

## APM 630 Regression Analysis

### Project #4 – Dummy Variable Regression

**Data:** Dummy.xls

Data on college enrollment (in millions) by sex were collected from 1975 to 1993. You are assigned to conduct a regression analysis to estimate the trend in college enrollment for both male and female students, and perform statistical tests to exam if the trend is different between male and female. The data contains calendar year (Year), X (recode year from 1 to 19), SEX (female or male), and college enrollment in millions (College, the response variable).

#### Assignment:

1. Re-code SEX by a dummy variable Z (female:  $Z = 0$  and male:  $Z = 1$ ), such as  
IF SEX='female' THEN  $Z=0$ ;  
ELSE  $Z=1$ ;
2. Compute descriptive statistics of College for each level of Z.
3. Draw a scatterplot of College and X using different symbols for the two levels of Z.
4. Fit an overall model  $\text{College} = \beta_0 + \beta_1 X + e$  for combining the data of male and female.
5. Fit the model  $\text{College} = \beta_0 + \beta_1 X + e$  by each level of Z (a separate model for male and female). Report the two models.
6. Then fit the following dummy variable model:

$$\text{College} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 (XZ) + e.$$

7. Manually derive the model  $\text{College} = \hat{\beta}_0 + \hat{\beta}_1 X$  by each level of Z (= 0 and 1) from the dummy variable model above. Compare them with the two models obtained in the step 5.
8. Test the regression line of the two levels of Z for (1) the same intercepts, (2) the same slopes, or (3) coincidence? What conclusion you can draw from the tests?