

0.1 Z Score

Transform Values

The following alternatives are available for transforming values:

- Z scores. Values are standardized to z scores, with a mean of 0 and a standard deviation of 1.
- Range -1 to 1. Each value for the item being standardized is divided by the range of the values.
- Range 0 to 1. The procedure subtracts the minimum value from each item being standardized and then divides by the range.
- Maximum magnitude of 1. The procedure divides each value for the item being standardized by the maximum of the values.
- Mean of 1. The procedure divides each value for the item being standardized by the mean of the values.
- Standard deviation of 1. The procedure divides each value for the variable or case being standardized by the standard deviation of the values.
- Additionally, you can choose how standardization is done. Alternatives are By variable or By case.

Transform Values

The following alternatives are available for transforming values:

- Z scores. Values are standardized to z scores, with a mean of 0 and a standard deviation of 1.
- Range -1 to 1. Each value for the item being standardized is divided by the range of the values.
- Range 0 to 1. The procedure subtracts the minimum value from each item being standardized and then divides by the range.
- Maximum magnitude of 1. The procedure divides each value for the item being standardized by the maximum of the values.
- Mean of 1. The procedure divides each value for the item being standardized by the mean of the values.
- Standard deviation of 1. The procedure divides each value for the variable or case being standardized by the standard deviation of the values.
- Additionally, you can choose how standardization is done. Alternatives are By variable or By case.

0.2 Standardizing the Variables

If variables are measured on different scales, variables with large values contribute more to the distance measure than variables with small values. In this example, both variables are measured on the same scale, so that's not much of a problem, assuming the judges use the scales similarly. But if you were looking at the distance between two people based on their IQs and incomes in dollars, you would probably find that the differences

in incomes would dominate any distance measures. (A difference of only \$100 when squared becomes 10,000, while a difference of 30 IQ points would be only 900. I'd go for the IQ points over the dollars!).

Variables that are measured in large numbers will contribute to the distance more than variables recorded in smaller numbers.

In the hierarchical clustering procedure in SPSS, you can standardize variables in different ways. You can compute standardized scores or divide by just the standard deviation, range, mean, or maximum. This results in all variables contributing more equally to the distance measurement. That's not necessarily always the best strategy, since variability of a measure can provide useful information.