

1. The result from data cleaning:

We checked the "new\_case" and "new\_death" in Arizona and Alabama, and clean the data by using the following steps:

- (1). Select AZ and AL data new\_case and new\_death from the database
- (2). Clean all Null data starting from 2020-05-01.
- (3). Sort the data based on date
- (4). Using Tukey's rule to detect and delete outlier

Problem: The new case and new death data have negative values. We deleted them.

Outlier detection results:

AL_new_case:	AZ_new_case:
low=-793.0	low=-1797.0
high=2063.0	high=4163.0
IQR=714	IQR=1490
outlier number:189	outlier number:123
AL_pnew_case:	AZ_pnew_case:
low=-196.5	low=-204.0
high=463.5	high=348.0
IQR=165.0	IQR=138.0
outlier number:215	outlier number:217
AL_new_death:	AZ_new_death:
low=-12.5	low=-32.0
high=39.5	high=56.0
IQR=13	IQR=22
outlier number:172	outlier number:172
AL_pnew_death:	AZ_pnew_death:
low=-3.5	low=-1.5
high=8.5	high=2.5
IQR=3.0	IQR=1.0
outlier number:195	outlier number:317

2. W,Z,T-test

(1). One sample test

=====Wald's test for Arizona new cases/death data=====

w\_new\_case=238.59066878817404, which is larger than 1.962. We need to reject H0

w\_new\_death=8.891039501678696, which is larger than 1.962. We need to reject H0

```

=====Wald's test for Alabama new cases/death data=====
w_new_case=95.52185087896518, which is larger than 1.962. We need to reject H0
w_new_death=32.622055662075816, which is larger than 1.962. We need to reject H0

=====z-test for Arizona new cases/death data=====
z_new_case=101.60887826363583, which is larger than 1.962. We need to reject H0
z_new_death=44.66446351274628, which is larger than 1.962. We need to reject H0

=====z-test for Alabama new cases/death data=====
z_new_case=61.22999387640326, which is larger than 1.962. We need to reject H0
z_new_death=100.4992908449547, which is larger than 1.962. We need to reject H0

=====t-test for Arizona new cases/death data=====
t_new_case=33.65963518825322, which is larger than 2.0518. We need to reject H0
t_new_death=5.132058610002958, which is larger than 2.0518. We need to reject H0

=====t-test for Alabama new cases/death data=====
t_new_case=24.017534104291745, which is larger than 2.0518. We need to reject H0
t_new_death=51.80123942468204, which is larger than 2.0518. We need to reject H0

```

## (2) Two-sample test

```

=====two samples Wald's test for Arizona new cases/death data=====
w_new_case=126.69228257809407, which is larger than 1.962. We need to reject H0
w_new_death=6.396342853612112, which is larger than 1.962. We need to reject H0

=====two samples Wald's test for Alabama new cases/death data=====
w_new_case=59.52176008581291, which is larger than 1.962. We need to reject H0
w_new_death=17.13679643557639, which is larger than 1.962. We need to reject H0

=====two samples t test for Arizona new cases/death data=====
t_new_case=30.656778936772874, which is larger than 1.962. We need to reject H0
t_new_death=4.78084824286444, which is larger than 1.962. We need to reject H0

```

## (3) Applicability

**Walt's test:** the estimator should be Asymptotically Normal. According to the “under 30 rule”, the size of our data set is 31, which may available to the 1 sample application. For the 2-sample case, we have 28 data points which are not enough, the W test is not applicable.

**Z-test:** Z test needs the sample data being normally distributed, however, it gives Poisson distribution instead, which makes the z-test not applicable to this case.

**t-test:** t-test can use a sample deviation instead of the true standard deviation. But the data is not normally distributed (see in the codes), and the t-test may also not be applicable.