How to report and interpret a line of best fit

Luc Hens

3 December 2014

1   Reporting the results

In a paper:

– don't include raw computer output, but report results in the standard way (see below);

– don't use x and y but the actual variable names (like price, income, quantity);

– report descriptive statistics (mean, standard deviation) of the variables;

– report the units of measurement (otherwise interpreting the results is impossible);

– include a scatter plot (exogenous variable on x axis, response on x axis) with the line of best fit;

– report the coefficient of correlation between the x and y variables (two decimals suffice), and interpret it. The coefficient of correlation r measures the strength of a linear relationship. See the textbook from your introductory statistics class (e.g., **?**, pp. 125-130): "Correlations are always between -1 and $+1$, but can take any values in between. A positive correlation means that the cloud slopes up; as one variable increases, so does the other. A negative correlation means that the cloud slopes down; as one variable increases, the other decreases." **?**, p. 128. To interpret the strength of the linear relationship, take the absolute value of r (that means: if r has a minus sign, omit it). If r in absolute value exceeds 0.80, the linear relationship is said to be strong; if r in absolute value is between 0.50 and 0.80, the linear relationship is said to be moderate; if r in absolute value is less than 0.50, the linear relationship is said to be weak.

– interpret the estimated slope coefficient (and the intercept, if relevant). Make sure you explain the meaning of the slope in a way relevant to the reader, that is, choose appropriate units for the marginal increase in x. Rule of thumb: discuss the effect of an increase of x in the same (or a similar) order of magnitude as the standard deviation of x. For example, if x is measured in miles and the standard deviation of x is 510 miles, discuss the marginal effect of an increase of x by, say, 1 000 miles (see example2.doc). Whether the effect is substantial or not is a matter

of judgement. In some cases you may want to express the slope as an elasticity (see below for details).

– interpret the coefficient of determination (R^2) If you include inferential statistics:

– report the 95% confidence intervals for the gradient and the y-intercept;

– interpret the p-value for the hypothesis that the slope coefficient is equal to zero. Note that statistical significance—measured by how small the p-value is—has nothing to do with substantive significance—measured by how large the estimated coefficient is.

2   The "test statistic bigger than 2" rule of thumb

Some authors (and most papers from the 1990s and before) unfortunately don't report p-values for testing the two-sided null hypothesis that a slope coefficient equals zero. When the sample size is sufficiently large (say, larger than 60), the p-value will be less than 5% when the test statistic takes a value of 2 or larger (verify using your TI-84 and normcdf() or tcdf()). You can then use the "test statistic bigger than 2" rule of thumb: *if the absolute value of the test statistic of a coefficient is bigger than 2, the null hypothesis that the coefficient is equal to zero can be rejected at the 5% level of significance.*

3   Expressing the gradient as elasticity

Sometimes, a slope coefficient is interpreted in terms of elasticity. That is especially (but not only) the case when the equation can be interpreted as a demand equation. The elasticity of y with regard to x measures by how many percent y changes if x increases by 1%:

$$\text{elasticity} = \frac{\text{percentage change in } y}{\text{percentage change in } x} = \frac{\Delta y/y}{\Delta x/x}$$

Economists often use price elasticity of demand:

$$\text{price elasticity of demand} = \frac{\Delta Q/Q}{\Delta P/P}$$

If price elasticity of demand is -0.5, this means that if the price (P) increases by 1%, the quantity demanded (Q) decreases by 0.5%. Similarly, income elasticity of demand is:

$$\text{income elasticity of demand} = \frac{\Delta Q/Q}{\Delta Y/Y}$$

The price elasticity of demand can be rewritten as:

$$\text{price elasticity of demand} = \frac{\Delta Q}{\Delta P} \times \frac{Q}{P}$$

Suppose you estimted demand as a straight line. Then the first factor ($\Delta Q/\Delta P$) is the estimated slope coefficient. The second factor (Q/P) will be different for

every point the demand line. Consequently, elasticity will be different at each point of the line. Therefore, estimates of the line of best fit usually report elasticity at the point of averages:

$$\text{price elasticity at sample mean} = \text{estimated slope coefficient} \times \frac{\text{ave of } Q}{\text{ave of } P}$$

In your report, write something like this:

> The estimated slope coefficient of -23.6 corresponds to a price elasticity at the sample mean of -0.52: this means that if the price increases by 1%, the quantity demanded is predicted to decrease by 0.52%.

Consult your principles of economics or intermediate microeconomics texts for more on elasticity.