# The Standard Error of Prediction

The standard error of prediction using simple linear regression has up to now been taken to be the residual standard deviation, on the basis that this was an estimate of the standard deviation of the "error process" which produced deviations of individual points away from the line. This ignores the fact that the residuals are deviations from the fitted line which is, in itself, subject to chance causes of variation. Thus, the residual standard deviation underestimates the standard error of prediction. It was suggested earlier that this underestimation could be ignored, especially for large values of n. In this appendix, we present the correction needed to allow for the uncertainty associated with the fitted line and study the extent of the underestimation involved in ignoring it.

As an aid to exposition, the material presented in this appendix is tied to the US Post Office example. However, the same development applies to any application of simple linear regression.

**Predicting Y when X = **

It is relatively easy to see what the standard error of prediction is if the anticipated volume coincides with the average volume in the data, that is, . In that case the predicted Manhours is given by



the average value of Manhours. , being based on data subject to chance variation, is itself subject to chance variation. Being a sample average, its standard error is given by the usual formula /√n. Thus,  estimates a point on the line, but that estimate is subject to error, the extent of which is measured by /√n.

Adding to the error of  as an estimate of a point on the line is the "error process" which produces deviations of individual points away from the line. The extent of this error is measured by . To combine the two sources of error[[1]](#footnote-1), we

square them to get variances,  and ,

add the variances, getting  and

take the square root, getting .

In this example, the estimated standard error of prediction is



The error involved in ignoring the factor 1/n in the standard error formula is relatively small in this example; the standard error is still 19, rounded to two significant figures. The magnitude of the error depends on the value of n; 1/n is smaller for larger values of n.

Exercise 1: Suppose that data had been collected over longer periods, e.g., 4 years, (n=52), 8 years, (n=104). Assuming that  is still 634 and  is still 19, calculate prediction intervals in each of these cases. Comment.

When the anticipated volume is some value X different from , the standard error of prediction has a more complicated formula:

.

The extra term under the square root sign, , reflects the fact that prediction is more uncertain when the anticipated volume differs from average; the level of uncertainty increases as X deviates from . However, the extra term will be dominated by its first factor of 1/n when n is large, in which case it may be ignored.

In general, therefore, the original simplified standard error of prediction may be used, provided the approximation is reasonable in the context of the problem in hand.

Exercise 2: Use a spreadsheet to calculate the standard error of prediction using the full formula and the simplified formula for each of the X values in the Post Office example. Make a table of the resulting values, the errors (both absolute and per cent) due to the simplification, and the corresponding X values. What are the minimum and maximum errors? What are the corresponding X values? Comment.

Exercise 3: Use a spreadsheet to make a graph which illustrates the formula for prediction error. Enter a range of Volume values in the first column; say 150 to 200 in steps of 1. In successive columns, enter the formulas for simple linear regression, regression + twice the standard error, regression – twice the standard error. Make line graphs of the last three columns against the first. Comment on the result.

1. Recall that standard deviations are not combined by addition, but variances are, in appropriate circumstances. See Statistical Analysis, Chapter 4, footnote 5, page 142. [↑](#footnote-ref-1)