** | **41** |

Normal levels of s100b are low (generally much lower than 0.1 microg/L), so let’s first use this threshold of s100b < 0.10 to produce the confusion matrix for using s100b as a diagnostic test for poor outcomes after ASH. Here the diagnostic test prediction is for higher levels of s100b (a positive test) to be able to classify patients after ASH into good and poor outcome groups. Here is this confusion matrix:

**Confusion Matrix 1 s100b<0.10**

|  |  |  |  |
| --- | --- | --- | --- |
| **s100b (microg/L)** |  | **Poor Outcome (1)** | **Good Outcome (0)** |
|  | **Prediction Values (Diagnostic Test Results)** | **Ground Truth**  **(Gold Standard, Actual Test Result, Disease State)** | |
| >= 0.1 | Positive | 34 | 44 |
| < 0.1 | Negative | 7 | 28 |

Answer the following questions:

1. Using confusion matrix 1, identify the four main elements of the matrix and briefly provide the definitions of the elements and discuss their use in the context of this diagnostic test.
2. Explain sensitivity and specificity and how these two values are used by definition and in the context of the diagnostic test represented by confusion matrix 1. Calculate the sensitivity and specificity in confusion matrix 1.
3. At each threshold value that can be chosen, a sensitivity and specificity can be calculated, and these can be plotted as an ROC curve. Discuss what a ROC curve is by definition and in the context of this example.
4. Use threshold values of <0.04, <0.10, <0.30, <0.50, and <0.60 to create 5 confusion matrices and 5 sets of sensitivity and specificity. Use these to create a simple ROC curve at these thresholds. For full marks, use python to graph a ROC curve and make sure to label all the graphical elements of a ROC curve.

**Problem 5: [10 points] Contingency Tables for Chi-Square Test**

1. Discuss what a Chi-square test of independence is used for and the assumptions necessary to use it. State the null hypothesis for a Chi-square test of independence.
2. Let’s look at using the 113 patients with the same test outcome as Problem 4 (good versus poor outcome after aSAH) but with another serum biomarker, nse (neuron-specific enolase). Using serum nse as a classifier for having a poor outcome after aSAH at the threshold values of nse<11.73 results in an accuracy of 56%.

Let’s compare this accuracy to that of using cerebrospinal fluid (CSF) nse levels. A separate study is done with 115 patients and the test accuracy, at the same threshold, is found to be 65%.

You can compare the accuracies between the tests as comparing the number of correct assessments to the number of total assessments between the two different tests, which can be shown as a contingency table. Using the contingency table below for the two diagnostic tests, compare their accuracies using a Chi-Squared Test of independence in python (using 95% confidence level). What do you conclude?

**Contingency table for 5.2: Comparing accuracies between 2 diagnostic tests.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Different diagnostic tests** | **Number of Correct Assessments (TP+TN)** | **Number of Incorrect Assessments (FP+FN)** | **Total number of all assessments (TP+TN+FP+FN)** |
| Serum nse | 63 | 50 | 113 |
| CSF nse | 75 | 40 | 115 |

1. 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. 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.

**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 |