How can we compute variables using rank transforms, also known as rank cases?

Authored by

stats writer

June 23, 2024

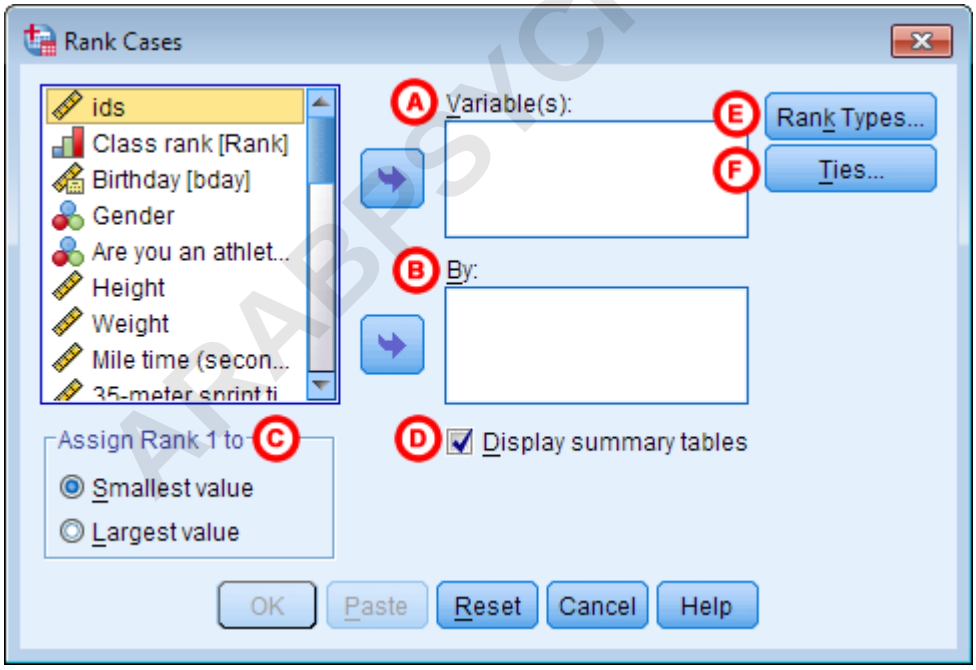
Rank transforms, also known as rank cases, refer to a method of computing variables by assigning ranks to data points based on their relative magnitude. This technique is commonly used in statistical analysis to transform non-normal data distributions into a more normal distribution, allowing for more accurate analysis and comparisons. The process involves ranking the data points from lowest to highest, with ties being assigned the average rank. These ranks are then used as the new values for the data points, allowing for the use of traditional statistical methods. By using rank transforms, we can effectively handle non-normal data and obtain more reliable results in our analysis.

Introduction

A *rank* variable represents the ordering of the values of a numeric variable from smallest to largest (or largest to smallest). Ranking is its own type of variable transformation, and is also useful when you want to convert a numeric variable into a categorical variable using percentiles.

Rank Cases

In SPSS, rank transforms and percentile groupings can be computed using the *Rank Cases* procedure. To open *Rank Cases*, click **Transform > Rank Cases**.



A**Variables:** The variables to compute rank transforms on. The new ranks will be saved to new variables (whose names will be automatically generated).

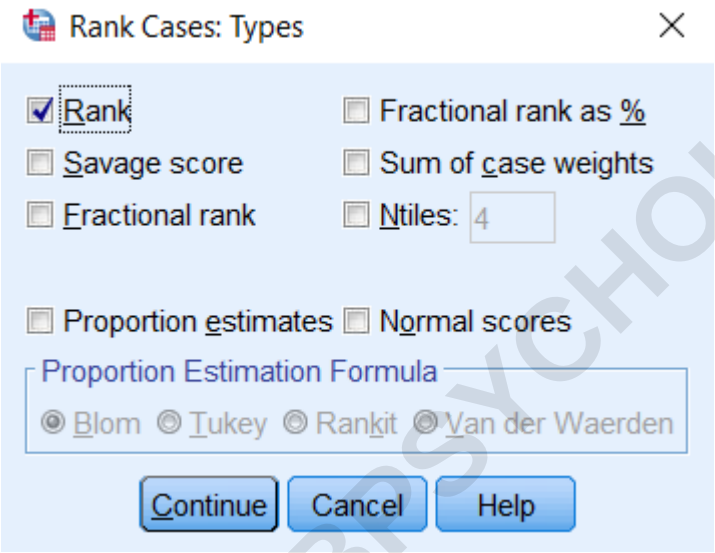
B**By:** (Optional) Assign ranks within groups. By variables should be nominal or ordinal, and have a

small number of categories.

C**Assign Rank 1 to:** Should ranks be assigned in increasing or decreasing order? By default, ranks are assigned by ordering the data values in ascending order (smallest to largest), then labeling the smallest value as rank 1. Alternatively, **Largest value** orders the data in descending order (largest to smallest), and assigns the largest value the rank of 1.

D**Display summary tables:** When checked, a summary of the new rank variables is printed to the Output window. The summary includes the original variables, the name of the new variables, the rank order, the ranking method, and the method used for ties. This option is on by default.

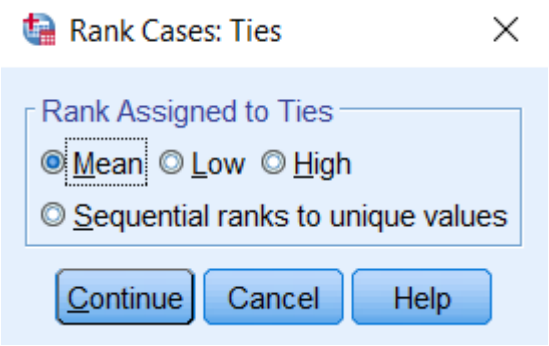
E**Rank types:** (Optional) Choose one or more formulas to compute the ranks. Each box you check on this screen will add another rank variable to your dataset.



By default, only the "Rank" option is selected; this computes simple ranks. The "Ntiles" option will produce percentile-based groupings: for example, Ntiles=2 will perform a median split; Ntiles=4 will produce quartiles; Ntiles=10 will produce decile groups.

For details about the other rank types and the proportion estimation formulas, please see the official SPSS documentation for Rank Cases. Note that the Proportion Estimation Formula options are inactive unless **Proportion estimates** and/or **Normal scores** are selected.

F**Ties:** How should ranks be assigned in the case of ties? (A *tie* occurs when two or more observations share the exact same value.) There are four options for how to resolve ties: **Mean**, **Low**, **High**, and **Sequential ranks to unique values**. By default, mean ranks are assigned to ties.



Mean - First, the observations are ordered and given unique, sequential ranks. Then, tied observations have their assigned ranks averaged together.**Low** - First, the observations are ordered and given unique, sequential ranks. Then, the ranks of any ties are re-assigned to the value of the smallest rank.**High** - First, the observations are ordered and given unique, sequential ranks. Then, the ranks of any ties are re-assigned to the value of the largest rank.**Sequential ranks to unique values** - First, the observations are ordered. Unique ranks are assigned in order until a tie is encountered. Ties receive the same rank until the next unique value appears. (The actual number of unique ranks assigned is therefore equal to the number of unique values.)

Example: Rank Transforms for Non-Normal Data

Many hypothesis tests require assumptions about the distribution of the data or residuals. A common way to adjust for non-normality is to perform a *transform* on that variable; for example, taking the log, square root, or square of a variable. *Rank transforms* are another type of transform. Suppose we want to perform a rank transform on a variable in the sample dataset that is non-normally distributed: MileMinDur.

Before the Procedure

Before we compute the ranks, let's check how many nonmissing values MileMinDur has. Let's also check the distribution of the mile run times graphically. The Frequencies procedure makes it easy to do both of these things at once:

Statistics

Mile run time

N	Valid	392
	Missing	43
Mean		0:08:09
Median		0:07:40
Std. Deviation		0:01:58.311
Minimum		0:05:01
Maximum		0:14:02
Percentiles	25	0:06:39
	50	0:07:40
	75	0:09:20