

# How to Perform Hierarchical Regression in Stata: A Step-by-Step Guide

Authored by  
**stats writer**

February 26, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Perform Hierarchical Regression in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=132782>

Hierarchical regression is a statistical technique used to examine the relationship between multiple independent variables and a single dependent variable. In Stata, this can be performed by using the "regress" command and specifying the order of entry for the independent variables. This can be done in two ways: stepwise regression, where variables are added or removed based on their significance, or block regression, where variables are added in predetermined blocks. The results of hierarchical regression can be interpreted through the regression coefficients, R-squared, and other diagnostic statistics. This method allows for the exploration of the unique contributions of each independent variable to the dependent variable, while controlling for the effects of other variables.

## **Perform Hierarchical Regression in Stata**

**Hierarchical regression is a technique we can use to compare several different linear models.**

**The basic idea is that we first fit a linear regression model with just one explanatory variable. Then we fit another regression model using an additional explanatory variable. If the R-squared (the proportion of variance in the response variable that can be explained by the explanatory variables) in the second model is significantly higher than the R-squared in the previous model, this means the second model is better.**

**We then repeat the process of fitting additional regression models with more explanatory variables and seeing if the newer models offer any improvement over the previous models.**

**This tutorial provides an example of how to perform hierarchical regression in Stata.**

**Example: Hierarchical Regression in Stata**

**We'll use a built-in dataset called *auto* to illustrate how to perform hierarchical regression in Stata. First, load the dataset by typing the following into the Command box:**

```
sysuse auto
```

**We can get a quick summary of the data by using the following command:**

```
summarize
```

```
. sysuse auto
(1978 Automobile Data)
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

**We can see that the dataset contains information about 12 different variables for 74 total cars.**

**We will fit the following three linear regression models and use hierarchical regression to see if each subsequent model provides a significant improvement to the previous model or not:**

**Model 1: price = intercept + mpg**

**Model 2: price = intercept + mpg + weight**

**Model 3: price = intercept + mpg + weight + gear ratio**

**In order to perform hierarchical regression in Stata, we**

will first need to install the hireg package. To do so, type the following into the Command box:

## findit hireg



```
search hireg, all x
+
search for hireg (manual: [R] search)

Search of official help files, FAQs, Examples, and Stata Journals
Search of web resources from Stata and other users
(contacting http://www.stata.com)
1 package found (Stata Journal and STB listed first)
-----
hireg from http://fmwww.bc.edu/RePEc/bocode/h
'HIREG': module for hierarchial regression / The hireg command conducts
hierarchical regressions. Users / enter blocks of independent variables
which are added to the / model in successive steps. R-squared change is
reported at each / step along with a summary table at the end. All options
(click here to return to the previous screen)
(end of search)
```

In the next window, click the link that says click here to install.

net describe hireg, from(http://f... x

+

Dialog ▾ Also see ▾ Jump to ▾

---

package **hireg** from <http://fmwww.bc.edu/RePEc/bocode/h>

---

**TITLE**  
'HIREG': module for hierarchial regression

**DESCRIPTION/AUTHOR(S)**

The hireg command conducts hierarchical regressions. Users enter blocks of independent variables which are added to the model in successive steps. R-squared change is reported at each step along with a summary table at the end. All options available on the regress command may also be used with hireg.

KW: regression  
KW: hierarchical regression

Requires: Stata version 8.0

Author: Paul H. Bern, Princeton University  
Support: email [pbern@princeton.edu](mailto:pbern@princeton.edu)

Distribution-Date: 20050723

**INSTALLATION FILES** [\(click here to install\)](#)  
hireg.ado  
[hireg.hlp](#)

---

[\(click here to return to the previous screen\)](#)

**The package will install in a matter of seconds. Next, to perform hierarchical regression we will use the following command:**

**hireg price (mpg) (weight) (gear\_ratio)**

**Here is what this tells Stata to do:**

**Perform hierarchical regression using price as the response variable in each model. For the first model,**

use mpg as the explanatory variable. For the second model, add in weight as an additional explanatory variable. For the third model, add in gear\_ratio as another explanatory variable.

Here is the output of the first model:

**Model 1:**  
Variables in Model:  
Adding : mpg

Source	SS	df	MS	Number of obs	=	74
Model	139449474	1	139449474	F(1, 72)	=	20.26
Residual	495615923	72	6883554.48	Prob > F	=	0.0000
Total	635065396	73	8699525.97	R-squared	=	0.2196
				Adj R-squared	=	0.2087
				Root MSE	=	2623.7

  

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mpg	-238.8943	53.07669	-4.50	0.000	-344.7008 -133.0879
_cons	11253.06	1170.813	9.61	0.000	8919.088 13587.03

We see that the R-squared of the model is 0.2196 and the overall p-value (Prob > F) for the model is 0.0000, which is statistically significant at  $\alpha = 0.05$ .

Next, we see the output of the second model:

**Model 2:**

Variables in Model: mpg  
 Adding : weight

Source	SS	df	MS	Number of obs	=	74
Model	<b>186321280</b>	<b>2</b>	<b>93160639.9</b>	F(2, 71)	=	<b>14.74</b>
Residual	<b>448744116</b>	<b>71</b>	<b>6320339.67</b>	Prob > F	=	<b>0.0000</b>
Total	<b>635065396</b>	<b>73</b>	<b>8699525.97</b>	R-squared	=	<b>0.2934</b>
				Adj R-squared	=	<b>0.2735</b>
				Root MSE	=	<b>2514</b>

  

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mpg	<b>-49.51222</b>	<b>86.15604</b>	<b>-0.57</b>	<b>0.567</b>	<b>-221.3025 122.278</b>
weight	<b>1.746559</b>	<b>.6413538</b>	<b>2.72</b>	<b>0.008</b>	<b>.467736 3.025382</b>
_cons	<b>1946.069</b>	<b>3597.05</b>	<b>0.54</b>	<b>0.590</b>	<b>-5226.245 9118.382</b>

R-Square Diff. Model 2 - Model 1 = **0.074** F(1,71) = **7.416** p = **0.008**

The R-squared of this model is 0.2934, which is larger than the first model. To determine if this difference is statistically significant, Stata performed an F-test which resulted in the following numbers at the bottom of the output:

R-squared difference between the two models = 0.074  
 F-statistic for the difference = 7.416  
 Corresponding p-value of the F-statistic = 0.008

Because the p-value is less than 0.05, we conclude that there is a statistically significant improvement in the second model compared to the first model.

## Lastly, we can see the output of the third model:

### Model 3:

Variables in Model: mpg weight  
Adding : gear\_ratio

Source	SS	df	MS	Number of obs	=	74
Model	200028459	3	66676153.1	F(3, 70)	=	10.73
Residual	435036937	70	6214813.38	Prob > F	=	0.0000
				R-squared	=	0.3150
				Adj R-squared	=	0.2856
Total	635065396	73	8699525.97	Root MSE	=	2493

  

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mpg	-50.60979	85.43696	-0.59	0.556	-221.0084	119.7889
weight	2.390466	.7697099	3.11	0.003	.8553283	3.925604
gear_ratio	1459.319	982.6304	1.49	0.142	-500.4757	3419.113
_cons	-4374.457	5552.978	-0.79	0.433	-15449.52	6700.609

R-Square Diff. Model 3 - Model 2 = 0.022 F(1,70) = 2.206 p = 0.142

The R-squared of this model is 0.3150, which is larger than the second model. To determine if this difference is statistically significant, Stata performed an F-test which resulted in the following numbers at the bottom of the output:

R-squared difference between the two models = 0.022  
F-statistic for the difference = 2.206  
Corresponding p-value of the F-statistic = 0.142

Because the p-value is not less than 0.05, we don't have

**sufficient evidence to say that the third model offers any improvement over the second model.**

**At the very end of the output we can see that Stata provides a summary of the results:**

Model	R2	F(df)	p	R2 change	F(df) change	p
1:	<b>0.220</b>	<b>20.258(1,72)</b>	<b>0.000</b>			
2:	<b>0.293</b>	<b>14.740(2,71)</b>	<b>0.000</b>	<b>0.074</b>	<b>7.416(1,71)</b>	<b>0.008</b>
3:	<b>0.315</b>	<b>10.729(3,70)</b>	<b>0.000</b>	<b>0.022</b>	<b>2.206(1,70)</b>	<b>0.142</b>

**In this particular example, we would conclude that model 2 offered a significant improvement over model 1, but model 3 did not offer a significant improvement over model 2.**