One-Sample t-Test

11/15/2021 15:38

# Abstract

Outcome variable: *year*  
Null mean: *2015*  
Alternative hypothesis: *two.sided*

A One-Sample t-Test is to check whether the mean of *year* is equal to 2015. According to the result, we can get the conclusion, two.sided, at the significance level of 0.05.

### One-Sample t-Test

One-sample t-test determines if the mean of the population equals the hypothesized value under the normality assumption.

# Descriptive Statistics

Table 1 gives the basic information of the analyzing data set. Observations with missing values are removed when calculating.

Completeness of data.

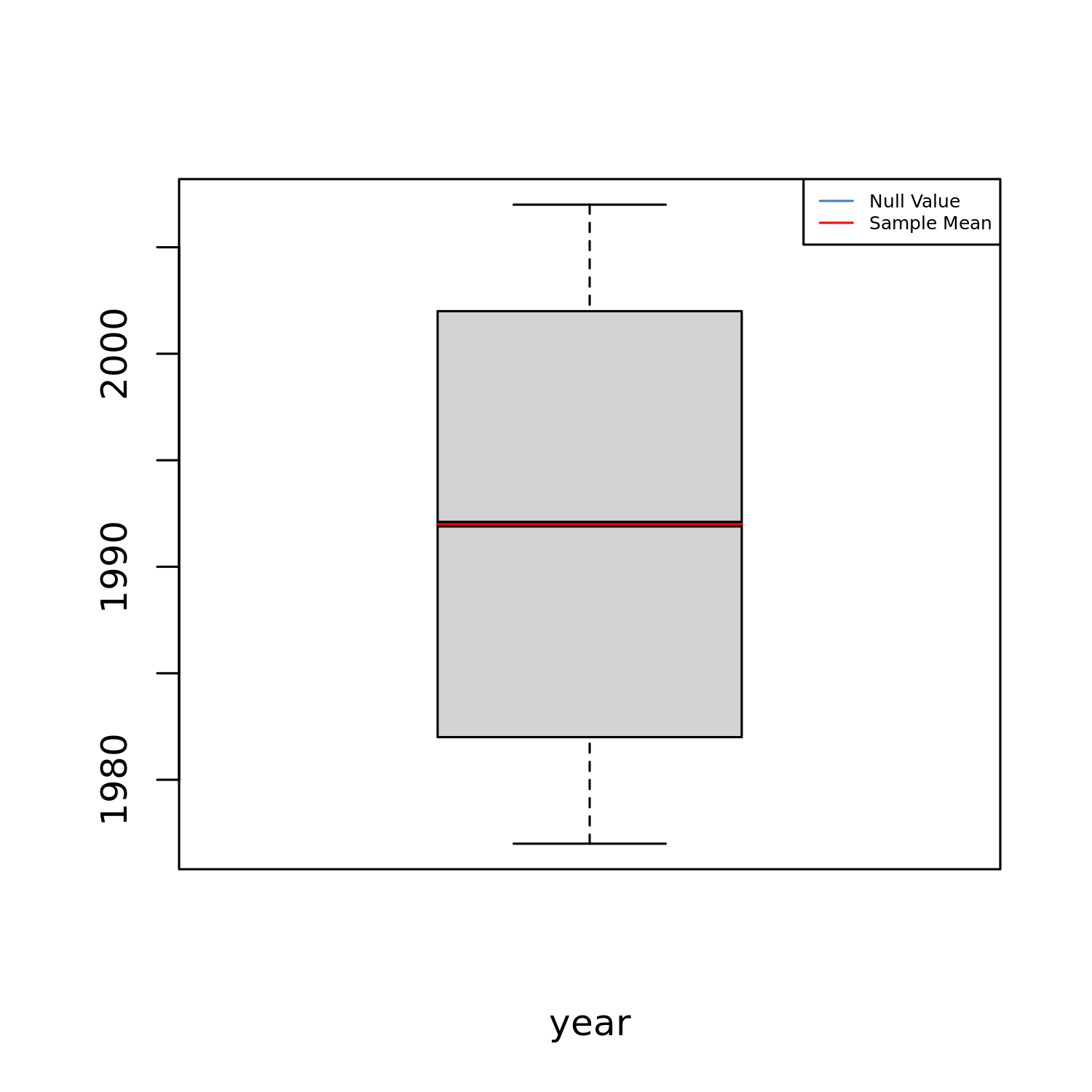
|  |  |  |
| --- | --- | --- |
|  | Observation | Incomplete Observation (not used) |
| Number | 994 | 0 |

Table 2 gives the descriptive statistics.

Descriptive statistics.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | S.D. |
| year | 1977 | 1982 | 1992 | 1992 | 2002 | 2007 | 10.005 |

The box plot for *year* is given by



Box plot for distribution of year.

The box shows the shape of the distribution and its central value. In a box, the ends of the box are the upper and lower quartiles. The median is the value separating the higher half from the lower half of the dataset. 1st quartile is the middle number between the smallest number and the median of the dataset. 3rd quartile is the middle number between the median and the largest number of the dataset. An outlier is a data point that differs significantly from other observations.

# Results

A t-test with an option “two.sided” is employed (Birnbaum, Tingey, and others 1951; Durbin 1973). The following table gives the results of the t-test, t statistic, degree of freedom, and p-value.

Results of the t-test.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Value | Alternative Hypothesis | t-Statistic | D.F. | p-value |
| 2015 | two.sided | -72.5 | 993 | 0 |

It is the result of the t-test. The most important is the *p*-value. If it is less than a given threshold (commonly used significance level is 0.05), the null hypothesis is rejected. In other words, the alternative hypothesis can be accepted. If it is greater than the threshold, the null hypothesis cannot be rejected. The null hypothesis can be accepted.

The test p-value is less than or equal to 0.05, indicating that the difference between the mean and the null value is statistically significant.

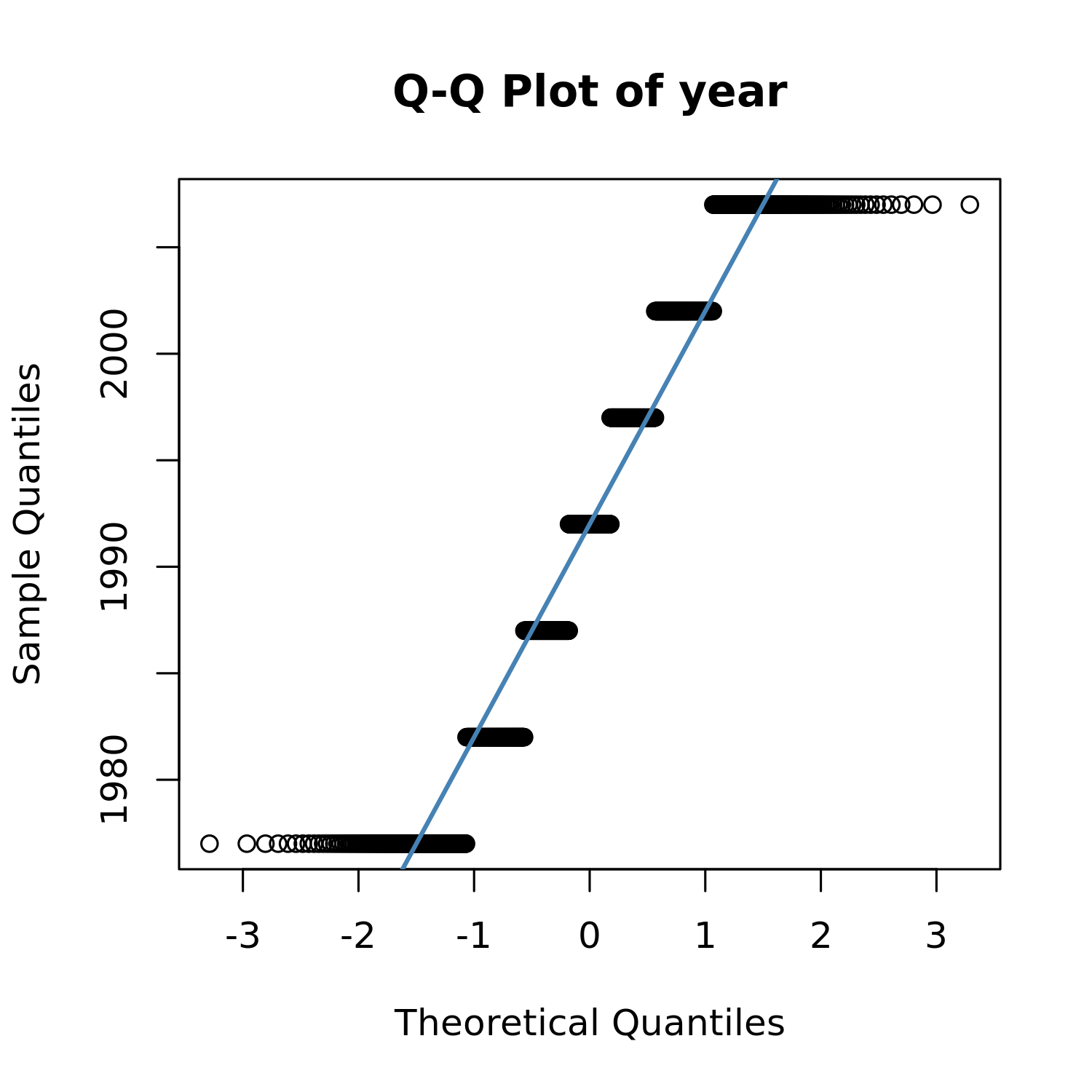
The mean estimation is given along with its 95% confidence interval.

Mean and 95% confidence interval.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Lower | Upper |
|  | 1990 | 1990 | 1990 |

## Assumption Checking

In this section, we will check the normal distribution assumptions of the model. The Q-Q plot is given below:



Normal Q-Q plot.

QQ plot is a graphical method for comparing two probability distributions by plotting their quantiles against each other. One is the estimated distribution of the observations and the other is the normal distribution. If the points are located around the diagonal line, then the population where the sample comes from is normally distributed.

If the observations are located around the diagonal line, then year is normally distributed. Tests are performed to check the normality distribution assumption. The result is given by

Results of the normality test.

|  |  |  |
| --- | --- | --- |
|  | Statistics | p-values |
| Kolmogorov-Smirnov test | 0.127 | 0 |
| Shapiro-Wilk test | 0.918 | 0 |

An important assumption of the t-test is that the population(s) being compared should follow a normal distribution(s). In this section, a Q-Q plot and two normality tests are employed to check this assumption.

The normality assumption doesn’t hold. We suggest making some data transformation to meet the requirements of the t-Test or use the following result of the Wilcoxon Signed-Rank test (a non-parametric analog of t-test) for further analysis.

Results of the Wilcoxon Signed-Rank test.

|  |  |  |  |
| --- | --- | --- | --- |
| Null Value (Location) | Alternative Hypothesis | V | p-value |
| 2015 | two.sided | 0 | 0 |

It is the result of the Wilcoxon Signed-Rank Test. The most important is the *p*-value. If it is less than a given threshold (commonly used significance level is 0.05), the null hypothesis is rejected. In other words, we can accept the alternative hypothesis.

# Conclusions

Based on the above results, we can get the following conclusions:

* Based on the result of the t-test, there is a statistically significant difference between the mean and the given value of 2015.
* The normal distribution assumption of the data doesn’t hold. We suggest making some data transformation before running the t-test or use the following result of the Wilcoxon Signed-Rank test.
  + The result of the Wilcoxon Signed-Rank test shows that there is a statistically significant difference between the median and the given value of 2015.

# Terminology

***confidence interval:*** In statistics, a confidence interval (CI) is a type of interval estimate, computed from the statistics of the observed data, that might contain the true value of an unknown population parameter.

***p-value:*** The p-value is, for a given statistical model, the probability that, when the null hypothesis is true, the statistical summary would be greater than or equal to the actual observed results.

***standard deviation:*** In statistics, the standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

***sample mean:*** In statistics, the sample mean is defined as the average of n observations from the sample.

***t-test statistic(t):*** A statistic, which constructed by the sample, follows a known distribution asymptotically under the null hypothesis. Hence, it can be used to test the hypothesis and define a *p*-value.

***normality:*** Normality means that the distribution of the test is bell-shaped.

***degrees of freedom:*** In statistics, the number of degrees of freedom is the number of values in the final calculation of a statistic that are free to vary.

# References

Birnbaum, ZW, Fred H Tingey, and others. 1951. “One-Sided Confidence Contours for Probability Distribution Functions.” *The Annals of Mathematical Statistics* 22 (4): 592–96.

Durbin, James. 1973. *Distribution Theory for Tests Based on the Sample Distribution Function*. Vol. 9. Siam.