Fox Module 10 variances correlations regression values practice problems

(The attached PDF file has better formatting.)

** Exercise 1.1: Variances, correlations, and regression output

A linear regression of Y on X from a bi-variate distribution with 100 observations has

- The sample variance of the observed X-values, or $\sigma^2(x)$, = 40.
- The sample variance of the observed Y-values, or $\sigma^2(y)$, = 60.
- The correlation of the observed X-values with the observed Y values is –50%.

Jacob: What does a bi-variate distribution mean?

Rachel: This linear regression is an observational study. Both X and Y are sampled from the population, so we compute the sample variance of each. The two random variables are correlated.

A. What is TSS, the total sum of squares?
B. What is the R^2 of the regression?
C. What is RegSS, the regression sum of squares?
D. What is RSS, the residual sum of squares?
E. What is B, the ordinary least squares estimate of β?
F. What is S_E^2 (or s^2), the least squares estimate of the variance of the error term σ^2_ε?
G. What is the variance of B, the ordinary least squares estimate of β?
H. What is the standard error of B, the ordinary least squares estimate of β?
I. What is the *t* value to test the null hypothesis that β = 0?
J. What is the *F* value to test the null hypothesis that β = 0?

Part A: The total sum of squares TSS is $\sigma^2(y) \times (n - 1) = 60 \times (100 - 1) = 5{,}940$.

Part B: R^2 is the square of ρ(x,y), the correlation between the explanatory variable and the response variable:

$$(-0.50)^2 = 0.25$$

Part C: The regression sum of squares RegSS is the TSS times the R^2:

$$5{,}940 \times 0.25 = 1{,}485$$

Part D: The residual sum of squares RSS is TSS – RegSS:

$$5{,}940 - 1{,}485 = 4{,}455$$

Part E: B, the ordinary least squares estimate of β, is $\sum(x_i - \overline{x})(y_i - \overline{y}) / \sum(x_i - \overline{x})^2$ =

$$\text{covariance}(x,y) / \text{variance}(x) =$$

$$\rho(x,y) \times (\sigma^2(x) \times \sigma^2(y))^{0.5} / \text{variance}(x) =$$

$$-0.50 \times (40 \times 60)^{0.5} / 40 = -0.612372$$

Part F: S_E^2 (or s^2), the least squares estimate of the variance of the error term σ^2_ε, is RSS / (n - k - 1) =

$$4{,}455 / (100 - 1 - 1) = 45.459184$$

Part G: The variance of B, the ordinary least squares estimate of β, is $S_E^2 / \sum(x_i - \overline{x})^2$ =

$$S_E^2 / (\sigma^2(x) \times (100 - 1)) =$$

$$45.459184 / (40 \times (100 - 1)) = 0.01147959$$

Part H: The standard error of B, the ordinary least squares estimate of β, is the square root of the variance:

$$0.01147959^{0.5} = 0.107126$$

Part I: The *t* value to test the null hypothesis that β = 0 is B / SE(B) =

$$-0.612372 / 0.107126 = -5.715486$$

Part H: The *F* value to test the null hypothesis that β = 0 is the square of the *t* value

$$(-5.715486)^2 = 32.6668$$

We can also compute the *F* value as (RegSS / 1) ÷ (RSS / (100 - k - 1) =

$$1{,}485 / (4{,}455 / 98) = 32.6667$$

[This example uses 6+ decimal places for some computations, so the two methods of computing the F-Ratio get the same figure. Final exam problems do not require 6 decimal place accuracy.]