Print

Detecting Influential Observations with DFBETAS

If you detect an influential observation, you want to know which predictor variable is being influenced. To help identify which parameter the observation might be influencing most, you can use DFBETAS, which stands for Difference in Betas.

DFBETAS measure the change in each parameter estimate, and one DFBETA is calculated per predictor variable per observation. This is similar to Cook's D, but it detects influence on individual parameters instead of on the entire set of parameters. Some researchers first use Cook's D to determine who is influential, and then use DFBETAS to determine which specific predictor is being influenced by that observation.

Each DFBETA is calculated by taking the estimated coefficient for a predictor variable, using all the data, and subtracting the estimated coefficient for that variable with the current observation removed. The difference in the betas is divided by its standard error. This calculation is repeated for all predictor variables and all observations.

Large DFBETAS indicate observations that are influential in estimating a given parameter. For DFBETAS, the general cutoff value is 2, and the size-adjusted cutoff is $2\sqrt{\frac{1}{n}}$, where n is the sample size.

Close