Econ 312: Problem Set #4

Solutions

1. (a) The estimated gender gap equals \$2.12/hour.

   (b) The hypothesis testing for the gender gap is $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. With a *t*-statistic

$$t^{act} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{2.12}{0.36} = 5.89,$$

   the *p*-value for the test is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-5.89) = 2 \times 0.0000 = 0.000 \text{ (to four decimal places)}$$

   The *p*-value is less than 0.01, so we can reject the null hypothesis that there is no gender gap at a 1% significance level.

   (c) The 95% confidence interval for the gender gap β_1 is $\{2.12 \pm 1.96 \times 0.36\}$, that is, $1.41 \leq \beta_1 \leq 2.83$.

   (d) The sample average wage of women is $\hat{\beta}_0 = \$12.52$/hour. The sample average wage of men is $\hat{\beta}_0 + \beta_1 = \$12.52 + \$2.12 = \$14.64$/hour.

   (e) The binary variable regression model relating wages to gender can be written as either

$$Wage = \beta_0 + \beta_1 Male + u_i,$$

   or

$$Wage = \gamma_0 + \gamma_1 Female + v_i.$$

   In the first regression equation, $Male$ equals 1 for men and 0 for women; β_0 is the population mean of wages for women and $\beta_0 + \beta_1$ is the population mean of wages for men. In the second regression equation, $Female$ equals 1 for women and 0 for men; γ_0 is the population mean of wages for men and $\gamma_0 + \gamma_1$ is the population mean of wages for women. We have the following relationship for the coefficients in the two regression equations:

$$\gamma_0 = \beta_0 + \beta_1,$$
$$\gamma_0 + \gamma_1 = \beta_0.$$

   Given the coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, we have

$$\hat{\gamma}_0 = \hat{\beta}_0 + \beta_1 = 14.64,$$
$$\hat{\gamma}_1 = \hat{\beta}_0 - \gamma_0 = -\beta_1 = -2.12.$$

2. Estimated Regressions

	Model	
Regressor	**a**	**b**
Age	0.45 (0.03)	0.44 (0.03)
Female		–3.17 (0.18)
Bachelor		6.87 (0.19)
Intercept	3.32 (0.97)	
SER	8.66	7.88
R^2	0.023	0.190
$\bar{R}^2$	0.022	0.190

(a) The estimated slope is 0.45

(b) The estimated marginal effect of *Age* on *AHE* is 0.44 dollars per year. The 95% confidence interval is $0.44 \pm 1.96 \times 0.03$ or 0.38 to 0.50.

(c) The results are quite similar. Evidently the regression in (a) does not suffer from important omitted variable bias.

(d) Bob's predicted average hourly earnings = $0.44 \times 26 - 3.17 \times 0 + 6.87 \times 0 + 3.32 = \$11.44$
Alexis's predicted average hourly earnings = $0.44 \times 30 - 3.17 \times 1 + 6.87 \times 1 + 3.32 = \$20.22$

(e) The regression in (b) fits the data much better. Gender and education are important predictors of earnings. The R^2 and $\bar{R}^2$ are similar because the sample size is large ($n = 7986$).

(f) Gender and education are important. The *F*-statistic is 752, which is (much) larger than the 1% critical value of 4.61.

(g) The omitted variables must have non-zero coefficients and must correlated with the included regressor. From (f) *Female* and *Bachelor* have non-zero coefficients; yet there does not seem to be important omitted variable bias, suggesting that the correlation of *Age* and *Female* and *Age* and *Bachelor* is small. (The sample correlations are *Cor* (*Age, Female*) = –0.03 and *Cor* (*Age,Bachelor*) = –0.00).