Distribution Test

11/14/2021 20:26

# Abstract

Outcome variable: *educ*  
Null distribution: *norm*

A distribution test is employed to check if the distribution of data is Normal. According to the result, we can reject the null distribution at the significance level of 0.05.

### Distribution Test

Distribution Test determines whether your data follows a hypothesized probability distribution.

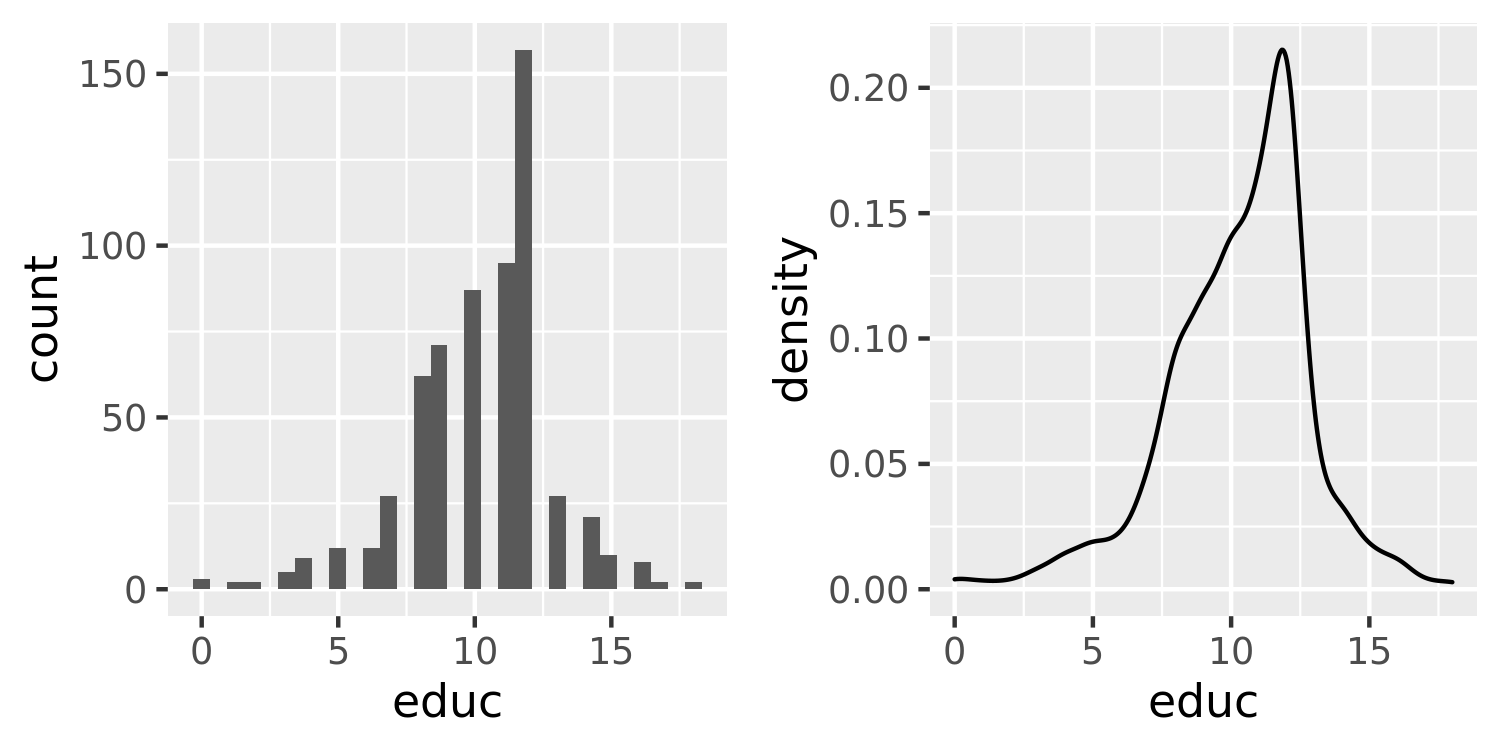
# Descriptive Statistics

Table 1 gives the basic information of analyzing data set. Observations with missing values are removed when calculating. It shows that there is no missing value in the data.

Completeness of Data.

|  |  |  |
| --- | --- | --- |
|  | Observation | Incomplete Observation (not used) |
| Number | 614 | 0 |

The histogram and kernel density estimate are given in Figure 1.



Histogram and kernel density estimate for *educ*.

Both histogram and smoothed density estimates are included in this graph. The smoothed density estimates are a useful alternative to the histogram for continuous data that comes from an underlying smooth distribution.

Table 2 gives the descriptive statistics.

Descriptive Statistics.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | S.D. |
| educ | 0 | 9 | 11 | 10.269 | 12 | 18 | 2.628 |

# Results

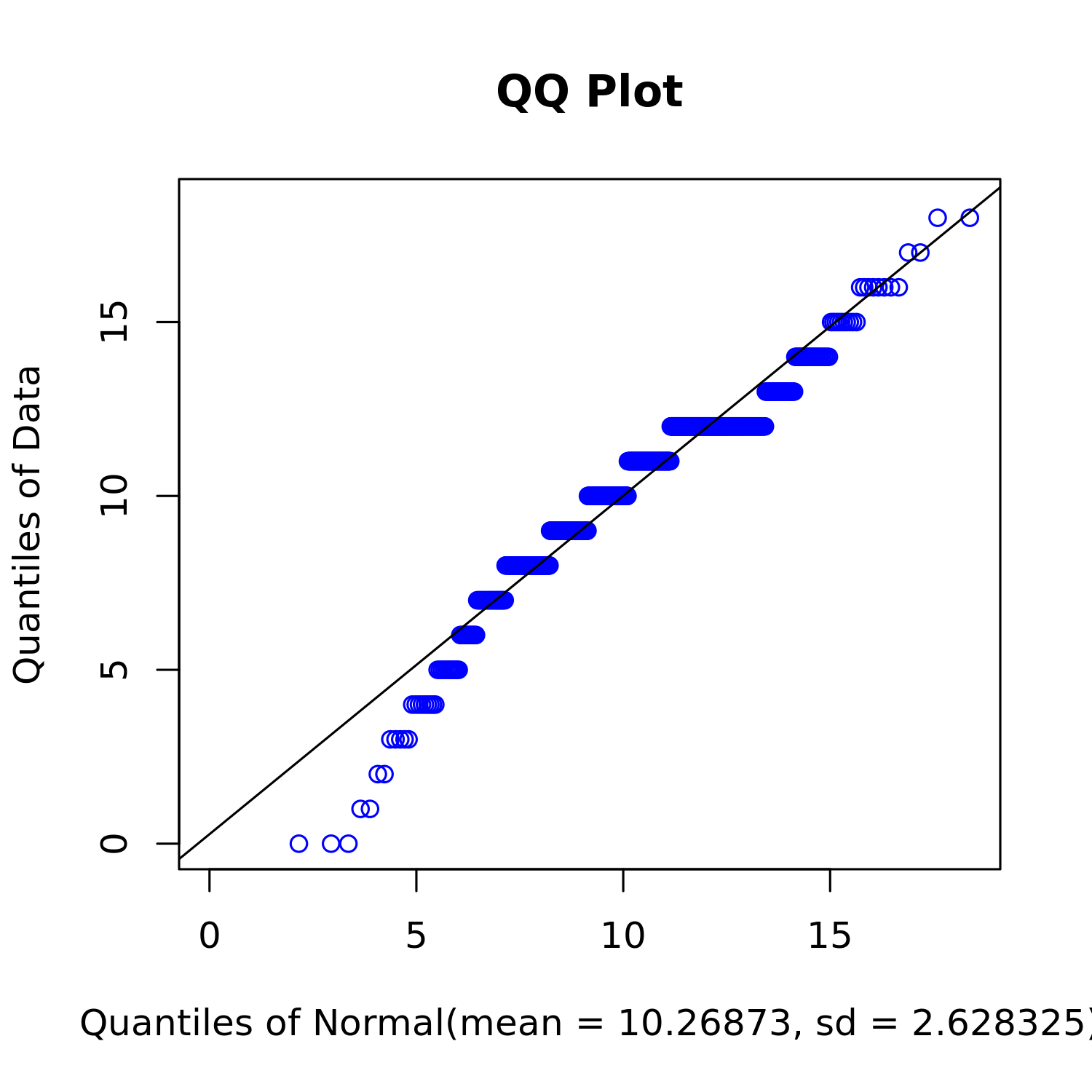
A Kolmogorov-Smirnov Goodness-of-Fit test is performed to determine whether the data set appears to come from Normal (Massey Jr 1951). Table 3 shows the results.

Result of Distribution Test.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hypothesized Distribution | Mean | S.D. | Test.Statistic | p.value |
| Normal | 10.3 | 2.63 | 0.141 | 0 |

Here, the p-value equals 4.9251603^{-11}, which means the data distribution is different from the hypothesized distribution at the significance level of 0.05.

Figure 2 gives the Q-Q plot.



Q-Q plot for *educ*.

Q-Q (quantile-quantile) plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. If the two distributions being compared are similar, the points in the Q-Q plot will approximately lie on the line *y* = *x*. If the distributions are linearly related, the points in the Q-Q plot will approximately lie on a line, but not necessarily on the line *y* = *x*.

After checking several commonly used distributions, we suggest using Nonparametric for further analysis at the significance level of 0.05.

# Conclusions

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

* The data distribution is different from Normal at the significance level of 0.05.
* We suggest using Nonparametric for further analysis.

# Terminologies

***Null hypothesis***: In inferential statistics, the null hypothesis is a general statement or default position that there is no relationship between two measured phenomena or no association among groups.

***p-value***: In statistical hypothesis testing, 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.

# References

Massey Jr, Frank J. 1951. “The Kolmogorov-Smirnov Test for Goodness of Fit.” *Journal of the American Statistical Association* 46 (253): 68–78.