Multiple Regression

Problem 1. Let a regression equation be:
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \,; \ t = 1,2,...,T$$
What are normal equations of OLS for the multiple linear regression in covariance form? Use sample variance and sample covariance functions to represent normal equations of OLS **without taking derivatives**.

Problem 2. Derive formulas for regression coefficients in the case of two regressors?
$$b_2 = \frac{\text{Cov}(X_2,Y)\text{Var}(X_3) - \text{Cov}(X_3,Y)\text{Cov}(X_2,X_3)}{\text{Var}(X_2)\text{Var}(X_3) - [\text{Cov}(X_2,X_3)]^2}$$
$$b_3 = \frac{\text{Cov}(X_3,Y)\text{Var}(X_2) - \text{Cov}(X_2,Y)\text{Cov}(X_2,X_3)}{\text{Var}(X_2)\text{Var}(X_3) - [\text{Cov}(X_2,X_3)]^2}$$

Problem 3. By definition $R^2_{adj} = \overline{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$. Compare it with the definition of conventional R^2, what is the difference? Comment on the meaning of $(n-k)$ and $(n-1)$ in this formula.

R2 adjusted

Problem 4. Show that $\overline{R}^2 = 1 - (1 - R^2)\dfrac{n-1}{n-k}$.

Problem 5. The $\overline{R}^2$ coefficient increases if and only if the absolute value of t-statistic of the added variable coefficient is greater than 1.

FINAL CHALLENGE
Problems from real UoL and ICEF exams

Problem 6 (UoL Exam). The following equation was estimated by Ordinary Least Squares using 37 annual observations of UK aggregate data. The dependent variable ($cloth_t$) is the log of expenditure on clothing at 1995 prices, yd_t is the log of aggregate disposable income at 1995 prices, pc_t is the log of the price of clothing relative to all consumer prices, ps_t is the log of the price of shoes relative to all consumer prices.
$$cloth_t = -3.256 + 1.021yd_t - 0.240pc_t - 0.429ps_t + e_t$$
$$\qquad\quad (1.531)\ \ (0.118)\qquad (0.132)\qquad (0.185)$$
standard errors in brackets, e_t is an OLS residual. $R^2 = 0.992$

(a) Test the hypothesis that the coefficient of yd_t is one.

(b) Construct a 95% confidence interval for the coefficient pc_t.

(c) Test the hypothesis that all slope coefficients in the equation above are zero. Give any assumptions which your results in b) and c) require.

1

Problem 7 (ICEF Exam). Two students A and B are trying to answer the following question: what is better for future earnings – to study or to work, and so to get working experience. They collect data on 28 people working in different companies on their schooling S_i (in years), their working experience W_i (also in years) and current hourly earnings $EARN_i$ (in dollars). They also calculate the total number of years of active life spent on work or study $A_i = S_i + W_i$. It is assumed that one can not work and study at the same time.

The student A runs the following equation

$$\widehat{EARN}_i = -22.96 + 2.44 \cdot S_i + 0.92 \cdot W_i \qquad R^2 = 0.26 \qquad \text{(1)}$$
$$(17.48)\ (0.89) \quad (0.85)$$

The student B using the same data estimates another equation

$$\widehat{EARN}_i = -22.96 + 2.44 \cdot A_i - 1.52 \cdot W_i \qquad R^2 = 0.26 \qquad \text{(2)}$$
$$(17.48)\ (0.89) \quad (0.68)$$

(a) . Give interpretation to the coefficients of equation (1) and to all coefficients except that of W_i in the second equation. Explain why the coefficient of S_i in equation **(1)** is equal to the coefficient of A_i in equation **(2)**. Explain why the constant terms in equations **(1)** and **(2)** are equal.

(b) The student A claims that spending one extra year to study is 2.5 times more useful for future earnings than spend it on work. Is he right? Give interpretation to the coefficient of W_i in the equation **(2)** and explain why it has negative sign. Is it possible to evaluate this coefficient using information from equation **(1)**?

(c) Evaluate the significance of the coefficients. Evaluate the significance of the equation **(1)** using F-statistic. Is it possible to get F-statistic for equation **(2)** without calculation? Why the equations **(1)** and **(2)** have the same R^2? Why the standard error of the coefficient of S_i in the equation **(1)** is equal to the standard error of the coefficient of A_i in the equation **(2)**?

Additional problems

Problem 8. Derive the formula of decomposition of the coefficient's estimator $b_3 = \hat{\beta}_3$

$$b_3 = \beta_3 + \left(\frac{\text{Cov}(X_3,u)\text{Var}(X_2) - \text{Cov}(X_2,u)\text{Cov}(X_3,X_2)}{\Delta} \right)$$

Write formula of decomposition for the coefficient's estimator $b_2 = \hat{\beta}_2$

Problem 9. Prove that OLS estimators b_2 of the coefficient β_2 and b_3 of the coefficient β_3 in multiple regression are unbiased. What assumptions are used in the proof?

Problem 10. Show that $\overline{R}^2 = R^2 - \dfrac{k-1}{n-k}(1-R^2) = \dfrac{ESS}{TSS} - \dfrac{k-1}{n-k} \cdot \dfrac{RSS}{TSS}$.