

# What are Clustered Standard Errors?

Authored by  
**stats writer**

April 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *What are Clustered Standard Errors?*. PSYCHOLOGICAL SCALES.  
Retrieved from <https://scales.arabpsychology.com/?p=140238>

Clustered standard errors refer to a statistical method used in regression analysis to account for potential correlation among observations within a dataset. This method recognizes that data points within a dataset may not be independent, and therefore adjusts the standard errors to accurately reflect the true variability in the data. This is particularly important when analyzing clustered or panel data, where observations may be grouped together based on certain characteristics. By taking into account the clustering of data, clustered standard errors provide more accurate estimates of the model's coefficients and can help to avoid biased results. This technique is commonly used in social sciences and economics research, where data often exhibits clustering.

## **What are Clustered Standard Errors? (Definition & Example)**

**Clustered standard errors are used in when some observations in a dataset are naturally "clustered" together or related in some way.**

**To understand when to use clustered standard errors, it helps to take a step back and understand the goal of regression analysis.**

**In statistics, regression models are used to quantify the relationship between one or more predictor variables and a .**

**Whenever you fit a regression model, your output will be displayed in a that looks like the following:**

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	17.175	12.556	1.368	0.205
Predictor Variable 1	6.384	1.087	5.874	0.000
Predictor Variable 2	0.486	0.179	2.709	0.024

**Here's how to interpret the values in the table:**

**Coefficient:** The average increase in the response variable associated with a one unit increase in a specific predictor variable, assuming all other predictor variables are held constant.**Standard Error:** A measure of the precision of the estimate of the coefficient.**t Stat:** The t-statistic for the predictor variable, calculated as  $\text{Coefficient} / \text{Standard Error}$ .**p-value:** The p-value associated with the t-statistic. If this value is less than a certain significance level (e.g. 0.05), we say that there is a statistically significant relationship between the predictor variable and the response variable.

One of the key assumptions of regression analysis is the . This assumptions states that each in the dataset should be independent of every other observation.

In practice, this assumption is sometimes violated.

For example, suppose a researcher wants to fit a

regression model using hours studied as the predictor variable and exam score as the response variable. He decides to collect data for 50 students spread across five different classrooms.

In this scenario, students are naturally clustered together into classrooms, which means the data collected for each student will not be independent.

For example, some classrooms may have an excellent teacher while other classrooms have a sub-par teacher who does a poor job of teaching their subject.

These students are in the same classroom

Student	Hours Studied	Exam Score
Student #1	3	77
Student #2	5	89
Student #3	4	92
Student #4	4	94
Student #5	3	84
Student #6	5	90
Student #7	2	68
Student #8	7	96
Student #9	2	88
Student #10	3	89
Student #11	3	94
Student #12	4	90
...	...	...
Student #50	6	88

If the researcher fits a regression model without

accounting for this clustered nature of the data, the standard errors of the regression coefficients will be smaller than they should be.

This will result in the following errors:

The t-statistics will be too large. The p-values will be too small. The confidence intervals will be too narrow.

Simply put, the results of the regression analysis will not be reliable.

For example, in Stata you can use the `cluster(variable name)` command to tell Stata to use clustered standard errors when fitting a regression model.

In practice, you can use the following syntax to fit a regression model in Stata with clustered standard errors:

```
regress x y, cluster(variable_name)
```

where:

**x:** The predictor variable  
**y:** The response variable  
**variable\_name:** The name of the variable that

**the data should be clustered based on**

**This will return a regression table with clustered standard errors.**

ARABPSYCHOLOGY.COM