LABORATORY EXERCISE 4: PARAMETRIC AND NON-PARAMETRIC TESTS ON TWO POPULATIONS

Name: Saint Joy A. Mandalinao Score:

Laboratory Section: J2L Date: 22/10/24

Objectives: At the end of this exercise, you should be able to:

1. Perform the appropriate hypothesis tests for two independent samples.

2. Perform the appropriate hypothesis tests for two related samples. Materials

1. Calculator/MS Excel 2. PAST Software tool Total: 50 pts

Problem Economics: Cost of Living Index. In the following data pairs, the cost-of-living index for housing and cost of living index for groceries are recorded for 36 metropolitan areas in the United States (Reference: Statistical Abstract of the United States, 121st edition). Is there a difference in the cost-of-living index for housing and groceries? Use a 5% level of significance.

a. The 15 data values included in your sample generated using simple random sampling. Use this sampled data to perform items b and c. (5 pts)

15 Random Numbers

7 27 2 22 23 6 33 26 5 30 12 3 20 35 14

Specs: This table of 15 random numbers was produced according to the following specifications: Numbers were randomly selected from within the range of 1 to 36. Duplicate numbers were not allowed. This table was generated on 10/20/2024.

Random number generated	Housing	Groceries	Difference
7	100	92	8
27	128	110	18
2	109	100	9
22	115	102	13
23	107	104	3
6	96	108	12
33	86	104	18
26	104	109	5
5	100	117	17
30	91	109	18
12	95	93	2
3	128	107	21
20	109	106	3
35	90	118	28
14	93	93	0

b. Summary of the assumptions tests performed, including screenshots of results from PAST or any free statistical software. (10pts)

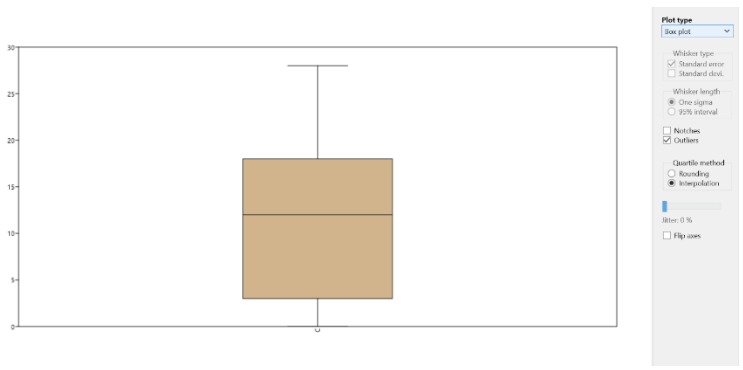
	A	B	C
1	100	92	8
2	128	110	18
3	109	100	9
4	115	102	13
5	107	104	3
6	96	108	12
7	86	104	18
8	104	109	5
9	100	117	17
10	91	109	18
11	95	93	2
12	128	107	21
13	109	106	3
14	90	118	28
15	93	93	0

C as the difference of A as Housing and B as groceries.

Checking for assumptions:

1. Samples are related
2. The cost-of-living indices can be measured on an interval or ratio scale
3. The pairs of observations are independent by randomization.

4. There are no significant outliers in the difference observed between the groups of housing and groceries.



5. Run the normality test.

Tests for normal distribution

	All
N	15
Shapiro-Wilk W	0.9473
p(normal)	0.4827
Anderson-Darling A	0.3403
p(normal)	0.4468
p(Monte Carlo)	0.4565
Lilliefors L	0.1414
p(normal)	0.571
p(Monte Carlo)	0.5716
Jarque-Bera JB	0.6887
p(normal)	0.7087
p(Monte Carlo)	0.5852

Using the Shapiro-Wilk test in assessing its normality. We can see that we obtained the results of 0.9473 which is significantly higher than 0.05. Therefore, we can conclude that it is approximately normal because the p-value is greater than 0.05.

Proceeding to Paired T-test since the assumptions are all satisfied.

Two-sample paired tests

A		***B***	
N:	15		
Mean:	103.4	**Mean:**	104.8
Median:	100	**Median:**	106
t test			
Mean difference:	1.4	**95% conf.:**	(-6.6545 9.4545)
***t*:**	-0.3728	***p* (same mean):**	0.71488
Exact:		***p* (same mean):**	0.72449
Sign test			
***r*:**	8	***p* (same median):**	0.79053
Wilcoxon test :			
***W*:**	56		
Normal appr. *z*:	0.21999	***p* (same median):**	0.82588
Monte Carlo (n=99999):		***p* (same median):**	0.84208
Exact:		***p* (same median):**	0.84314

C. The hypothesis test (Hypotheses, Decision Rule, Computation, Decision, Conclusion) including screenshots of results from PAST or any free statistical software. (10p pts)

Null Hypothesis ($H0$): There is no significant variation in the cost-of-living index for housing compared to groceries within the same metropolitan areas.

Alternative Hypothesis (Ha): There is a significant variation in the cost-of-living index for housing and groceries.

1. Hypotheses: $H0$: μd=0 | Ha: $\mu d \neq 0$

2. Choosing level of significance and formulating decision rule. Decision Rule: If p-value is < 0.05, reject null Hypothesis or $H0$. If p-value is > 0.05, accept the null hypothesis or the HO.

3. Computation: the p-value that we obtained using PAST is p-value = 0.71488

4. Decision: Since p-value = 0.71488 therefore we fail to reject the null hypothesis.

5. Conclusion: Therefore, at the 5% level of significance, we see that there is no statistically significant difference in the cost-of-living index between housing and groceries in the sampled metropolitan areas. Thus, we can conclude that the cost-of-living indices for housing and groceries are statistically similar

ACTIVITY 2

Generate your own sample by following the instruction below. Check the assumptions and perform the appropriate test. This activity has a total of 25 points, and can be achieved by following the instructions indicated in the sub-items (a,b and c)

Problem

Wildlife: Fox Rabies. A study of fox rabies in southern Germany gave the following information about different regions and the occurrence of rabies in each region (Reference:

B. Sayers, et. al., "A Pattern Analysis Study of a Wildlife Rabies Epizootic," Medical Informatics 2:11-34).

Based on the information from this article, a random sample of 10 locations in region I gave the following information about the number of cases of fox rabies near that location:

Region 1:

1	3	2	3	7
5	8	4	(Random) 1	(Random) 8

A second random sample of 8 locations in region II gave the following information about the number of cases of fox rabies near that location.

Region 2

1	4	1	2
8	6	(Random) 9	(Random) 5

Does this information indicate that there is a difference in the mean number of cases of fox rabies between two regions? Use a 5% level of significance.

Your answer should have the following:

a. The values you generated, two values for Region I and two values for Region II, using simple random sampling. Use this sampled data to perform items b and c. (5 pts)

   Values generated are the following.

   For region 1 – 7th (1) and 1st (8)
   For region 2 – 8th (9) and 7th (5)

b. Summary of the assumptions tests performed, including screenshots of results from PAST or any free statistical software. (10pts)

Step 1

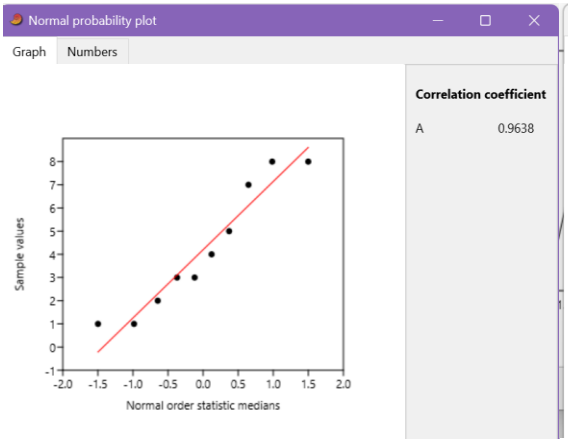
Check Assumptions:

- Independent samples which we get from two different regions
- Variable is measured at the ratio level.
- Independence is met by virtue of randomization being performed.
- There are no significant outliers in the differences in box plot.

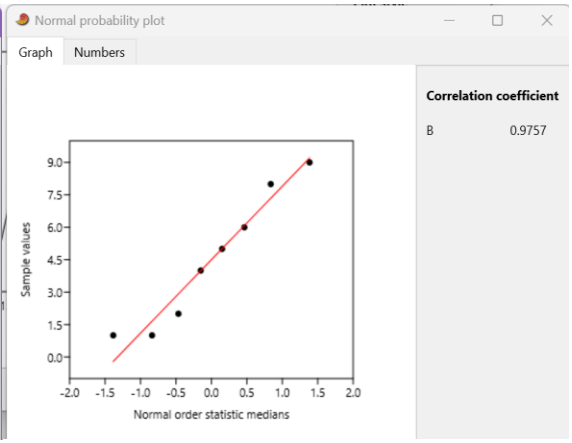
A represents Region 1 and B Represents Region 2.

- 

- Checking for normality



Region 1



Region 2

Tests for normal distribution

	A	B
N	10	8
Shapiro-Wilk W	0.8982	0.9245
p(normal)	0.2095	0.4678
Anderson-Darling A	0.3838	0.2561
p(normal)	0.3231	0.6151
p(Monte Carlo)	0.3438	0.6682
Lilliefors L	0.1717	0.1672
p(normal)	0.5451	0.7295
p(Monte Carlo)	0.5421	0.7262
Jarque-Bera JB	0.9065	0.6456
p(normal)	0.6356	0.7241
p(Monte Carlo)	0.3156	0.4939

Note that A represents Region 1 and B Represents Region 2.

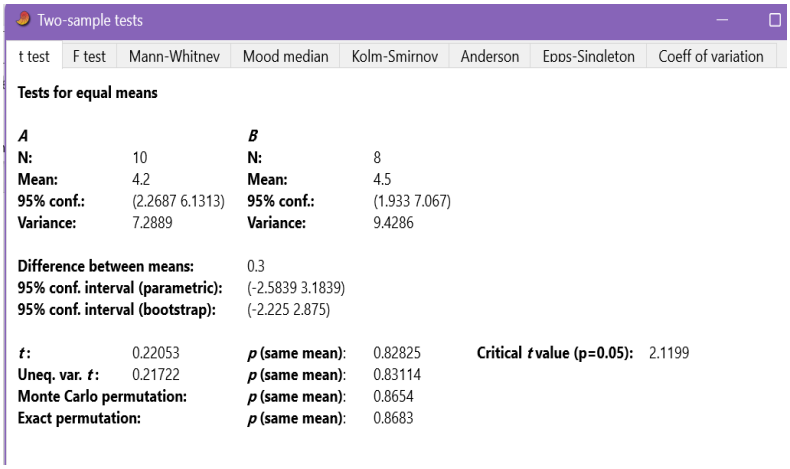
C. The hypothesis test (Hypotheses, Decision Rule, Computation, Decision, Conclusion) including screenshots of results from PAST or any free statistical software. (10p pts)

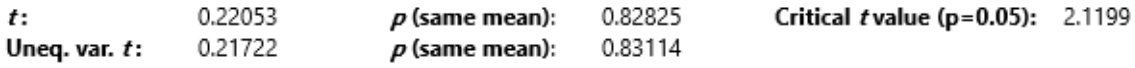
Null Hypothesis (H_o): There is no significant difference in the occurrence of fox rabies between Region I and Region II.

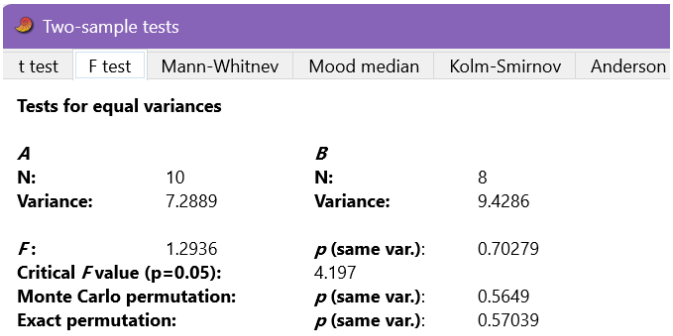
Alternative Hypothesis (H_1): There is a significant difference in the occurrence of fox rabies between Region I and Region II.

STEP 2

PERFORM T-TEST







The variance is 0.70279, since it is greater than 0.05, this means that the variance is approximately equal, and so we proceed with Student's t-test.

STEP 3

SUMMARIZE THE RESULTS

- Hypotheses: $H_0: \mu_A = \mu_B$ | $H_a: \mu_A \neq \mu_B$
- Decision Rule: If p-value is < 0.05, reject H_0.
- Computation:
    a.) F-test suggests that the variances are approximately equal since the p-value that we obtained from the f-test is p=70279 which is greater than 0.05. therefore, we used the students t-test because it has sufficed the condition that if the test for equality of two variances is greater than the level of significance, which in this case is 0.05, then the variances are equal.

    b.) Student t-test p-value = 0.82825

- Decision: Since p-value = 0.82825 > 0.05, we fail to reject H_0.

- In conclusion, at the 5% significance level, we did not find sufficient evidence to suggest a significant difference in the average number of fox rabies cases between Region I and Region II. The data indicate that there is no statistically significant variation in rabies cases between the two regions.