THE ANALYSIS FACTOR

(https://www.theanalysisfactor.com/)

Eight Ways to Detect Multicollinearity

by Karen Grace-Martin (https://www.theanalysisfactor.com/author/kgm_admin/)

Multicollinearity can affect any regression model with more than one predictor. It occurs when two or more predictor variables overlap so much in what they measure that their effects are indistinguishable.

When the model tries to estimate their unique effects, it goes wonky (yes, that's a technical term).

So for example, you may be interested in understanding the separate effects of altitude and temperature on the growth of a certain species of mountain tree.

Altitude and temperature are distinct concepts, but the mean temperature is so correlated with the altitude at which the tree is growing that there is no way to separate out their effects.

But it's not always easy to tell that the wonkiness (https://www.theanalysisfactor.com/steps-to-take-when-your-regression-or-other-statistical-results-just-lookwrong/) in your model comes from multicollinearity.

One popular detection method is based on the bivariate correlation (https://www.theanalysisfactor.com/7-guidelines-model-building/) between two predictor variables. If it's above .8 (or .7 or .9 or some other high number), the rule of thumb says you have multicollinearity.

And it is certainly true that a high correlation between two predictors is an indicator of multicollinearity. But there are two problems with treating this rule of thumb as a rule.

First, how high that correlation has to be before you're finding inflated variances depends on the sample size. There is no one good cut off number.

Second, it's possible that while no two variables are highly correlated, three or more together are multicollinear. Weird idea, I know. But it happens.

You'll completely miss the multicollinearity in that situation if you're just looking at bivariate correlations.

So like a lot of things in statistics, when you're checking for multicollinearity, you have to check multiple indicators and look for patterns among them. Sometimes just one is all it takes and sometimes you need to see patterns among a few.

Seven more ways to detect multicollinearity

1. Very high standard errors for regression coefficients

When standard errors are orders of magnitude higher than their coefficients, that's an indicator.

2. The overall model is significant, but none of the coefficients are

Remember that a p-value for a coefficient tests whether the unique effect of that predictor on Y is zero. If all predictors overlap in what they measure, there is little unique effect, even if the predictors as a group have an effect on Y.

3. Large changes in coefficients when adding predictors

If the predictors are completely independent of each other, their coefficients won't change at all when you add or remove one. But the more they overlap, the more drastically their coefficients will change.

4. Coefficients have signs opposite what you'd expect from theory

Be careful here as you don't want to disregard an unexpected finding as problematic. Not all effects opposite theory indicate a problem with the model. That said, it could be multicollinearity and warrants taking a second look at other indicators.

5. Coefficients on different samples are wildly different

If you have a large enough sample, split the sample in half and run the model separately on each half. Wildly different coefficients in the two models could be a sign of multicollinearity.

6. High Variance Inflation Factor (VIF) and Low Tolerance

These two useful statistics are reciprocals of each other. So either a high VIF or a low tolerance is indicative of multicollinearity. VIF is a direct measure of how much the variance of the coefficient (ie. its standard error) is being inflated due to multicollinearity (https://www.theanalysisfactor.com/linear-regression-analysis-3-common-causes-of-multicollinearity-and-what-do-to-about-them/).

7. High Condition Indices

Condition indices are a bit strange. The basic idea is to run a Principal Components Analysis (https://www.theanalysisfactor.com/pca-for-reducing-predictors-and-response-variables/) on all predictors. If they have a lot of shared information, the first Principal Component will be much higher than the last. Their ratio, the Condition Index, will be high if multicollinearity is present.

The Analysis Factor
Free Webinar Series

Fixed and Random Factors in Mixed Models

One of the hardest parts of mixed models is understanding which factors to make fixed and which to make random. Learn the important criteria to help you decide.

(https://thecraftofstatisticalanalysis.com/webinar-recording-signup/?
cosid=555)

Comments

kassawmar (http://bdu.edu.et) says
October 29, 2019 at 10:23 am (https://www.theanalysisfactor.com/eight-ways-to-detect-multicollinearity/#comment-297031)

Can we use VIF for nonlinear variables to check the multicollinearity effect? If So how? and what are the alternative methods?

Reply.

F says
January 10, 2022 at 12:06 am (https://www.theanalysisfactor.com/eight-ways-to-detect-multicollinearity/#comment-645104)

If VIF values are less than 5 some researchers suggested even 10, implying that the CMB was not a problem for evaluating the structural model. Hence no issue about multicollinearity

Reply

Seren says

June 5, 2019 at 8:02 am (https://www.theanalysisfactor.com/eight-ways-to-detect-multicollinearity/#comment-270834)

Could you use a Chi square test to identify multicollinearity?

For instance if a Chi square test gave a Cramer's V effect size that indicated that the two variables were probably measuring the same concept (Redundant) is this evidence for multicollinearity in regression with those two variables as predictors?

Thanks very much for the stats help!

Reply

Karen Grace-Martin (https://www.theanalysisfactor.com/) says
June 7, 2019 at 4:50 pm (https://www.theanalysisfactor.com/eight-ways-to-detect-multicollinearity/#comment-271028)

Hi Seren,

Sure, if both variables were categorical. It's essentially the same concept as a correlation.

Reply

Seren says

June 14, 2019 at 7:36 am (https://www.theanalysisfactor.com/eight-ways-to-detect-multicollinearity/#comment-271582)

Amazing, thank you!

Reply

Please note that, due to the large number of comments submitted, any questions on problems related to a personal study/project **will not** be answered. We suggest joining Statistically Speaking (https://www.theanalysisfactor.com/membership/), where you have access to a private forum and more resources 24/7.

The Analysis Factor uses cookies to ensure that we give you the best experience of our website. If you continue we assume that you consent to receive cookies on all websites from The Analysis Factor.

Continue   Privacy Policy (https://www.theanalysisfactor.com/privacy-policy/)