Exercise Session - Problem Set 3

The Simple Linear Regression Model

Problem 1 (Wooldridge, Example 2.2, p. 61)

The following table contains the ACT (achievement examination for college admissions in the US) scores and the GPA (grade point average) for 8 college students. Grade point average is based on a four–point scale and has been rounded to one digit after the decimal.

<table>
<thead>
<tr>
<th>Student</th>
<th>GPA</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>3.4</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>3.5</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>3.6</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>3.0</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>2.7</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>3.7</td>
<td>30</td>
</tr>
</tbody>
</table>

a) Estimate the relationship between GPA and ACT using OLS, i.e., obtain the OLS estimates of the equation

\[ \hat{\text{GPA}} = \hat{\beta}_0 + \hat{\beta}_1 \text{ACT}. \]  

(1)

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? How much higher is GPA predicted to be, if the ACT score is increased by 5 points?

b) What is the predicted value of GPA when ACT = 20?

c) How much of the variation in GPA for these 8 students is explained by ACT? Explain.

Problem 2 (Wooldridge, Example 2.1, p. 61)

In the simple linear regression model \( y = \beta_0 + \beta_1 x + u \), suppose \( E(u) = \alpha_0 \neq 0 \). Show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has zero mean.

Problem 3

In the simple linear model, what happens to the OLS intercept and slope estimates if each observation of the dependent (independent) variable is multiplied by a constant \( c \)? (Economically, this amounts to a change of the unit of measurement of the variables in question.)
Problem 4 (Wooldridge, Problem 2.3, page 62)

Let \textit{kids} denote the number of children ever born to a woman, and let \textit{edu} denote years of education for the woman. A simple model relating fertility to years of education is

\[ \text{kids} = \beta_0 + \beta_1 \cdot \text{edu} + u, \]  

(2)

where \( u \) is the unobserved error.

a) What kind of factors are contained in \( u \)? Are these likely to be correlated with level of education?

b) Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

Problem 5

Consider a random sample of size \( n \), i.e., \( \{(x_i, y_i), i = 1, \ldots, n\} \), from the simple linear model \( y = \beta_0 + \beta_1 x + u \) with the assumptions \( E(u|x) = 0 \) and \( \text{Var}(u|x) = \sigma^2 \).

i) Find \( \text{Var}(\hat{\beta}_0) \), \( \text{Var}(\hat{\beta}_1) \), and \( \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \).

ii) Show that

\[ \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2 \]  

(3)

is an unbiased estimator of \( \sigma^2 \). You may, for example, first show

\[ \text{Var}(\hat{y}_i) = \frac{\sigma^2}{n} \left[ 1 + \frac{(x_i - \bar{x})^2}{s_x^2} \right], \quad s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2, \]  

(4)

and then use \( \hat{u}_i = y_i - \hat{y}_i = \epsilon_i - [\hat{y}_i - E(\hat{y}_i)] \), where \( \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \).