Module 12: Multiple regression variance of least squares estimators practice problems

(The attached PDF file has better formatting.)

** Exercise 12.1: Multiple regression variance of least squares estimators

A regression equation with two explanatory variables is $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$, with

- $X_1 = (1, 2, 3, 4, 5) \Rrightarrow \sum(x_i - \overline{x})^2 = 10$
- RSS (the residual sum of squares) = 9.3333
- N (number of observations) = 5
- $\rho(X_1, X_2)$, the correlation of X_1 and X_2, is -0.7

A. What is s^2, the ordinary least squares estimator for σ^2_ε?
B. What is $\sum(x_i - \overline{x})^2$?
C. What is $R^2_{1,2}$, the R^2 of X_1 regressed on X_2?
D. What is $\sigma^2(B_1)$, the variance of B_1, where B_1 is the ordinary least squares estimator of β_1?
E. What is $\sigma(B_1)$, the standard error of B_1?

Part A: s^2, the ordinary least squares estimator for σ^2_ε (the square of the standard error of the regression) = the residual sum of squares (RSS) divided by the degrees of freedom, or N–k–1 (the number of observations – the number of explanatory variables – 1) = 0.93333 / (5 - 2 - 1) = 4.6667.

Part B: $\overline{x} = (1+2+3+4+5) / 5 = 3$, so $\sum(x_i - \overline{x})^2 = (-2)^2 + (-1)^2 + (0)^2 + (+1)^2 + (+2)^2 = 10$.

Part C: $\rho(X_1, X_2) = -0.7$, so $R^2 (X_1, X_2) = 0.49$. The values of $\rho(X_1, Y) = 0.1$ and $\rho(X_2, Y)$ are not relevant.

Part D: $\sigma^2(B_1)$, the variance of B_1, = $[\ s^2 / \sum(x_i - \overline{x})^2\ ] / [\ 1 - R^2 (X_1, X_2)\ ] = 0.91503$. The estimated value of B_1 is not relevant.

Part E: $\sigma(B_1)$, the standard error of B_1, is the square root of $\sigma^2(B_1)$, the variance of B_1: $0.91503^{0.5} = 0.95657$.

With two explanatory variables, the square of the correlation of X_1 with X_2 is the same as the R^2 value of X_1 regressed on X_2. With more than two explanatory variables, we must use the R^2 of the regression, not the correlations among the explanatory variables.

** Exercise 12.2: Degrees of freedom of F-statistic

A regression model has *N* data points, *k* explanatory variables (ß's), and an intercept.

A. An F-test for the null hypothesis that **q** slopes are 0 has how many degrees of freedom in the numerator?
B. This F-test has how many degrees of freedom in the denominator?

Part A: The F-test says: "How much additional predictive power does the model under review have compared to what we would otherwise use, as a ratio to the total predictive power of the model under review?" Each part of this ratio is adjusted for the degrees of freedom.

The degrees of freedom in the numerator adjusts for the extra predictive power of the model under review stemming from additional explanatory variables. If the model under review has one extra explanatory variable, it predicts better even if this extra explanatory variable has no actual correlation with the response variable. The degrees of freedom is the number of extra explanatory variables, or **q**.

If the F-test has a p-value of P% with q degrees of freedom in the numerator, its p-value is more than P% with q+1 degrees of freedom in the numerator. A higher p-value means that it is more likely that the observed increase in predictive power reflects the spurious effects of additional explanatory variables.

Part B: The degrees of freedom for the model under review is N – *k* – 1; this is the degrees of freedom in the denominator of the F-ratio. As N increases but no other parameters change, the additional predictive power of the model under review is less likely to be spurious (more likely to be real), so the p-value decreases