

## 625.661 Statistical Models and Regression

### Module 5 Discussion Question

H.M. James Hung

Let  $y$  be a continuous random variable,  $x_1$  be a categorical variable that has 3 levels (L1, L2, L3),  $x_2$  be a categorical variable that has 2 levels ("yes" or "no"),  $x_3$  be a continuous variable. Construct a multiple linear regression model such that we can study the effect of  $x_3$  on  $y$  and study whether the effect of  $x_3$  on  $y$  is equal across all levels of  $x_1$  and  $x_2$ .

For  $x_1$ , select L1 to be the reference level. For  $x_2$ , select "no" to be the reference level.

Define  $x_{12} = 1$  if L2 and 0 otherwise. Define  $x_{13} = 1$  if L3 and 0 otherwise. Immediately, we have coding in the following table.

	$x_{12}$	$x_{13}$
L1	0	0
L2	1	0
L3	0	1

Likewise, for  $x_2$ ,

	$x_2$
yes	1
no	0

The regression model needed is

$$E(y | x_3, x_{12}, x_{13}, x_2) = \beta_0 + \beta_3 x_3 + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_2 x_2 + \gamma_{312} x_3 x_{12} + \gamma_{313} x_3 x_{13} + \gamma_{212} x_2 x_{12} + \gamma_{213} x_2 x_{13} + \gamma_{32} x_3 x_2 + \lambda_1 x_3 x_{12} x_2 + \lambda_2 x_3 x_{13} x_2$$

That is, the regression model needs to contain two-way interactions ( $\gamma$ 's coefficients) and three-way interactions ( $\lambda$ 's coefficients).