



FIGURE 7.1
Me playing with my ding-a-ling in the Holimarine Talent Show. Note the groupies queuing up at the front

7.1. What will this chapter tell me? ①

Although none of us can know the future, predicting it is so important that organisms are hard wired to learn about predictable events in their environment. We saw in the previous chapter that I received a guitar for Christmas when I was 8. My first foray into public performance was a weekly talent show at a holiday camp called ‘Holimarine’ in Wales (it doesn’t exist anymore because I am old and this was 1981). I sang a Chuck Berry song called ‘My ding-a-ling’¹ and to my absolute amazement I won the competition.² Suddenly other 8 year olds across the land (well, a ballroom in Wales) worshipped me (I made lots of friends after the competition). I had tasted success, it tasted like praline chocolate, and so I wanted to enter the competition in the second week of our holiday. To ensure success, I needed to know why I had won in the first week. One way to do this would have been to collect data and to use these data to predict people’s evaluations of children’s performances in the contest

¹ It appears that even then I had a passion for lowering the tone of things that should be taken seriously.

² I have a very grainy video of this performance recorded by my dad’s friend on a video camera the size of a medium-sized dog that had to be accompanied at all times by a ‘battery pack’ the size and weight of a tank. Maybe I’ll put it up on the companion website ...

from certain variables: the age of the performer, what type of performance they gave (singing, telling a joke, magic tricks), and maybe how cute they looked. A regression analysis on these data would enable us to predict future evaluations (success in next week's competition) based on values of the predictor variables. If, for example, singing was an important factor in getting a good audience evaluation, then I could sing again the following week; however, if jokers tended to do better then I could switch to a comedy routine. When I was 8 I wasn't the sad geek that I am today, so I didn't know about regression analysis (nor did I wish to know); however, my dad thought that success was due to the winning combination of a cherub-looking 8 year old singing songs that can be interpreted in a filthy way. He wrote me a song to sing in the competition about the keyboard player in the Holimarine Band 'messing about with his organ', and first place was mine again. There's no accounting for taste.

7.2. An introduction to regression ①

In the previous chapter we looked at how to measure relationships between two variables. These correlations can be very useful but we can take this process a step further and predict one variable from another. A simple example might be to try to predict levels of stress from the amount of time until you have to give a talk. You'd expect this to be a negative relationship (the smaller the amount of time until the talk, the larger the anxiety). We could then extend this basic relationship to answer a question such as 'if there's 10 minutes to go until someone has to give a talk, how anxious will they be?' This is the essence of regression analysis: we fit a model to our data and use it to predict values of the dependent variable from one or more independent variables.³ Regression analysis is a way of predicting an **outcome variable** from one **predictor variable (simple regression)** or several predictor variables (**multiple regression**). This tool is incredibly useful because it allows us to go a step beyond the data that we collected.

In section 2.4.3 I introduced you to the idea that we can predict any data using the following general equation:

$$\text{outcome}_i = (\text{model}) + \text{error}_i \quad (7.1)$$

This just means that the outcome we're interested in for a particular person can be predicted by whatever model we fit to the data plus some kind of error. In regression, the model we fit is linear, which means that we summarize a data set with a straight line (think back to Jane Superbrain Box 2.1). As such, the word 'model' in the equation above simply gets replaced by 'things that define the line that we fit to the data' (see the next section).

With any data set there are several lines that could be used to summarize the general trend and so we need a way to decide which of many possible lines to choose. For the sake of making accurate predictions we want to fit a model that *best* describes the data. The simplest way to do this would be to use your eye to gauge a line that looks as though it summarizes the data well. You don't need to be a genius to realize that the 'eyeball' method is very subjective and so offers no assurance that the model is the best one that could have been chosen. Instead, we use a mathematical technique called the *method of least squares* to establish the line that best describes the data collected.

How do I fit a straight line to my data?



³ I want to remind you here of something I discussed in Chapter 1: SAS refers to regression variables as dependent and independent variables (as in controlled experiments). However, correlational research by its nature seldom controls the independent variables to measure the effect on a dependent variable and so I will talk about 'independent variables' as *predictors*, and the 'dependent variable' as the *outcome*.

7.2.1. Some important information about straight lines ①

I mentioned above that in our general equation the word ‘model’ gets replaced by ‘things that define the line that we fit to the data’. In fact, any straight line can be defined by two things: (1) the slope (or gradient) of the line (usually denoted by b_1); and (2) the point at which the line crosses the vertical axis of the graph (known as the *intercept* of the line, b_0). In fact, our general model becomes equation (7.2) below in which Y_i is the outcome that we want to predict and X_i is the i th participant’s score on the predictor variable.⁴ Here b_1 is the gradient of the straight line fitted to the data and b_0 is the intercept of that line. These parameters b_1 and b_0 are known as the *regression coefficients* and will crop up time and time again in this book, where you may see them referred to generally as b (without any subscript) or b_i (meaning the b associated with variable i). There is a residual term, ε_i , which represents the difference between the score predicted by the line for participant i and the score that participant i actually obtained. The equation is often conceptualized without this residual term (so ignore it if it’s upsetting you); however, it is worth knowing that this term represents the fact that our model will not fit the data collected perfectly:

$$Y_i = (b_0 + b_1 X_i) + \varepsilon_i \quad (7.2)$$

A particular line has a specific intercept and gradient. Figure 7.2 shows a set of lines that have the same intercept but different gradients, and a set of lines that have the same gradient but different intercepts. Figure 7.2 also illustrates another useful point: the gradient of the line tells us something about the nature of the relationship being described. In Chapter 6 we saw how relationships can be either positive or negative (and I don’t mean the difference between getting on well with your girlfriend and arguing all the time!). A line that has a gradient with a positive value describes a positive relationship, whereas a line with a negative gradient describes a negative relationship. So, if you look at the graph in Figure 7.2

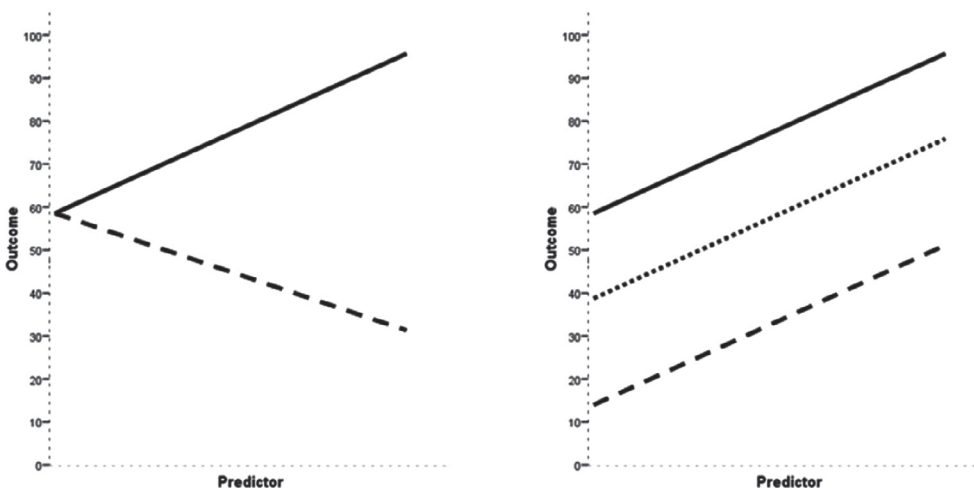


FIGURE 7.2 Lines with the same gradients but different intercepts, and lines that share the same intercept but have different gradients

⁴ You’ll sometimes see this equation written as:

$$Y_i = (\beta_0 + \beta_1 X_i) + \varepsilon_i$$

The only difference is that this equation has got β s in it instead of b s and in fact both versions are the same thing, they just use different letters to represent the coefficients.

in which the gradients differ but the intercepts are the same, then the dashed line describes a positive relationship whereas the solid line describes a negative relationship. Basically, then, the gradient (b_1) tells us what the model looks like (its shape) and the intercept (b_0) tells us where the model is (its location in geometric space).

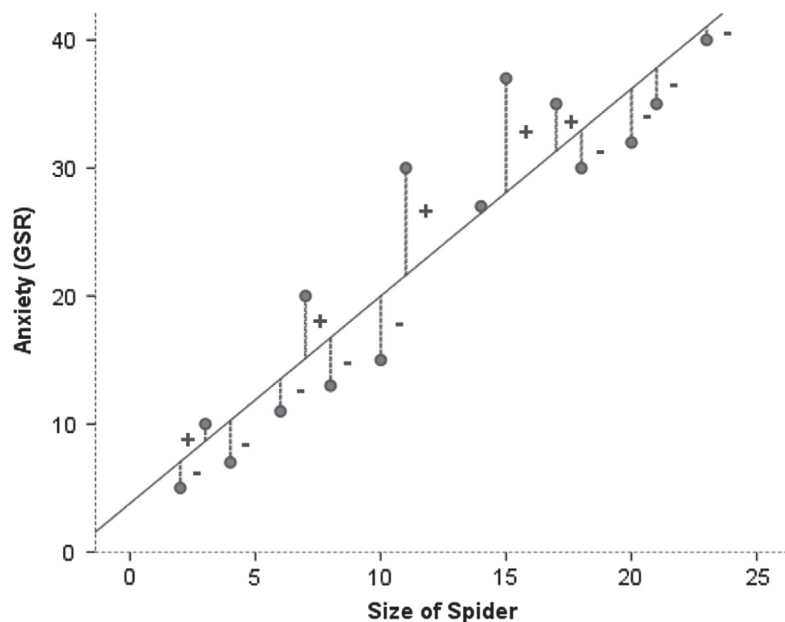
If it is possible to describe a line knowing only the gradient and the intercept of that line, then we can use these values to describe our model (because in linear regression the model we use is a straight line). So, the model that we fit to our data in linear regression can be conceptualized as a straight line that can be described mathematically by equation (7.2). With regression we strive to find the line that best describes the data collected, then estimate the gradient and intercept of that line. Having defined these values, we can insert different values of our predictor variable into the model to estimate the value of the outcome variable.

7.2.2. The method of least squares ①

I have already mentioned that the method of least squares is a way of finding the line that best fits the data (i.e. finding a line that goes through, or as close to, as many of the data points as possible). This 'line of best fit' is found by ascertaining which line, of all of the possible lines that could be drawn, results in the least amount of difference between the observed data points and the line. Figure 7.3 shows that when any line is fitted to a set of data, there will be small differences between the values predicted by the line and the data that were actually observed.

Back in Chapter 2 we saw that we could assess the fit of a model (the example we used was the mean) by looking at the deviations between the model and the actual data collected. These deviations were the vertical distances between what the model predicted and each data point that was actually observed. We can do exactly the same to assess the fit of a regression line (which, like the mean, is a statistical model). So, again we are interested in the vertical differences between the line and the actual data because the line is our model: we use it to predict values of Y from values of the X variable. In regression these differences are usually called *residuals* rather than deviations, but they are the same thing. As with the mean, data points fall both above (the model underestimates their value) and

FIGURE 7.3
This graph shows a scatterplot of some data with a line representing the general trend. The vertical lines (dotted) represent the differences (or residuals) between the line and the actual data



below (the model overestimates their value) the line, yielding both positive and negative differences. In the discussion of variance in section 2.4.2 I explained that if we sum positive and negative differences then they cancel each other out and that to circumvent this problem we square the differences before adding them up. We do the same thing here. The resulting squared differences provide a gauge of how well a particular line fits the data: if the squared differences are large, the line is not representative of the data; if the squared differences are small, the line is representative.

You could, if you were particularly bored, calculate the sum of squared differences (or SS for short) for every possible line that is fitted to your data and then compare these ‘goodness-of-fit’ measures. The one with the lowest SS is the line of best fit. Fortunately we don’t have to do this because the method of least squares does it for us: it selects the line that has the lowest sum of squared differences (i.e. the line that best represents the observed data). How exactly it does this is by using a mathematical technique for finding maxima and minima and this technique is used to find the line that minimizes the sum of squared differences. I don’t really know much more about it than that, to be honest, so I tend to think of the process as a little bearded wizard called Nephwick the Line Finder who just magically finds lines of best fit. Yes, he lives inside your computer. The end result is that Nephwick estimates the value of the slope and intercept of the ‘line of best fit’ for you. We tend to call this line of best fit a *regression line*.

7.2.3. Assessing the goodness of fit: sums of squares, R and R^2 ①

Once Nephwick the Line Finder has found the line of best fit it is important that we assess how well this line fits the actual data (we assess the **goodness of fit** of the model). We do this because even though this line is the best one available, it can still be a lousy fit to the data! In section 2.4.2 we saw that one measure of the adequacy of a model is the sum of squared differences (or more generally we assess models using equation (7.3) below). If we want to assess the line of best fit, we need to compare it against something, and the thing we choose is the most basic model we can find. So we use equation (7.3) to calculate the fit of the most basic model, and then the fit of the best model (the line of best fit), and basically if the best model is any good then it should fit the data significantly better than our basic model:

$$\text{deviation} = \sum (\text{observed} - \text{model})^2 \quad (7.3)$$

This is all quite abstract so let’s look at an example. Imagine that I was interested in predicting record sales (Y) from the amount of money spent advertising that record (X). One day my boss came into my office and said ‘Andy, I know you wanted to be a rock star and you’ve ended up working as my stats-monkey, but how many records will we sell if we spend £100,000 on advertising?’ If I didn’t have an accurate model of the relationship between record sales and advertising, what would my best guess be? Well, probably the best answer I could give would be the mean number of record sales (say, 200,000) because on average that’s how many records we expect to sell. This response might well satisfy a brainless record company executive (who didn’t offer my band a record contract). However, what if he had asked ‘How many records will we sell if we spend £1 on advertising?’ Again, in the absence of any accurate information, my best guess would be to give the average number of sales (200,000). There is a problem: whatever amount of money is spent on advertising I



always predict the same level of sales. As such, the mean is a model of ‘no relationship’ at all between the variables. It should be pretty clear then that the mean is fairly useless as a model of a relationship between two variables – but it is the simplest model available.

So, as a basic strategy for predicting the outcome, we might choose to use the mean, because on average it will be a fairly good guess of an outcome, but that’s all. Using the mean as a model, we can calculate the difference between the observed values, and the values predicted by the mean (equation (7.3)). We saw in section 2.4.1 that we square all of these differences to give us the sum of squared differences. This sum of squared differences is known as the **total sum of squares** (denoted SS_T) because it is the total amount of differences present when the most basic model is applied to the data. This value represents how good the mean is as a model of the observed data. Now, if we fit the more sophisticated model to the data, such as a line of best fit, we can again work out the differences between this new model and the observed data (again using equation (7.3)). In the previous section we saw that the method of least squares finds the best possible line to describe a set of data by minimizing the difference between the model fitted to the data and the data themselves. However, even with this optimal model there is still some inaccuracy, which is represented by the differences between each observed data point and the value predicted by the regression line. As before, these differences are squared before they are added up so that the directions of the differences do not cancel out. The result is known as the **sum of squared residuals** or **residual sum of squares** (SS_R). This value represents the degree of inaccuracy when the best model is fitted to the data. We can use these two values to calculate how much better the regression line (the line of best fit) is than just using the mean as a model (i.e. how much better is the best possible model than the worst model?). The improvement in prediction resulting from using the regression model rather than the mean is calculated by calculating the difference between SS_T and SS_R . This difference shows us the reduction in the inaccuracy of the model resulting from fitting the regression model to the data. This improvement is the *model sum of squares* (SS_M). Figure 7.4 shows each sum of squares graphically.

If the value of SS_M is large then the regression model is very different from using the mean to predict the outcome variable. This implies that the regression model has made a big improvement to how well the outcome variable can be predicted. However, if SS_M is small then using the regression model is little better than using the mean (i.e. the regression model is no better than taking our ‘best guess’). A useful measure arising from these sums of squares is the proportion of improvement due to the model. This is easily calculated by dividing the sum of squares for the model by the total sum of squares. The resulting value is called R^2 and to express this value as a percentage you should multiply it by 100. R^2 represents the amount of variance in the outcome explained by the model (SS_M) relative to how much variation there was to explain in the first place (SS_T). Therefore, as a percentage, it represents the percentage of the variation in the outcome that can be explained by the model:

$$R^2 = \frac{SS_M}{SS_T} \quad (7.4)$$

This R^2 is the same as the one we met in Chapter 6 (section 6.5.2.3) and you might have noticed that it is interpreted in the same way. Therefore, in simple regression we can take the square root of this value to obtain Pearson’s correlation coefficient. As such, the correlation coefficient provides us with a good estimate of the overall fit of the regression model, and R^2 provides us with a good gauge of the substantive size of the relationship.

A second use of the sums of squares in assessing the model is through the F -test. I mentioned way back in Chapter 2 that test statistics (like F) are usually the amount of systematic variance divided by the amount of unsystematic variance, or, put another way, the model compared against the error in the model. This is true here: F is based upon the ratio of the improvement due to the model (SS_M) and the difference between the model and the observed data (SS_R). Actually, because the sums of squares depend on the number

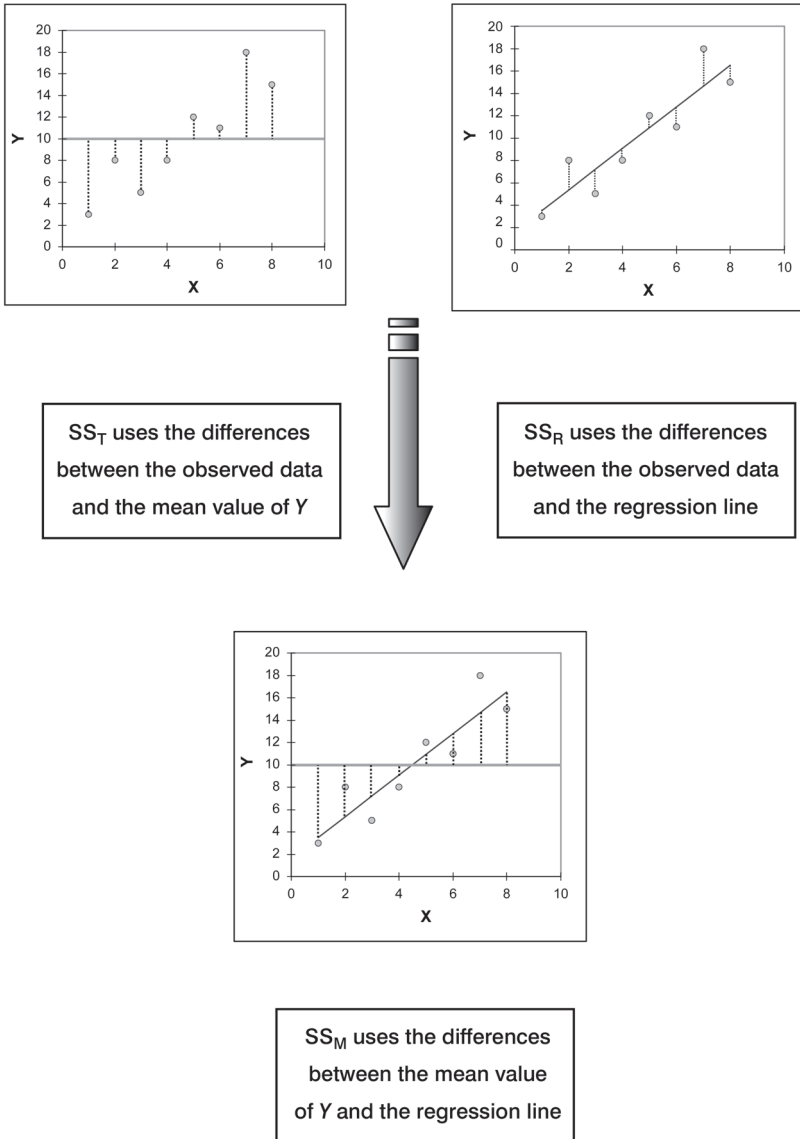


FIGURE 7.4
Diagram showing from where the regression sums of squares derive

of differences that we have added up, we use the average sums of squares (referred to as the **mean squares** or MS). To work out the mean sums of squares we divide by the degrees of freedom (this is comparable to calculating the variance from the sums of squares – see section 2.4.2). For SS_M the degrees of freedom are simply the number of variables in the model, and for SS_R they are the number of observations minus the number of parameters being estimated (i.e. the number of beta coefficients including the constant). The result is the mean squares for the model (MS_M) and the residual mean squares (MS_R). At this stage it isn't essential that you understand how the mean squares are derived (it is explained in Chapter 10). However, it is important that you understand that the **F-ratio** (equation (7.5)) is a measure of how much the model has improved the prediction of the outcome compared to the level of inaccuracy of the model:

$$F = \frac{MS_M}{MS_R} \tag{7.5}$$

If a model is good, then we expect the improvement in prediction due to the model to be large (so MS_M will be large) and the difference between the model and the observed data to be small (so MS_R will be small). In short, a good model should have a large F -ratio (greater than 1 at least) because the top of equation (7.5) will be bigger than the bottom. The exact magnitude of this F -ratio can be assessed using critical values for the corresponding degrees of freedom (as in the Appendix).

7.2.4. Assessing individual predictors ①

We've seen that the predictor in a regression model has a coefficient (b_1), which in simple regression represents the gradient of the regression line. The value of b represents the change in the outcome resulting from a unit change in the predictor. If the model was useless at predicting the outcome, then if the value of the predictor changes, what might we expect the change in the outcome to be? Well, if the model is very bad then we would expect the change in the outcome to be zero. Think back to Figure 7.4 (see the panel representing SS_T) in which we saw that using the mean was a very bad way of predicting the outcome. In fact, the line representing the mean is flat, which means that as the predictor variable changes, the value of the outcome does *not* change (because for each level of the predictor variable, we predict that the outcome will equal the mean value). The important point here is that a bad model (such as the mean) will have regression coefficients of 0 for the predictors. A regression coefficient of 0 means: (1) a unit change in the predictor variable results in no change in the predicted value of the outcome (the predicted value of the outcome does not change at all); and (2) the gradient of the regression line is 0, meaning that the regression line is flat. Hopefully, you'll see that it logically follows that if a variable significantly predicts an outcome, then it should have a b -value significantly different from zero. This hypothesis is tested using a t -test (see Chapter 9). The **t-statistic** tests the null hypothesis that the value of b is 0: therefore, if it is significant we gain confidence in the hypothesis that the b -value is significantly different from 0 and that the predictor variable contributes significantly to our ability to estimate values of the outcome.

Like F , the t -statistic is also based on the ratio of explained variance to unexplained variance or error. Well, actually, what we're interested in here is not so much variance but whether the b we have is big compared to the amount of error in that estimate. To estimate how much error we could expect to find in b we use the standard error. The standard error tells us something about how different b -values would be across different samples. We could take lots and lots of samples of data regarding record sales and advertising budgets and calculate the b -values for each sample. We could plot a frequency distribution of these samples to discover whether the b -values from all samples would be relatively similar, or whether they would be very different (think back to section 2.5.1). We can use the standard deviation of this distribution (known as the *standard error*) as a measure of the similarity of b -values across samples. If the standard error is very small, then it means that most samples are likely to have a b -value similar to the one in our sample (because there is little variation across samples). The t -test tells us whether the b -value is different from 0 relative to the variation in b -values across samples. When the standard error is small even a small deviation from zero can reflect a meaningful difference because b is representative of the majority of possible samples.

Equation (7.6) shows how the t -test is calculated and you'll find a general version of this equation in Chapter 9 (equation (9.1)). The b_{expected} is simply the value of b that we would expect to obtain if the null hypothesis were true. I mentioned earlier that the null

hypothesis is that b is 0 and so this value can be replaced by 0. The equation simplifies to become the observed value of b divided by the standard error with which it is associated:

$$\begin{aligned} t &= \frac{b_{\text{observed}} - b_{\text{expected}}}{SE_b} \\ &= \frac{b_{\text{observed}}}{SE_b} \end{aligned} \tag{7.6}$$

The values of t have a special distribution that differs according to the degrees of freedom for the test. In regression, the degrees of freedom are $N - p - 1$, where N is the total sample size and p is the number of predictors. In simple regression when we have only one predictor, this reduces down to $N - 2$. Having established which t -distribution needs to be used, the observed value of t can then be compared to the values that we would expect to find if there was no effect (i.e. $b = 0$): if t is very large then it is unlikely to have occurred when there is no effect (these values can be found in the Appendix). SAS provides the exact probability that the observed value (or a larger one) of t would occur if the value of b was, in fact, 0. As a general rule, if this observed significance is less than .05, then scientists assume that b is significantly different from 0; put another way, the predictor makes a significant contribution to predicting the outcome.

7.3. Doing simple regression on SAS ①

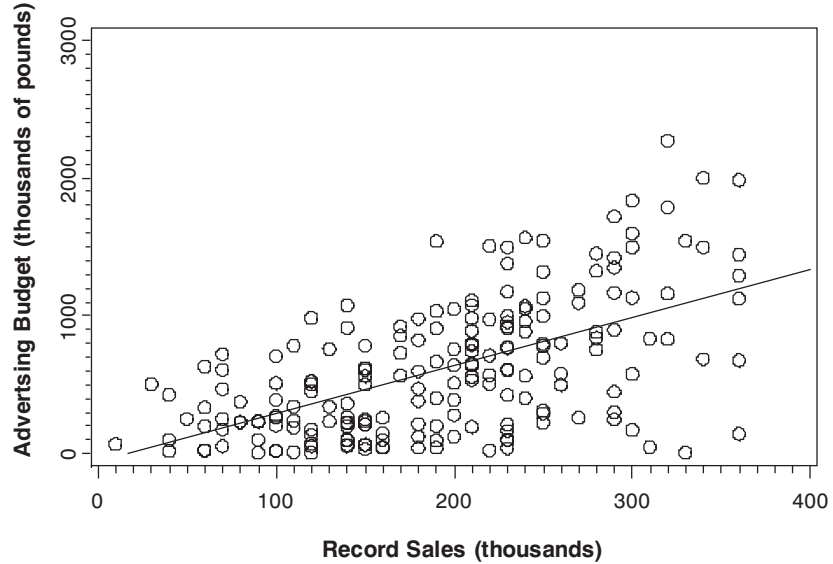
So far, we have seen a little of the theory behind regression, albeit restricted to the situation in which there is only one predictor. To help clarify what we have learnt so far, we will go through an example of a simple regression on SAS. Earlier on I asked you to imagine that I worked for a record company and that my boss was interested in predicting record sales from advertising. There are some data for this example in the file **Record1.sas7bdat**. This data file has 200 rows, each one representing a different record. There are also two columns, one representing the sales of each record in the week after release and the other representing the amount (in pounds) spent promoting the record before release. This is the format for entering regression data: the outcome variable and any predictors should be entered in different columns, and each row should represent independent values of those variables.

The pattern of the data is shown in Figure 7.5 and it should be clear that a positive relationship exists: so, the more money spent advertising the record, the more it is likely to sell. Of course there are some records that sell well regardless of advertising (top left of scatterplot), but there are none that sell badly when advertising levels are high (bottom right of scatterplot). The scatterplot also shows the line of best fit for these data: bearing in mind that the mean would be represented by a flat line at around the 200,000 sales mark, the regression line is noticeably different.

To find out the parameters that describe the regression line, and to see whether this line is a useful model, we need to run a regression analysis. To do the analysis you need to use PROC REG.

PROC REG syntax is very straightforward, and is shown in SAS Syntax 7.1. The MODEL statement is written in the form of the equation: we think that sales are a function of adverts, so we write `sales = adverts`. Notice that PROC REG is slightly different to other procedures, because we need to write `QUIT;` after the `RUN` statement.

FIGURE 7.5
Scatterplot showing the relationship between record sales and the amount spent promoting the record



```
PROC REG DATA=chapter6.record1;
  MODEL sales = adverts;
  RUN;
QUIT;
SAS Syntax 7.1
```

7.4. Interpreting a simple regression ①

7.4.1. Overall fit of the model ①

The output from the regression is shown in SAS Output 7.1. The first part of the output reports an analysis of variance (ANOVA – see Chapter 10). The summary table shows the various sums of squares described in Figure 7.4 and the degrees of freedom associated with each. From these two values, the average sums of squares (the mean squares) can be calculated by dividing the sums of squares by the associated degrees of freedom. The most important part of the table is the F -ratio, which is calculated using equation (7.5), and the associated significance value of that F -ratio. For these data, F is 99.59, which is significant at $p < .001$ (because the value in the column labelled Sig. is less than .001). Researchers usually don't report p -values below 0.001 though. This result tells us that there is less than a 0.1% chance that an F -ratio at least this large would happen if the null hypothesis were true. Therefore, we can conclude that our regression model results in significantly better prediction of record sales than if we used the mean value of record sales. In short, the regression model overall predicts record sales significantly well.

The second part of the output is a summary of the model. This summary table provides the value of R and R^2 for the model that has been derived (as well as some other things we are not going to worry about for now). For these data, R has a value of .578 and because there is only one predictor, this value represents the simple correlation between advertising

The REG Procedure
 Model: MODEL1
 Dependent Variable: SALES Record Sales (thousands)

SAS OUTPUT 7.1

Number of Observations Read	200
Number of Observations Used	200

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	433688	433688	99.59	<.0001
Error	198	862264	4354.86953		
Corrected Total	199	1295952			

Root MSE	65.99144	R-Square	0.3346
Dependent Mean	193.20000	Adj R-Sq	0.3313
Coeff Var	34.15706		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	134.13994	7.53657	17.80	<.0001
ADVERTS	Advertising Budget (thousands of pounds)	1	0.09612	0.00963	9.98	<.0001

and record sales (you can confirm this by running a correlation using what you were taught in Chapter 6). The value of R^2 is .335, which tells us that advertising expenditure can account for 33.5% of the variation in record sales. In other words, if we are trying to explain why some records sell more than others, we can look at the variation in sales of different records. There might be many factors that can explain this variation, but our model, which includes only advertising expenditure, can explain approximately 33% of it. This means that 67% of the variation in record sales cannot be explained by advertising alone. Therefore, there must be other variables that have an influence also.

7.4.2. Model parameters ①

The ANOVA tells us whether the model, overall, results in a significantly good degree of prediction of the outcome variable. However, the ANOVA doesn't tell us about the individual contribution of variables in the model (although in this simple case there is only one variable in the model and so we can infer that this variable is a good predictor. The third part of the output provides details of the model parameter estimates (the beta values) and the significance of these values. We saw in equation (7.2) that b_0 was the Y intercept and this value is the value labelled *Parameter Estimate* (in the SAS output) for the constant. So, from the table, we

How do I interpret b values?



can say that b_0 is 134.14, and this can be interpreted as meaning that when no money is spent on advertising (when $X = 0$), the model predicts that 134,140 records will be sold (remember that our unit of measurement was thousands of records). We can also read off the value of b_1 from the table and this value represents the gradient of the regression line. It is 0.096. Although this value is the slope of the regression line, it is more useful to think of this value as representing *the change in the outcome associated with a unit change in the predictor*. Therefore, if our predictor variable is increased by one unit (if the advertising budget is increased by 1), then our model predicts that 0.096 extra records will be sold. Our units of measurement were thousands of pounds and thousands of records sold, so we can say that for an increase in advertising of £1000 the model predicts 96 ($0.096 \times 1000 = 96$) extra record sales. As you might imagine, this investment is pretty bad for the record company: it invests £1000 and gets only 96 extra sales! Fortunately, as we already know, advertising accounts for only one-third of record sales.

We saw earlier that, in general, values of the regression coefficient b represent the change in the outcome resulting from a unit change in the predictor and that if a predictor is having a significant impact on our ability to predict the outcome then this b should be different from 0 (and big relative to its standard error). We also saw that the t -test tells us whether the b -value is different from 0. SAS provides the exact probability that the observed value of t would occur if the value of b in the population were 0. If this observed significance is less than .05, then scientists agree that the result reflects a genuine effect (see Chapter 2). For these two values, the probabilities are $<.0001$ and so we can say that the probability of these t values (or larger) occurring if the values of b in the population were 0 is less than .0001. Therefore, the b s are different from 0 and we can conclude that the advertising budget makes a significant contribution ($p < .0001$) to predicting record sales.



SELF-TEST How is the t in SAS Output 7.1 calculated? Use the values in the table to see if you can get the same value as SAS.

7.4.3. Using the model ①

So far, we have discovered that we have a useful model, one that significantly improves our ability to predict record sales. However, the next stage is often to use that model to make some predictions. The first stage is to define the model by replacing the b -values in equation (7.2) with the values from SAS Output 7.1. In addition, we can replace the X and Y with the variable names so that the model becomes:

$$\begin{aligned} \text{record sales}_i &= b_0 + b_1 \text{advertising budget}_i \\ &= 134.14 + (0.096 \times \text{advertising budget}_i) \end{aligned} \quad (7.7)$$

It is now possible to make a prediction about record sales, by replacing the advertising budget with a value of interest. For example, imagine a record executive wanted to spend £100,000 on advertising a new record. Remembering that our units are already in thousands of pounds; we can simply replace the advertising budget with 100. He would discover that record sales should be around 144,000 for the first week of sales: