Multicollinearity - How related predictors are to each other?
(covariates, independent variables)

Related to questions like:

* What is the relative importance of the effects of different predictors?
* What is the magnitude of the effect of a given predictor variable on the response?
* Can we "drop" a predictor variable because it has little or no effect on the response?
* Should any predictor variables that are not yet included in the model be added?

Answers depend on whether or not predictor variables are correlated

usually in real life predictors are in between

* If uncorrelated - things are easy!
* If they are perfectly correlated - things are difficult.

In the perfectly correlated example,

$$X_{i2} = 5 + 0.5 X_{i1} \text{ for } i = 1, \dots, 4$$

$\Rightarrow$ the data don't contain a random error

$\Rightarrow$ can't get unique estimates b_1 and b_2 of β_1 and β_2

$\Rightarrow$ no meaningful interpret

$\not\Rightarrow$ we can't fit the data well, in fact,
we can fit the data perfectly

Multicollinearity:

* Doesn't mean we can't fit the data well
* Might mean that $\beta_1, \dots, \beta_{p-1}$ are imprecisely estimated
* Might mean that interpretation of b_k for some k isn't really applicable/doesn't really make sense

What does this really mean in practice?

We saw that can't get unique estimates of $\beta_1, \dots, \beta_{p-1}$ when covariates are perfectly correlated

$\Longrightarrow$ Can't get unique estimates of $\beta_1, \dots, \beta_{p-1}$ when $n \leq p$

Regression Models for Quantitative and Qualitative Predictors

ways to use multiple linear regression to fit complex models

first order terms

second order terms

Polynomial Models for Response

If $p=2$, $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_{11} x_{i1}^2 + \varepsilon_i$, where $x_{i1} = X_{i1} - \bar{X}_1$

second order model with one predictor

If $p=3$, $Y_i = \beta_0 + \underline{\beta_1 x_{i1}} + \underline{\beta_2 x_{i2}} + \underline{\beta_{12} x_{i1} x_{i2}} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \varepsilon_i$

second order model with two predictors

$x_{ik} = X_{ik} - \bar{X}_k$

second order terms are always used <u>with</u> first order terms