3.1 Alternate Variable Types and Interactions

Dr. Bean - Stat 5100

Describe the difference between multicollinearity and interactions.

Multicollinearity only deals with relationships among the X variables and has NOTHING to do with Y. Interactions have EVERYTHING to do with Y and the way that X_k's relationship with Y is influenced by the other X variables.

Identify whether each of the following variables are qualitative or quantitative:

- hours worked quantitative
- shirt color qualitative
- a person's shoe size quantitative
- systolic blood pressure quantitative
- blood type qualitative
- college major qualitative

Ignoring the significance of the model coefficients and assuming that assumptions regarding residuals are satisfied, please write the estimated regression equation corresponding to the following SAS output (not that Y represents Oxygen Intake Rate).

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	107.29565	43.38302	2.47	0.0202
Age	Age (in years)	1	-0.46161	0.87289	-0.53	0.6014
Weight	Weight (in kilograms)	1	-0.23470	0.54681	-0.43	0.6713
RunTime	Time to Run 1.5 Miles (in minutes)	1	-3.15204	0.37523	-8.40	<.0001
ageWeight	Age*Weight	1	0.00370	0.01115	0.33	0.7427

$$\hat{Y} = 107.296 - 0.461 * Age - 0.235 * Weight - 3.152 * RunTime + 0.004 * Age * Weight$$

What is the expected change in the average of Y when a person ages by one year?

$$-0.461 + 0.004 * Weight$$

Suppose you are trying to predict a person's happiness and you suspect that country of origin is a significant predictor of happiness. You take a sample of 100 people to test your hypothesis. In this scenario, what issue will you run into trying to use country of origin as a predictor variable?

We need $q-1$ dummy variables to represent q countries. There are more than 100 countries in the world, so our model would require more degrees of freedom than is feasible with our sample size.