# How can robust regression be used to improve the accuracy of regression analysis results?

Authored by

**stats writer**

June 29, 2024

Robust regression is a statistical technique that can be used to improve the accuracy of regression analysis results. It is particularly useful in situations where the data being analyzed may contain outliers or other anomalies that can significantly affect the results of traditional regression methods. By using robust regression, these outliers are given less weight, allowing for a more accurate estimation of the relationship between variables. Additionally, robust regression is less sensitive to violations of assumptions, such as non-normality or heteroscedasticity, which can also impact the accuracy of traditional regression models. Overall, robust regression can help to produce more reliable and robust results, making it a valuable tool for improving the accuracy of regression analysis.

## Robust Regression | Stata Annotated Output

## Ordinary least squares (OLS) regression is an extremely useful, easily interpretable statistical method. However, it is not perfect. When running an OLS regression, you want to be aware of its sensitivity to outliers. By "sensitivity to outliers", we mean that an OLS regression model can at times be highly affected by a few records in the dataset and can then yield results that do not accurately reflect the relationship between the outcome variable and the predictor variables seen in the rest of the records. Robust regression offers an alternative to OLS regression that is less sensitive to outliers and still defines a linear relationship between

the outcome and the
predictors. Note that robust regression does not address leverage.

This page shows an example of robust
regression analysis in Stata with footnotes explaining the output. We
will use the crime data set. This dataset appears in Statistical
Methods for Social Sciences, Third Edition by Alan Agresti and Barbara
Finlay (Prentice Hall, 1997). The variables are state id (sid), state
name (state), violent crimes per 100,000 people (crime), murders
per 1,000,000 people(murder), the percent of the population living in
metropolitan areas (pctmetro), the percent of the population that is
white (pctwhite), percent of population with a high school education or
above (pcths), percent of population living under poverty line (poverty),
and percent of population that are single parents

(single). We will drop
the observation for Washington, D.C. (sid=51) because it is not a state.

```
use
https://stats.idre.ucla.edu/stat/stata/webbooks/reg/crime, clear
drop if sid == 51
```

To determine if a robust regression model would be appropriate, OLS
regression is a good starting point. After running the regression, postestimation
graphing techniques and an examination of the model residuals can be implemented to determine if there are any points in
the data that might influence the regression results disproportionately.
The commands for an OLS regression, predicting crime with poverty and
single, and a postestimation graph appear below. Details for
interpreting this graph and other methods for detecting high influence points can

be found in the Robust Regression Data
Analysis Example. We will be interested in the residuals from this
regression when looking at our robust regression, so we have added a predict
command and generated a variable containing the absolute value of the OLS
residuals.

```
regress crime poverty single

lvr2plot, mlabel(state)

predict r1, rstandard
gen absr1 = abs(r1)
```

The same model can be run as a robust regression. Robust
regression works by first fitting the OLS regression model from above and
identifying the records that have a Cook's distance greater than 1. Then,
a regression is run in which those records with Cook's distance greater than 1

are given zero weight. From this model, weights are assigned to records according to the absolute difference between the predicted and actual values (the absolute residual). The records with small absolute residuals are weighted more heavily than the records with large absolute residuals. Then, another regression is run using these newly assigned weights, and then new weights are generated from this regression. This process of regressing and reweighting is iterated until the differences in weights before and after a regression is sufficiently close to zero. For a detailed illustration of this process, see Chapter Six of Regression with Graphics.

The Stata command for robust regression is rreg. The model portion of the command is identical to an OLS regression: outcome variable followed by predictors. We have added gen(weight) to the command so that we will be

able to examine the final weights used in the model.

rreg crime poverty single, gen(weight)

Huber iteration 1: maximum difference in weights = .66846346
Huber iteration 2: maximum difference in weights = .11288069
Huber iteration 3: maximum difference in weights = .01810715
Biweight iteration 4: maximum difference in weights = .29167992
Biweight iteration 5: maximum difference in weights = .10354281
Biweight iteration 6: maximum difference in weights = .01421094
Biweight iteration 7: maximum difference in weights = .0033545

Robust regression Number of obs = 50
F( 2, 47) = 31.15
Prob > F = 0.0000

---

crime | Coef. Std. Err. t P>|t|

**----------+----------**

**poverty | 10.36971 7.629288 1.36 0.181 -4.978432 25.71786**

**single | 142.6339 22.17042 6.43 0.000 98.03276 187.235**

**_cons | -1160.931 224.2564 -5.18 0.000 -1612.076 -709.7849**

**----------**

**We can see that large residuals correspond to low weights in robust**

**regression.**

**sort weight**

**li sid state weight absr1 in 1/10**

**+----------+**

**| sid state weight absr1 |**

**|----------|**

**1. | 25 ms .02638862 3.158753 |**

**2. | 9 fl .11772218 3.023632 |**

**3. | 46 vt .59144513 1.831356 |**

**4. | 26 mt .66441582 1.588843 |**

**5. | 20 md .67960728 1.62075 |**

**|----------------------------------|**

**6. | 14 il .69124917 1.550569 |**

**7. | 21 me .69766511 1.578434 |**

**8. | 31 nj .74574796 1.193654 |**

**9. | 19 ma .75392127 1.288611 |**

**10. | 5 ca .80179038 1.401128 |**

**+----------------------------------+**

**Here we can see that, generally, small weights are given to cases with large**
**absolute residuals.**

### Stata Output

**Huber iterationa1: maximum difference in weights = .66846346**

**Huber iteration 2: maximum difference in weights = .11288069**

**Huber iteration 3: maximum difference in weights = .01810715**

**Biweight iterationb4: maximum difference in weights = .29167992**

**Biweight iteration 5: maximum difference in weights = .10354281**

**Biweight iteration 6: maximum difference in weights =**

.01421094
Biweight iteration 7: maximum difference in weights = .0033545

Robust regression Number of obsc = 50
F( 2, 47)d = 31.15
Prob > Fe = 0.0000

---

crime | Coef.f Std. Err.g th P>|t|i j
---+---

poverty | 10.36971 7.629288 1.36 0.181 -4.978432 25.71786
single | 142.6339 22.17042 6.43 0.000 98.03276 187.235
_cons | -1160.931 224.2564 -5.18 0.000 -1612.076 -709.7849

---

a. Huber iteration - These are iterations in which Huber weightings
are implemented. In Huber weighting, the larger the residual, the smaller the
weight. These weights are used until they are nearly unchanged from iteration to

iteration. In this example, three iterations were necessary for the model to converge using Huber weights. The converged model is then weighted using biweights (see superscript b). Both weighting methods are used because both have problems when used alone: Huber weights can work poorly with extreme outliers and biweights do not always converge.

b. Biweight iteration - These are iterations in which biweights are implemented. To see the precise functions that define biweights and Huber weights, consult the Stata manual. Biweight iterations continue until the biweights are nearly unchanged from iteration to iteration. In this example, four iterations were required for convergence. The model to which the biweight iterations converge is considered the final model.

c. Number of obs - This is the number of observations in our dataset.

**Our dataset started with 51 cases, and we dropped the record corresponding to
Washington, D.C., leaving us with 50 cases in our analysis.**

**d. F(2, 47) - This is the model F-statistic. It is the test statistic
used in evaluating the null hypothesis that all of the model coefficients are
equal to zero. Under the null hypothesis, our predictors have no linear
relationship to the outcome variable. The numbers in parenthesis are
degrees of freedom. The model degrees of freedom is equal to the number of predictors and the error degrees of freedom
is calculated as (number of observations - (number of predictors+1)). This statistic follows an F
distribution with df1 = 2, df2 = 47.**

**e.
Prob > F - This is the probability of getting an F statistic test
statistic as extreme as, or more so, than the observed**

**statistic under the null hypothesis; the null hypothesis is that all of the regression coefficients are simultaneously equal to zero. In other words, this is the probability of obtaining this F statistic (31.15) or one more extreme if there is in fact no effect of the predictor variables. This p-value is compared to a specified alpha level, our willingness to accept a type I error, which is typically set at 0.05 or 0.01. The small p-value, <0.0001, would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero.**

**f. Coef. - These are the values for the regression equation for predicting the dependent variable from the independent variable. The regression equation is presented in many different ways, for example:**

**Y(predicted) = b0 + b1*x1 + b2*x2.**

**The column of estimates provides the values for b0, b1 and b2 for this equation. Expressed in terms of the variables used in this example, the regression equation is**

**crime(predicted) = -1160.931 + 10.36971*poverty + 142.6339*single.**

**These estimates tell you about the relationship between the predictor variables and the outcome variable. These estimates indicate the amount of increase in crime that would be predicted by a 1 unit increase in the predictor variable.**

**poverty - The coefficient for poverty is 10.36971. For every unit increase in poverty, a 10.36971 unit increase in crime is predicted, holding all other variables constant.**

**single - The coefficient for single is 142.6339. For every unit increase in single, a 142.6339 unit**

**increase in crime
is predicted, holding all other variables constant.**

**g. Std. Err. - These are the standard errors associated with the
coefficients. The standard error is used for testing whether the parameter is
significantly different from 0 by dividing the parameter estimate by the
standard error to obtain a t-value (see superscripts h and i).
The standard errors can also be used to form a confidence interval for the
parameter, as shown in the last two columns of this table.**

**h. t - The test statistic t is the ratio of the Coef. to
the Std. Err. of the respective predictor. The t value follows a t-distribution
which is used to test against a two-sided alternative hypothesis that the
Coef. is not equal to zero.**

**poverty - The t test statistic for the
predictor poverty is (10.36971 / 7.629288) = 1.36 with an**

associated
p-value of 0.181. If we set our alpha level to 0.05, we would fail to reject the
null hypothesis and conclude that the regression coefficient for poverty
has not been found to be statistically different from zero given that single
is in the model.

single -The t test statistic for the predictor single
is (142.6339 / 22.17042) = 6.43 with an associated p-value of < 0.001. If we set
our alpha level to 0.05, we would reject the null hypothesis and conclude that
the regression coefficient for single has been found to be statistically
different from zero given that poverty is in the model.

_cons - The t test statistic for the intercept, _cons,
is (-1160.931 / 224.2564) = -5.18 with an associated p-value of < 0.001. If we
set our alpha level at 0.05, we would reject the null hypothesis and conclude
that _cons has been found to be statistically different

**from zero given
poverty and single are in the model and evaluated at zero.**

**i. P>|t| - This is the probability the t test statistic (or a more extreme test statistic) would be observed under the null hypothesis that
a particular predictor's regression coefficient is zero, given that the rest of
the predictors are in the model. For a given alpha level, P>|t|
determines whether of not the null hypothesis can be rejected. If P>|t|
is less than alpha, then the null hypothesis can be rejected and the parameter
estimate is considered to be statistically significant at that alpha level.**

**j.
- This is the Confidence Interval (CI) for an individual coefficient given that the other predictors are in the model. For a
given predictor with a level of 95% confidence, we'd say that we are 95%**

**confident that the "true" coefficient lies between the lower and upper limit of
the interval. It is calculated as the Coef. (zα/2)*(Std.Err.), where zα/2 is a critical value on the standard normal distribution. The CI is equivalent to the t test statistic: if the CI includes zero,
we'd fail to reject the null hypothesis that a particular regression coefficient
is zero given the other predictors are in the model. An advantage of a CI is
that it is illustrative; it provides a range where the "true" parameter may
lie.**