Tran Bui

June 10, 2021

1 Theorem

Let x and y denote the numbers of successes observed in two independent sets of n and m Bernoulli trials, respectively, where p_x and p_y are the true success probabilities associated with each set of trials. Let $p_e = \frac{x+y}{n+m}$ and define

$$z = \frac{\frac{x}{n} - \frac{y}{m}}{\sqrt{\frac{p_e(1-p_e)}{n} + \frac{p_e(1-p_e)}{m}}} \tag{1}$$

a. To test $H_0 : p_x = p_y$ versus $H_1 : p_x > p_y$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.

b. To test $H_0 : p_x = p_y$ versus $H_1 : p_x < p_y$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.

2 Hypothesis Testing

2.1 Case study

The onset of pandemic had a massive impact on crime in the U.S., with large drops in almost all types of crime. There is a widespread immediate drop in both criminal incidents and arrests most heavily pronounced among **drug crimes**, theft, residential burglaries, and most violent crime. - By David S.Abrams, University of Pennsylvania

2.1.1 Construct a hypothesis test

Let x be the number of drug crimes in 2019, so x = 1614.

Let y be the number of drug crimes in 2020, so y = 1000.

Let p_x be the true probability of the number of drug crimes in 2019.

Let p_y be the true probability of the number of drug crimes in 2020.

Step 1. State the null and alternative hypotheses

The hypotheses to be tested are

$H_0 : p_x = p_y$ (The proportions of drug crimes in 2019 and 2020 are the same)

versus

$H_1 : p_x > p_y$ (The proportion of drug crimes decreases in 2020)

Step 2. Choose a significant level α

By convention, we choose $\alpha = 0.05$.

When we let $\alpha = 0.05$, which means that there is 5% chance that we will accept your alternative hypothesis when your null hypothesis is actually true. In other words, we are about 95% confident that we have made the right decision.

Step 3. Compute the test statistic

By Theorem (1)

$$p_e = \frac{x+y}{n+m} = \frac{1641+1000}{70289+75790} = \frac{641}{146079} = 0.01807926$$

The sample proportion of the number of drug crimes in 2019 is

$$\frac{x}{m} = \frac{1641}{70289} = 0.02334647$$

The sample proportion of the number of drug crimes in 2020 is

$$\frac{y}{m} = \frac{1000}{75790} = 0.01319435$$

According to theorem (1), then, the test statistic is

$$z = \frac{\frac{x}{n} - \frac{y}{m}}{\sqrt{\frac{p_e(1-p_e)}{n} + \frac{p_e(1-p_e)}{m}}} = \frac{0.02334647 - 0.01319435}{\sqrt{\frac{p_e(1-p_e)}{n} + \frac{p_e(1-p_e)}{m}}} \approx 14.55$$

Step 4. Find the critical values and decision rule, based on both the form of the hypotheses and the significance level

Rejection region for upper-tailed z test ($H_1 : p_x > p_y$) with $\alpha = 0.05$.

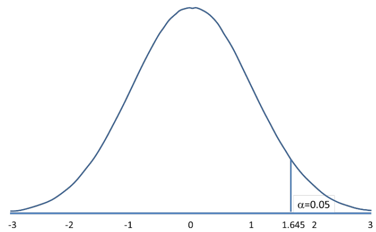

Figure 1: The z-score of $\alpha = 0.05$

Using the z-table to find the probabilities corresponding to each z-value.

We have $z_\alpha = z_{0.05} = 1.645$.

By Theorem (1), the decision rule is

Reject H_0 if $z \geq 1.645$.

Step 5. Decide to reject or not reject the null hypothesis by comparing the test statistic to the critical value

By Theorem (1),

To test $H_0 : p_x = p_y$ versus $H_1 : p_x > p_y$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.

Since $z \geq z_{0.05}$ $(14.55 \geq 1.645)$ we should reject the null hypothesis.

From here, we can go through the same procedure, constructing similar hypothesis tests for data of the preceding years.

By using software, we get the following z-score values:

	z-score
2017 vs 2020	24.75
2018 vs 2020	21.82

Since the z-score values are all greater than $z_\alpha = z_{0.05} = 1.645$, we should reject the null hypothesis.

Step 6. Interpret results with reference to the original question

We conclude that these data provide convincing evidence that the number of drug crimes in 2020 declined compared to past years due to the COVID-19.

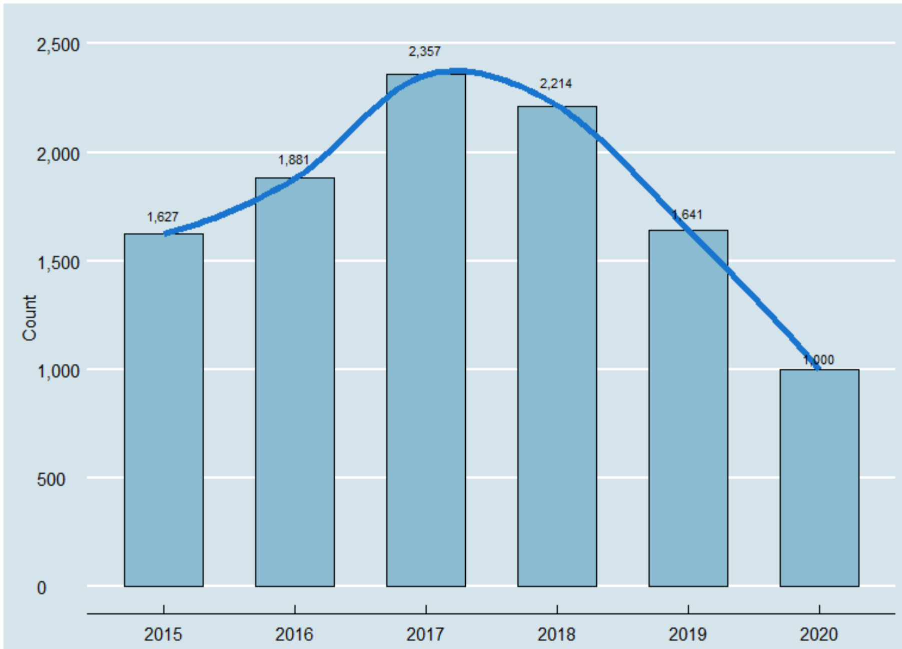

Figure 2: Drug Crimes in Seattle 2015 - 2020