Homework 3 - Answers.

1. (i) Several factors can be contained in u. For example,

1. age
2. marriage status
3. race
4. Income

Some of these variables could be correlated with education: Income or marriage status

(ii) Probably not. Additional variables should be added to the regression model.

2.
$$y = \beta_0 + \beta_1 x + u - \alpha_0 + \alpha_0$$
$$= (\beta_0 + \alpha_0) + \beta_1 x + (u - \alpha_0)$$
$$= \gamma + \beta_1 x + v$$

Now, $E(v) = E(u - \alpha_0) = \alpha_0 - \alpha_0 = 0$.

The new intercept is $\gamma = \beta_0 + \alpha_0$

4. (i) When $\begin{cases} cigs = 0, & \widehat{bwght} = 119.77 \\ cigs = 20, & \widehat{bwght} = 109.49 \end{cases}$

Smoking 20 cigarettes per day reduces birth weight by 10.28 ounces.

(ii) No. Captures a correlation between the two variables.

(iii) $125 = 119.77 - 0.514\, cigs \Rightarrow cigs = -10.17$

Negative consumption of cigarettes, which is impossible.

(iv) Yes. For 85% of the sample we have $(bwght_i, 0)$ where $cigs_i = 0$. If a random sample of bwghts is used a good predictor at $cigs = 0$ is the average of the birthweights for

9\. (i) $\tilde{\beta}_1 = \dfrac{\sum_{i=1}^{n}(c_2 x_i - c_2\bar{x})(c_1 y_i - c_1\bar{y})}{\sum_{i=1}^{n}(c_2 x_i - c_2\bar{x})^2} = \dfrac{c_1}{c_2}\hat{\beta}_1$

$$\tilde{\beta}_0 = c_1\bar{Y} - c_2\bar{X}\frac{c_1}{c_2}\hat{\beta}_1 = c_1\hat{\beta}_0$$

(ii) $\tilde{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(c_1 + y_i)}{\sum_{i=1}^{n}\left([c_2 + x_i] - \frac{1}{n}\sum_{i=1}^{n}(c_2 + x_i)\right)^2} = \dfrac{\sum(x_i - \bar{x})y_i}{\sum(c_2 + x_i - c_2 - \bar{x})^2} = \hat{\beta}_1$

$$\tilde{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n}(c_1 + y_i) - \frac{1}{n}\sum_{i=1}^{n}(c_2 + x_i)\hat{\beta}_1$$

$$= c_1 + \bar{y} - \hat{\beta}_1 c_2 - \bar{x}\hat{\beta}_1 = c_1 + \hat{\beta}_0 - \hat{\beta}_1 c_2$$

(iii) $$\tilde{\beta}_1 = \frac{\Sigma (x_i - \bar{x}) \log c_1 y_i}{\Sigma (x_i - \bar{x})^2}$$

$$= \frac{\Sigma (x_i - \bar{x}) (\log c_1 + \log y_i)}{\Sigma (x_i - \bar{x})^2}$$

$$= \frac{\Sigma (x_i - \bar{x}) \log y_i}{\Sigma (x_i - \bar{x})^2} = \hat{\beta}_1$$

$$\tilde{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} \log (c_1 y_i) - \bar{x} \hat{\beta}_1$$

$$= \log c_1 + \frac{1}{n} \sum_{i=1}^{n} \log y_i - \bar{x} \hat{\beta}_1$$

$$= \log c_1 + \hat{\beta}_0$$

(iv) $$\tilde{\beta}_1 = \frac{\Sigma_{i=1}^{n} \left(\log c_2 + \log x_i - \left(\frac{1}{n} \Sigma_{i=1}^{n} (\log c_2 + \log x_i) \right) \right) y_i}{\Sigma_{i=1}^{n} \left(\log c_2 + \log x_i - \frac{1}{n} \Sigma (\log c_2 + \log x_i) \right)^2}$$

$$= \hat{\beta}_1$$

$$\tilde{\beta}_0 = \bar{y} - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^{n} (\log c_2 + \log x_i)$$

$$= \bar{y} - \hat{\beta}_1 \left(\log c_2 + \frac{1}{n} \Sigma_{i=1}^{n} \log x_i \right)$$

$$= \hat{\beta}_0 - \hat{\beta}_1 \log c_2$$

12. (i) Let $S_n(b_0) = \sum_{i=1}^{n} (y_i - b_0)^2$. Then,

$$\frac{d S_n(b_0)}{b_0} = \sum_{i=1}^{n} 2(y_i - b_0)(-1).$$

Setting $\frac{d S_n}{d b_0}(\hat{b}_0) = 0$ gives $\sum_{i=1}^{n} (y_i - \hat{b}_0) = 0 \Leftrightarrow$

$\Sigma y_i = n\hat{b}_0 \Leftrightarrow \hat{b}_0 = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$. So $\tilde{\beta}_0 = \bar{y}$.

(ii) $\tilde{u}_i = y_i - \bar{y}$

$$\sum_{i=1}^{n} \tilde{u}_i = \Sigma y_i - n\bar{y} = 0.$$