PROBLEM SET I: ANSWERKEY

1 Problem 1

The dependent variable is **luwe** (log weekly wage) and the independent variables are **educ** (years of education), **exper** (years of experience) and **exper** squared.

1. Estimate a linear regression model for log wages on education, experience, and experience squared. Report the estimates, the standard errors of the estimates, the R^2 and the standard deviation of the random error.

   Table 1: ESTIMATES AND STANDARD ERRORS

	intercept	educ	exper	expersquared	$\hat{\sigma}$
estimate	4.0163	0.0923	0.0791	-0.0020	.410
s.e.	(0.2223)	(0.0076)	(0.0249)	(0.0009)	

   Further $R^2 = .143$.

2. Predict the effect on average log wage of increasing everybody's education level by one year.

   Let the regression be

   $$\log(\text{wage})_i = \beta_0 + \beta_1 \times \text{educ}_i + \beta_2 \times \text{exper}_i + \beta_3 \times \text{exper}_i^2 + \varepsilon_i.$$

   For individual i with current level of education educ_i and level of experience exper_i, the effect of increasing their education level by one year is

   $$\theta_i = \beta_1 - \beta_2 - \beta_3 \cdot (2 \cdot \text{exper}_i - 1).$$

   The effect we are interested in is

   $$\theta = \beta_1 - \beta_2 - \beta_3 \cdot (2 \cdot \overline{\text{exper}} - 1).$$

Therefore the vector c in Lecture 5 is

$$c = (0\ 1\ \ -1\ 2\overline{\text{exper}} - 1)'$$

We find an effect of 0.0647, with a standard error of 0.0062.

3. Can you obtain the above effect by running a regression with a redefined set of covariates? How?

   Writing

$$\beta_1 = \theta + \beta_2 + \beta_3 \cdot (2 \cdot \overline{\text{exper}} - 1),$$

   we can write the regression function as

$$\log(\text{earnings})_i = \beta_0 + (\theta + \beta_2 + \beta_3 \cdot (2 \cdot \overline{\text{exper}} - 1)) \times \text{educ}_i$$

$$+\beta_2 \times \text{exper}_i + \beta_3 \times \text{exper}_i^2 + \varepsilon_i,$$

   or

$$\log(\text{earnings})_i = \beta_0 + \theta \times \text{educ}_i + \beta_2 \times (\text{exper}_i + \text{educ}_i)$$

$$+\beta_3 \times (\text{exper}_i^2 + (2 \cdot \overline{\text{exper}} - 1) \cdot \text{educ}_i) + \varepsilon_i.$$

   Hence we do ols with an intercept, education, education plus experience and experience squared plus education times two times average education minus one. The results are in the next table. So, the coefficient of interest is the coefficient on education, equal

Table 2: Estimates and Standard Errors

	intercept	educ	$\text{exper}_i + \text{educ}$	$\text{exper}_i^2 + (2 \cdot \overline{\text{exper}} - 1) \cdot \text{educ}_i$	$\hat{\sigma}$
estimate	4.0163	0.0647	0.0791	-0.0020	0.410
s.e.	(0.2223)	(0.0062)	(0.0249)	(0.0009)	

   to 0.0647, with a standard error of 0.0062.

4. Predict the effect on the average level of earnings of the following policy: increase the level of education for those who currently have earnings below 12 years of education to 12, and leave the level of education for others unchanged.

   The parameter of interest is the expected difference in average earnings where we change the education level of those with less than twelve years of education to exactly twelve years of education. The expected level of earnings (use the formula for the mean of a lognrmal distribution) for someone with education level educ and experience exper is

   $$\exp(\beta_0 + \beta_1 \cdot \text{educ} + \beta_2 \cdot \text{exper} + \beta_3 \cdot \text{exper}^2 + \sigma^2/2).$$

   Hence the parameter of interest is

   $$\theta = \frac{1}{N} \sum_{i|\text{educ}_i<12} \left(\exp\left(\beta_0 + 12\beta_1 + \beta_2(\text{exper}_i - 12 + \text{educ}_i) + \beta_3(\text{exper}_i - 12 + \text{educ}_i)^2 + \frac{\sigma^2}{2} \right) \right.$$
   $$\left. - \exp\left(\beta_0 + \beta_1\text{educ}_i + \beta_2\text{exper}_i + \beta_3\text{exper}_i^2 + \frac{\sigma^2}{2} \right) \right)$$

   With the estimates from Table 1, the estimated coefficient is 4.4187 (in dollars per week). (Note that only 88 out of the 935 individuals are affected by this policy because they have education levels less than 12.)

2 Problem 2

Prove partial regression formula. Consider the model

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

The normal equations are given by

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} \tag{1}$$

Solving for β_1 we get

$$\beta_1 = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2\beta_2 = (X_1'X_1)^{-1}X_1'(y - X_2\beta_2)$$

Using the second equation in (1) we get

$$X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y.$$

Substituting the expression for β_1 we get

$$X_2'X_1(X_1'X_1)^{-1}X_1'y - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 + X_2'X_2\beta_2 = X_2'y,$$

which, after some rearrangement, yields

$$\beta_2 = [X_2'(I - X_1(X_1'X_1)^{-1}X_1'X_2]^{-1}[X_2'(I - X_1(X_1'X_1)^{-1}X_1')y] = (X_2'M_1X_2)^{-1}(X_2'M_1y)$$

Hence, we can write

$$\beta_2 = (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}y^*,$$

where $X_2^* = M_1X_2$ and $y^* = M_1y$.

For the second part of the question, note that

$$\begin{aligned}
\beta_2 &= (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}y^* = (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}M_1y &(2)\\
&= (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}(I - X_1'(X_1'X_1)^{-1}X_1)y &(3)\\
&= (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}y - (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}X_1'(X_1'X_1)^{-1}X_1y &(4)\\
&= (X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}y &(5)
\end{aligned}$$

because $(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}X_1'(X_1'X_1)^{-1}X_1y = 0$ since $X_2^{*\prime} = M_1X_2$ is the projection of X_2 onto the space orthogonal to X_1. Hence, $X_2^{*\prime}X_1 = 0$.