Mahalanobis Distance

Mahalanobis distance provides a way to measure how similar some set of conditions is to a known set of conditions. It accounts the covariance among variables.

It is calculated as:

$$D^2 = (x - m)^T C^{-1} (x - m)$$

- D^2 Mahalanobis distance

- x Vector of data

- m Vector of mean values of independent variables

- C^{-1} Inverse Covariance matrix of independent variables

- T Indicates vector should be transposed

- This page provides a detailed explanation (with examples from landscape analysis)

- Mahalanobis distance is used to find outliers in a set of data. I don't know what field you are in, but in psychology it is used to identify cases that do not "fit" in with what is expected given the norms for the data set.

- For example, if your sample is composed of individuals with low levels of depression and you have one or two individuals with very high levels of depression, then they would have Mahalanobis distances which are greater than the expected critical value.

- You would probably want to remove these cases if they are very extreme AND if you feel that they do not fit in with your data set. (Using provided example, your sample is composed of individuals with low levels of depression, therefore those one or two individuals with high levels of depression do not fit in with the rest).

- Identifying outliers is very important because many statistical analyses have an "assumption of normality" i.e. an expectation that your data a normally distributed. Outliers can also contribute to skewed data, and for this reason they should also be removed. (Unless you transform the entire variable and this corrects the skew).

- Many stats programs such as SPSS allow you to calculate the M distances and the probability associated with each score to identify outliers. I can provide you with SPSS instructions, but I don't know if you are using SPSS.