



Confusing Stats Terms Explained: Heteroscedasticity (Heteroskedasticity)

Jeremy J. Taylor • 68 Comments

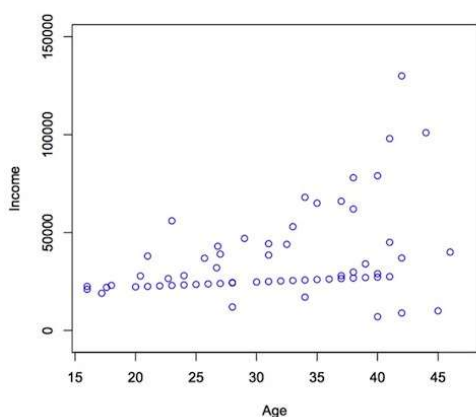
Heteroscedasticity is a hard word to pronounce, but it doesn't need to be a difficult concept to understand. Put simply, heteroscedasticity (also spelled heteroskedasticity) refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it.

A scatterplot of these variables will often create a cone-like shape, as the scatter (or variability) of the dependent variable (DV) widens or narrows as the value of the independent variable (IV) increases. The inverse of heteroscedasticity is homoscedasticity, which indicates that a DV's variability is equal across values of an IV.

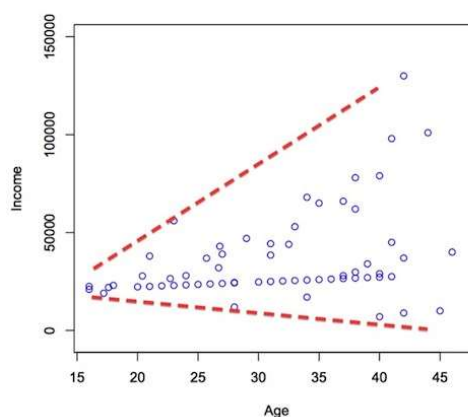
For example: annual income might be a heteroscedastic variable when predicted by age, because most teens aren't flying around in G6 jets that they bought from

their own income. More commonly, teen workers earn close to the minimum wage, so there isn't a lot of variability during the teen years. However, as teens turn into 20-somethings, and 20-somethings into 30-somethings, some will tend to shoot-up the tax brackets, while others will increase more gradually (or perhaps not at all, unfortunately). Put simply, the gap between the "haves" and the "have-nots" is likely to widen with age.

If the above were true and I had a random sample of earners across all ages, a plot of the association between age and income would demonstrate heteroscedasticity, like this:



Plot No. 1 demonstrating heteroscedasticity (heteroskedasticity)



Plot No. 2 demonstrating heteroscedasticity (heteroskedasticity)

By the way, I have no real data behind this example; this is just a hypothetical

situation, though it does seem logical.

Heteroscedasticity is most frequently discussed in terms of the assumption of parametric analyses (e.g. linear regression). More specifically, it is assumed that the error (a.k.a residual) of a regression model is homoscedastic across all values of the predicted value of the DV. Put more simply, a test of homoscedasticity of error terms determines whether a regression model's ability to predict a DV is consistent across all values of that DV. If a regression model is consistently accurate when it predicts low values of the DV, but highly inconsistent in accuracy when it predicts high values, then the results of that regression should not be trusted.

I want to re-iterate that the concern about heteroscedasticity, in the context of regression and other parametric analyses, is specifically related to error terms and NOT between two individual variables (as in the example of income and age). This is a common misconception, similar to the misconception about normality (IVs or DVs need not be normally distributed, as long as the residuals of the regression model are normally distributed). Now that you know what heteroscedasticity means, now try saying it five times fast!

I hope you found this helpful. What stats terms do you find confusing?



Previous

R Is Not So Hard! A Tutorial, Part 4 (repost)

Stats Make Me Cry Blog Entries

Next

How to Create APA Style Graphs and Then Teach SPSS to Do it Automatically!

Stats Make Me Cry Blog Entries

GET HELP TODAY

Powered by [Squarespace](#)