5 Regression Models

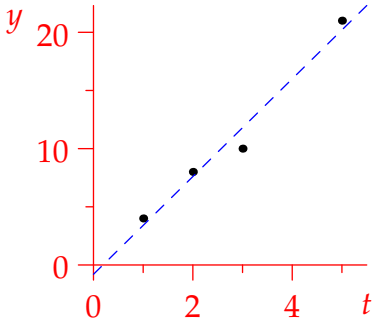
We've studied several types of function, and seen how to spot whether a particular model might suit a given data set. To get further with this analysis, we need a method for comparing and quantifying *how well a model fit given data*.

5.1 Best-fitting Lines and Linear Regression

We start with an example of some data which appears reasonably linear.

Example 5.1. At t p.m., a trail-runner's GPS locator says they've traveled y miles along a trail

t_i	1	2	3	5
y_i	4	8	10	21



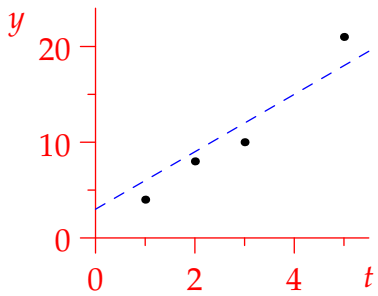
Where is the trail-runner at 6 p.m.?

To answer this, we'd like a simple model for how far the runner has traveled as a function of t.

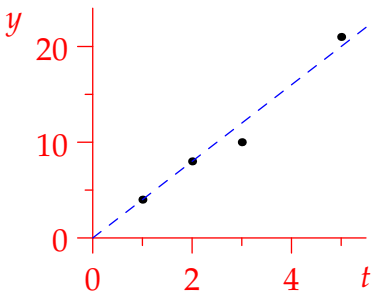
The data looks to be approximately[16] linear: $y \approx mt + c$. What is the *best* choice of line, and how should we find its coefficients m, c?

What might be good criteria for choosing our line? The points should plainly be *close* to the line, but how should this be measured?

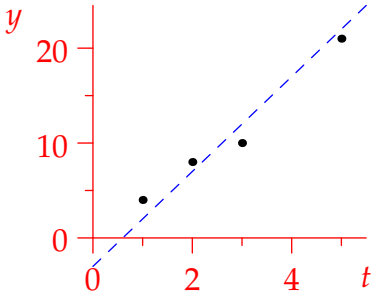
Here are three candidates: of the options, which line seems best and why?



$y = 3t + 3$



$y = 4t$



$y = 5t - 3$

To properly answer the question, consider what use we want to make of the approximating line. The model should accurately *predict* the runner's location $y \approx \hat{y} = mt + c$ at a given time t. As such, it should *minimize vertical errors* $\hat{y}_i - y_i$.

The vertical errors for the three candidates are shown in the table; since a positive error is as bad as a negative, we've made the errors positive. The second line certainly seems the best choice of the three.

But can we do better?

	t_i	1	2	3	5
	y_i	4	8	10	21
$y = 3t + 3$	$\lvert\hat{y}_i - y_i\rvert$	2	1	2	3
$y = 4t$	$\lvert\hat{y}_i - y_i\rvert$	0	0	2	1
$y = 5t - 3$	$\lvert\hat{y}_i - y_i\rvert$	2	1	2	1

Least-Squares Minimization

As seen in the example, if we want to use our model $\hat{y}(t)$ to *predict* y given t, it makes sense for the model to minimize the errors $|\hat{y}_i - y_i|$. One possibility is to minimize their *sum*:

$$\sum_{i=1}^{n} |\hat{y}_i - y_i|$$

For several reasons (computational simplicity, mathematical cleanliness, statistical interpretation, to discourage large individual errors), we *don't* do this! Instead, the standard approach is to minimize the sum of the *squared* errors.

Definition 5.2. Let (t_i, y_i) be data points with at least two distinct t-values. Let $\hat{y} = mt + c$ be a linear predictor (model) for y given t.

- The i^{th} *error* in the model is the difference $e_i := \hat{y}_i - y_i = mt_i + c - y_i$.
- The *regression line* or *best-fitting least-squares line* is the function $\hat{y} = mt + c$ which minimizes the sum of the squared-errors:

$$S := \sum e_i^2 = \sum (\hat{y}_i - y_i)^2$$

As we'll see later, having at least two distinct t-values guarantees the regression line is unique.

Example (5.1, cont). Suppose the predictor was $\hat{y} = mt + c$. We expand the table

t_i	1	2	3	5
y_i	4	8	10	21
$\hat{y}_i$	$m + c$	$2m + c$	$3m + c$	$5m + c$
e_i	$m + c - 4$	$2m + c - 8$	$3m + c - 10$	$5m + c - 21$

Our goal is to minimize the function

$$S(m, c) = \sum e_i^2 = (m + c - 4)^2 + (2m + c - 8)^2 + (3m + c - 10)^2 + (5m + c - 21)^2$$

This is easy if we invoke some calculus. If (m, c) minimizes S, then the first derivative tests says that the (partial) derivatives of S must be zero.

- Keep c constant and differentiate with respect to m:

$$\begin{aligned} \frac{\partial S}{\partial m} &= 2(m + c - 4) + 4(2m + c - 8) + 6(3m + c - 10) + 10(5m + c - 21) \\ &= 2\Big[39m + 11c - 155\Big] \end{aligned}$$

[16] Why should we not expect the distance traveled by the hiker to be perfectly linear?

- Now keep m constant and differentiate with respect to c:

$$\begin{aligned}\frac{\partial S}{\partial c} &= (m + c - 4) + (2m + c - 8) + (3m + c - 10) + (5m + c - 21) \\ &= 11m + 4c - 43\end{aligned}$$

The regression line is found by solving a pair of simultaneous linear equations

$$\begin{cases} 39m + 11c = 155 \\ 11m + 4c = 43 \end{cases} \implies m = \frac{21}{5},\ c = -\frac{4}{5} \implies \hat{y}(t) = \frac{1}{5}(21t - 4)$$

By 6 p.m., we predict that the runner would have covered $\hat{y}(6) = 24.4$ miles.

The sum of the squared errors for our regression line is $S = \sum e_i^2 = \sum |\hat{y}_i - y_i|^2 = 4.4$, compared to 18, 5 and 10 for our earlier options.

To obtain the general result for n data points, we return to our computations of the partial derivatives:

$$\frac{\partial S}{\partial m} = \sum \frac{\partial}{\partial m}(mt_i + c - y_i)^2 = 2\sum t_i(mt_i + c - y_i) = 2\Big[\Big(\sum t_i^2\Big)m + \Big(\sum t_i\Big)c - \sum t_i y_i\Big]$$
$$\frac{\partial S}{\partial c} = \sum \frac{\partial}{\partial c}(mt_i + c - y_i)^2 = 2\sum (mt_i + c - y_i) = 2\Big[\Big(\sum t_i\Big)m + nc - \sum y_i\Big]$$

These expressions are often written using a short-hand for *average*:

$$\overline{t} = \frac{1}{n}\sum_{i=1}^{n} t_i, \qquad \overline{t^2} = \frac{1}{n}\sum_{i=1}^{n} t_i^2, \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i^2, \qquad \overline{ty} = \frac{1}{n}\sum_{i=1}^{n} t_i y_i$$

Theorem 5.3 (Linear Regression). *Given n data points (t_i, y_i) with at least two distinct t-values, the best-fitting least-squares line has equation $\hat{y} = mt + c$, where m, c satisfy*

$$\begin{cases} \Big(\sum t_i^2\Big)m + \Big(\sum t_i\Big)c = \sum t_i y_i \\ \Big(\sum t_i\Big)m + nc = \sum y_i \end{cases} \quad \leftrightsquigarrow \quad \begin{cases} \overline{t^2}m + \overline{t}c = \overline{ty} \\ \overline{t}m + c = \overline{y} \end{cases}$$

These equations have solutions

$$m = \frac{\overline{ty} - \overline{t}\,\overline{y}}{\overline{t^2} - \overline{t}^2}, \qquad c = \overline{y} - m\overline{t}$$

In the next section, we'll see that having at least two distinct t-values guarantees a non-zero denominator $\overline{t^2} - \overline{t}^2$. The expression for c shows that the regression line passes through the data set's *center of mass* $(\overline{t}, \overline{y})$.

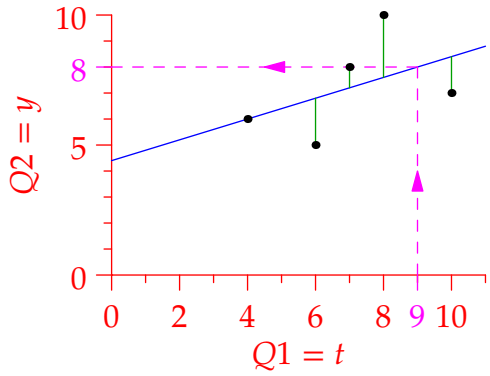
Example 5.4. Five students' scores on two quizzes are given.

Quiz 1	8	10	6	7	4
Quiz 2	10	7	5	8	6

If a student scores 9/10 on the first quiz, what might we expect them to score on the second?

To put the question in standard form, suppose Quiz 1 is the t-data and Quiz 2 the y-data. It is helpful to rewrite and add lines to the table to more easily compute the needed ingredients.

Data						$\sum$	Average
t_i	8	10	6	7	4	35	7
y_i	10	7	5	8	6	36	7.2
t_i^2	64	100	36	49	16	265	53
$t_i y_i$	80	70	30	56	24	260	52



$$m = \frac{52 - 7 \times 7.2}{53 - 7^2} = \frac{1.6}{4} = 0.4, \quad c = 7.2 - 0.4 \times 7 = 4.4$$

$$\implies \hat{y}(t) = \frac{2}{5}(t + 11)$$

This line minimizes the sum of the squares of the vertical deviations. The hypothetical student is predicted to score $\hat{y}(9) = \frac{2}{5} \cdot 20 = 8$ on Quiz 2.

Note that the predictor isn't symmetric: if we reverse the roles of t, y we don't get the same line!

Exercises 5.1. *Key concepts: Best-fitting line minimizes* $S = \sum |\hat{y}_i - y_i|^2$, *Computing* $\hat{y} = mt + c$

1. Compute the sum of the absolute errors $\sum |\hat{y}_i - y_i|$ for the regression line and compare it to the sum of the absolute errors for $\hat{y} = 4t$: what do you notice?

2. Let $\hat{y} = mt + c$ be a linear predictor for the given data.

t_i	0	1	2	3
y_i	1	2	2	3

   (a) Compute the sum of squared-errors

   $$S(m, c) = \sum e_i^2 = \sum |\hat{y}_i - y_i|^2$$

   as a function of m and c.

   (b) Compute the partial derivatives $\frac{\partial S}{\partial m}$ and $\frac{\partial S}{\partial c}$.

   (c) Find m and c by setting both partial derivatives to zero. Hence find the equation of the regression line for these data.

   (d) Compare the sum of square errors S for the regression line with the errors if we use the simple predictor $y(t) = 1 + \frac{2}{3}t$ which passes through the first an last data points.

3. Consider Example 5.4.

   (a) Compute the sum of square-errors $S = \sum e_i^2 = \sum |\hat{y}_i - y_i|^2$ for the regression line.

   (b) Consider an alternative model $\hat{y}(t) = t$ where students are expected to score *exactly the same* on both quizzes. What is the sum of squared-errors in this case?

   (c) Use linear regression to find the best-fitting predictor for *Quiz 1* given Quiz 2. If a student scores 8 points on *Quiz 2*, what is their predicted score on Quiz 1.

   (*Warning: the answer is NOT* $\frac{5}{2} \cdot 8 - 11 = 9$)

4. The heights in inches of ten children are measured on their first and second birthdays. The data was as follows.

1^{st} birthday	28	28	29	29	29	30	30	32	32	33
2^{nd} birthday	30	32	31	34	35	33	36	37	35	37

   Find the linear regression model and use it to predict the height at 2 years of a child who measures 32 inches at age 1.

   (*It is acceptable—and encouraged—to use a spreadsheet to find the necessary ingredients. You can do this by hand if you like, but the numbers are large…*)

5. (a) Let a, b be given. Find the value of y which minimizes the sum of squares

$$(y-a)^2 + (y-b)^2$$

   (b) Given the data set $\{(t,y)\} = \{(1,1),(2,1),(2,3)\}$, find the unique least-squares linear model for predicting y given t.

   (*What does this have to do with part* (a)*?*)

   (c) For the data set in part (b), show that there are *infinitely many* lines $\hat{y} = mt + c$ which minimize the sum of the absolute errors $\sum_{i=1}^{3} |\hat{y}_i - y_i|$.

6. Suppose $\hat{y}(t) = mt + c$ is the regression line for a data set. Verify that the average of the predicted outputs $\hat{y}_i = \hat{y}(t_i)$ equals that of the original outputs y: that is

$$\overline{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}$$

5.2 The Coefficient of Determination

In the sense that it minimizes the sum of the squared errors $S = \sum e_i^2$, the linear regression model is as good as it can be—but *how* good? We could use S as a *quantitative* measure of the model's accuracy, but this doesn't do a good job at comparing the accuracy of models for *different* data sets. The standard approach to this problem relies the concept of variance, a measure of how a data set deviates from being constant.

Definition 5.5. The *variance* of a data set $\{y_1, \dots, y_n\}$ is the average of the squared deviations from the mean $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$,

$$\operatorname{Var} y := \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

The *standard deviation* is $\sigma_y := \sqrt{\operatorname{Var} y}$.

Example 5.6. Suppose $(y_i) = (1, 2, 5, 4)$. Then

$$\overline{y} = \frac{1+2+5+4}{4} = 3 \qquad \operatorname{Var} y = \frac{(-2)^2 + (-1)^2 + 2^2 + 1^2}{4} = \frac{5}{2} \qquad \sigma_y = \frac{\sqrt{10}}{2}$$

The square-root means that σ_y has the same units as y. Loosely speaking, a typical data value is about twice as likely as not to lie within $\sigma_y = \frac{1}{2}\sqrt{10} \approx 1.58$ of the mean $\overline{y} = 3$.

To obtain a measure for how well a regression line fits given data (t_i, y_i), we ask what *fraction* of the variance in y is explained by the model.

Definition 5.7. The *coefficient of determination* of a model $\hat{y} = mt + c$ is the ratio

$$R^2 := \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} y}$$

Examples 5.8. We start by considering two extreme examples.

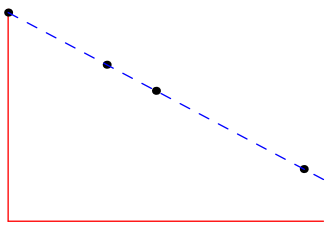
1. Suppose data is perfectly linear: $y_i = mt_i + c$ for all i. It is both intuitive and easy to verify that the regression line also has equation

   $$\hat{y} = mt + c$$

   Since $\hat{y}_i = y_i$ for all i, the variances are identical and the coefficient of determination is therefore

   $$R^2 = \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} y} = 1$$

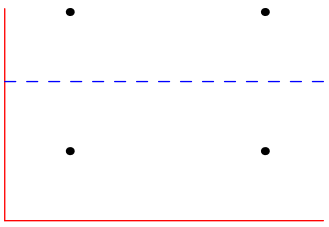
   All variance in the output y is explained by the model.



Perfect model $R^2 = 1$

2. Consider the data in the table. We work out all necessary details to find the regression line.

	data				average
t_i	1	1	4	4	$\bar{t} = 2.5$
y_i	1	3	1	3	$\bar{y} = 2$
t_i^2	1	1	16	16	$\overline{t^2} = 8.5$
$t_i y_i$	1	3	4	12	$\overline{ty} = 5$



Useless model $R^2 = 0$

$$m = \frac{\overline{ty} - \bar{t}\bar{y}}{\overline{t^2} - \bar{t}^2} = \frac{5 - 2.5 \cdot 2}{8.5 - 2.5^2} = 0, \quad c = \bar{y} - m\bar{t} = 2$$

The regression line is therefore the *constant*

$$\hat{y}(t) = mt + c = 2$$

Since $\hat{y}$ has *no variance*, the coefficient of determination is $R^2 = 0$. The t-values have no impact on the predicted y-values.

These extreme examples are just that.

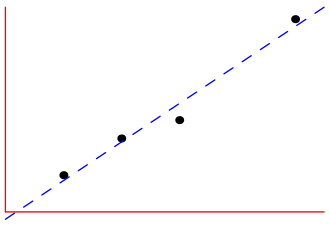
Theorem 5.9. *If $\{(t_i, y_i)\}$ has at least two distinct t-values, then the coefficient of determination satisfies $0 \leq R^2 \leq 1$. Moreover:*

- *$R^2 = 1$ if and only if the data is precisely linear $y_i = mt_i + c$.*
- *$R^2 = 0$ if and only if the regression line has slope $m = 0$.*

A proof is in Exercise 8. In practice, therefore, $0 < R^2 < 1$. We now revisit two earlier examples. Note how Exercise 5.1.6 makes computing the variance of $\hat{y}$ easier.

Example 5.1. Recall that $\hat{y} = \frac{1}{5}(21t - 4)$. Everything necessary is in the table

	data				average
t_i	1	2	3	5	$\bar{t} = 2.75$
y_i	4	8	10	21	$\bar{y} = 10.75$
$\hat{y}_i$	3.4	7.6	11.8	20.2	$\overline{ty} = 10.75$



Good model $R^2 \approx 97\%$

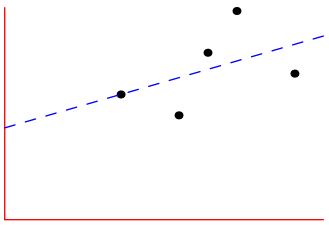
$$\operatorname{Var} y = \frac{6.75^2 + 2.75^2 + 0.75^2 + 10.25^2}{4} \approx 39.69$$

$$\operatorname{Var} \hat{y} = \frac{7.35^2 + 3.15^2 + 1.05^2 + 9.45^2}{4} \approx 38.59$$

from which $R^2 = \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} y} \approx 97\%$. The data is very close to being linear with approximately 97% of the variance of y explained by the model.

Example 5.4. This time $\hat{y} = \frac{2}{5}(t + 11)$.

	data					average
t_i	8	10	6	7	4	$\overline{t} = 7$
y_i	10	7	5	8	6	$\overline{y} = 7.2$
$\hat{y}_i$	7.6	8.4	6.8	7.2	6	$\overline{\hat{y}} = 7.2$



Poor model $R^2 \approx 22\%$

$$\operatorname{Var} y = \frac{2.8^2 + 0.2^2 + 2.2^2 + 0.8^2 + 1.2^2}{5} = 2.96$$
$$\operatorname{Var} \hat{y} = \frac{0.4^2 + 1.2^2 + 0.4^2 + 0^2 + 1.2^2}{5} = 0.64$$

from which $R^2 = \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} y} = \frac{8}{37} \approx 22\%$. In this case the coefficient of determination is small, indicating that the model does not explain much of the variation in the output.

Efficient computation of R^2

Our current process for computing R^2 is lengthy and awkward. To obtain a more efficient process, we first consider an alternative expression for the variance of any data set $\{x_i\}$:

$$\begin{aligned}\operatorname{Var} x &= \frac{1}{n}\sum (x_i - \overline{x})^2 = \frac{1}{n}\sum \left(x_i^2 - x_i\overline{x} + \overline{x}^2\right) = \frac{1}{n}\sum x_i^2 - \frac{2\overline{x}}{n}\sum x_i + \frac{\overline{x}}{n}\sum x_i \\ &= \overline{x^2} - \overline{x}^2\end{aligned}$$

Recalling Theorem 5.3, we see that the regression line has slope

$$m = \frac{\overline{ty} - \overline{t}\,\overline{y}}{\operatorname{Var} t}$$

Since $\operatorname{Var} t \geq 0$, with equality if and only if t is constant, this justifies the claim of Theorem 5.3, that the regression line exists and is unique, provided there are at least two distinct t-values.[17] Now expand the variance of the predicted outputs:

$$\begin{aligned}\operatorname{Var} \hat{y} &= \frac{1}{n}\sum \left(\hat{y}_i - \overline{y}\right)^2 = \frac{1}{n}\sum \left(mt_i + c - (m\overline{t} + c)\right)^2 = \frac{m^2}{n}\sum \left(t_i - \overline{t}\right)^2 \\ &= m^2 \operatorname{Var} t\end{aligned}$$

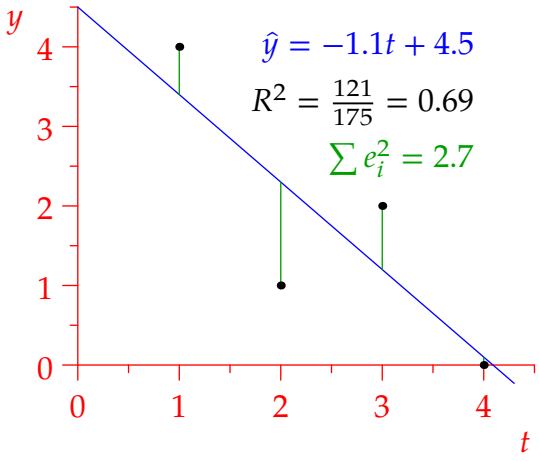
Theorem 5.10. *Putting everything together results in several equivalent expressions for the coefficient of determination:*

$$R^2 = \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} y} = m^2 \frac{\operatorname{Var} t}{\operatorname{Var} y} = m^2 \frac{\overline{t^2} - \overline{t}^2}{\overline{y^2} - \overline{y}^2} = \frac{\left(\overline{ty} - \overline{t}\,\overline{y}\right)^2}{\left(\overline{t^2} - \overline{t}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}$$

[17] Strictly, the system of equations we solve for m has a unique solution if and only if $\operatorname{Var} t \neq 0$.

Example 5.11. Consider the data set in the table.

	data				average
t_i	1	2	3	4	$\bar{t} = \frac{10}{4}$
y_i	4	1	2	0	$\bar{y} = \frac{7}{4}$
t_i^2	1	4	9	16	$\overline{t^2} = \frac{15}{2}$
y_i^2	16	1	4	0	$\overline{y^2} = \frac{21}{4}$
$t_i y_i$	4	2	6	0	$\overline{ty} = 3$



$$m = \frac{\overline{ty} - \bar{t}\bar{y}}{\overline{t^2} - \bar{t}^2} = \frac{3 - \frac{70}{4^2}}{\frac{15}{2} - \frac{100}{4^2}} = -\frac{11}{10} = -1.1$$

$$c = \bar{y} - m\bar{t} = \frac{7}{4} + \frac{11 \cdot 10}{10 \cdot 4} = \frac{9}{2} = 4.5$$

The regression line is $\hat{y} = -\frac{11}{10}t + \frac{9}{2} = -1.1t + 4.5$, and the coefficient of determination is

$$R^2 = m^2 \frac{\overline{t^2} - \bar{t}^2}{\overline{y^2} - \bar{y}^2} = \frac{121}{100} \cdot \frac{\frac{15}{2} - \frac{100}{4^2}}{\frac{21}{4} - \frac{49}{4^2}} = \frac{121}{100} \cdot \frac{20}{35} = \frac{121}{175} \approx 69\%$$

The minimized square error is also easily computed:

$$S = \sum e_i^2 = \sum (\hat{y}_i - y_i)^2 = (3.4 - 4)^2 + (2.3 - 1)^2 + (1.2 - 2)^2 + (0.1 - 0)^2 = 2.7$$

Reversion to the Mean & Correlation

By Theorem 5.10, the regression model may be re-written in terms of the standard-deviation and R^2:

$$\hat{y}(t) = mt + c = \bar{y} + m(t - \bar{t}) = \bar{y} + \sqrt{R^2}\frac{\sigma_y}{\sigma_t}(t - \bar{t}) \implies \hat{y}(\bar{t} + \lambda\sigma_t) = \bar{y} + \lambda\sqrt{R^2}\sigma_y$$

Definition 5.12. The *correlation coefficient* is $r := \pm\sqrt{R^2}$, whose sign is equal to that of m.

An input λ standard-deviations from the mean ($t = \bar{t} + \lambda\sigma_t$) results in a prediction λr standard-deviations from the mean ($\hat{y} = \bar{y} + \lambda r\sigma_y$). Unless data is perfectly linear, the correlation coefficient satisfies $|r| < 1$, whence a prediction $\hat{y}$ is *closer to the mean* than the input t, relative to the 'neutral' measure given by the standard-deviation

$$\frac{|\hat{y}(t) - \bar{y}|}{\sigma_y} = |r|\frac{|\hat{t} - \bar{t}|}{\sigma_t} < \frac{|\hat{t} - \bar{t}|}{\sigma_t}$$

Example (5.11, cont). The correlation coefficient is $r = -\sqrt{R^2} \approx -0.832$. We say that the data is *negatively correlated*: the output y tends to *decrease* as t increases. The standard deviations may be read from the table:

$$\sigma_t = \sqrt{\operatorname{Var} t} = \sqrt{\overline{t^2} - \overline{t}^2} = \frac{\sqrt{5}}{2} \approx 1.12, \qquad \sigma_y = \sqrt{\operatorname{Var} y} = \sqrt{\overline{y^2} - \overline{y}^2} = \frac{\sqrt{35}}{4} \approx 1.48$$

The predictor may therefore be written (approximately)

$$\hat{y}(\overline{t} + \lambda\sigma_t) = \hat{y}(2.5 + 1.12\lambda) = \overline{y} + \lambda r \sigma_y = 1.75 - 1.23\lambda$$

As a sanity check, the input $t = 2.5 + 1.12 = 3.62$ is one standard-deviation above the mean; its predicted output

$$\hat{y}(3.62) = -1.1 \times (-3.62) + 4.5 = 0.52 = 1.75 - 1.23$$

is *less than one standard-deviation below* the mean.

Weaknesses of Linear Regression

There are two obvious issues:

- Outliers massively influence the regression line. Dealing with this problem is complicated and a variety of approaches can be used. It is important to remember that any approach to modeling requires some subjective choice.
- If the data is not very linear then the regression model will produce a weak predictor. There are several simple ways around this as we'll see in the remaining sections: higher-degree polynomial regression can be performed, and data sometimes becomes more linear after some manipulation, say by an exponential or logarithmic function.

Exercises 5.2. *Key Concepts: R^2, its meaning and computation, Reversion to the Mean*

1. Suppose $(z_i) = (2, 4, 10, 8)$ is *double* the data set in Example 5.6. Find $\overline{z}$, $\operatorname{Var} z$ and σ_z. Why are you not surprised?
2. Use a spreadsheet to find R^2 for the predictor in Exercise 5.1.4. How confident do you feel in your prediction?
3. For the data in both Examples 5.1 and 5.4, find the standard deviations and correlation coefficients.
4. The adult heights of men and women in a given population satisfy the following:

   Men: average 69.5 in, $\sigma = 3.2$ in. Women: average 63.7 in, $\sigma = 2.5$ in.

   The height of a father and his adult daughter have correlation coefficient $r = 0.35$. If all you know is that a father's height is 72 in, how tall do you expect his daughter to grow up to be?

5. Suppose a data set is perfectly linear: for each i, we have $y_i = mt_i + c$. Prove that the regression line really is $\hat{y} = mt + c$.

6. Consider the data set

$$\{(t_i, y_i)\} = \{(3,1), (3,5), (3,6)\}$$

Find the equations of *all* lines $\hat{y}(t) = mt + c$ that minimize the sum of the squares of the vertical errors $S = \sum(\hat{y}_i - y_i)^2$. What is going on?

7. Suppose R^2 is the coefficient of determination for a linear regression model $\hat{y} = mt + c$. Use one of the alternative expressions for R^2 (Theorem 5.10) to find the coefficient of determination for the reversed predictor $\hat{t}(y)$? Are you surprised?

8. Suppose that a data set $\{(t_i, y_i)\}_{1 \le i \le n}$ has at least two distinct t-values (some $t_i \neq t_j$), that it has regression line $\hat{y} = mt + c$, and coefficient of determination R^2.

   (a) Show that $R^2 = 0 \iff m = 0$.

   (b) (Hard) Prove that the sum of squared errors equals

$$S = \sum_{i=1}^{n} e_i^2 = n(\operatorname{Var} y - \operatorname{Var} \hat{y})$$

   (c) Suppose additionally that $\operatorname{Var} y \neq 0$ (at least two distinct y-values). Obtain the alternative expression

$$R^2 = 1 - \frac{S}{n \operatorname{Var} y}$$

Hence conclude that $R^2 \le 1$, with equality if and only if the original data set is perfectly linear.

5.3 Polynomial Regression & Matrices

In this section we consider how to find a best-fitting least-squares polynomial for given data. The structure is almost identical to what we did in Section 5.1, though the algebra is nastier.

Suppose, for instance, that we want to find a *quadratic* polynomial predictor $\hat{y}(t) = at^2 + bt + c$ for a data set $\{(t_i, y_i) : 1 \le i \le n\}$. As before, our criteria is that this model minimizes the sum of the squared vertical errors

$$S(a,b,c) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (at_i^2 + bt_i + c - y_i)^2$$

This might look terrifying, but can be attacked exactly as before using differentiation: to minimize S, we require that its derivatives with respect to the coefficients a, b, c be zero.

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum at_i^4 + bt_i^3 + ct_i^2 - t_i^2 y_i = 0 \\ \dfrac{\partial S}{\partial b} = 2\sum at_i^3 + bt_i^2 + ct_i - t_i y_i = 0 \\ \dfrac{\partial S}{\partial c} = 2\sum at_i^2 + bt_i + c - y_i = 0 \end{cases} \iff \begin{cases} \left(\sum t_i^4\right) a + \left(\sum t_i^3\right) b + \left(\sum t_i^2\right) c = \sum t_i^2 y_i \\ \left(\sum t_i^3\right) a + \left(\sum t_i^2\right) b + \left(\sum t_i\right) c = \sum t_i y_i \\ \left(\sum t_i^2\right) a + \left(\sum t_i\right) b + nc = \sum y_i \end{cases}$$

This is just a linear system of equations for a, b, c. It might look nasty, but for a given example it is not hard to solve. The idea extends naturally. If you want a *cubic* predictor $\hat{y}(t) = at^3 + bt^2 + ct + d$, then you'll need to solve a system of *four* equations.

Example 5.13. Consider the data set shown.

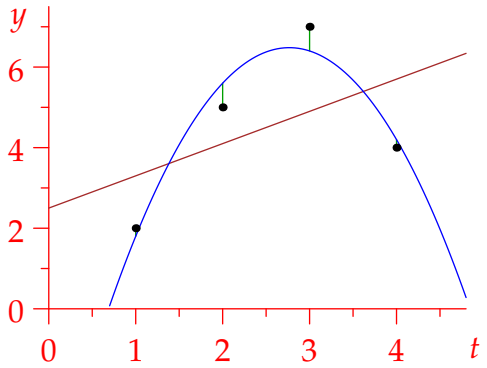
t	1	2	3	4
y	2	5	7	4

The system of equations satisfied by the coefficients of the least-squares quadratic predictor amounts to

$$\begin{cases} 354a + 100b + 30c = 149 \\ 100a + 30b + 10c = 49 \\ 30a + 10b + 4c = 18 \end{cases}$$

This has solution $a = -1.5$, $b = 8.3$, $c = -5$, yielding the quadratic model

$$\hat{y}(t) = -1.5t^2 + 8.3t - 5$$



This certainly seems to fit the data more accurately than the linear regression model. Indeed, one may compute the coefficient of determination in the usual way and compare:

$$R^2_{\text{quad}} = \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} y} \approx 94\% \qquad R^2_{\text{linear}} \approx 25\%$$

Extending the problem, it can be shown that there is a unique *cubic* function passing through the four data points: since $\hat{y} = y$, such a model necessarily has $R^2 = 1$.

Polynomial Regression using Matrices

Even for a small data set, finding a quadratic predictor by hand is ugly. Instead of continuing with this approach, we rephrase the problem in terms of matrices in the hope of finding a cleaner, more unifying, exposition.[18]

Start by observing that the system of equations in Theorem 5.3 can be written in as a 2×2 matrix problem. For a data set with n pairs, the coefficients m, c satisfy

$$\begin{cases} \left(\sum t_i^2\right) m + \left(\sum t_i\right) c = \sum t_i y_i \\ \left(\sum t_i\right) m + nc = \sum y_i \end{cases} \quad \leftrightarrow \quad \begin{pmatrix} \sum t_i^2 & \sum t_i \\ \sum t_i & n \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} \sum t_i y_i \\ \sum y_i \end{pmatrix}$$

The square matrix on the left can be decomposed as the product of a simple $n \times 2$ matrix P and its transpose P^T;

$$\begin{pmatrix} \sum t_i^2 & \sum t_i \\ \sum t_i & n \end{pmatrix} = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix} =: P^T P$$

The right side is moreover the product of P^T and the column vector of outputs $\mathbf{y}$:

$$\begin{pmatrix} \sum t_i y_i \\ \sum y_i \end{pmatrix} = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} =: P^T \mathbf{y}$$

If *at least two* of the t_i are distinct, then the 2×2 matrix $P^T P$ is invertible;[19] in which case there is a *unique* regression line whose coefficients may be found by taking the matrix inverse

$$\begin{pmatrix} m \\ c \end{pmatrix} = (P^T P)^{-1} P^T \mathbf{y} \implies \hat{y} = mt + c = \begin{pmatrix} t & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} t & 1 \end{pmatrix} (P^T P)^{-1} P^T \mathbf{y}$$

We can also easily compute the vector of predicted values $\hat{y}_i = \hat{y}(t_i)$:

$$\hat{\mathbf{y}} = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = P(P^T P)^{-1} P^T \mathbf{y}$$

and the squared error $\sum e_i^2 = \sum |\hat{y}_i - y_i|^2 = \|\hat{\mathbf{y}} - \mathbf{y}\|^2$, which leads to an alternative expression for the coefficient of determination

$$R^2 = \frac{\|\hat{\mathbf{y}}\|^2 - n\overline{y}^2}{\|\mathbf{y}\|^2 - n\overline{y}^2}$$

where $\|\mathbf{y}\|$ is the *length* of a vector.

---

[18] Matrix computations are non-examinable. Our purpose is to be see how regression may easily be extended by computer and to understand a little of how a spreadsheet calculates best-fitting curves of different types.

[19] For those who've studied linear algebra, P and $P^T P$ have the same nullspace:

$$P\mathbf{x} = \mathbf{0} \implies P^T P\mathbf{x} = \mathbf{0} \quad \text{and} \quad P^T P\mathbf{x} = \mathbf{0} \implies \mathbf{x}^T P^T P\mathbf{x} = \mathbf{0} \implies \|P\mathbf{x}\| = \mathbf{0} \implies P\mathbf{x} = \mathbf{0}$$

By the rank-nullity theorem, they have the same rank. For linear regression, having at least two distinct t_i values means rank $P = 2$, whence $P^T P$ (being rank 2) is invertible.

Example (5.11, cont.). As an illustration, we revisit an old example in this language.

$$P = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \implies P^T P = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

The coefficients of the regression model are as we found previously:

$$\begin{pmatrix} m \\ c \end{pmatrix} = (P^T P)^{-1} P^T \mathbf{y} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \\ 2 \\ 0 \end{pmatrix}$$

$$= \frac{1}{30 \cdot 4 - 10^2} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix} \begin{pmatrix} 12 \\ 7 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} -11 \\ 45 \end{pmatrix}$$

Moreover, the prediction vector $\hat{\mathbf{y}}$ given inputs t_i is

$$\hat{\mathbf{y}} = P \begin{pmatrix} m \\ c \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -11 \\ 45 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 34 \\ 23 \\ 12 \\ 1 \end{pmatrix}$$

from which the coefficient of determination is, as before

$$R^2 = \frac{\|\hat{\mathbf{y}}\|^2 - 4\overline{y}^2}{\|\mathbf{y}\|^2 - 4\overline{y}^2} = \frac{\frac{1}{100}(34^2 + 23^2 + 12^2 + 1^2) - 4 \cdot \frac{7^2}{4^2}}{(4^2 + 1^1 + 2^2 + 0^2) - 4 \cdot \frac{7^2}{4^2}} = \frac{121}{175} \qquad (*)$$

It is unnecessary ever to use the matrix approach, though it does have some advantages.

- Computers store and manipulate data in matrix format, so this approach is computer-ready.
- Suppose you repeat an experiment several times, taking measurements y_i at times t_i. Since the matrix P depends only on the t-data, you need only compute the matrix $(P^T P)^{-1} P^T$ *once*, making computation of the regression line for repeat experiments very efficient.
- The method generalizes to a larger class of least-squares problems. Indeed, the same method produces best-fitting least-squares *polynomial* models. Indeed the system of equations on page 77 can also be written in matrix format:

$$\begin{pmatrix} \sum t_i^4 & \sum t_i^3 & \sum t_i^2 \\ \sum t_i^3 & \sum t_i^2 & \sum t_i \\ \sum t_i^2 & \sum t_i & cn \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum t_i^2 y_i \\ \sum t_i y_i \\ \sum y_i \end{pmatrix} \rightsquigarrow P^T P \begin{pmatrix} a \\ b \\ c \end{pmatrix} = P^T \mathbf{y} \quad \text{where} \quad P = \begin{pmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_n^2 & t_n & 1 \end{pmatrix}$$

  This is guaranteed to have a unique solution provided we have *at least three* distinct t-values. Everything else proceeds as above. For cubic, or higher-degree, polynomial regression, simply add another column of powers of t_i to the start of P...

Example (5.13, cont.). We compute regression models for the data $\{(1,2),(2,5),(3,7),(4,4)\}$.

1. For the linear model we use the same P (and thus P^TP) from the previous example:

$$\begin{pmatrix} m \\ c \end{pmatrix} = (P^TP)^{-1}P^T\mathbf{y} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 5 \\ 7 \\ 4 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 2 & -5 \\ -5 & 15 \end{pmatrix}\begin{pmatrix} 49 \\ 18 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 2.5 \end{pmatrix}$$

   which yields $\hat{y}(t) = 0.8t + 2.5$. The predicted values and coefficient of determination are then

$$\hat{\mathbf{y}} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0.8 \\ 2.5 \end{pmatrix} = \begin{pmatrix} 3.3 \\ 4.1 \\ 4.9 \\ 5.7 \end{pmatrix} \qquad R^2 = \frac{\|\hat{\mathbf{y}}\|^2 - 4\overline{y}^2}{\|\mathbf{y}\|^2 - 4\overline{y}^2} = \frac{84.2 - 81}{94 - 81} \approx 25\%$$

2. For the quadratic model, all that changes is the matrix P:

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} \implies P^TP = \begin{pmatrix} 1 & 4 & 9 & 16 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 4 \end{pmatrix}$$

$$\implies \begin{pmatrix} a \\ b \\ c \end{pmatrix} = (P^TP)^{-1}P^T\begin{pmatrix} 2 \\ 5 \\ 7 \\ 4 \end{pmatrix} = \begin{pmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 4 \end{pmatrix}^{-1}\begin{pmatrix} 149 \\ 49 \\ 18 \end{pmatrix} = \begin{pmatrix} -1.5 \\ 8.3 \\ -5 \end{pmatrix}$$

   from which $\hat{y} = -1.5t^2 + 8.3t - 5$. The predicted outputs are

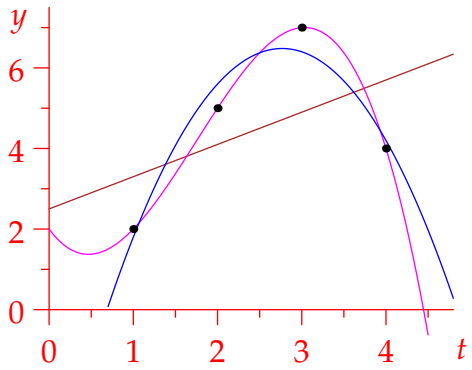
$$\hat{\mathbf{y}} = P\begin{pmatrix} -1.5 \\ 8.3 \\ -5 \end{pmatrix} = \begin{pmatrix} 1.8 \\ 5.6 \\ 6.4 \\ 4.2 \end{pmatrix} \qquad R^2 = \frac{93.2 - 81}{94 - 81} \approx 94\%$$

3. For the cubic model, P is now a 4×4 matrix! Omitting the details, we obtain

$$\hat{y} = \frac{1}{6}(-4t^3 + 21t^2 - 17t + 12)$$

   This cubic passes through all four data points; there is *no error*, and $R^2 = 1$.

   For real-world data this is likely *less useful* than the quadratic model. Not only does it take longer to find, but the original y-data likely contained experimental error— using a cubic to match the data perfectly is, in effect, modeling *noise*.



Exercises 5.3. *Key Concepts: Higher-order polynomial regression, Rephrasing as a matrix problem, Considering whether the effort is worth it*

1. Recall Example 4.2, with the following *almost* linear data set.

x	0	2	4	6	8	10
y	3	23	41	59	77	93

   Find the least-squares linear model for the data, then use a spreadsheet to find the best-fitting quadratic. Is the extra effort worth it?

2. Consider the data in Example 5.11. State the corresponding matrices P for the least-squares quadratic and cubic models for this data. In each case, use a spreadsheet to compute the models and the coefficients of determination.

3. You are given the following data consisting of measurements from an experiment recorded at times t_i seconds.

t_i	1	2	3	4	5	6	7	8	9	10
y_i	7	5	3	2	3	5	6	9	8	12

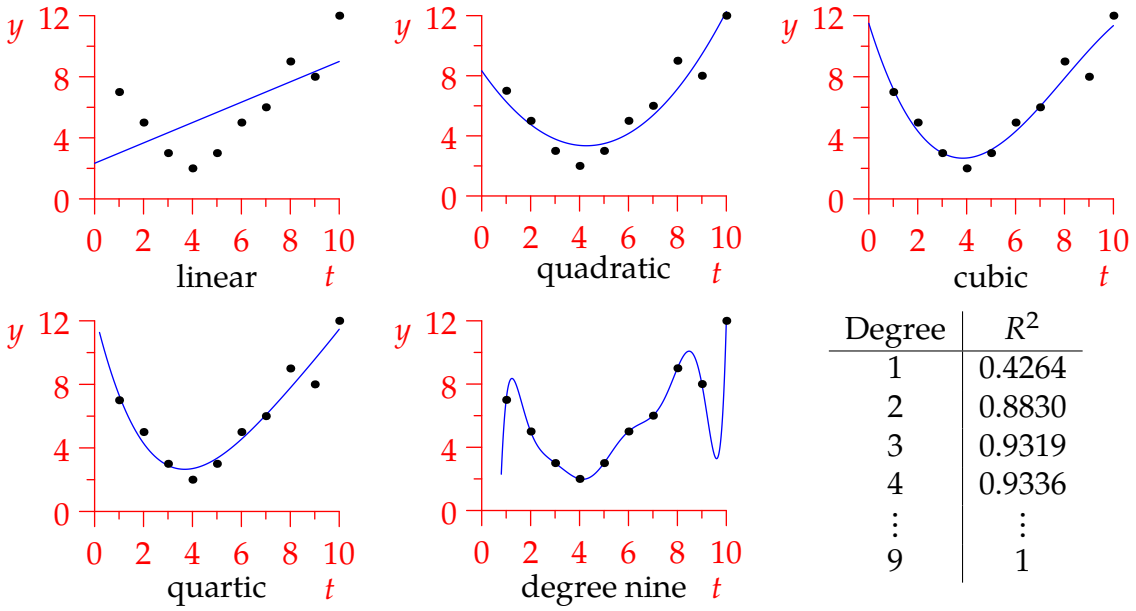
   (a) Given the values

   $$\sum t_i = 55, \quad \sum t_i^2 = 385, \quad \sum y_i = 60, \quad \sum t_i y_i = 385$$

   find the least-squares linear model for this data, and use it to predict $\hat{y}(13)$.

   (b) Find the least-squares quadratic model for the data: feel free to use a spreadsheet!

   (c) The graphs below show the least-squares linear, quadratic, cubic, quartic, and ninth-degree models and their coefficients of determination.



Degree	R^2
1	0.4264
2	0.8830
3	0.9319
4	0.9336
⋮	⋮
9	1

   Which of these models would *you* choose for this data and why? What considerations would you take into account?

5.4 Exponential & Power Regression Models

If you suspect that a non-polynomial function would better fit your data, there are several things you can try. Minimizing the sum of squared-errors is likely very difficult for non-polynomial functions (there is likely no simple tie-in with linear equations/algebra).[20]

Log Plots

The most common approach when trying to fit an exponential model $\hat{y} = e^{mt+c}$ to data is to use a log plot: taking logarithms of both sides results in

$$\ln \hat{y} = mt + c$$

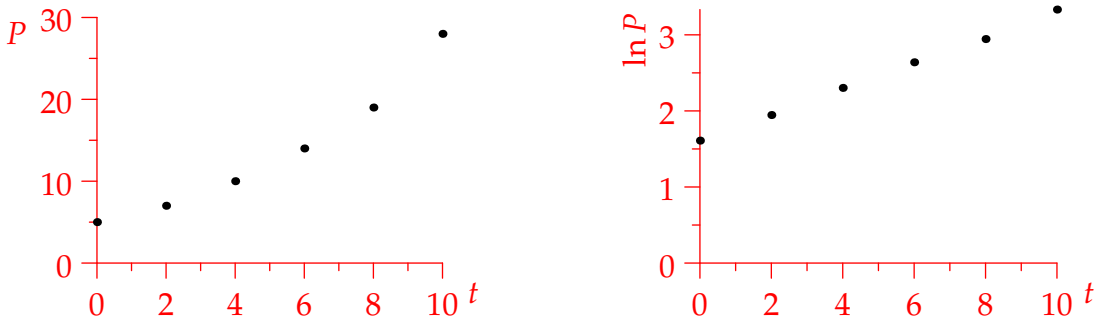
If we take $\hat{Y} := \ln \hat{y}$ as a new variable, the model is now a *straight-line*! The idea is then to use linear regression to find the coefficients $m, \ln a$.

Example (4.4, cont). Recall our earlier rabbit-population $P(t)$, repeated in the table below. We previously considered modelling this with an exponential function for two reasons:

1. We were told we had population data!
2. The *t-differences* are constant while the *P-ratios* are approximately so (≈ 1.41).

t_i	0	2	4	6	8	10
P_i	5	7	10	14	19	28
$\ln P_i$	1.61	1.95	2.30	2.64	2.94	3.33

After constructing a log-plot, the relationship is much clearer:



Since the relationship between t and $\ln P$ is close to linear, we perform a linear regression calculation to find the least-squares line for the $(t_i, \ln P_i)$ data.

[20] Attempting this is likely to result in a horrible *non-linear* system for your coefficients which is difficult to analyze either theoretically or using a computer. As an example, suppose you want to minimize the sum of square-errors for data (t_i, y_i) using an exponential model $\hat{y}(t) = ae^{kt}$. The coefficients of our model, a, k should minimize

$$S(a,k) = \sum_{i=1}^{n} \left(ae^{kt_i} - y_i\right)^2$$

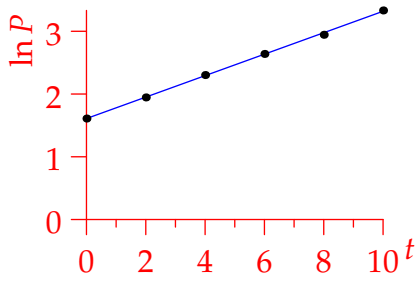
Differentiating this with respect to a, k and setting equal to zero results in

$$\begin{cases} \frac{\partial S}{\partial a} = 2\sum e^{kt_i}(ae^{kt_i} - y_i) = 0 \\ \frac{\partial S}{\partial k} = 2a\sum t_i e^{kt_i}(ae^{kt_i} - y_i) = 0 \end{cases} \implies \left(\sum y_i e^{kt_i}\right)\left(\sum t_i e^{2kt_i}\right) = \left(\sum e^{2kt_i}\right)\left(\sum t_i y_i e^{kt_i}\right)$$

where we substituted for a to obtain the last equation. Remember that this is an equation for k; if you think you can solve this easily, think again!

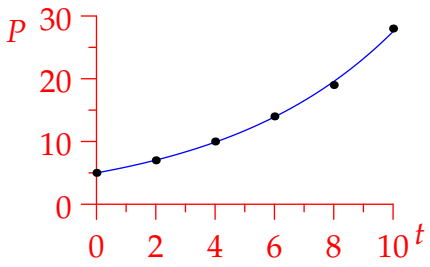
Everything necessary comes from extending the table.

	data						average
t_i	0	2	4	6	8	10	5
P_i	5	7	10	14	19	28	13.83
$\ln P_i$	1.61	1.95	2.30	2.64	2.94	3.33	2.46
t_i^2	0	4	16	36	64	100	36.67
$t_i \ln P_i$	0	3.89	9.21	15.83	23.56	33.32	14.30



$$m = \frac{\overline{t \ln P} - \bar{t} \cdot \overline{\ln P}}{\overline{t^2} - \bar{t}^2} = \frac{14.30 - 5 \cdot 2.46}{36.67 - 5^2} = 0.171$$

$$c = \overline{\ln P} - m\bar{t} = 3.46 - 0.171 \cdot 5 = 1.609$$



which yields the exponential model

$$\hat{P}(t) = e^{0.171t + 1.609} = 4.998(1.186)^t$$

This is very close to the model $(5(1.188)^t)$ we obtained previously by pure guesswork. The approximate doubling time T for the population satisfies

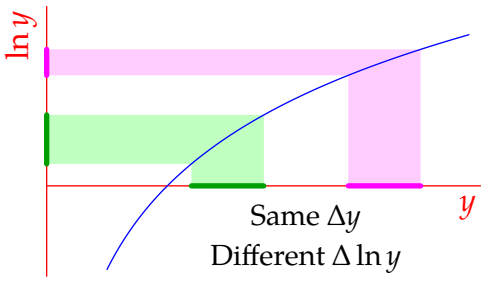
$$e^{mT} = 2 \iff T = \frac{\ln 2}{m} = 4.06 \text{ months}$$

Coefficient of Determination for Log Plots Interpreting errors and the goodness of fit of a model is a little more difficult when using a log plot. Typically one computes the coefficient of determination R^2 of the *underlying linear model*: in our example,[21]

$$R^2 = \frac{\operatorname{Var} \hat{y}}{\operatorname{Var} \ln P} = m^2 \frac{\operatorname{Var} t}{\operatorname{Var} \ln P} \approx 99\%$$

The log plot method does not treat all errors equally. As the output y increasing, $\ln y$ increases more slowly, reducing the effect of errors. This should be clear from the picture, and more formally by the mean value theorem: if $1 \le y_1 < y_2$, then there is some $\xi \in (y_1, y_2)$ for which

$$\ln y_2 - \ln y_1 = \frac{1}{\xi}(y_2 - y_1) < y_2 - y_1$$



The log plot approach places a higher emphasis on accurately matching data when the output y is *small*. This isn't such a bad thing since our intuitive view of error depends on the size of the data. For instance, misplacing a $100 bill is annoying, but a $100 mistake in escrow when buying a house is unlikely to concern you very much!

[21] This needs more decimal places of accuracy for the log-values than what's in our table!

Log-Log Plots

If you suspect a *power function model* $\hat{y} = at^m$, then taking logarithms of both sides

$$\ln \hat{y} = m \ln t + \ln a$$

results in a linear model between $\hat{Y} := \ln \hat{y}$ and $T := \ln t$. Similarly to log plots, we apply linear regression to the data $\{(\ln t, \ln y)\}$ to find a linear model. The goodness of fit is again described by the coefficient of determination of the underlying linear model

$$R^2 = \frac{\operatorname{Var} \ln \hat{y}}{\operatorname{Var} \ln y}$$

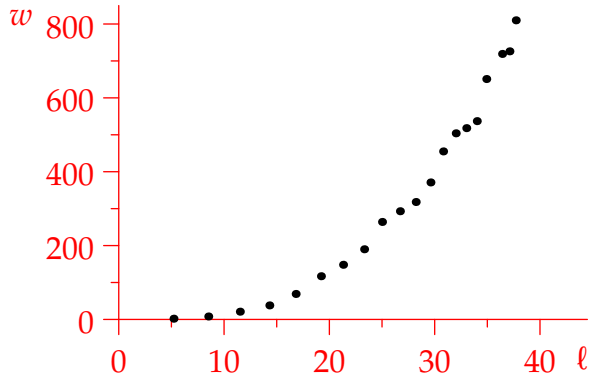
rather than the power model. Remember that regression is still trying to minimize (squared) output-errors: just like an exponential model, the log-log plot approach puts more emphasis on minimizing errors when y is *small*.

Exercises 5.4. *Key concepts: Log plots & Log-log plots*

1. You suspect a *logarithmic* model for a data set. Describe how you would approach the search for such a model in the context of this section.

2. The average weight and length of a population of fish is measured at different ages.

Age (years)	Length(cm)	Weight (g)
1	5.2	2
2	8.5	8
3	11.5	21
4	14.3	38
5	16.8	69
6	19.2	117
7	21.3	148
8	23.3	190
9	25.0	264
10	26.7	293
11	28.2	318
12	29.6	371
13	30.8	455
14	32.0	504
15	33.0	518
16	34.0	537
17	34.9	651
18	36.4	719
19	37.1	726
20	37.7	810



(a) Do you think an exponential model is a good fit for this data? Take logarithms of the weight values and use a spreadsheet to obtain a model

$$\hat{w}(\ell) = ae^{m\ell}$$

where ℓ, w are, respectively, the length and weight.

(b) What happens if you try a log-log plot? Given what we're measuring, why might you expect a power model to be more accurate?

3. Population data for Long Beach, CA is given.

   Using a spreadsheet or otherwise, find linear, quadratic, exponential and logarithmic regression models for this data.

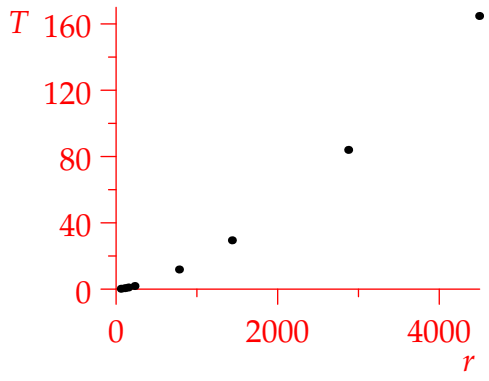
   Which of these models seems to fit the data best, and which would you trust to best predict the population in 2020?

   Look up the population of Long Beach in 2020; does it confirm your suspicions? What do you think is going on?

Year	Years since 1900	Population
1900	0	2,252
1910	10	17,809
1920	20	55,593
1930	30	142,032
1940	40	164,271
1950	50	250,767
1960	60	334,168
1970	70	358,879
1980	80	361,498
1990	90	429,433
2000	100	461,522
2010	110	462,257

4. In the early 1600s, Johannes Kepler used observational data to derive his *laws of planetary motion*. His third law relates the orbital period T of a planet (how long it takes to go round the sun) to its (approximate) distance r from the sun.

Planet	T (years)	r (millions km)
Mercury	0.24	58
Venus	0.61	110
Earth	1	150
Mars	1.88	230
Jupiter	11.9	780
Saturn	29.5	1400
Uranus	84	2900
Neptune	165	4500



   The table shows the data for all the planets. Use a spreadsheet to analyze this data and find a model relating T to r.

   Kepler did not known about Uranus and Neptune and only had relative distances for the planets. Research the correct statement of Kepler's third law and compare it with your findings.