Figure 13.9. SIMPLE LINEAR REGRESSION: Case Study 2

The case study in this Figure is concerned with a machining process in the manufacture of crankshafts, a major component of normal automobile engines; it involves the following terminology and abbreviations:

- *journal* – a cylindrical bearing on a crankshaft;
- *microfinish* – the microscopic smoothness of the journal surface (see also the top of the first side of Figure 11.3a);
- *nominal* – the target value (in this instance, for the journal diameter);
- *micron* – a unit of length, one millionth of a metre, one thousandth of a millimetre;
- *USL* – upper specification limit;
- *LSL* – lower specification limit.

Example 13.9.1: In a journal machining process, a final *lapping* step improves the microfinish of the journal surface; the lapping process slightly decreases the journal diameter. The specifications for final journal diameter (expressed as deviation from nominal) are 0 ± 3 microns.

To control *final* journal diameter, the diameter of parts coming into the lapping process must be controlled. In an investigation to determine appropriate specifications for the *incoming* parts, 30 journals are measured before (x) and after (y) lapping; the data are tabulated (in order of decreasing incoming size) at the right.

(a) Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of **Y** on **X**.

(b) Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of σ for these data.

(c) Assess how well the regression model fits the data; give your reasons completely but concisely.

(d) Test the hypothesis the lapping process removes a *constant* amount of material regardless of the incoming part diameter.

(e) Find a 99% *confidence* interval for the *average* final diameter if the incoming diameter is $\bar{x}$; indicate briefly what this interval suggests about the *target* value for incoming diameter.

(f) Describe briefly conditions under which it would *not* be desirable to centre the diameters of the incoming parts on the target value.

(g) Find a 99% *prediction* interval for the final diameter of a journal for the two cases where the incoming part iameter is –1 micron and 4 microns; indicate briefly the information these two intervals convey.

(h) Indicate where you would place the specifications for *in*coming diameter; justify your answer.

Diameter (microns) Incoming	Final
2.8	0.7
2.6	0.5
2.5	0.9
2.3	0.6
2.3	0.8
2.2	0.7
2.1	0.9
2.0	0.1
2.0	0.4
1.9	0.3
1.4	−0.4
1.3	−0.3
1.3	−0.8
1.2	−0.6
1.2	−0.8
1.1	−0.4
1.1	−0.9
0.9	−0.4
0.7	−0.7
0.5	−1.1
0.5	−1.3
0.2	−1.0
0.1	−1.1
−0.3	−1.7
−0.3	−2.1
−0.4	−1.9
−0.4	−2.1
−0.7	−2.1
−1.1	−2.9
−2.2	−3.3

Solution: (a) From the n = 30 observations (x_j, y_j) given in the statement of the question, we find the following:

$\Sigma x_j = 28.8$, $\Sigma x_j^2 = 71.82$;

$\Sigma y_j = -20$, $\Sigma y_j^2 = 51.76$;

$\Sigma x_j y_j = 20.93$;

$\bar{x} = 0.96$, $\bar{y} = -0.\dot{6}$,

$SS_{xy} = 40.13$,

$SS_x = 44.172$,

$SS_y = 38.42\dot{6}$;

hence, the *estimates* of β_1 (the slope) and β_0 (the intercept) of the regression of **Y** on **X** are:

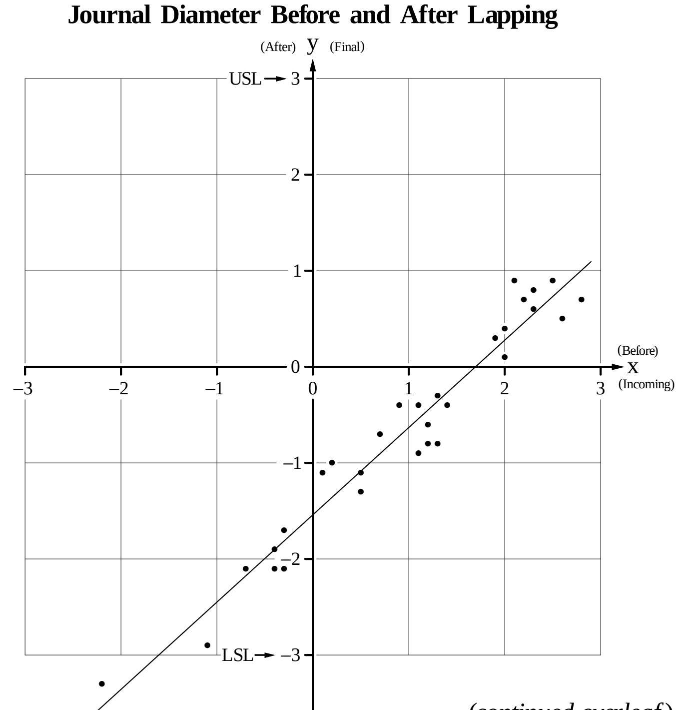
$b_1 = 0.908\,494\,068$,

$b_0 = -1.538\,820\,973$,

so the equation of the straight-line model is:

$$_{\text{reg}}\bar{y} = -1.5388 + 0.9085x$$
$$= -0.\dot{6} + 0.9085(x - 0.96).$$

Journal Diameter Before and After Lapping



(*continued overleaf*)

Example 13.9.1: (continued) (a) The scatter diagram of the data for the journal diameters before and after lapping, with the estimated regression of Y on X superimposed on it, is shown overleaf on page 13.47 at the lower right.

(b) The ANOVA table for these data is:

SOURCE	SUM of SQUARES	*Df*	MEAN SQUARE	*F*-RATIO
Model	$b_1 SS_{xy} = 36.457\,866\,975$	1	$MSM = 36.457\,866\,975$	$\frac{MSM}{MSE} = \frac{MSM}{\hat{\sigma}^2} \simeq 518.5$
Estimated residual	$SSE = 1.968\,799\,692$	28	$MSE = 0.070\,314\,275$	
Total	$SS_y = 38.42\dot{6}$	29	$(s_y^2 = 1.325\,057\,471)$	-----

Also, the coefficient of determination, the correlation coefficient and the estimate of σ are:

$r^2 = 0.948\,764\,754$, $r = 0.974\,045\,560$, $\hat{\sigma} = \sqrt{MSE} = 0.265\,168\,389\,364$.

(c) *Three* matters provide an assessment of how well the regression model fits the data:

- Visual inspection of the scatter diagram with the estimated regression of Y on X superimposed on it shows all the points lie reasonably close to the line – the largest estimated residual is for the journal with an incoming diameter of 2.1 microns;
  - the points appear to be scattered without obvious pattern on both sides of the line;
  - there does not appear to be any systematic change in the magnitude of the estimated residuals with increasing values of x, which is consistent with the assumption of constant σ.
- The *F*-ratio (518.5) is very high and so provides highly statistically significant evidence against the hypothesis $\beta_1 = 0$, indicating a meaningful regression.
- The value of the coefficient of determination shows that nearly 95% of the variation of the y_js about their average has been accounted for by the estimated regression of Y on X.

These matters show the straight-line model is a good fit to the data.

(d) We want to determine whether the *difference* between the final and incoming diameters is *constant*; *i.e.*, we want to check if: $Y - X = \text{constant}$ or: $Y = \text{constant} + X$;
hence, we want to test the hypothesis: $\beta_1 = 1$,
so the relevant test statistic is: $\dfrac{B_1 - \beta_1}{\hat{s.d.}(B_1)} \sim t_{n-2}$.

We have: $\hat{\sigma} = 0.265\,168\,389\,364$, $b_1 = 0.908\,494\,068$, $\Pr[-2.04841 \le t_{28} \le 2.04841] = 0.95$,

so that: $\hat{s.d.}(B_1) = \hat{\sigma}\sqrt{\dfrac{1}{SS_x}} = \hat{\sigma}\sqrt{\dfrac{1}{44.172}} \simeq 0.039\,897\,733$;

hence, under H: $\beta_1 = 1$, the value of the test statistic is: $\dfrac{0.908\,494 - 1}{0.039\,898} \simeq -2.293\,512$,

so the *P*-value is: $\Pr[|t_{28}| \le -2.293\,512] = 2 \times \Pr[t_{28} \ge 2.293\,512] \simeq 2 \times 0.014\,766 \simeq 0.029\,532 \simeq 0.03$.

We thus find that the data provide statistically significant evidence against H: $\beta_1 = 1$ and so conclude the lapping process does *not* remove a constant amount of material regardless of incoming part diameter.

NOTE: 1. A 95% CI for β_1 is: $0.908\,494 \pm 0.081\,726\,915 \implies (0.826\,767,\ 0.990\,221)$
or about $(0.827,\ 0.990)$ microns after/microns before;
in agreement with the result of the test of significance, the value $\beta_1 = 1$ lies *out*side this 95% CI.

(e) We want a 99% *confidence* interval (CI) for the *mean* $\mu_Y(x_j = 0.96)$, representing the *average* of the response variate Y when the study population, specified by the explanatory variate X, is journals with an incoming diameter of $\bar{x} = 0.96$ microns above target.

We have: $\hat{\sigma} = 0.265\,168\,389\,364$, $\hat{\mu}_Y(x_j = 0.96) = -0.\dot{6}$, $\Pr[-2.76326 \le t_{28} \le 2.76326] = 0.99$,

so that: $\hat{\sigma}\sqrt{\dfrac{(x_j - \bar{x})^2}{SS_x} + \dfrac{1}{n}} = \hat{\sigma}\sqrt{\dfrac{(0.96 - 0.96)^2}{44.172} + \dfrac{1}{30}} \simeq 0.048\,412\,903$;

hence, a 99% CI for $\mu_Y(x_j = 0.96)$ is: $-0.\dot{6} \pm 0.133\,777\,438 \implies (-0.800\,444,\ -0.532\,889)$
or about $(-0.800,\ -0.533)$ microns (below target).

Because this CI covers only values appreciably *below* zero, it indicates that the process is centred *below* the target for final journal diameter.

NOTE: 2. $\hat{\mu}_Y(x_j = \bar{x}) = \bar{y}$ and the form of the CI is actually: $\bar{y} \pm {}_{\alpha}t^*_{n-2} \times \hat{s.d.}(\bar{Y}) = \bar{y} \pm {}_{\alpha}t^*_{n-2} \times \hat{\sigma}\sqrt{\frac{1}{n}}$

Figure 13.9. SIMPLE LINEAR REGRESSION: Case Study 2 (continued 1)

Example 13.9.1: (continued) (f) Conditions under which it would *not* be desirable to centre the diameters of the incoming parts on the target value are when the process that produces the journals is not *capable* of meeting the specifications. It would then be better to centre the process somewhat *above* the target, because journals *above* the USL can be lapped a second time to make them smaller (and within specifications) whereas journals *below* the LSL can only be discarded as scrap; the latter is the *greater* of these two costs of poor quality. The much better alternative, of course, is to have a process capable of producing an acceptably high proportion of journals within specifications after *one* lapping operation.

(g) We want 99% *prediction* intervals (PIs) for the *random variables* $Y(x_j = -1)$ and $Y(x_j = 4)$, representing the response variate **Y** for an *individual* randomly-selected journal when the study populations, specified by the explanatory variate **X**, are journals with incoming diameters of –1 micron and of 4 microns.

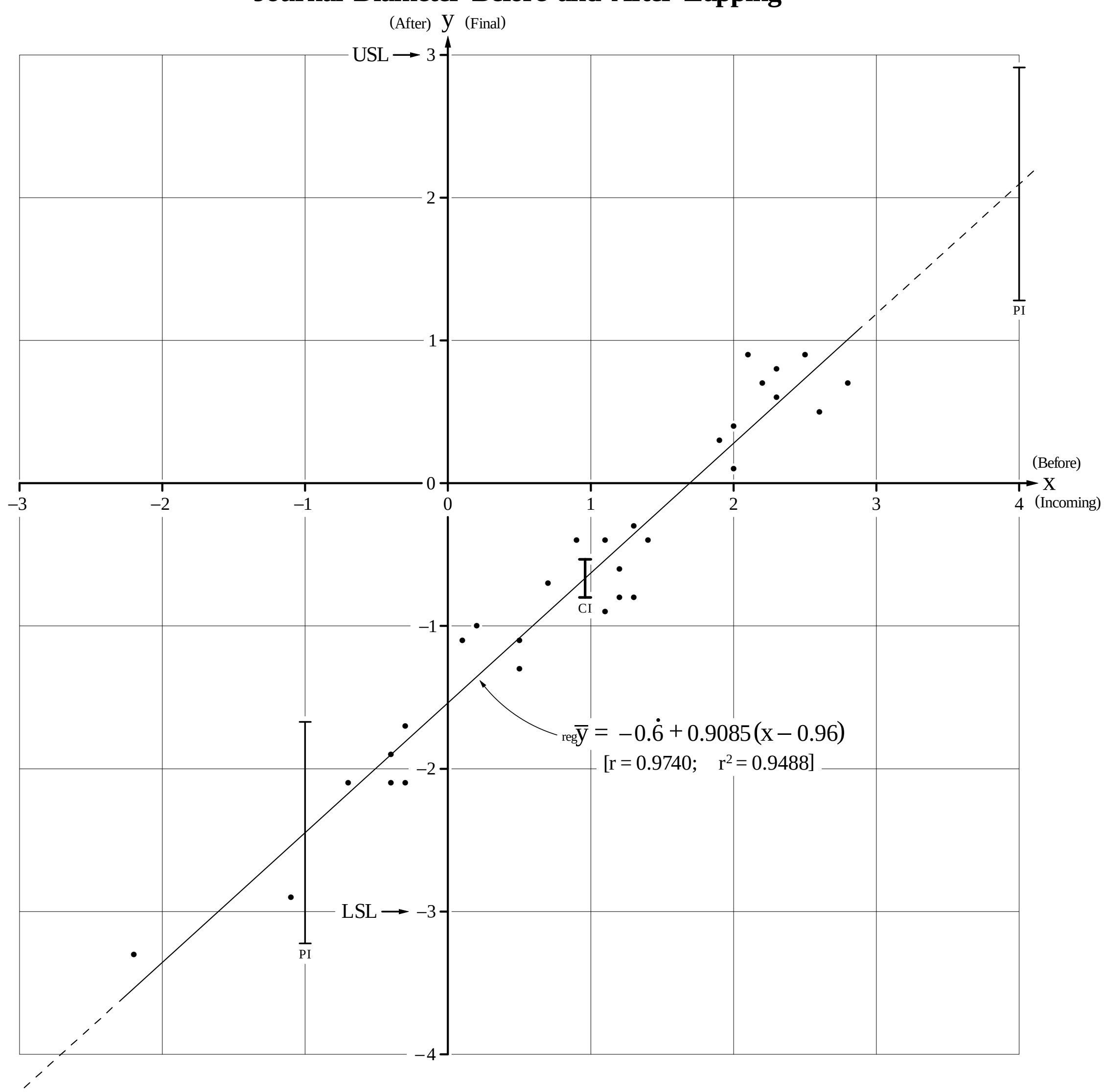
We have: $\hat{\sigma} = 0.265\,168\,389\,364$, $\hat{\mu}_Y(x_j = -1) = -2.447\,315\,041$, $\Pr[-2.76326 \le t_{28} \le 2.76326] = 0.99$,

so that: $\hat{\sigma}\sqrt{\dfrac{(x_j - \bar{x})^2}{SS_x} + \dfrac{1}{n} + 1} = \hat{\sigma}\sqrt{\dfrac{(-1 - 0.96)^2}{44.172} + \dfrac{1}{30} + 1} \simeq 0.280\,665\,734;$

hence, a 99% PI for $Y(x_j = -1)$ is: $-2.447\,315\,041 \pm 0.775\,552\,398 \Rightarrow (-3.222\,867, -1.671\,763)$
or about (–3.22, –1.67) microns.

[This PI, together with the one from overleaf on page 13.50 and the CI from (e) on the facing page 13.48, are shown on the diagram below.]

Journal Diameter Before and After Lapping



(*continued overleaf*)

Example 13.9.1: (continued) (g) And: $\hat{\sigma} = 0.265\,168\,389\,364$, $\hat{\mu}_Y(x_j = 4) = 2.095\,155\,299$, $\Pr[-2.76326 \le t_{28} \le 2.76326] = 0.99$,

so that: $\hat{\sigma}\sqrt{\frac{(x_j-\bar{x})^2}{SS_x} + \frac{1}{n} + 1} = \hat{\sigma}\sqrt{\frac{(4-0.96)^2}{44.172} + \frac{1}{30} + 1} \simeq 0.295\,582\,698$;

hence, a 99% PI for $Y(x_j = 4)$ is: $2.095\,155\,299 \pm 0.816\,771\,847 \implies (1.278\,383,\ 2.911\,927)$ or about $(1.28,\ 2.91)$ microns.

NOTE: 3. Two factors make the PIs in (g) so much *wider* than the CI in (e):

- a PI is an interval estimate for a random variable representing an *individual*, whereas a CI is an interval estimate for a *mean* representing a study population *average*;
- in this Example, the CI is for $x_j = \bar{x}$ in the *centre* of the data (where Answers are most *precise*), whereas the two PIs are for values of x_j towards or at the *ends* of the interval of observation (where Answers are *less* precise).

(h) The specifications for incoming diameter should be such that an appropriately high percentage of the (individual) journals *after* lapping are within the specifications of 0 ± 3 microns deviation from nominal;

- at the *left* of the scatter diagram overleaf on page 13.49, this means we want a PI whose *lower* limit is as close as possible to the LSL of −3 microns;
- at the *right* of the diagram, we want a PI whose *upper* limit is as close as possible to the USL of 3 microns.

For 99% within specifications for journals with incoming diameters at the lower and upper limits, by trial and error we find the relevant PIs are those for $\mathbf{X} = -0.76$ microns and and $\mathbf{X} = 4.09$ microns;

lower: $\hat{\sigma} = 0.265\,168\,389\,364$, $\hat{\mu}_Y(x_j = -0.76) = -2.229\,276\,465$, $\Pr[-2.76326 \le t_{28} \le 2.76326] = 0.99$,

so that: $\hat{\sigma}\sqrt{\frac{(x_j-\bar{x})^2}{SS_x} + \frac{1}{n} + 1} = \hat{\sigma}\sqrt{\frac{(-0.76-0.96)^2}{44.172} + \frac{1}{30} + 1} \simeq 0.278\,149\,872$;

hence, a 99% PI for $Y(x_j = -0.76)$ is: $-2.229\,276\,465 \pm 0.768\,600\,414 \implies (-2.997\,877,\ -1.460\,676)$ or about $(-3.00,\ -1.46)$ microns;

upper: $\hat{\sigma} = 0.265\,168\,389\,364$, $\hat{\mu}_Y(x_j = 4.09) = 2.176\,919\,765$, $\Pr[-2.76326 \le t_{28} \le 2.76326] = 0.99$,

so that: $\hat{\sigma}\sqrt{\frac{(x_j-\bar{x})^2}{SS_x} + \frac{1}{n} + 1} = \hat{\sigma}\sqrt{\frac{(4.09-0.96)^2}{44.172} + \frac{1}{30} + 1} \simeq 0.297\,074\,190$;

hence, a 99% PI for $Y(x_j = 4.09)$ is: $2.176\,919\,765 \pm 0.820\,893\,226 \implies (1.356\,027,\ 2.997\,813)$ or about $(1.36,\ 3.00)$ microns.

NOTE: 4. If we require a probability level *higher* than 99% within specifications for journals at the lower and upper limits, we could use 99.9% or 99.99%, for which:

$$\Pr[-3.67391 \le t_{28} \le 3.67391] = 0.999, \qquad \Pr[-4.53047 \le t_{28} \le 4.53047] = 0.9999.$$

The results of trial and error calculations for the relevant 99.9% and 99.99% PIs are as shown in the following table, which also contains the results for the 99% PIs given above:

Probability Level	Incoming Diameter			Lower Prediction Interval		Upper Prediction Interval	
	Lower	Upper	Difference	End points	Width	End points	Width
99%	−0.76	4.09	4.85	(−3.00, −1.46)	1.54	(1.36, 3.00)	1.64
99.9%	−0.49	3.81	4.30	(−3.00, −0.97)	2.03	(0.85, 3.00)	2.15
99.99%	−0.24	3.56	3.81	(−3.00, −0.52)	2.48	(0.39, 3.00)	2.61 .

We note three matters from the information in the table:

- even the *lowest* probability level (99%) is *higher* than the *overall* proportion (*ca*. 96.7%) of journals in the sample within specifications; [Why?]
- as the probability level *in*creases, we require a *narrower* range of incoming journal diameters;
- as the probability level *in*creases, the PIs become *wider*, reminding us that, for the *fixed* information content of a given set of data, an increased probability (or confidence) level is obtained only at the cost of a *wider* interval.

NOTE: 5. Of particular interest in this case study are:

- the interpretation of the *slope* of the regression line and the test of significance [in (d) on the second side (page 13.48) of the Figure];
- the use of *prediction intervals* in (h) in managing the lapping process to make it meet specifications.

ACKNOWLEDGEMENT: The context and data for this Figure were kindly provided by Professor R.J. Mackay.