Module 4: Multiple Linear Regression

Jeffrey Woo

MSDS, University of Virginia

Word of Caution in MLR

Be careful of notation from different books/websites. In our textbook:

- k: number of predictors in MLR model (assuming only first order additive terms)
- p: number of parameters in MLR model

In some other books, p is used to denote the number of predictors, k is used to denote the number of parameters.
Recommendation: think in terms of number of parameters.

SLR vs MLR

- If you compare formulas between SLR and MLR, you realize that the formula for SLR is the same as MLR with $p = 2$ or $k = 1$. E.g. Compare (3.17) with (2.19) in textbook.
- To assess if regression assumptions are met, a residual plot must be used, and not a scatterplot of the response against the predictor (with multiple predictors, which do we use?)
- There are some "minor" differences in interpretation.

MLR Interpretation

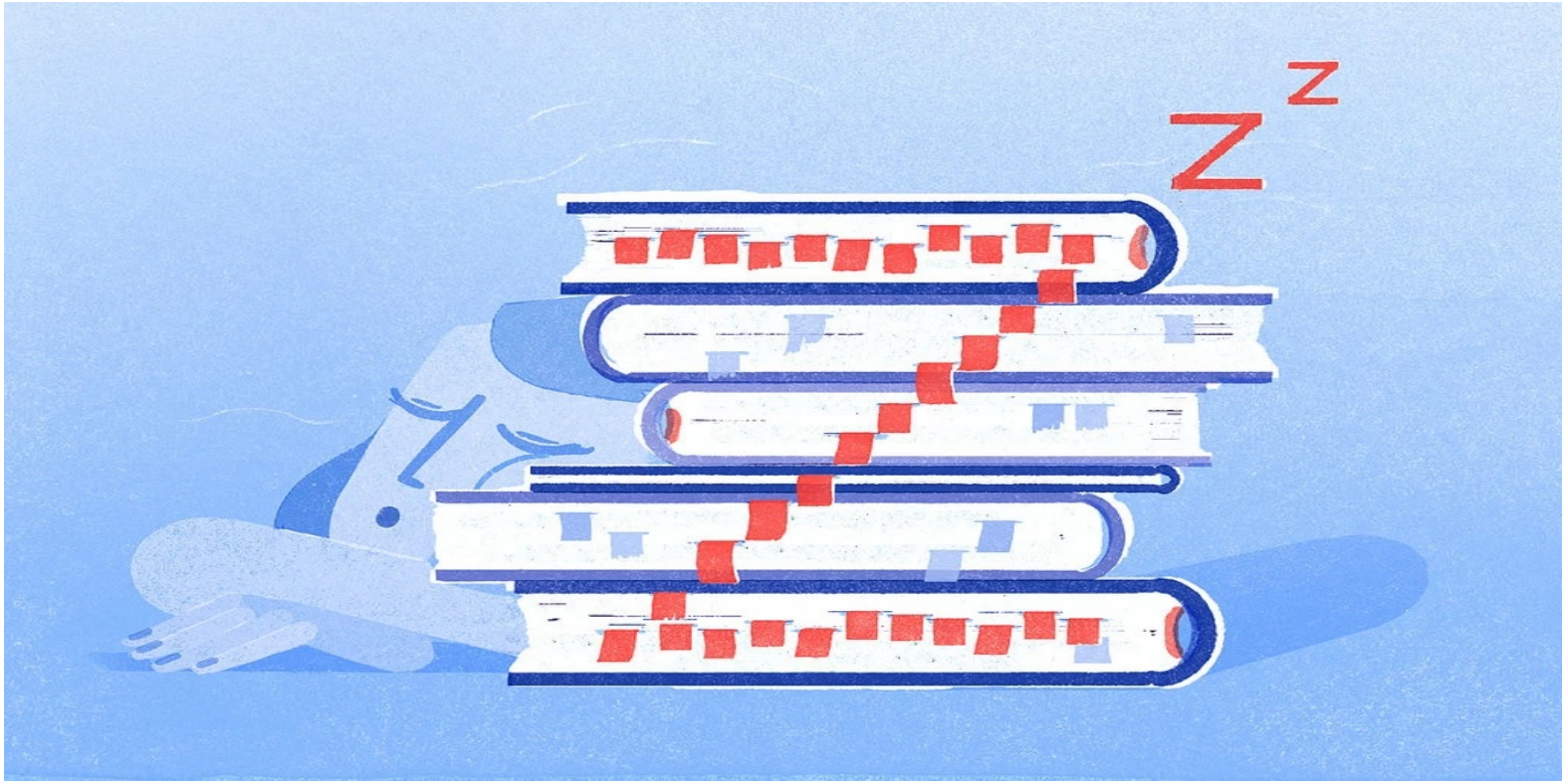

Figure: An Underappreciated Key to College Success: Sleep. NY Times, Aug 13 2018.

MLR Interpretation

"The study, by Monica E. Hartmann and Dr. Prichard of the University of St. Thomas in St. Paul, Minn., found that for each additional day of sleep disturbance a college student experienced each week, the likelihood of dropping a course rose by 10 percent and grade point average fell by 0.02, even **when most other factors known to influence academic success were taken into account.**"

MLR Interpretation

Other ways of stating the bolded statement:

- when holding the other predictors constant.
- when controlling for the effect of other predictors.
- after adjusting for the effect of other predictors.

MLR Interpretation: Textbook

From page 68: "The parameter β_j represents the expected change in the response y per unit change in x_j when all of the remaining regressor variables x_i $(i \neq j)$ are held constant."

ANOVA F Test for MLR

- $H_0 : \beta_1 = \cdots \beta_k = 0$. In words, our model is not useful in predicting the response.
- H_a : At least one of the βs in H_0 is not zero. In words, our model is useful in predicting the response.

Question: How does the null hypothesis simplify in SLR?
Opinion: The ANOVA F test is not very useful in MLR, except in some settings. Will explore one such setting in module 5.

t Test for MLR

- $H_0 : \beta_j = 0$. In words, we **can drop** predictor x_j from the model in the presence of the other predictors.
- $H_a : \beta_j \neq 0$. In words, we **cannot drop** predictor x_j from the model in the presence of the other predictors.
- An insignificant t test for a coefficient β_j in MLR indicates that predictor x_j can be removed from the model (and leave the other predictors in). It is not needed in the presence of the other predictors.
- If x_j is highly correlated with at least one of the other predictors, or is a linear combination of a number of other predictors, x_j will probably be insignificant as the addition of x_j doesn't help in improving the model.
- x_j itself **may still** be linearly related to the response variable.
- If your goal is to assess if x_j is linearly related to the response, need to use SLR.

SLR vs MLR

- Scatterplot, Correlation, SLR: relationship between two variables.
- MLR: relationship between response and multiple predictors (simultaneously).