Due 11/06 at 11:59 PM

Assignment 3 (65 pts)

Where relevant, please include code and output needed to interpret the answers.

Unless otherwise specified, use alpha (α)=0.05 and 2-sided tests.

1. Use the *pbc\_Mayo Clinic 312 pats.csv* file to answer the following questions.

Between 1974 and 1984, the Mayo Clinic conducted a clinical trial for a new drug (D-penicillamine) for patients with primary biliary cirrhosis (PBC) of the liver. In this time frame, 312 patients were randomized to receive the drug or a placebo.

Low serum albumin is a sign of liver disease and decreased liver function, so it was an important variable to collect in these patients. In the general population, the average serum albumin level is 4.4 g/dL.

* 1. Plot a histogram of serum albumin in our sample (variable name: albumin). Describe the shape of the distribution of serum albumin and provide the appropriate measures of center and spread (5 pts)

Table

Description automatically generatedGraphical user interface, text

Description automatically generated

Chart, histogram

Description automatically generated

The measure of center/spread is mean/Standard Deviation

My data seems to be roughly symmetric and no outliners. There is zero skewness in my data. Therefore, I choose mean/Standard deviation (center/spread)

Mean: 3.5200000

Standard Deviation: 0.4198920

* 1. We want to know whether the mean serum albumin for the population of PBC patients differs from that of the general population. Conduct a 1 sample t-test and interpret the results. Make sure to write out the null and alternative hypotheses, report the p-value, decision, and your conclusion in your answer. (10 pts)

**Null Hypothesis:** Population mean serum albumin level for PBS patients is equal to 4.4

**Alternate Hypothesis:** Population mean serum albumin level for PBS patients is not equal to 4.4

H0: μ=4.4

H1: μ≠4.4

**One Sample T-Test:**

**Table

Description automatically generated**

**Text

Description automatically generated with medium confidence**

P-value=<0.0001

**Interpretation**

There is 0.1% chance of observing our sample mean of 4.4 if the population of PBC patients had a mean albumin count of 4.4

**Decision**

Since this result is very unlikely if the null hypothesis were true (P-value<0.05), we reject the null hypothesis. We conclude that the mean albumin count for the population of PBS patients is statistically significantly lower than 4.4

**Conclusion**

In the given question

* The total number of observations are 312 where n>30
* We can apply Central Limit Theorem, use the properties of normal distribution and t-test.
* We have one group of interest, and we are comparing it to a known population mean, with n>30, where Central Limit Theorem applies. Hence, we will conduct a Parametric 1 sample t-test.
* Our P-value which is <0.0001 is less than 0.05 and we can reject null hypothesis.
* Therefore, I conclude that the mean albumin count for the population of PBS patients is statistically significantly lower than 4.4
  1. If serum albumin were different between the treatment and placebo groups, any differences in outcomes could be due to this difference and not due solely to the drug. Which type of validity would be threatened by unaccounted for baseline differences in the treatment groups, and why? (Recall week 1) (5 pts)

Baseline characteristics plays a crucial role in judging the validity of a trial.

Internal validity would be threatened by unaccounted for baseline differences in treatment groups.

Reason: Internal validity is all about inside of the study. Was the research done right?

So, when there are baseline differences in the treatment groups, we do not measure exactly what we wanted to measure. The study results would not be appropriate, and we are not measuring what we want to be measure.

* 1. Table

     Description automatically generatedPlot histograms of serum albumin, by treatment group (variable name: trt). Provide the appropriate measures of center and spread for each group. (5 pts)

Graphical user interface, text, application

Description automatically generatedChart, histogram

Description automatically generated

**Group one:** There is minor negative skew, but the distribution is relatively symmetric, so the **mean** and **standard deviation** are appropriate measures for center/spread.

**Mean:** 3.516266

**Standard deviation:** 0.44331

|  |  |  |  |
| --- | --- | --- | --- |
| **Mean** | 3.516266 | **Std Deviation** | 0.44331 |
| **Median** | 3.565000 |  |  |

**Group two:** There is minor negative skew, but the distribution is relatively symmetric, so the **mean** and **standard deviation** are appropriate measures for center/spread

**Mean:** 3.523831

**Standard Deviation:** 0.39584

|  |  |  |  |
| --- | --- | --- | --- |
| **Mean** | 3.523831 | **Std Deviation** | 0.39584 |
| **Median** | 3.545000 |  |  |

* 1. Conduct a 2-sample t-test to compare the baseline serum albumin in the drug and placebo groups and interpret the results. Write out the hypotheses, report the p-value, decision, and conclusion in your answer. Make sure to indicate the correct p-value based on the test for the equality of variances (10 pts).

**Null Hypothesis**: Mean Albumin count for population of PBC patients who received D-penicillamine (treatment group) is equal to Mean Albumin count for population of PBC patients who did not receive D-penicillamine (placebo group)

OR

The difference in the mean albumin count of patients who received D-penicillamine (treatment group) and who did not receive D-penicillamine (placebo group) is equal to zero

**Alternate Hypothesis**: Mean Albumin count for population of PBC patients who received D-penicillamine (treatment group) is not equal to Mean Albumin count for population of PBC patients who did not receive D-penicillamine (placebo group)

OR

The difference in the mean albumin count of patients who received D-penicillamine (treatment group) and who did not receive D-penicillamine (placebo group) is not equal to zero

H0: μ1= μ2 (μ1-μ2 =0)

H1: μ1 ≠μ2 (μ1-μ2 ≠ 0)

Table

Description automatically generated

**Two Sample T-Test:**

Graphical user interface

Description automatically generated with medium confidence

* P-Value for equality of Variances= 0.1602
* Here our P-Value is >0.05
* If P-Value is greater than 0.05 we can assume that our Variance is equal, and we will use the Pooled method
* Therefore P-Value=0.8730

**Interpretation**

There is 87.3% chance of observing our sample means of 3.51 and 3.52 if the population of PBC patients in treatment group and Placebo group has the same mean albumin county

**Decision**

Since the result is not that unlikely if the null hypothesis were true (P-Value>0.05), we do not reject the null hypothesis. We don’t have the sufficient evidence to conclude that the Mean Albumin count for PBC patients who received D-penicillamine (treatment group) and who did not receive D-penicillamine (placebo group) are different.

**Conclusion**

In the given question

* The total number of observations in treatment group 158 and placebo group are 154 where n>30
* We can apply Central Limit Theorem, use the properties of normal distribution and t-test.
* In this scenario we have two groups of unknown mean and testing the hypothesis that they are equal to each other, with n>30, where Central Limit Theorem applies. Hence, we will conduct a Parametric Two Sample T-Test.
* Our P-value which is 0.8730 is greater than 0.05 and we cannot reject null hypothesis, stating that the Mean Albumin count for population of PBC patients with treatment group and placebo group are different.
* Therefore, we don’t have the sufficient evidence to conclude that the Mean Albumin count for PBC patients who received D-penicillamine (treatment group) and who did not receive D-penicillamine (placebo group) are different
  1. Chart, histogram

     Description automatically generatedSuppose instead of serum albumin, we were interested in assessing whether bilirubin differed between the treatment groups. See the distribution of bilirubin in our sample below (the sample size remains the same as above). Would you conduct a 2-sample t-test or would you use a Wilcoxon rank sum test? Why? (5 pts)

Table

Description automatically generated

I will conduct **Two Sample T-Test**

* From the above histograms we can notice that the distribution is relatively symmetric, more of my data is concentrated on the left side and long tail goes to the right side and the data is skewed for both the groups. Data in both groups is positively skewed.
* Number of observations in group one is 158 and group 2 is 154 which is above 30
* For the study where n>30, we can apply Central Limit Theorem, we can use the properties of normal distribution and conduct T-test.
* When Central Limit Theorem applies and if we have two groups of interest, we can conduct Parametric, Two Sample T-test.

1. A doctor suspects that a new drug is helpful in lowering systolic blood pressure. In order to test this, they measure blood pressure for 75 patients at baseline, and again after 2 weeks of taking the drug.
   1. Based on the information provided, what hypothesis test is most appropriate here to assess whether blood pressure changed after taking the drug and why? Write out the null and alternative hypothesis (5 pts)

Based on the information provided **Parametric Paired T-Test** is the most appropriate to assess whether Blood Pressure Changed after taking the drug.

I chose Parametric Paired T-Test because:

* The sample size is 75 which is greater than 30. If n greater than 30, Central Limit Theorem applies, and we can use properties of normal distribution and conduct t-test. Hence, I chose Parametric Testing first
* Later in the given sample we are measuring each person against their own baseline, i.e., we are measuring BP for the same person before and after taking the drugs. Here we are interested in whether BP changes after administering the new drug. Hence, I chose Paired T-Testing

Therefore, Parametric Paired T-Test is most appropriate

**Paired differences:** For paired data we create differences between pre and post treatment values for each patient and create mean for those differences. We subtract differences for each patient (pre-post), and we calculate mean of those differences.

**Null Hypothesis:** The mean of paired differences is equal to zero

**Alternate Hypothesis:** The mean of paired differences is not equal to zero

H0: μd= 0

H1: μd ≠0

* 1. Suppose instead that only 8 patients were measured, and there was no information available about the distribution of the change in systolic blood pressure. What test would you recommend using and why? Write out the null and alternative hypothesis (5 pts)

Based on the provided information **Non-Parametric,** **Wilcoxon Signed Rank Test** is most appropriate which is an analog for Paired T-Test

I chose Wilcoxon Rank Test because

* The sample size is 8 which is less than 30. If n less than 30, Central Limit Theorem doesn’t apply, and we cannot use properties of normal distribution and cannot conduct t-test. Hence, I chose Non-Parametric Testing first
* Moreover, even if there is no information available on the change, the data is paired because each person was given new drug and we are comparing each person to their own baseline. Hence, I chose Wilcoxon Signed Rank Test which is analog for Paired T-Test.

Therefore, Non-Parametric, Wilcoxon Signed Rank Test is most appropriate

**Null Hypothesis:** The median of paired differences is equal to zero

**Alternate Hypothesis:** The median of paired differences is not equal to zero

H0: Median difference= 0

H1: Median difference ≠0

1. Use the *asthma.csv* dataset to answer the following questions. A Phase II clinical trial is being conducted to investigate whether a new therapy reduces symptom burden of asthma in children. Fifteen children are randomized to the drug or placebo groups and asked to record the number of episodes of shortness of breath over the week following the receipt of treatment. We are interested in assessing whether the number of episodes differs for placebo and drug groups.
   1. Plot the histograms of number of episodes by treatment group. Based on this, what test do you think is most appropriate and why? (5 pts)

Chart, histogram

Description automatically generatedGraphical user interface, text, application

Description automatically generated

Table

Description automatically generated

Distribution of episodes of shortness of breath is skewed in both placebo and drug group. Data is concentrated on the left and the long tail goes towards the right. Data is positively skewed and has outliners. Hence center/spread is median/IQR.

**Non-Parametric, Wilcoxon Rank Sum Test** is the appropriate test to test and interpret the results. I choose this test because

* The total size is 15 which is less than 30. Moreover, number of observations in drug group is 7, and placebo group is 8 which is less than 30. If n less than 30, Central Limit Theorem doesn’t apply, and we cannot use properties of normal distribution and cannot conduct t-test. Hence, I chose Non-Parametric Testing first
* In this study we are comparing two groups the placebo and the drug group, and we are interested in assessing whether the number of episodes differs for placebo and drug groups.
* While comparing groups where n<30, we use Non-Parametric Wilcoxon Rank Sum Test to find P-Value.
  1. Conduct the appropriate hypothesis test and interpret your results. Make sure to include your hypotheses, p-value, decision, and conclusion in your answer (10 pts).

**Null Hypothesis**: Median of number of episodes of shortness of breath of asthama.csv population who received drug (drug group) is equal to Median of number of episodes of shortness of breath of asthama.csv population who did not received drug (placebo group)

**OR**

The difference in the median of number of episodes of shortness of breath among patients who received drug (treatment group) and who did not drug (placebo group) is equal to zero

**Alternate Hypothesis**: Median of number of episodes of shortness of breath of asthama.csv population who received drug (drug group) is not equal to Median of number of episodes of shortness of breath of asthama.csv population who did not received drug (placebo group)

**OR**

The difference in the median of number of episodes of shortness of breath among patients who received drug (treatment group) and who did not drug (placebo group) is not equal to zero

H0: Median1= Median2 (Mediandifference=0)

H1: Median1 ≠Median2 (Median difference≠0)

Graphical user interface, application, table

Description automatically generatedGraphical user interface, text

Description automatically generated**Non-Parametric, Wilcoxon Rank Sum Test**

* P-Value=0.0103

**Interpretation**

There is 1.03% chance of observing the distribution of sample if the population distributions were actually same

**Decision**

Since this result is very unlikely if the null hypothesis were true (P-value<0.05), we reject the null hypothesis. We have sufficient evidence to conclude that number of episodes of shortness of breath at baseline is different for 2 groups.

**Conclusion**

In the given question

* The total number of observations are 15 where n<30
* We cannot apply Central Limit Theorem, cannot use the properties of normal distribution and t-test.
* In this scenario we have two groups of interest placebo and drug group and testing the hypothesis that they are equal to each other, with n<30, where Central Limit Theorem cannot be applied. Hence, we will conduct a Non-Parametric Wilcoxon Rank Sum Test.
* Our P-value which is 0.0103 which is less than 0.05 and we can reject null hypothesis.
* Therefore, we have sufficient evidence to conclude that number of episodes of shortness of breath at baseline is different for 2 groups