Midterm 2 - Solution Set

Matthew Strathearn

March 15, 2015

Question 1

(a)

Lets denote wage as w_i, education as e_i and female as f_i. The results of the regressions are as follows;

$$\log \hat{w}_i = 1.609 + 0.090e_i \tag{1}$$
$$\log \hat{w}_i = 1.654 + 0.096e_i - 0.024f_i \tag{2}$$

Figure 1: Regression Table of Values

$\log \hat{w}_{i1}$	$\log \hat{w}_{i2}$	e_i	f_i
1.699	1.750	1	0
1.789	1.603	2	1
1.879	1.942	3	0
1.969	1.795	4	1
2.059	1.981	5	1

Figure 2: Regression # 1

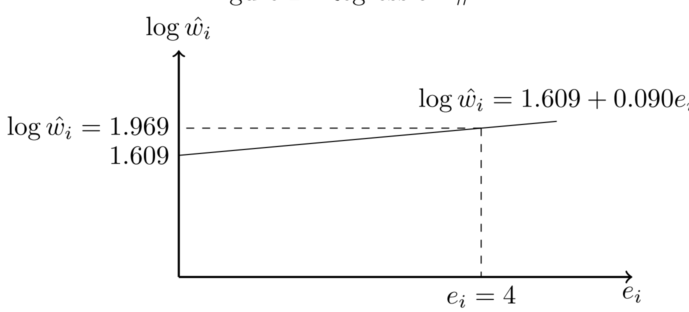
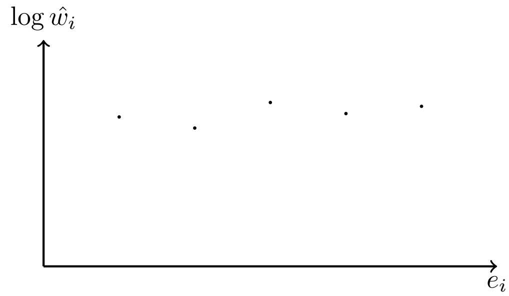


Figure 3: Regression # 2



(b)

The point forecasts for a female ($f_i = 1$), with 10 years of education ($e_i = 10$) are calculated as follows;

$$\log \hat{w}_i = 1.609 + 0.090(10) = 2.509 \tag{3}$$

$$\log \hat{w}_i = 1.654 + 0.096(10) - 0.243 = 2.371 \tag{4}$$

(c)

The marginal effect for model 1 is

$$\left.\frac{\partial w_i}{\partial e_i}\right|_{e_i=5} = 0.09 * \exp(1.609 + 0.09 * 5).$$

The marginal effect for model 2 is

$$\left.\frac{\partial w_i}{\partial e_i}\right|_{e_i=5} = 0.096 * \exp(1.654 + 0.096 * 5 - 0.0243 * f_i).$$

Thus, if this worker is female, then the marginal effect is equal to $0.096 * \exp(1.654 + 0.096 * 5 - 0.0243)$.

(d)

Notice that the marginal effect of another year of education on log wages is independent of the current level of education. We can construct the 95% CI as follows;

Model 1

$$P\left(\hat{\beta} - t_{0.975,998}\text{s.e}(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{0.975,998}\text{s.e}(\hat{\beta})\right) = 95\% \tag{5}$$

$$P\left(0.090 - 1.962(0.006) \leq \beta \leq 0.090 + 1.962(0.006)\right) = 95\% \tag{6}$$

$$P\left(0.090 - 0.012 \leq \beta \leq 0.090 + 0.012\right) = 95\% \tag{7}$$

$$P\left(0.078 \leq \beta \leq 0.102\right) = 95\% \tag{8}$$

Therefore, 95% of the time $\beta \in [0.078, 0.102]$. Because the interval does not include zero, we can conclude that $\hat{\beta}$ is statistically different from zero at the 5% level.

Model 2

$$P\left(\hat{\beta} - t_{0.975,998}\text{s.e}(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{0.975,998}\text{s.e}(\hat{\beta})\right) = 95\% \tag{9}$$

$$P\left(0.096 - 1.962(0.006) \leq \beta \leq 0.096 + 1.962(0.006)\right) = 95\% \tag{10}$$

$$P\left(0.096 - 0.012 \leq \beta \leq 0.096 + 0.012\right) = 95\% \tag{11}$$

$$P\left(0.084 \leq \beta \leq 0.108\right) = 95\% \tag{12}$$

Therefore, 95% of the time $\beta \in [0.084, 0.108]$. Because the interval does not include zero, we can conclude that $\hat{\beta}$ is statistically different from zero at the 5% level.

Question 2

(a)

We can calculate the estimator as follows;

$$\begin{aligned}\hat{\beta} &= \frac{\sum y_i x_i}{\sum x_i^2} \\ &= \frac{64}{30} \\ &= 2.133\end{aligned} \tag{13}$$

Figure 4: Coefficient Calculation

y_i	x_i	$y_i x_i$	x_i^2	$\hat{y}_i$	$\hat{\epsilon}_i$
2.5	1	2.5	1	2.133	0.367
5	2	10	4	4.267	0.733
6.5	3	19.5	9	6.400	0.100
8	4	32	16	8.533	-0.533
Σ		64	30		0.966

(b)

We can calculate the predicted y_i using, $\hat{y}_i = 2.133x_i$, and we can calculate the residual using $y_i - \hat{y}_i = \hat{\epsilon}_i$. Refer to part (a).

(c)

We can compute the R^2 as follows;

$$\begin{aligned} R^2 &= 1 - \frac{\sum \hat{\epsilon}_i}{\sum y_i^2} \\ &= 1 - \frac{0.966}{137.500} \\ &= 0.993 \end{aligned} \tag{14}$$

R^2 captures how well the data fits the model. Typically R^2 is measured as the Sum of Squares Regression over the Sum of Squares Total, and since $SST = SSR + SSE$ it is bounded between zero and one, with a perfectly fit model having an R^2 of 1. It represents how well the regression explains the variation in the dependent variable. However, in a regression through the origin we must use the formula above to yield a meaningful interpretation. The R^2 is usually inflated when the regression runs through the origin.

(d)

$$SST = \sum (y_i - \bar{y})^2 = \sum (y_i - 5.5)^2 = 16.5 \tag{15}$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2 = \sum (\hat{y}_i - 5.5)^2 = 22.866 \tag{16}$$

$$SSE = \sum \hat{\epsilon}_i = 0.993 \tag{17}$$

The equality does not hold.