

1. We have a NYC house sale dataset from 01/01/2016 to 12/31/2021. We want to find the correlation between the number of daily sales and COVID cases/vaccines and the correlation between the average daily sale price and COVID cases/vaccines. This is useful and important because one of the biggest things to happen during the covid pandemic era was that investments went digital and a general trend where the poor get poorer, and the rich get richer. Therefore, the wealthy may start to put their assets into investments like real estate. We hypothesize that as covid cases rise, the number of sales for houses and the average sale price will also increase. Moreover, vaccinations have started to roll out, and daily life is slowly returning back to normal. We infer that rather than continuously investing their assets, people will start putting their money into expensive real estate that they can enjoy. Therefore, we use a Pearson correlation test to measure our inferences. Our results are summarized in table 1, from which we can give **four inferences** which weakly suggest that there may be a slight positive correlation between the number of new covid cases/administered vaccine and the number of daily house sale/average daily sale price. This can help people who want to buy house during COVID pandemic.

Table 1. The Pearson correlation coefficient results.

	The number of new cases	The number of new vaccines
The number daily sales	0.122	0.105
The average daily price	0.125	0.138

2. We then check if the COVID19 cases data changed after the NYS reopen rule on July 16 2020.

assume the average new cases before reopen is  $\mu_{bf}$   
the average new cases after reopen is  $\mu_{af}$

The null hypothesis ~~is~~ is that  $\mu_{bf} \leq \mu_{af}$ .  
The alternative hypothesis is  $\mu_{bf} > \mu_{af}$

we then perform a one-tailed z-test (unpaired)

$$H_0: \mu = \mu_{bf} - \mu_{af} \leq \mu_0 = 0$$

$$H_1: \mu > \mu_0$$

$$Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

if  $Z > Z_{\alpha}$ , we reject  $H_0$ , here we set  $\alpha = 0.05$

From the calculation,  $Z = 5.087 > Z_{0.05} = 1.645$

We reject the null hypothesis, so the average new cases before reopen is larger than the average new cases after reopen.

This inference is useful because it tells us that the reopen rule leads to a increase of covid cases, to some extent. We need to be more careful when we are outside.