

How do you perform the Goldfeld-Quandt Test in R?

Authored by
stats writer

April 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you perform the Goldfeld-Quandt Test in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138407>

The Goldfeld-Quandt Test is a statistical test used to determine if the variance of two groups is equal or not. In order to perform this test in R, one must first load the "car" package and then use the "gqtest" function, specifying the two groups to be compared. The test will then provide a p-value, which can be used to determine if the variance of the two groups is significantly different or not. This test is commonly used in regression analysis to assess the homoscedasticity assumption.

Perform the Goldfeld-Quandt Test in R

The Goldfeld-Quandt test is used to determine if is present in a regression model.

Heteroscedasticity refers to the unequal scatter of at different levels of a in a regression model.

If heteroscedasticity is present, this violates one of the key that the residuals are equally scattered at each level of the response variable.

This tutorial provides a step-by-step example of how to perform the Goldfeld-Quandt test in R to determine whether or not heteroscedasticity is present in a given regression model.

Step 1: Build a Regression Model

First, we'll build a using the built-in mtcars dataset in R:

#fit a regression model

```
model <- lm(mpg~disp+hp, data=mtcars)
```

```
#view model summary
```

```
summary(model)
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 30.735904 1.331566 23.083 < 2e-16 ***
```

```
disp -0.030346 0.007405 -4.098 0.000306 ***
```

```
hp -0.024840 0.013385 -1.856 0.073679 .
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.127 on 29 degrees of freedom
```

```
Multiple R-squared: 0.7482, Adjusted R-squared: 0.7309
```

```
F-statistic: 43.09 on 2 and 29 DF, p-value: 2.062e-09
```

Step 2: Perform the Goldfeld-Quandt test

Next, we will use the `gqtest()` function from the `lmtest` package to perform the Goldfeld-Quandt test to determine if heteroscedasticity is present.

This function uses the following syntax:

```
gqtest(model, order.by, data, fraction)
```

where:

model: The linear regression model created by the `lm()` command.
order.by: The predictor variable(s) in the model.
data: The name of the dataset.
fraction*: The number of central observations to remove from the dataset.

*The Goldfeld-Quandt test works by removing some number of observations located in the center of the dataset, then testing to see if the spread of residuals is different from the resulting two datasets that are on either side of the central observations.

Typically we choose to remove around 20% of the total observations. In this case, `mtcars` has 32 total observations so we can choose to remove the central 7 observations:

```
#load lmtest library
```

```
library(lmtest)
```

```
#perform the Goldfeld Quandt test
```

```
gqtest(model, order.by = ~disp+hp, data = mtcars,
```

fraction = 7)

Goldfeld-Quandt test

data: model

GQ = 1.0316, df1 = 10, df2 = 9, p-value = 0.486

alternative hypothesis: variance increases from segment 1 to 2

Here is how to interpret the output:

The test statistic is 1.0316. The corresponding p-value is 0.486.

The Goldfeld-Quandt test uses the following null and alternative hypotheses:

Null (H0): Homoscedasticity is present. Alternative (HA): Heteroscedasticity is present.

Since the p-value is not less than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that heteroscedasticity is present in the regression model.

What To Do Next

If you fail to reject the null hypothesis of the Goldfeld-Quandt test then heteroscedasticity is not present and you can proceed to interpret the output of the original regression.

However, if you reject the null hypothesis, this means heteroscedasticity is present in the data. In this case, the standard errors that are shown in the output table of the regression may be unreliable.

There are a couple common ways that you can fix this issue, including:

1. Transform the response variable.

You can try performing a transformation on the response variable, such as taking of the response variable. Typically this can cause heteroscedasticity to go away.

2. Use weighted regression.

Weighted regression assigns a weight to each data point based on the variance of its fitted value.

Essentially, this gives small weights to data points that have higher variances, which shrinks their squared residuals.

When the proper weights are used, weighted regression can eliminate the problem of heteroscedasticity.

ARABPSYCHOLOGY.COM