Hypothesis, the travel times in Delhi city decreased from 2016 to 2019:

The data set: The three quarters aggregated monthly means of each of the origin-destination id pairs with a recorded value across all the years.

Hypothesis 1.1: The distribution of the travel times changed from 2016 to 2017, 2017 to 2018, 2018 to 2019.  
Test used: Kolmogorov Smirnov test

Results:

|  |  |  |
| --- | --- | --- |
| Years | P-value | Inference |
| 16 -17 | < 2.2e-16 | Distributions differ |
| 17 -18 | 2.824e-09 | Distributions differ |
| 18 - 19 | p-value < 2.2e-16 | Distributions differ |

Conclusions: the underlying distribution of the travel times did not stay uniform over the years. To find the directionality of the change, we shall use the paired sample Wilcoxon test.

By KS test and Jarque-Bera test for normality, the data samples a significantly non-normal. (p-values < 2.2e-16).

Hence, the paired sample Wilcoxon test shall be used to clarify the directionality of the changes.

Let us first re-analyse whether the change in distributions is significant by the paired sample Wilcoxon test with alternative hypothesis that true location shift is not equal to zero.

Results:

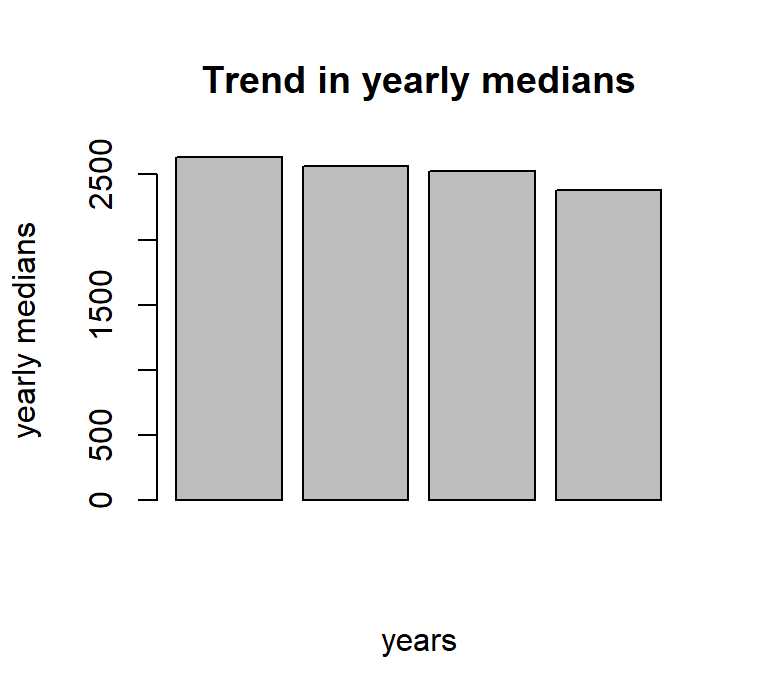
|  |  |  |
| --- | --- | --- |
| Years | P-value | Inference |
| 16 -17 | < 2.2e-16 | True shift not equal to 0 |
| 17 -18 | < 2.2e-16 | True shift not equal to 0 |
| 18 - 19 | < 2.2e-16 | True shift not equal to 0 |

Hence, this test gives full indication that the distributions have changed. Now for the directionality, our hypothesis is that the values have shifted downwards, I. e. the true location shift in the median of the dataset is less than 0.

TBC.

|  |  |  |
| --- | --- | --- |
| Years | P-value | Inference |
| 16 -17 | 1 | True shift is less than 0. |
| 17 -18 | 1 | True shift is less than 0. |
| 18 - 19 | 1 | True shift is less than 0. |

This is predictable as the medians of the data are indeed decreasing.



It would be much more conclusive if a confidence interval could be established for the drops in travel times across years. Unfortunately, that shall require us to assume a distribution for the travel times. Hence, let us assume that the samples follow a loosely normal distribution. \* This makes computation much simpler.

Even though we have shown that the samples do not follow the normal distributions, we can take the assumption since, the distributions do follow a bell-like curve and the skew for the distributions is well within the acceptable range.

\*\*skewness of the data

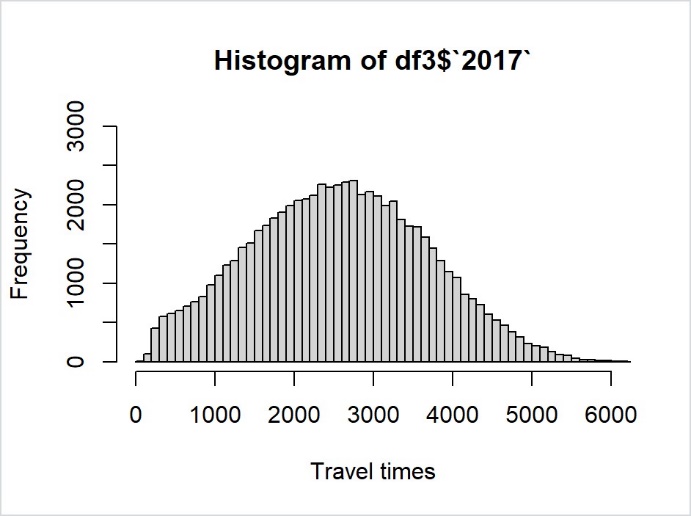
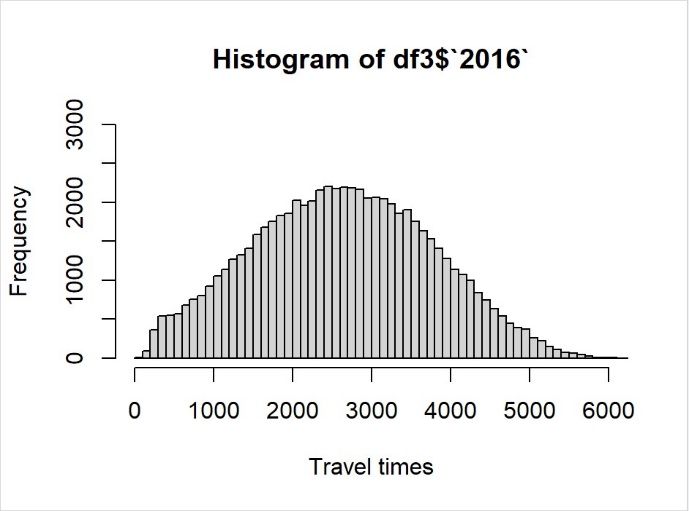
2016: 0.01235329 < 0.4

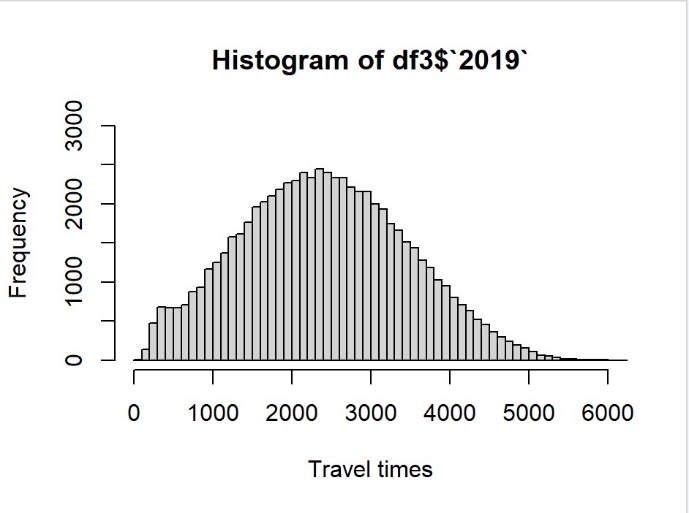
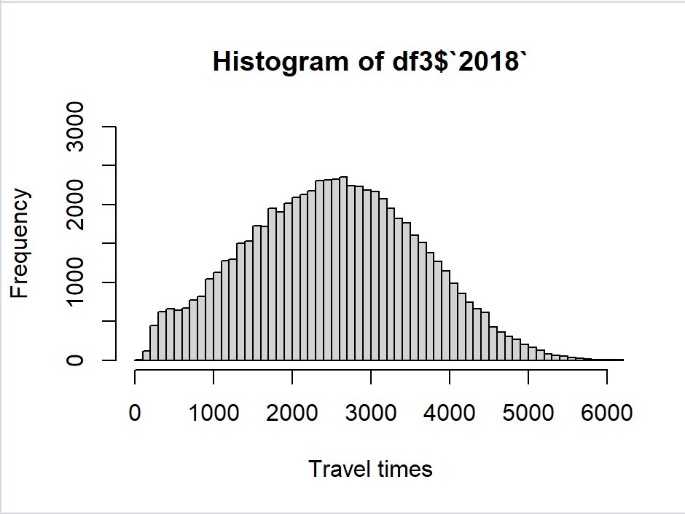
2017: 0.009026067 < 0.4

2018: -0.0001721346 > -0.4

2019: 0.06678994 < 0.4

Visual plausibility for the assumption:





Hence, with this assumption we can perform the t-tests for the pairs of these data-sets.

|  |  |  |
| --- | --- | --- |
| Year pairs | 95% confidence intervals | Mean Difference |
| 16 – 17 | [72.91330, 75.12967] | 74.02149 |
| 17 – 18 | [39.49212, 41.39311] | 40.44262 |
| 18 – 19 | [117.848, 120.875] | 119.3615 |

Hence, we can conclude that the travel times in the Delhi city decreased significantly across the years from 2016 to 2019. Here, the exact numbers of decrease cannot be pinpointed and the confidence intervals are only approximate due to the normality assumption. But, all the statistical tests point unequivocally towards decreasing travel times.