Lesson 15:
Multiple Linear Regression

SEARCH

RESOURCES

CONCEPTS

✓ 9. Text: Dummy Variables

✓ 10. Quiz: Dummy Variables

✓ 11. Screencast: Dummy Variables

✓ 12. Notebook + Quiz: Dummy Vari...

✓ 13. Video: Dummy Variables Recap

✓ 14. [Optional] Notebook + Quiz: O...

✓ 15. Video: Potential Problems

✓ 16. [Optional] Text: Linear Model ...

✓ 17. Screencast: Multicollinearity & ...

✓ 18. Video: Multicollinearity & VIFs

✓ 19. Notebook + Quiz: Multicollinea...

Mentor Help
Ask a mentor on our Q&A platform

Peer Chat 2
Chat with peers and alumni

Text: Dummy Variables

The Math Behind Dummy Variables

In the last video, you were introduced to the idea the way that categorical varia **dummy variables** in order to be added to your linear models.

Then, you will need to drop one of the **dummy columns** in order to make your

If you remember back to the closed form solution for the coefficients in regres estimated by $(X'X)^{-1}X'y$.

In order to take the inverse of $(X'X)$, the matrix X must be full rank. That is, must be linearly independent.

If you do not drop one of the columns (from the model, not from the datafram dummy variables, your solution is unstable and results from Python are unrelia example of what happens if you do not drop one of the dummy columns in the

The takeaway is ... **when you create dummy variables using 0, 1 encodings, drop one of the columns from the model to make sure your matrices are solutions are reliable from Python).**

The reason for this is linear algebra. Specifically, in order to invert matrices, a n (that is, all the columns need to be linearly independent). Therefore, you need dummy columns, to create linearly independent columns (and a full rank matri