**Covid outbreak with a focus on the impact of workers in the tourist industry**

**Section 1: Introduction**

Tourism industry has been one of the fastest and most top economy as compare to rest of the sectors. Tourism in countries will be temporarily suspended from April 2020 due to the coronavirus disease (COVID-19) outbreak. The tourism sector significantly generates the GDP of the economy. Additionally, it employs specialists and professionals and welcomes women, migrants, students, and elderly workers; yet, it primarily employs lower-skilled individuals. After, COVID-19 has had a negative influence on tourism employment or workers, resulting in rising unemployment, decrease in labour quality, and a shift toward more informality.

**Section 2: Description of the each of the 3 datasets**

For this analysis, there is a given dataset which is divided into three datasets i.e., (dataset 1, dataset 2, and dataset3). There are two different variables for each dataset.

|  |  |  |
| --- | --- | --- |
| Variable | Description | Type |
| Are most customers international? | Customer belongs to international or national | Categorical or Dummy Variable |
| Does covid reduce work hours? | Covid word hours reduced or not. | Categorical or Dummy Variable |
| Income before covid? | Worker’s income before covid-19 | Quantitative |
| Income after covid? | Worker’s income after covid-19 | Quantitative |
| Based in NSW? | Answer to an online survey which is based on NSW or not. | Categorical or Dummy Variable |
| Reduction in hours? | hours reduction | Quantitative |

**Section 3: Relationship between the variables in each of the datasets**

1. Dataset 1

The relationship can be defined if the most customers belong to the international or national then does covid reduce work hours or not?

1. Dataset 2

The below chart shows that the strongly positive relationship between the income after covid and income before covid. Also, the correlation (r) is 0.956, which also indicates that the strength is strong between the variables.

1. Dataset 3

The relationship can be defined if the business belongs to work at in NSW or not, then how much has the covid crisis decreased your hours of work each week?

**Section 4: Confidence intervals**

**a)** The total count of reduced hours (x) = 63

Sample size (n) = 100

To find 95% confidence interval for the proportion of people with reduced hours

So, p(hat) = x/n = 63/100 = 0.63 = 63%

q(hat) = 1-p(hat) = 1-0.63 = 0.37 = 37%

Now, confidence intervals for population proportions

C.I. =

C.I. =

C.I. = 0.63 1.960(0.0483)

Confidence Interval (0.5354, 0.7246)

**b)** Average income of before covid = $75,036

Standard Deviation (Before) = $15,090

And, average income of after covid = $52,524

Standard Deviation (After) = $11,109

Now,

Confidence Interval (Before) =

After putting the value in the above formula, and we get

95% Confidence Interval: 75036 ± 2960

(72100 to 78000)

"With 95% confidence the population mean is between 72100 and 78000, based on 100 samples."

And,

Confidence Interval (After) =

After putting the value in the above formula, and we get

95% Confidence Interval: 52524 ± 2180

(50300 to 54700)

"With 95% confidence the population mean is between 50300 and 54700, based on 100 samples."

**Section 5: Hypothesis tests for the relationship between two variables**

**a)** Let = X1 be the are most customers international?

and X2 be the does covid reduce work hours?

Null hypothesis (H0): X1 = X2

Alternative Hypothesis (H1): X1 X2

If the p-value is less than the level of significance (0.05), then we can reject null hypothesis, otherwise not.

**b)** Null Hypothesis (H0): Income before covid = Income after covid

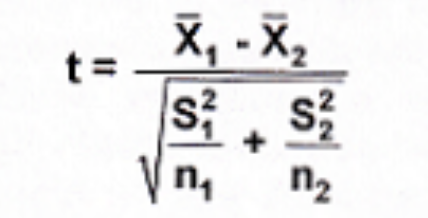
Alternative Hypothesis (H1): Income before covid Income after covid

Mean (X1) =75,036, Mean (X2) = 52,524

Standard Deviation (S1) = 15,090, Standard Deviation (S2) = 11,109

n1 = 100, n2 = 100

Using the formula as



t = (75,036 – 52,524)/SQRT(((15,090^2)/100) + ((11,109^2)/100))

t (calculated) = 12.014

t (tabulated) = 2.132 (where df = 198, and alpha = 0.05)

Hence, t (calculated) > t (tabulated) is greater. So, we can reject the null hypothesis and support H1. Therefore, income before covid and income after covid is different.

c)Null hypothesis (H0): Hours reduced (X) = 0

Alternative Hypothesis (H1): X 0

Mean = 4.81

s = 6

Z score = (x – μ)/σ

= (4.81 – 0)/6

Z (calculated) = 0.80167

Zaplha/2 (0.05) = 1.96

Hence, z (calculated) < z (tabulated) is greater. So, we cannot reject the null hypothesis and support H1. Therefore, the covid crisis decreased your hours of work each week is zero.

**Section 6: Conclusion**

In conclusion, we know about the covid outbreak impact of workers in the tourist industry. According to the dataset 2, both the variables are quantitative, the relationship between the variables are positively strong and the correlation (r) is 0.95 which supports the relationship between the income after and before covid. Further, in dataset 1, both variables are categorical variable, so we cannot find relationship between them. Moreover, based on the hypothesis testing, we found that the income before covid and income after covid is different.

**Section 7: Graphs**

Based on the dataset 2, variables are numeric, so we can find the relationship easily between the variables. The below scatter graph shows that the relationship between the income before and after covid. The dots are moving in the same direction or upward which indicates that the variables interconnection is positive.

Further, in the below graph, each bar represents the business you work at in NSW or not. The height of the bar tells that the frequency of occurrence. Thus, the tallest bar shows that the most frequency in dataset. Therefore, based in NSW is more frequency of reduction in hours as compared to not based in NSW.

