

# Project Part 5

## Context

Suicide is a serious public health problem that invokes detrimental effects on individuals, families, and communities. Each suicidal event ensues a tragedy or loss of their loved ones. In this research, the suicide mortality rates in 2019 and 2020 will be investigated in the scope of 50 states in the United States (excluding the District of Columbia) The goal is to find whether there exists a significant difference in suicide mortality rate between 2019 and 2020.

## Research Question

Is there a significant difference over the mean suicide mortality rate (the number of deaths per 100,000 total population) over the 50 states in America from 2019 to 2020?

## Data Description (description & sources & appropriateness)

The data is gathered from certain sampling methods implemented by CDC, so the data is essentially a sample. There are three variables in the data set after filtering and cleaning the data: the suicide mortality rate in 2020, the suicide mortality rate in 2019, and the difference between the suicide mortality rate between 2020 and 2019. The suicide mortality rate is in the number of deaths per 100,000 total population. It has been adjusted for differences in age-distribution and population size. The variables that we are seeking for will be found in primary existing data from 1: CDC (Centers for Disease Control and Prevention) It is from the latest published statistics on suicide from the Centers for Disease Control and Prevention (CDC) Data & Statistics Fatal Injury Report for 2020. 2: National Institute of Mental Health. It is based on data from the WISQARS Fatal Injury Data Visualization Tool. The overall data is sufficient and appropriate to conduct research analysis. The authenticity and reliability of the data set are valid since the sources are from authoritative agencies and websites. Since both variables are in number of deaths per 100,000 of the total population of the United States, it is plausible to subtract the data to get the difference in 2020 and

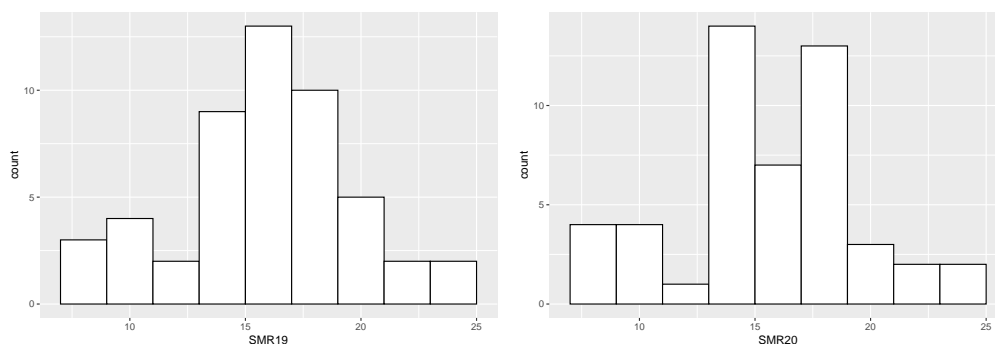
2019. As there are 50 states being investigated, indicating that it is sufficient for the central limit theorem to hold. Hence, the data is adequate to answer the research question.

## Test Identification & Appropriateness & Characteristic & Explanation

A paired t-test will be implemented to test whether the mean difference between pairs of measurements is zero or not. The paired t-test is appropriate to use. In a paired sample t-test, each subject or entity is measured twice, resulting in pairs of observations. In this research, the suicide mortality rate is measured twice for 2019 and 2020. In both cases, we have the same subjects/items in both groups, which is the 50 states of America (excluding the district of Columbia) Each subject has a pair of measurements. A paired t-test determines whether the mean difference of these pairs equals zero (no effect). Therefore, it is valid to operate the paired t-test. The variables are quantitative, so significant tests are necessitated, such as t-test, f-test, z-test, chi square test, etc. It is first of all a t-test because the population standard deviation is unknown. The data is gathered through samples implemented by CDC and the National Institute of Mental Health. It is a paired t-test because the goal of the research question is to compare the means of two measurements taken from the same individual, object, or related units.

## Test Assumptions & Validity

First, subjects must be independent. Measurements of the suicide mortality rate for each state do not affect measurements for other states. Second, each of the paired measurements is obtained from the same state. For example, the 2019-and-2020 suicide mortality rate is measured for the same state individually. Third, the dependent variable must be continuous (interval/ratio). The suicide mortality rate is continuous through time. Fourth, the measured suicide mortality rate for both years are approximately normally distributed based on the histograms of distribution. Also, the sample size is greater than 30, CLT holds.



## Test Hypothesis

H0:  $\mu_d = 0$  the true mean difference ( $\mu_d$ ) is equal to zero

HA:  $\mu_d \neq 0$  the true mean difference ( $\mu_d$ ) is not equal to zero

```
t.test(data$SMR20, data$SMR19, mu=0, alternative="two.sided", paired=TRUE)
```

```
##
## Paired t-test
##
## data: data$SMR20 and data$SMR19
## t = -1.9357, df = 49, p-value = 0.05869
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.09245857 0.02045857
## sample estimates:
## mean of the differences
## -0.536
```

## Test Result & Conclusion & Generalization

The p-value of the two-sided paired t-test is 0.02818, which is greater than the significance level at  $\alpha = 0.05$ . As we rejected the null hypothesis, there is sufficient evidence that there is a significant difference over the mean suicide mortality rate (the number of deaths per 100,000 total population) over the 50 states in America from 2019 to 2020.

According to the above test, there is not an essential difference over the mean suicide mortality rate (the number of deaths per 100,000 total population) over the 50 states in America from 2019 to 2020. The advent of COVID-19 will not necessarily lead to more or less suicidal events.

To better generalize the result, researchers can try to draw the data from the population rather than a sample. It will significantly reduce variability so that it will be sufficient enough to generalize beyond the data. In future studies, other researchers could take in more parameters into consideration and build stronger model.

## References

1. <https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm>
2. [https://www.nimh.nih.gov/health/statistics/suicide#part\\_2585](https://www.nimh.nih.gov/health/statistics/suicide#part_2585)
3. <https://afsp.org/suicide-statistics/>
4. <https://libguides.library.kent.edu/spss/pairedsamplesttest>