This question refers to the estimated regressions in table 1 computed using data for 1988 from the U.S. Current Population Survey. The data set consists of information on 4000 full-time full-year workers. The highest educational achievement for each worker was either a high school diploma or a bachelor's degree. The worker's ages ranged from 25 to 34 years. The dataset also contained information on the region of the country where the person lived, marital status, and number of children. For the purposes of these exercises let

AHE = average hourly earnings (in 1998 dollars)

College = binary variable (1 if college, 0 if high school)

Female = binary variable (1 if female, 0 if male) Age = age (in years)

Ntheast = binary variable (1 if Region = Northeast, 0 otherwise)

Midwest = binary variable (1 if Region = Midwest, 0 otherwise)

South = binary variable (1 if Region = South, 0 otherwise)

West = binary variable (1 if Region = West, 0 otherwise)

You have data (above) on average hourly earnings, gender, and where a person lives, represented as dummy variables for West, East, North and South 4 regions.

1. Write a regression equation that uses these variables to predict average hourly earnings and be sure to define the Gender variable.
2. Explain how you avoid the dummy variable trap (Hint: why you omitted one regional variable and what happen if it is included).
3. Explain the important meanings of regional difference (compare to reference group).