Exercise 2.4.2

$$e_i = 0.03 - 0.06\text{EL2}_i - 0.09\text{EL3}_i + 0.06\text{EL4}_i + \text{res}_i$$
$$R^2 = 0.04$$

(a) Give an intuitive interpretation of the four regression coefficients.

For $EL1 = 1$, $e_i = 0.03$. Actual wage is about 3% higher than predicted by the current model for education level 1.
For $EL2 = 1$, $e_i = 0.03 - 0.06 = -0.03$. Actual wage is about 3% lower than predicted by the current model for education level 2.
For $EL3 = 1$, $e_i = 0.03 - 0.09 = -0.06$. Actual wage is about 6% lower than predicted by the current model for education level 3.
For $EL4 = 1$, $e_i = 0.03 + 0.06 = 0.09$. Actual wage is about 9% higher than predicted by the current model for education levele 4.

(b) Test if the three dummy coefficients are jointly significant, by means of the F-test

$$F = \frac{(R_1^2 - R_0^2)/g}{(1 - R_1^2)/(n-k)}$$

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$
$$e_i = \beta_1 + \text{res}_i$$
$$e = X\beta_1 + \text{res}$$
$$\text{where } x = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (x \times 1) \text{ vector}$$

$$
\begin{aligned}
b &= (X^T X)^{-1} X^T y \\
\hat{\beta}_1 &= (X^T X)^{-1} X^T e \\
&= \left((1\ 1\ ...\ 1) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right)^{-1} (1\ 1\ ...\ 1) \begin{pmatrix} e_1 \\ e_0 \\ \vdots \\ e_n \end{pmatrix} \\
&= n^{-1} (1\ 1\ ...\ 1) \begin{pmatrix} e_1 \\ e_0 \\ \vdots \\ e_n \end{pmatrix} \\
&= \frac{1}{n} \sum_{i=1}^{n} e_i = \bar{e}
\end{aligned}
$$

$$
\begin{aligned}
\text{res}_i &= e_i - \hat{\beta}_i \\
&= e_i - \bar{e}
\end{aligned}
$$

$$
\begin{aligned}
SSR &= \sum \text{res}_i^2 = \sum (e_i - \bar{e})^2 \\
SST &= \sum (y_i - \bar{y})^2 = \sum (e_i - \bar{e})^2 \\
R_0^2 &= 1 - \frac{\text{res}_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{1}{1} = 0
\end{aligned}
$$

F-test :
$R_1^2 = 0.04$ (given)
$R_0^2 = 0$
$g = 3 \ \ (\beta_2, \beta_3, \beta_4)$
$n = 500$ (given)
$k = 4 \ \ (\beta_1, \beta_2, \beta_3, \beta_4)$
Critical value $f_{0.025,3,496} = 2.6$

$$
F = \frac{(R_1^2 - R_0^2)/g}{(1 - R_1^2)/(n-k)} = \frac{(0.04 - 0)/3}{(1 - 0.04)/(496)} = 6.89
$$

$F = 6.89 > 2.6$, reject H_0

(c) Give an economic interpretation of the result in part (b).

$$\begin{aligned} \text{Original} \;\; \log(\text{Wage})_i &= 3.05 - 0.04\text{Female}_i + 0.03\text{Age}_i + 0.23\text{Educ}_i \\ &\quad - 0.37\text{Parttime}_i + \epsilon_i \\ \text{New} \;\; \log(\text{Wage})_i &= 3.32 - 0.03\text{Female}_i + 0.03\text{Age}_i + 0.17\text{EL2}_i + 0.38\text{EL3}_i \\ &\quad + 0.78\text{EL4}_i - 0.37\text{Parttime}_i + \epsilon_i \end{aligned}$$

Model with fixed education level effects gives systematically biased wage forecast for each education level.

For example:
In the original model, for education level 2, wage is $e^{0.23} = 26\%$ higher. In the new model, wage is only $e^{0.17} = 19\%$ higher.

(d) Above, it was stated that the residuals res_i have sample mean zero for each of the four education levels. Can you prove this result?

$$\begin{aligned} & X^T e = 0 \\ & X^T \text{res} = 0 \\ & \sum \text{res}_i = \sum_{\text{Educ}_i=1} \text{res}_i + \sum_{\text{Educ}_i=2} \text{res}_i + \sum_{\text{Educ}_i=3} \text{res}_i + \sum_{\text{Educ}_i=4} \text{res}_i = 0 \\ \text{Regression property} \;\; & \sum_{\text{Educ}_i=2} \text{res}_i = \sum_{\text{Educ}_i=3} \text{res}_i = \sum_{\text{Educ}_i=4} \text{res}_i = 0 \\ & \therefore \sum_{\text{Educ}_i=1} \text{res}_i = 0 \end{aligned}$$

Since the sum of residual is zero, the sample mean for each education level must be zero.