Dummy/Indicator Variables and Analysis of Covariance

transform qualitative variables to incorporate into linear regression model, e.g. sex
set up a dummy variable for each class of the qualitative variable
qualitative variable with c classes represented by $c - 1$ binomial dummy variables

Y = speed that an insurance innovation is adopted
X_1 = total assets of the firm
X_2 = type of firm, stock or mutual company

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$X_1 = size\ of\ firm$$

$$X_2 = \begin{cases} 1 & if\ stock\ company \\ 0 & otherwise \end{cases}$$

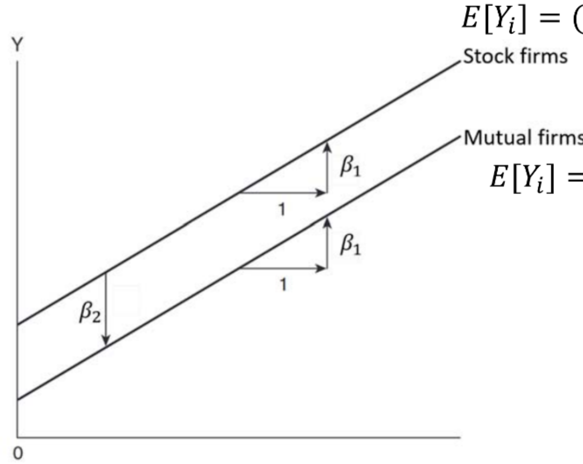
mutual firms $\quad E[Y] = \beta_0 + \beta_1 X_1 + \beta_2(0) = \beta_0 + \beta_1 X_1$
stock firms $\quad E[Y] = \beta_0 + \beta_1 X_1 + \beta_2(1) = (\beta_0 + \beta_2) + \beta_1 X_1$

$E[Y]$ = linear function of the size of the firm, X_1
slope for both firms is the same for any given size of the firm

β_1 = mean time elapsed before the innovation is adopted
β_2 = difference in intercept between the response variable for stock firms than mutual firms for any given size of the firm

$$E[Y_i] = (\beta_0 + \beta_2) + \beta_1 X_{i1}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$



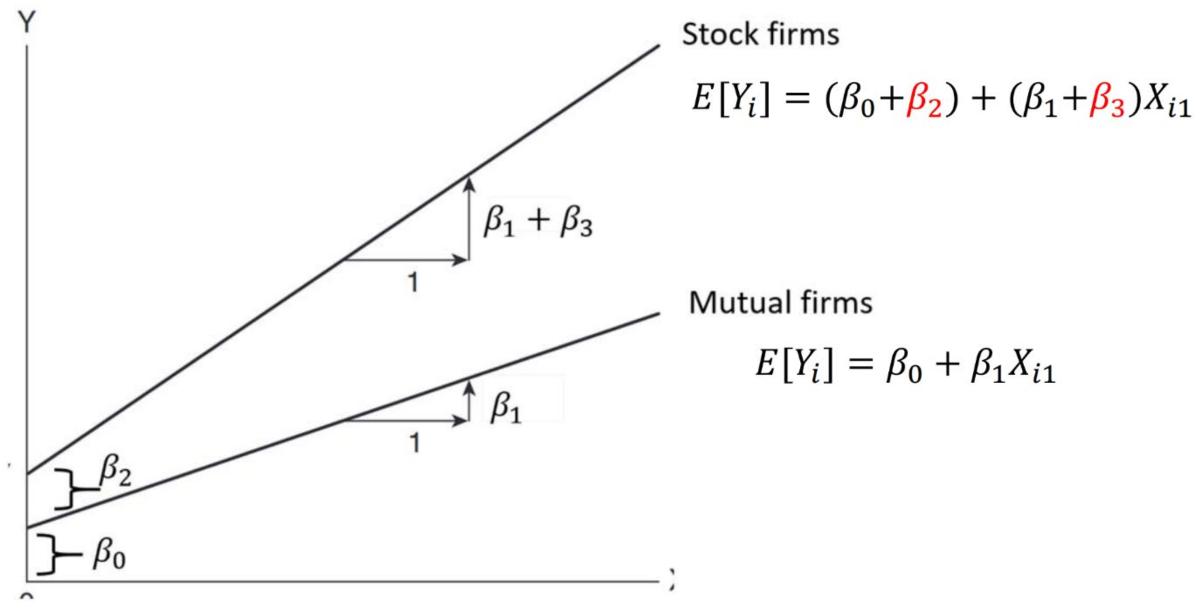
$$E[Y_i] = (\beta_0) + \beta_1 X_{i1}$$

Interaction Effects

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

$\beta_3 X_1 X_2$ = interaction term between size of firm and type of firm
β_1 = mean time elapsed before the innovation is adopted
β_2 = difference in intercept between the response variable for stock firms than mutual firms
β_3 = difference in slope/effect size between the response variable for stock firms than mutual firms for any given size of the firm
if there is no interaction between the two variables, the slope for both firms will remain the same

mutual firms $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2(0) + \beta_3 X_1(0) = \beta_0 + \beta_1 X_1$
stock firms $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1(1) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1$



interaction term is a second order polynomial
drop higher order terms if they're not significant

Least Square Means/Adjusted Means

raw mean = an average of the observations without considering other covariates
least square means = mean estimated from a linear regression adjusted for other covariates
least square means are theoretical estimates of the true population mean

Qualitative Variable with > 2 Classes

Y = test score
X_1 = number of hours spent studying
X_2 = level of education: high school, undergraduate, master's, doctorate

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i$$

$$X_1 = test\ score$$

$$X_2 = \begin{cases} 1 & if\ high\ school \\ 0 & otherwise \end{cases}$$

$$X_3 = \begin{cases} 1 & if\ undergradute \\ 0 & otherwise \end{cases}$$

$$X_4 = \begin{cases} 1 & if\ master's \\ 0 & otherwise \end{cases}$$

***X* Matrix**					
Level of Education	X_0	X_1	X_2	X_3	X_4
high school	1	X_{i1}	1	0	0
undergraduate	1	X_{i1}	0	1	0
master's	1	X_{i1}	0	0	1
doctorate	1	X_{i1}	0	0	0

doctorate $E[Y] = \beta_0 + \beta_1 X_1$
high school $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2$
undergraduate $E[Y] = \beta_0 + \beta_1 X_1 + \beta_3$
master's $E[Y] = \beta_0 + \beta_1 X_1 + \beta_4$

$E[Y]$ = linear function of the number of hours spend studying, X_1
slope for all levels of education is the same for any given number of hours spent studying

β_1 = mean test score
β_2 = difference in intercept between the response variable for high school than doctorate for any given number of hours spent studying
β_3 = difference in intercept between the response variable for undergraduate than doctorate for any given number of hours spent studying
β_4 = difference in intercept between the response variable for master's than doctorate for any given number of hours spent studying

ANOVA and ANCOVA

ANOVA models association between a continuous response variable and continuous quantitative predictor variables

ANCOVA models association between a continuous response variable and a categorical qualitative variable, adjusting for a continuous covariate

e.g. modeling how long two types of companies take to adopt a new innovation, adjusting for size of the company

covariance models = chief independent variables of interest are qualitative and quantitative independent variables are introduced primarily to reduce the variance of error terms