Data analysis

PengYangYu

2022-11-14

0.0.1 Research question 1:

Is there any correlation between ozone concentration and daily AQI in 2022?

The null hypothesis: ozone concentration has no correlation with daily AQI

The alternative hypothesis: ozone concentration has correlation with daily AQI

```
##
## Call:
## lm(formula = Concentration ~ DailyAQI, data = EPAair_O3_22)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042585 -0.000866  0.000372  0.001434  0.003672
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.350e-03  1.110e-04   84.25   <2e-16 ***
## DailyAQI    8.231e-04  2.688e-06  306.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002882 on 9287 degrees of freedom
## Multiple R-squared:  0.9099, Adjusted R-squared:  0.9099
## F-statistic: 9.378e+04 on 1 and 9287 DF,  p-value: < 2.2e-16
```

The R-squared of this linear model is 0.9099, which means that 90.99% variability in daily air quality index (AQI) is explained by changes in ozone concentration. The degrees of freedom of the model is 9287 = 9289-2 since the number of observations is 9289 and the number of parameters is 2. According to the p-value of slope, which is smaller than 0.05, we reject the null hypothesis, so the correlation between ozone concentration and daily AQI in 2022 is significant.

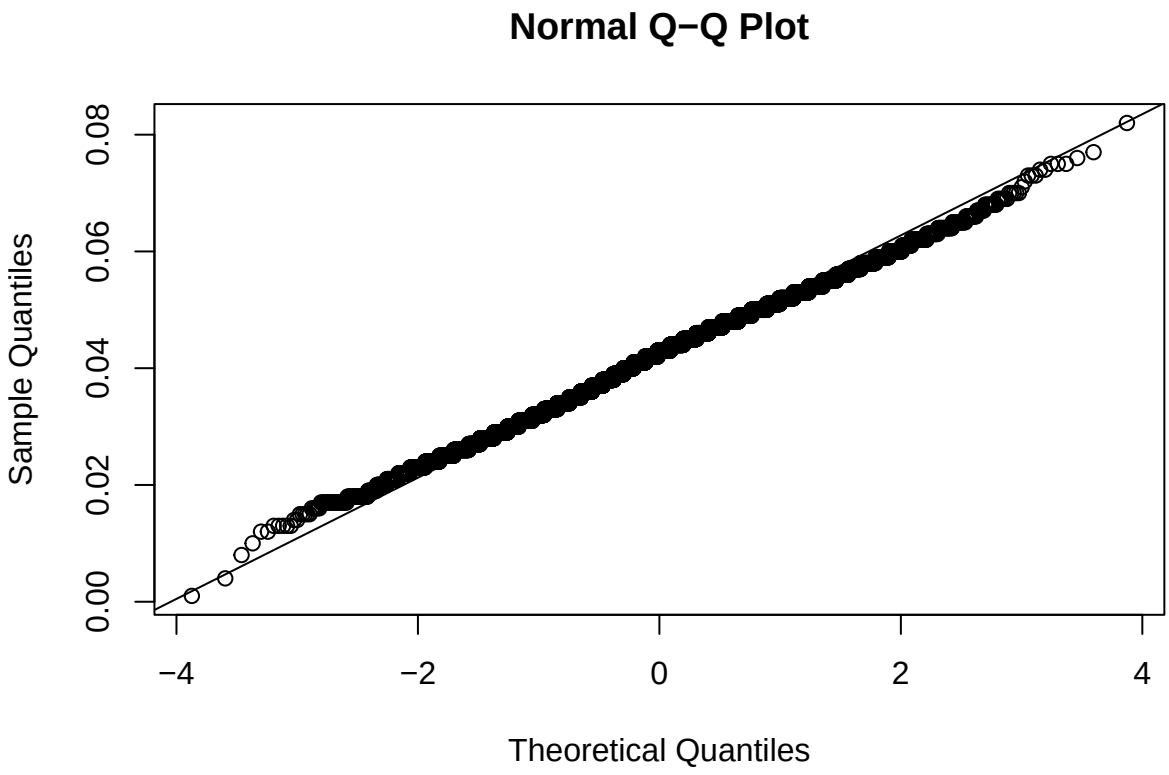
0.0.2 Research question 2:

Do different sites have equal mean of ozone concentrations in 2022?

The null hypothesis: they have equal mean of concentrations

The alternative hypothesis: they do not have equal mean of concentrations

```
## [1] "There are 38 sites in total and 28 of them follow a normal distribution."
```




```
##
##  Bartlett test of homogeneity of variances
##
## data:  EPAair_O3_22$Concentration and EPAair_O3_22$SiteName
## Bartlett's K-squared = 114.06, df = 37, p-value = 8.769e-10
```

```
##               Df Sum Sq   Mean Sq F value Pr(>F)
## SiteName      37 0.0546 0.0014746   17.02 <2e-16 ***
## Residuals   9251 0.8016 0.0000867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the Shapiro-Wilk test result and normal Q-Q plot, we find that most of the sites are conform to normal population distribution assumption. However, the Bartlett's test result shows that the null hypothesis that the variances in each sites are the same is rejected. Since ANOVA is robust against departures from equal variance, we can still apply one-way ANOVA on our dataset.

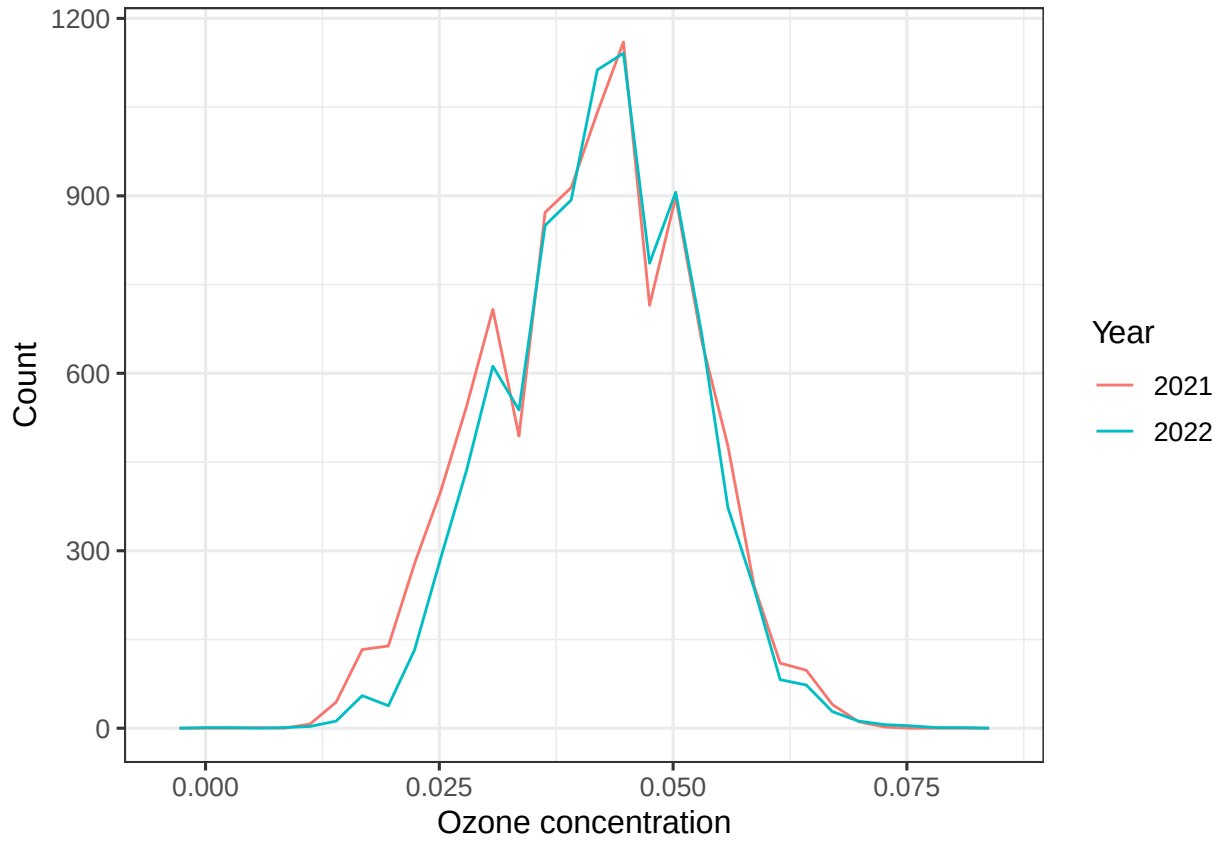
The p-value of the ANOVA is smaller than 0.05, so we reject the null hypothesis. Therefore, the mean of ozone concentrations in 2022 significantly differ among sites.

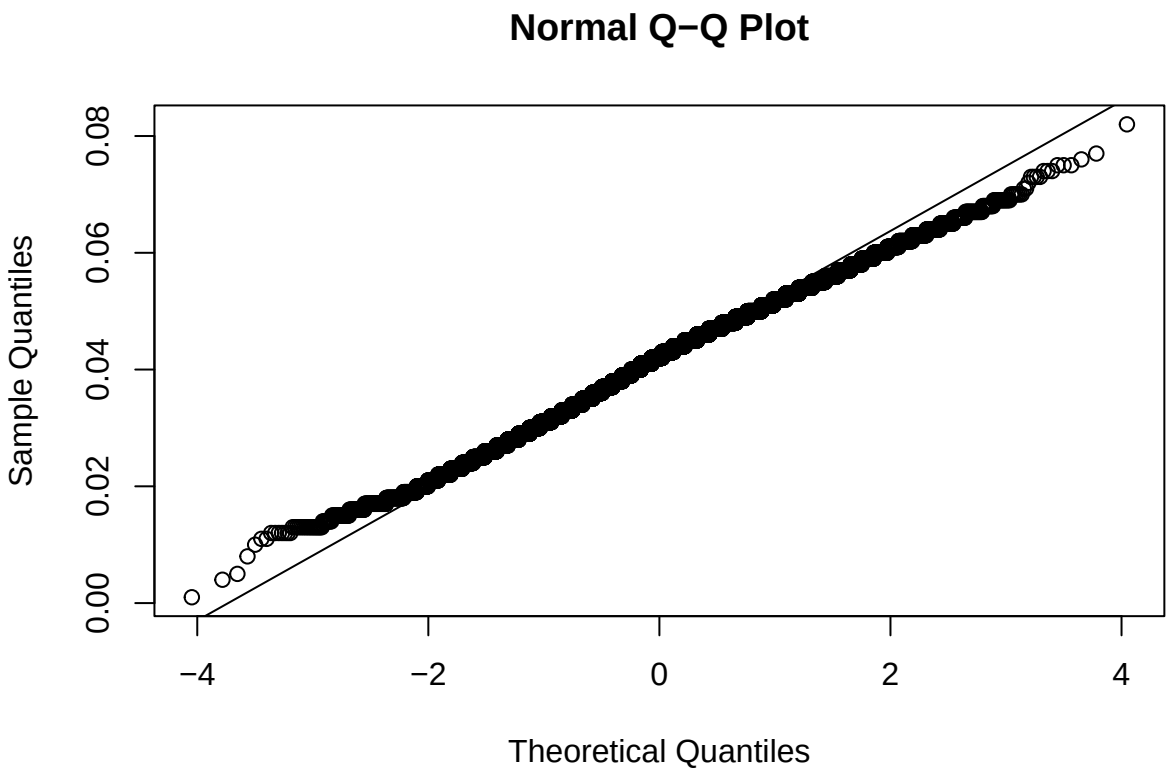
0.0.3 Research question 3:

Is the mean of ozone concentrations in 2021 and 2022 equivalent?

The null hypothesis: the mean between 2021 and 2022 is equivalent

The alternative hypothesis: the mean between 2021 and 2022 is not equivalent



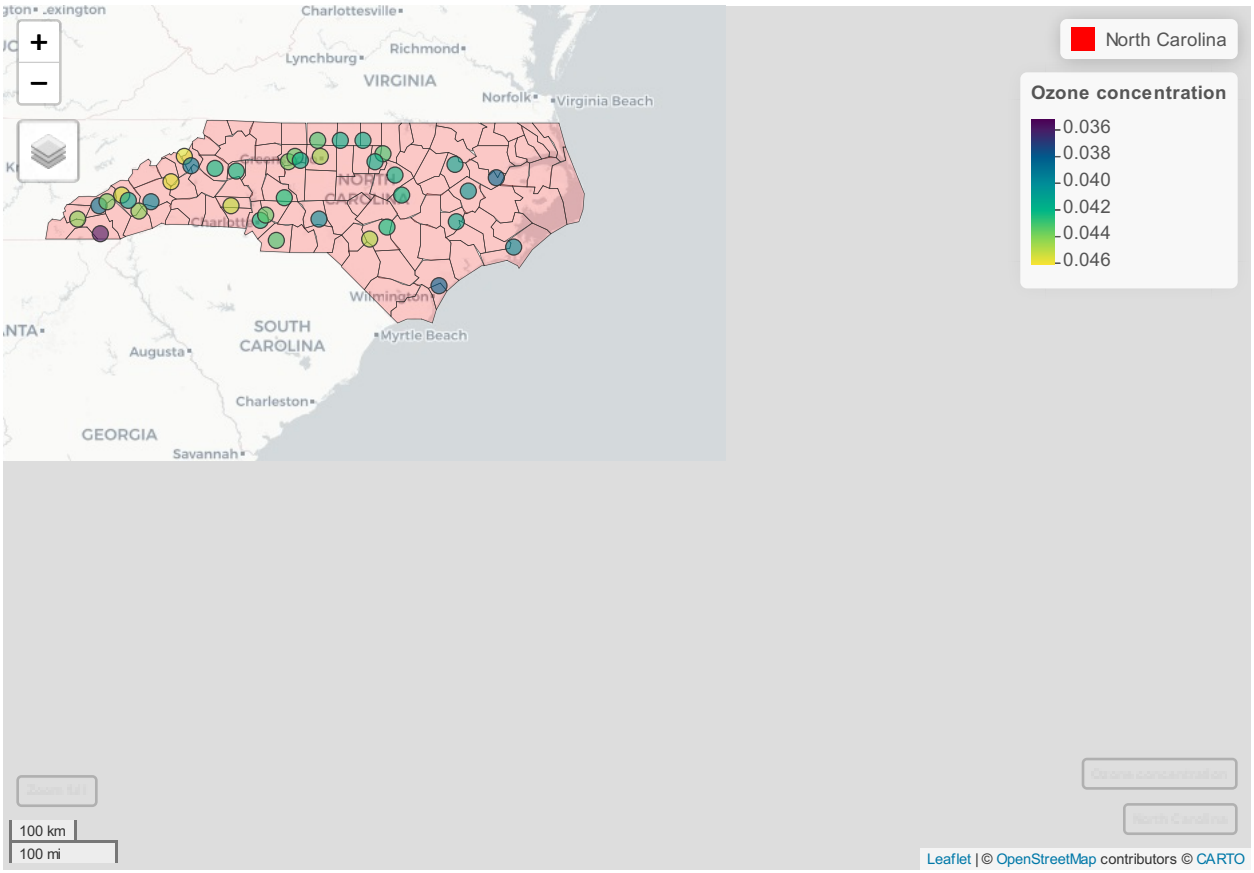


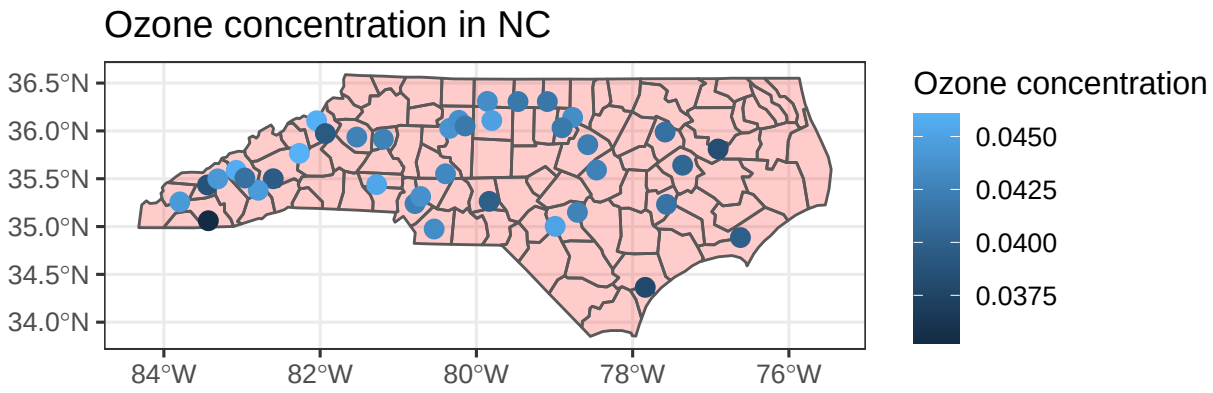

```
##
##  Bartlett test of homogeneity of variances
##
## data:  EPAair_O3_2122$Concentration and EPAair_O3_2122$Year
## Bartlett's K-squared = 103.52, df = 1, p-value < 2.2e-16

##
##  Welch Two Sample t-test
##
## data:  Concentration by Year
## t = -7.0983, df = 19249, p-value = 1.307e-12
## alternative hypothesis: true difference in means between group 2021 and group 2022 is
## 95 percent confidence interval:
##  -0.0013217999 -0.0007497675
## sample estimates:
## mean in group 2021 mean in group 2022
##         0.04104097         0.04207676
```

The normal Q-Q plot shows that the data has small deviations from normal distribution. The Bartlett's test result shows that the variances between 2021 and 2022 are different. Again, t-test is robust to these. The T-test suggests that the mean of ozone concentrations in 2021 and 2022 is not equivalent with p-value smaller than 0.05.

0.0.4 Research question 4:

Is there any trend of ozone concentrations in space?





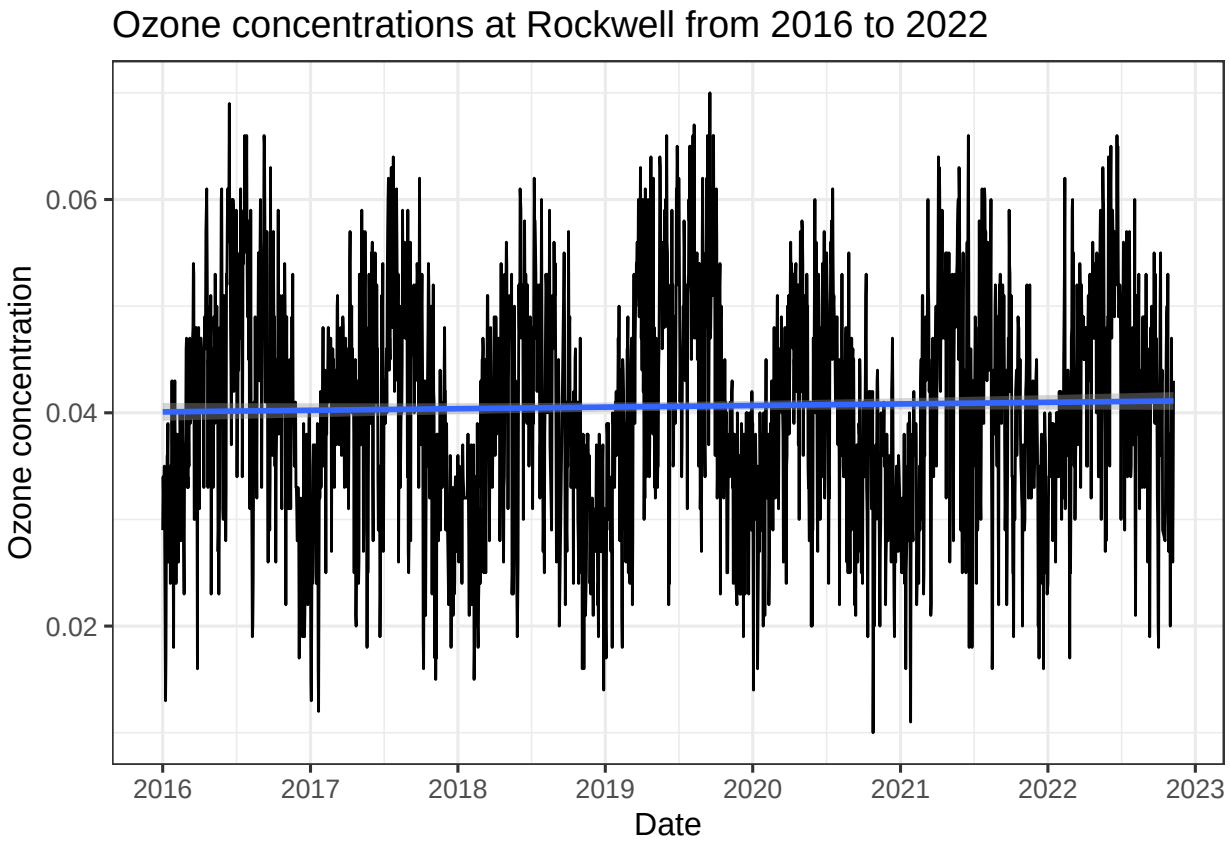
It seems there is no obvious trend between the location of sites and the mean ozone concentration collected from the sites.

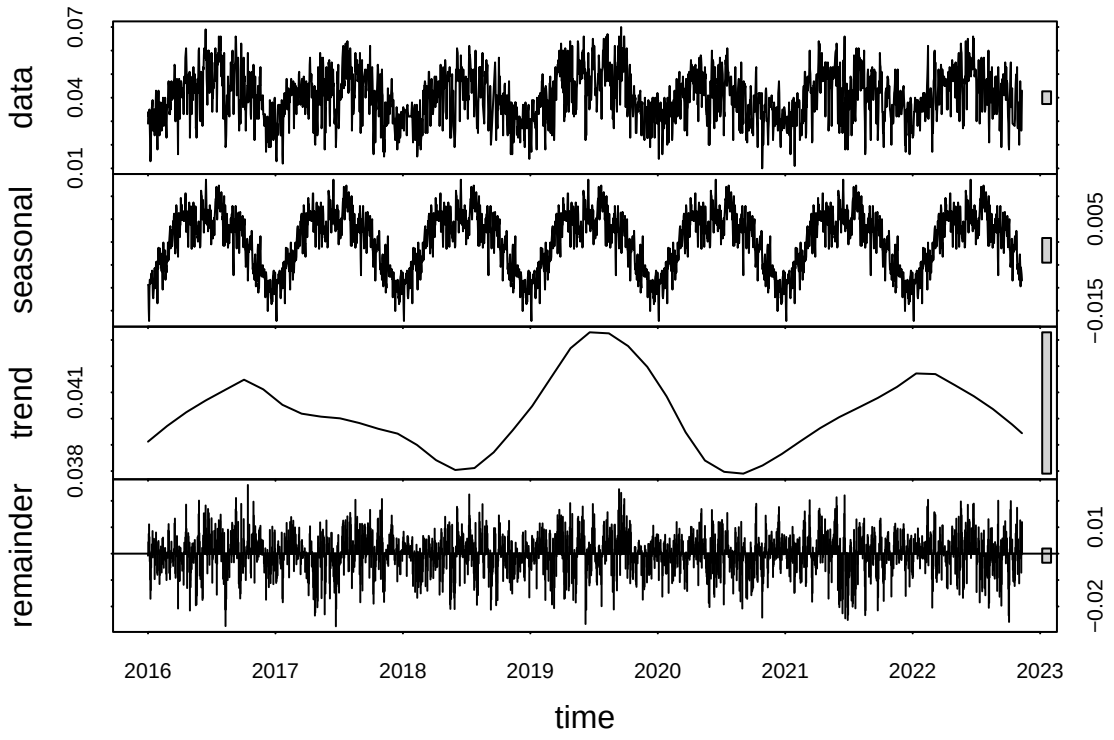
0.0.5 Research question 5:

Have ozone concentrations changed from 2016 to 2022 at Rockwell?

The null hypothesis: the ozone concentration is stationary over time

The alternative hypothesis: the ozone concentration change over time


Ozone concentrations at Rockwell from 2016 to 2022
Ozone concentration
0.06
0.04
0.02
2016
2017
2018
2019
2020
2021
2022
2023
Date




```
## Score =  87 , Var(Score) = 15119
## denominator =  7239.089
## tau = 0.012, 2-sided pvalue =0.47922
```

The Seasonal Mann-Kendall test is chosen to test monotonic trend because the decomposed figure shows that the time series object has a strong seasonal component. From the result, we accept the null hypothesis since the p-value is greater than 0.05. Therefore, the ozone concentration at Rockwell is stationary from 2016 to 2022.