Fox Module 12 Statistical inference for multiple regression

Fox’s analysis of Duncan’s prestige data (pages 105-110)

(The attached PDF file has better formatting.)

Fox regresses prestige on income and education (by occupation). Fox shows the results of each step; this posting provides an Excel workbook with the data and the cell functions. (Fox uses R, not Excel.)

On the attached workbook:

- Column B is the occupation.
- Column C is the type of occupation (professional, white collar, blue collar), not used in this regression.
- Column D is income, column E is education, and column F is prestige level. These three variables are scaled from 0 to 100. In later modules, we use logit transformations; this module uses the raw figures.

Average income for all occupations is 41.86667; average education is 52.55556 (row 50).

Column H shows the squared deviation for income, and column H shows the squared deviation for education.

Illustration: For accountants, $(62 – 41.86667)^2$ = 405.35, and $(86 – 52.55556)^2$ = 1,118.53.

We use the squared deviations to derive the standard errors of the beta coefficients.

The correlation of income and education (row 52) is 0.724512, using Excel’s CORREL built-in function.

The R^2 of income and education (row 53) is 0.724512^2 = 0.524918.

The multiple regression is computed here by the regression option in the data analysis tools. If you use Excel (instead of other statistical packages), you use the data analysis tools for your student project. The estimated parameters are

- A = –6.06466
- B (for income) = 0.598733
- B (for education) = 0.545834

The regression add-in computes the standard errors of the regression coefficients. This posting shows how the values are computes.

The predicted prestige values in the RESIDUAL OUTPUT are A + B_1 × income + B_2 × education. The residual is the actual prestige value minus the predicted prestige value.

Illustration: For accountants:

- The predicted prestige is –6.06466 + 0.598733 × 62 + 0.545834 × 86 = 77.9985.
- The residual is 82 – 77.9985 = 4.0015.
- The squared residual is 4.0015^2 = 16.0121.

[To fit the precise values in the Excel sheet, you need more decimal places.]

- The sum of squared residuals is 7,506.699.
- The degrees of freedom = 45 data points – 3 parameters = 42.
- The estimate of the standard error of the regression is 7,506.699 / 42 = 178.7309.

We compute the standard errors of the beta coefficients. The correction factor is 1 / √(1-0.524918) = 1.45083.

- Income: 1.45083 × √178.7309 / √26,271.20 = 0.11967.
- Education: 1.45083 × √178.7309 / √38,971.11 = 0.09825.

The standard errors are shown in the output of the *REGRESSION* add-in.

Jacob: How do we derive the *F*-statistic for the multiple regression analysis?

Rachel: The procedure is the same as for linear regression. The regression sum of squares (RegSS) is the square of (the predicted prestige – the average prestige). These figures are in Column Q. The *F*-statistic is 18090.47 / 178.7309 = 101.216.

Jacob: Do the final exam problems ask to work out regression coefficients for multiple regression?

Rachel: On the final exam, you work out regression coefficients for a simple linear regression. You may also be given sums of squares and correlations of the explanatory variables and asked to work out the standard errors of the regression coefficients in multiple regression. If you understand the examples in Fox's textbook, you should be able to solve the final exam problems.

Jacob: This posting just repeats the material in the textbook; is there anything new here?

Rachel: Nothing is new here. This posting simply shows all the figures in an Excel workbook.