Figure 13.2. SIMPLE LINEAR REGRESSION: Background and Review

The two videos whose contents are summarized in this Figure provide background information and review for our discussion of modelling relationships using simple linear regression. Program 7 (described overleaf on page 13.16) will be shown in class; Pro- gram 6 (described below) can be viewed privately for self-study at the Audiovisual Centre, room E2 1309.

Time Series

Program 6 in: *Against All Odds: Inside Statistics*

This program shows how to examine data that are measurements of the same variable at regular intervals of time. You will learn that such *time series* can show long-term *trends*, more or less regular up and down *cycles*, and regular *seasonal variation* as well as irregular fluctuations. You will also learn that these patterns can be made clearer by averaging over many individuals at one time or by averaging over time. Averaging in these ways is called *smoothing* the data.

The video begins by looking at circadian rhythms, the daily cycles of biological activity apparent in the behaviour of many living things. You will see how experiments with people kept isolated from the light of day help uncover the working of our biological clocks. Plots of time series of measurements are a basic tool in such studies.

A second biological example will look at the patterns of brain activity induced by a sentence with a nonsense ending. The overall pattern becomes visible only when we average over the responses to many sentences. This illustrates the effectiveness of smoothing.

A similar kind of smoothing is used to create stock market indexes, which are averages of the prices of many individual stocks. You will hear the author of a stock market newsletter argue that the market has regular cycles that can be predicted, and Dean Burton Malkiel of Yale's business school say that there is no solid evidence for regular market cycles. You will learn to see patterns over time more clearly by using *running medians* of three or five consecutive observations.

(*continued overleaf*)

Models for Growth

Program 7 in: *Against All Odds: Inside Statistics*

Like Program 6, this program looks at changes in a variable over *time*. Program 6 concerned general time series. Our attention now turns to data that show a regular pattern of growth over time, so regular that it is described by a simple mathematical model. To see the pattern of growth, plot the variable of interest y on the vertical axis of a scatterplot against time t on the horizontal axis. We are interested in both the *overall* pattern of growth and in *deviations* from the pattern.

Linear growth *adds* a fixed amount in each equal time period. A variable y that shows exact linear growth is described by the equation of a straight line, $y = a + bt$. The *slope b* is the amount added in one unit of time.

Real data will rarely show *exact* linear growth. If the plot against time is *approximately* straight, the overall pattern of growth is linear. To describe this pattern, *fit a line* to the data; that is, pass a line as closely as possible through the points. You can do this by eye, or you can use the *least squares* method; the details of how to calculate a least squares line are given in the next Program; only the *idea* appears in this Program. Linear growth is illustrated in the video by the increase in the height of young children.

The deviations from the overall pattern of linear growth are described by the *residuals.* A plot of the *estimated* residuals against time magnifies the deviations for easier inspection. Look for outlying points and for *systematic* patterns in the residuals.

If you are satisfied that linear growth describes the overall pattern of the data, you can use the fitted line for *prediction*. Prediction *out*side the time interval for which you have data is called *extrapolation*. The results of extrapolation are often inaccurate.

Exponential growth *multiplies* by a fixed amount in each equal time period. You can see the shape of the *exponential curves* by multiplying repeatedly by the same number and plotting the results. Compound interest produces exponential growth. Real data rarely show *exact* exponential growth, but may show a pattern of *approximate* exponential growth. The example used in the video is the number of acres defoliated by gypsy moths in successive years of an infestation.

You can check whether a variable y has the overall pattern of exponential growth by taking the *logarithm of each value.* Use a calculator to find logarithms and plot the logarithms against time. If the logarithms of y show an approximate *linear* pattern, y itself is approximately exponential. You can fit a straight line to the logarithms to describe exponential growth. The estimated residuals from this line describe the deviations from the exponential pattern (or model). If you are satisfied that the line fits the logarithms adequately, you can use it for prediction. The video briefly discusses an example that is presented in more detail in the text (third edition pages 181-188): the growth in world oil production over time. (The relevant data are tabulated below.)

Text Table 2.13: Annual world crude oil production, 1880-1994 [in barrels $\times 10^6$ (Mbbl)]

Year	Mbbl	Year	Mbbl	Year	Mbbl	Year	Mbbl	Year	Mbbl	Year	Mbbl
1880	30	1920	689	1950	3,803	1968	14,104	1980	21,722	1992	22,028
1890	77	1925	1,069	1955	5,626	1970	16,690	1982	19,411	1994	22,234
1900	149	1930	1,412	1960	7,674	1972	18,584	1984	19,837		
1905	215	1935	1,655	1962	8,882	1974	20,389	1986	20,246		
1910	328	1940	2,150	1964	10,310	1976	20,188	1988	21,338		
1915	432	1945	2,595	1966	12,016	1978	21,922	1990	22,100		