7.2 Using the regression results in column (1):

a. Is the college–high school earnings difference estimated from this regression statistically significant at the 5% level? Construct a 95% confidence interval of the difference.

b. Is the male–female earnings difference estimated from this regression statistically significant at the 5% level? Construct a 95% confidence interval for the difference.

a) t-stat = $8.31 / 0.23 = 36.13$; since it is greater than cval. of 1.96, we reject null hypothesis.

C.F. interval = $8.31 \pm 1.96 \times 0.23$

b) On avg. men earn 3.85 \$/h more than women.
∴ Male-female diff. @ 5% C.F. level is

$3.85 / 0.23 = 16.74$. Since this is greater than 1.96, we reject null hypothesis. at the 5% level.

C.F. Interval = $3.85 \pm 1.96 \times 0.23$

7.4 Using the regression results in column (3):

a. Do there appear to be important regional differences? Use an appropriate hypothesis test to explain your answer.

a) Northeast earns 0.18 \$/hr more than the west.
Midwest earns 1.23 \$/hr less than the west.
South earns 0.43 \$/hr less than the west.
the Fstat val. test the coeff. on the regional regressors are zero is 7.38. the Fstat cv at 1% significance is 3.78. Since Fstat > Fcritical val, the regional diff. are significant at 1% significance level.

b. Juanita is a 28-year-old female college graduate from the South. Molly is a 28-year-old female college graduate from the West. Jennifer is a 28-year-old female college graduate from the Midwest.

i. Construct a 95% confidence interval for the difference in expected earnings between Juanita and Molly.

ii. Explain how you would construct a 95% confidence interval for the difference in expected earnings between Juanita and Jennifer. (*Hint:* What would happen if you included *West* and excluded *Midwest* from the regression?)

i) Molly makes 0.43 \$/hr more than Juanita
∴ @ 95% C.I. int. we get :-

$$0.43 \pm 1.96 \times 0.30$$

ii) Expected difference bet. Juanita and Jennifer

$$(X_{6\,Juanita} - X_{6\,Jennifer}) \times \beta_6 - (X_{5\,Juanita} - X_{5\,Jennifer}) \times \beta_5$$

$$= \beta_6 + \beta_5$$

In this case, if we omit midwest and replaced it with X_5 = west. It would imply that the coeff. for south would be measuring difference in wages bet. south and midwest. Making computing the 95% C.I. interval in a more simplified manner.

7.6 Evaluate the following statement: "In all of the regressions, the coefficient on *Female* is negative, large, and statistically significant. This provides strong statistical evidence of gender discrimination in the U.S. labor market."

→ Since the estimated coeff. suggest a strong correlation between gender and earnings, however the aforementioned statement cannot be interpreted as causal unless there is assurance of an absence of omitted variable bias. For eg: we may believe that level of education has a direct impact on earnings and gender discrimination. In this case, ~~so~~ a lower level of education might result in lower wages, causing an upward bias on gender discrimination in the U.S. labor force.

7.9 Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Use Approach #2 from Section 7.3 to transform the regression so that you can use a t-statistic to test

a. $\beta_1 = \beta_2$.

b. $\beta_1 + 2\beta_2 = 0$.

c. $\beta_1 + \beta_2 = 1$. (*Hint:* You must redefine the dependent variable in the regression.)

a) Using approach #2 ⇒ we add & subtract $\beta_2 X_{1i}$

$$\beta_1 X_{1i} + \beta_2 X_{2i} = \beta_1 X_{1i} + \beta_2 X_{1i} - \beta_2 X_{1i} + \beta_2 X_{2i}$$

$$= X_{1i}(\beta_1 - \beta_2) + \beta_2 (X_{2i} + X_{1i})$$

$\gamma_1 = \beta_1 - \beta_2$, we re-arrange eqn. as

$$Y_1 = \beta_0 + X_{1i}\gamma_1 + \beta_2 (X_{2i} + X_{1i}) + u_i$$

we test $H_0: \gamma_1 = 0$, $H_1: \gamma_1 \neq 0$

b) $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

Assume $\gamma_1 = \beta_1 + 2\beta_2$

$$\beta_1 X_{1i} + \beta_2 X_{2i} = \beta_1 X_{1i} + 2\beta_2 X_{1i} - 2\beta_2 X_{1i} + \beta_2 X_{2i}$$

$$= (\beta_1 + 2\beta_2) X_{1i} - \beta_2 (2X_{1i} - X_{2i})$$

$$\therefore Y_1 = \beta_0 + \gamma_1 X_{1i} - \beta_2 (2X_{1i} - X_{2i}) + u_i$$

We estimate $\gamma_1 = 0$

$H_0: \gamma_1 = 0$; $H_1: \gamma_1 \neq 0$

c) Assume $\gamma_1 = \beta_1 + \beta_2 - 1$

$$\beta_1 X_{1i} + \beta_2 X_{2i} = \beta_1 X_{1i} + \beta_2 X_{1i} - \beta_2 X_{1i} + X_{1i} - X_{1i} + \beta_2 X_{2i}$$

$$\Rightarrow \cancel{\beta_2} X_{1i}(\beta_1 + \beta_2 - 1) + \beta_2 (X_{2i} - X_{1i}) + X_{1i}$$

$$Y_1 = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i = \beta_0 + X_{1i}\gamma_1 + \beta_2 (X_{2i} - X_{1i}) + X_{1i} + u_i$$

Assume $\gamma_1 = 0$

t-Test $H_0: \gamma_1 = 0$; $H_1: \gamma_1 \neq 0$; redefining the dependent variable

we get $Y_1 - X_{1i} = \beta_0 + X_{1i}\gamma_1 + \beta_2 (X_{2i} - X_{1i}) + u_i$