Nikhil Wora, Allison Nielsen, Damon Halvis, Amit Broca

April 27th, 2020

Professor Kim

Project Part 3:

Come up with at least 2 hypotheses relevant to your data:

1. Assessment of relations between all patients previously infected with COVID-19 and those that were infected by those previously infected patients.
2. Assessment of relations between those that got tested for COVID-19 and tested negative.
3. Testing for significant difference in the averages of patients who have tested positive with COVID-19, and patients who have tested negative.

What statistical tests did you run and why?

1. Correlation test to determine the correlation between the samples.
2. Correlation test to determine the correlation between the samples.
3. T-Test to determine significant differences in averages.

Operationalize the hypothesis. State the null and alternative hypothesis

1. Null (Ho): True correlation is equal to zero.
   Alternative (Ha): True correlation is not equal to zero.
2. Null (Ho): True correlation is equal to zero.
   Alternative (Ha): True correlation is not equal to zero.
3. Null (Ho): True difference in means is equal to zero.
   Alternative (Ha): True difference in means is not equal to zero.

What is the population, sample, parameter, and statistic

1. Population = n/a; Sample = 0.7219733; Parameter = 68; Statistic = 8.6044

```
> cor.test(patient$id, patient$infected_by)

        Pearson's product-moment correlation

data:  patient$id and patient$infected_by
t = 8.6044, df = 68, p-value = 1.753e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5864953 0.8181520
sample estimates:
      cor
0.7219733
```

2.

```
        Pearson's product-moment correlation

data:  time$test and time$negative
t = 116.94, df = 69, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9959604 0.9984369
sample estimates:
      cor
0.9974868
```

3. Population = ; Sample = mean of x = 5034.196, mean of y = 188171.511;

Parameter = 91.099; Statistic = -9.1198

```
> t.test(Time$confirmed, Time$negative)

        Welch Two Sample t-test

data:  Time$confirmed and Time$negative
t = -9.1198, df = 91.099, p-value = 1.757e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -223025.5 -143249.1
sample estimates:
 mean of x  mean of y
  5034.196 188171.511
```

What did you conclude? Did you accept or reject the null hypothesis?

1. The p-value of 1.753e-12 is significantly less than 0.05, so we reject the null hypothesis, rendering that there is no true correlation.
2. P-value of 2.2e-16 is 0.05 so we reject the null hypothesis.
3. The p-value of 1.757e-14 is significantly less than 0.05, so we reject the null hypothesis, so that there is a significant difference in means.

What are the weaknesses in your analysis?

1. Not all patients have been infected by a previously infected patient in this particular dataset.
2. Correlations between these variables are hard to determine.
3. The data is collected from a representative, randomly selected portion of the total population.