

# How is mean centering used to compute variables?

Authored by  
**stats writer**

June 23, 2024

## RECOMMENDED CITATION

stats writer (2024). *How is mean centering used to compute variables?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149238>

Mean centering is a statistical technique used to adjust variables by subtracting the overall mean value from each individual value. This process is often used to standardize data and remove any potential bias from the mean value. By centering the data, it becomes easier to compare and analyze different variables on a common scale. This method is commonly used in statistical analysis, such as regression models, to improve the accuracy and interpretation of results. Mean centering is a simple yet effective approach to compute variables and can provide valuable insights in various fields, including economics, psychology, and social sciences.

## Introduction

**Mean centering** refers to a type of variable transformation wherein the average of a variable is subtracted from every observation of that variable. The result of this transformation is a new variable that retains its original units, but whose mean value is 0. Mean centering is often used when fitting linear regression models because of how it affects the interpretation of the intercept term. With an uncentered predictor variable, the intercept of a linear regression represents the expected value of  $y$  when the  $x$  variable is zero. With a centered predictor variable, the intercept of that linear regression model represents the expected value of  $y$  when  $x$  is at its average.

In this tutorial, we will show how to mean-center a variable in SPSS Statistics by combining the **Aggregate** procedure and the **Compute Variables** procedure. We recommend this approach because it is **reproducible**: the value of the grand mean used for centering is saved in your dataset, making it easier to audit. Additionally, the Aggregate + Compute Variables approach adapts well if your sample (and therefore the mean value you will use to center) changes due to new observations and/or new filtering criteria.

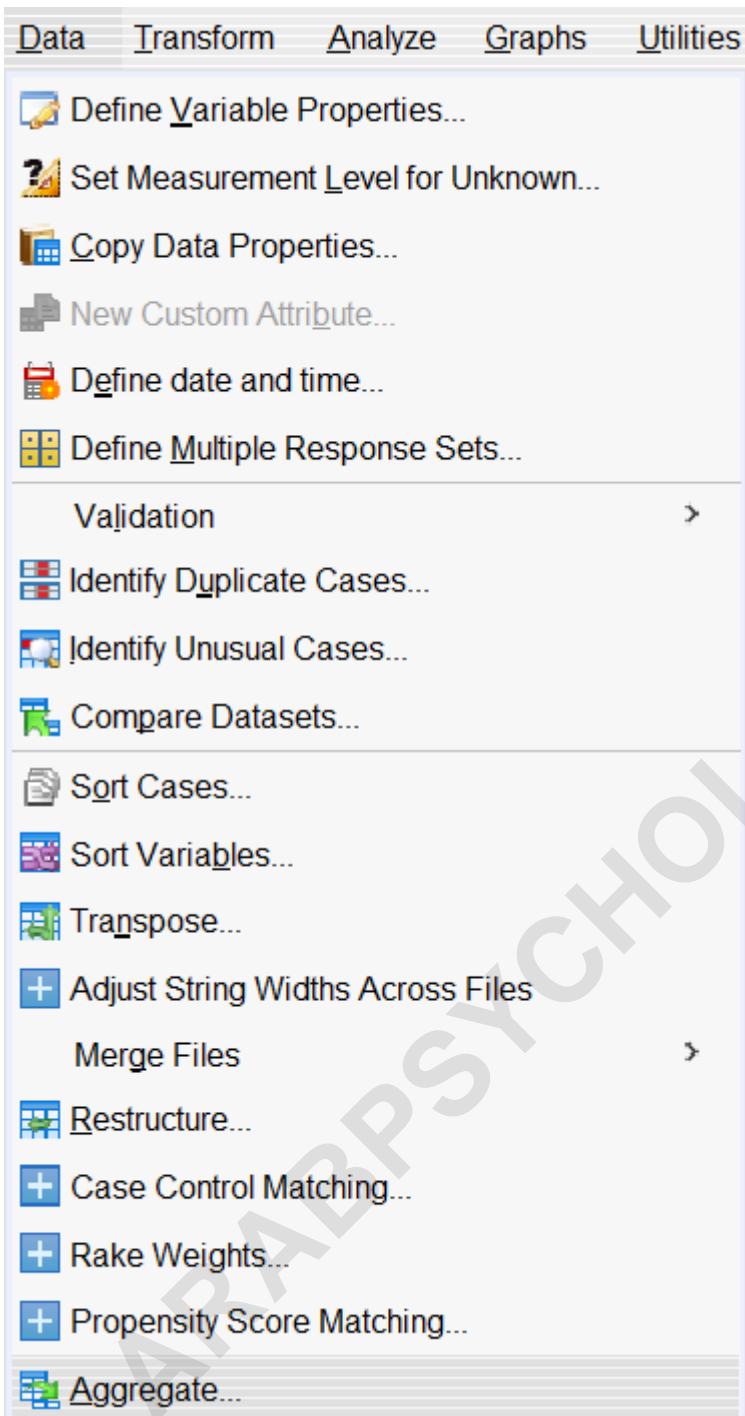
## Aggregating Data

In SPSS, the Aggregate procedure can be used to compute new variables based on summary statistics, or to compute new variables by group. This process can be used to add variables to an existing dataset, or it can be used to create a new dataset containing only the compressed, aggregated information.

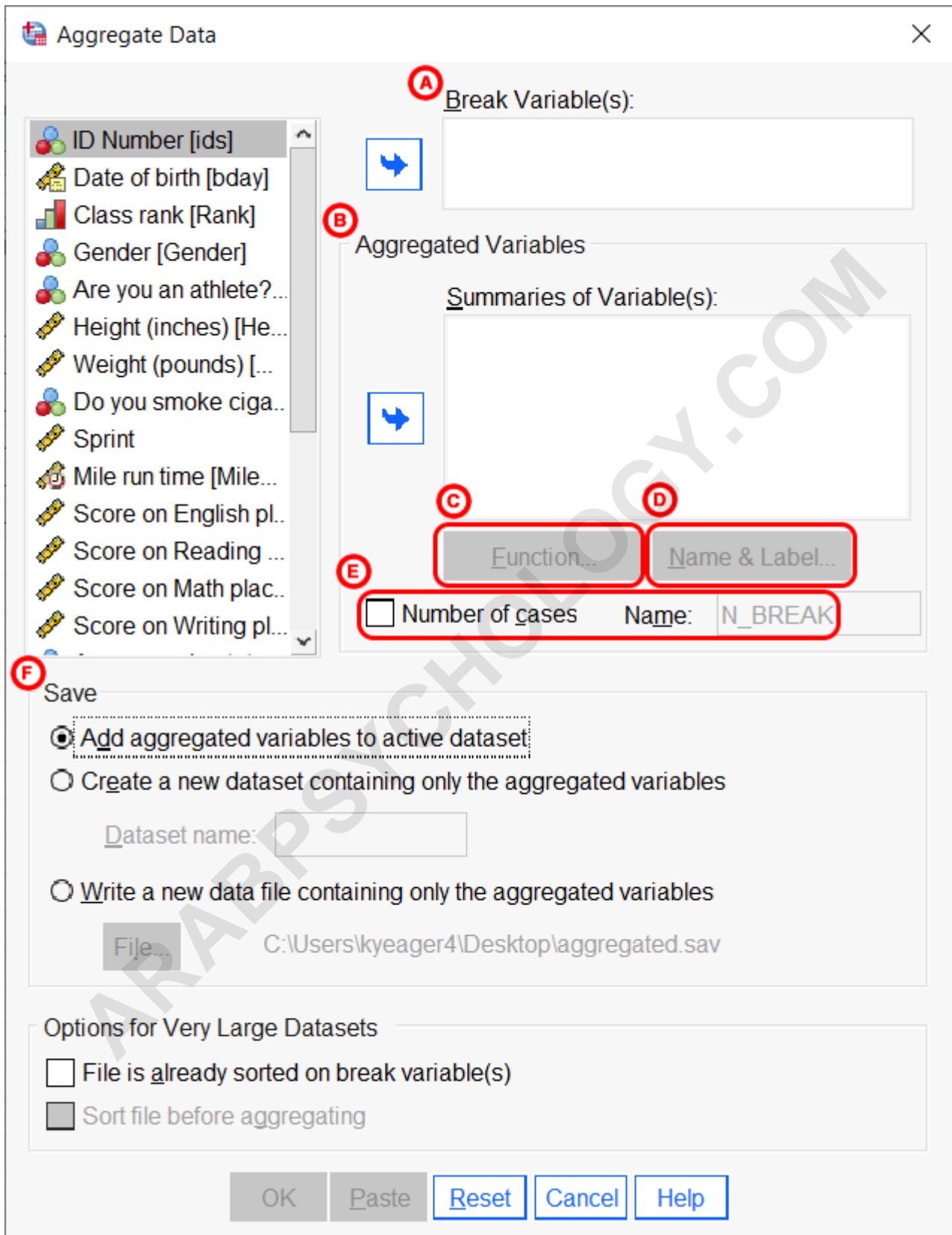
This may look like:

Computing a new variable containing the grand mean of a predictor variable, which can then be used to center the predictor. Extracting the distinct cases in a dataset. Creating a dataset containing summary statistics for groups.

To launch the Aggregate Data procedure, click **Data > Aggregate**.



The Aggregate Data window opens.



**ABreak Variable:** The variable(s) that will determine the grouping structure for the aggregation. The output of the Aggregate procedure will have summary statistics reported for each of the

groups identified by this variable/these variables. These are typically categorical variables, or in the case of longitudinal data, could be ID variables.

**BAggregated Variables/Summaries of Variables:** These represent the variables that will be summarized within each group. These will typically be numeric variables.

**CFunction:** Options to change the function used to compute the new aggregate variables. This button will only be clickable if you have added one or more variables to the "Summaries of Variables" box and have one of those variables selected.

**Aggregate Data: Aggregate Function**

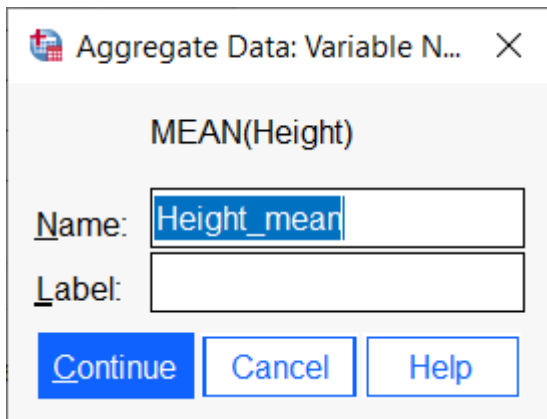
Summary Statistics	Specific Values	Number of cases
<input checked="" type="radio"/> Mean	<input type="radio"/> First	<input type="radio"/> Weighted
<input type="radio"/> Median	<input type="radio"/> Last	<input type="radio"/> Weighted missing
<input type="radio"/> Sum	<input type="radio"/> Minimum	<input type="radio"/> Unweighted
<input type="radio"/> Standard Deviation	<input type="radio"/> Maximum	<input type="radio"/> Unweighted missing

Percentages, Fractions, Counts

Percentages     Fractions     Counts

Above  
 Below    Value:   
 Inside  
 Outside    Low:     High:

**DName & Label:** Options to change the variable names and variable labels for the new variables being created. This button will only be clickable if you have added one or more variables to the "Summaries of Variables" box and have one of those variable formulas selected.



**ENumber of Cases/Name:** Option to create an additional aggregate variable that counts the number of rows for a given combination of break variables. The name entered in the "Name" field will become the name of the new variable.

**FSave:** How to structure the output of the Aggregate procedure. You must select one of the three options:

**Add aggregated variables to active dataset:** The dataset in the active Data View window will have column(s) added to it.  
**Create a new dataset containing only the aggregated variables:** The output dataset will be created in a Data View window in your SPSS instance.

**Dataset name:** A nickname for the data view window to be created. Cannot contain spaces or special characters. The name you apply in this box does not correspond to a file name.  
**Write a new data file containing only the aggregated variables:** The output dataset will be saved as a new SPSS data file (\*.sav format) in a location of your choice. The new dataset will not automatically open in your instance of SPSS Statistics.

## Example: Mean-centering a variable about the grand mean

### Problem Statement

Suppose we plan to fit a linear regression model to predict students' weight based on their height. We wish to center the independent variable, Height, about its mean, in order to improve the interpretability of our regression coefficients.

Our starting dataset should contain at least one continuous numeric variable, and each row should represent an independent observation.

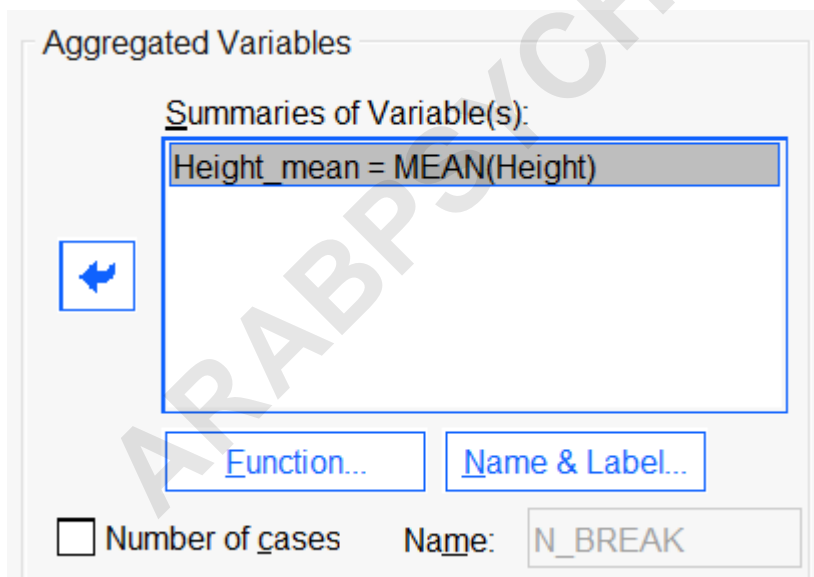
ids	Height
20183	66.92
20230	80.11
20243	65.99
20248	61.32
20255	65.75
20278	70.66
20389	70.68

## Running the Procedure

### Using the Dialog Windows

Click **Data > Aggregate**.

Add the variable you want to center -- in this case, Height -- to the **Aggregated Variables/ Summaries of Variables** box.



While the variable is selected, the Function and Name & Label buttons will be clickable:

Click **Function**. Under Summary Statistics, select the Mean option. Then click **Continue**. Click **Name & Label**. Enter a name and label for the new aggregated variable, then click **Continue**. In this example, we will use the name Mean\_Height for the new variable. In the Save area, make sure **Add aggregated variables to active dataset** is selected. When finished, click **OK**.

This calculates the mean of the variable, and adds it as a separate variable to the data set:

ids	Height	Mean_Height
20183	66.92	68.03
20230	80.11	68.03
20243	65.99	68.03
20248	61.32	68.03
20255	65.75	68.03
20278	70.66	68.03
20389	70.68	68.03

Now that we have a column containing the grand average, we can perform the mean centering by using the Compute Variables procedure:

Click **Transform > Compute Variable**. In the Target Variable box, choose a name for the new centered variable. In this example, we'll call the new variable: `Centered_Height` In the Numeric Expression box, subtract the calculated mean variable from the original variable. The formula should look like:

`Height - Mean_Height` When finished, click **OK**.

After this step, we will finally have our mean-centered variable:

ids	Height	Mean_Height	Centered_Height
20183	66.92	68.03	-1.11
20230	80.11	68.03	12.08
20243	65.99	68.03	-2.04
20248	61.32	68.03	-6.71
20255	65.75	68.03	-2.28
20278	70.66	68.03	2.63
20389	70.68	68.03	2.65

## Using Syntax

```
/*Use the Aggregate procedure to compute the grand mean as a new variable and
add it to the existing dataset.*/
```

```
AGGREGATE
```

```
/OUTFILE=* MODE=ADDVARIABLES
```

```
/Mean_Height=MEAN(Height) .  
  
/*Compute the mean-centered values.*/  
COMPUTE Centered_Height = Height - Mean_Height.  
EXECUTE.  
  
/*Check results.*/  
DESCRIPTIVES Height Centered_Height.
```

## Checking our Work

While not a requirement, we can check the descriptive statistics for the original variable and the mean-centered version of the variable using the **Descriptives** command.

The box below shows the mean, standard deviation, minimum, and maximum for the Centered\_Height variable and the original, uncentered variable Height:

	N	Minimum	Maximum	Mean	Std. Deviation
Height	408	55.00	84.41	68.0318	5.32566
Centered_Height	408	-13.03	16.38	.0000	5.32566
Valid N (listwise)	408				

Notice that the original variable has a mean of 68.03, which matches the Aggregate procedure, and the centered variable has a mean of zero. The number of observations and the standard deviation stay the same.

## Tutorial Feedback