1.2 Quantitative Aspects

Interpreting probabilities

Ecologists need to know, as do any scientists dealing with sets of data, what conclusions can be drawn from those data. Imagine we are interested in determining whether high abundances of a pest insect in summer are associated with high temperatures the previous spring, and imagine we have data on summer insect abundances and mean spring temperatures for each of a number of years. How do we use statistical analysis to conclude, with a stated degree of confidence, either than there is or is not a relationship between the spring temperature and summer insect numbers?

Null hypotheses and *P*-values

To carry out a statistical test we first need a *null hypothesis*, which simply means in this case that there is *no* association; that is, no association between insect abundance and temperature. The statistical test (stated simply) then generates a probability (a *P*-value) of getting a data set like ours if the null hypothesis is correct.

Suppose the data were like those in Figure 1.7a. The probability generated by a statistical test of association on these data is $P = 0.5$ (equivalently 50%). This means that, if the null hypothesis really was correct (no association), then 50% of studies like ours should generate just such a data set, or one even further from the null hypothesis. We therefore could have no confidence in any claim that there *was* an association.

Suppose, however, that the data were like those in Figure 1.7b, where the *P*-value is 0.001 (0.1%). This would mean that such a data set (or one even further from the null hypothesis) could be expected in only 0.1% of similar studies if there was really no association. In other words, either something very improbable has occurred, or there *was* an association between insect abundance and spring temperature. Thus, since we do not expect highly improbable events to occur, we can have a high degree of confidence in the claim that there *was* an association between abundance and temperature.

Significance testing

Both 50% and 0.01%, though, make things easy for us. Where, between the two, do we draw the line? There is no absolute answer to this, but scientists and statisticians have established a convention in *significance testing*, which says that if *P* is less than 0.05 (5%), written $P < 0.05$ (e.g., Figure 1.7d), then results are described as 'statistically significant' and confidence can be placed in the effect being examined; whereas if $P > 0.05$, then there is no statistical foundation for claiming the effect exists (e.g., Figure 1.7c). A further elaboration of the convention often describes results with $P < 0.01$ as 'highly significant.'

'Insignificant' results?

Some effects are naturally strong (there is a powerful association between people's weight and their height) and others are weak (the association between people's weight and their risk of heart disease is real but weak, since weight is only one of many important factors). More data are needed to establish support for a weak effect than for a strong one. Hence a *P*-value of greater than 0.05 (lack of statistical significance) may mean one of two things in an ecological study:

1. There really is no effect of ecological importance.
2. The data are simply not good enough, or there are not enough of them, to support the effect even though it exists, possibly because the effect itself is real but weak.

Throughout this book, then, studies of a wide range of types are described, and their results often have *P*-values attached to them. Remember that statements like $P < 0.05$ and $P < 0.01$ mean that these are studies where: (i) sufficient data have been collected to establish a conclusion in which we can be confident; (ii) that confidence has been established by agreed means (statistical testing); and (iii) confidence is being measured on an agreed and interpretable scale.

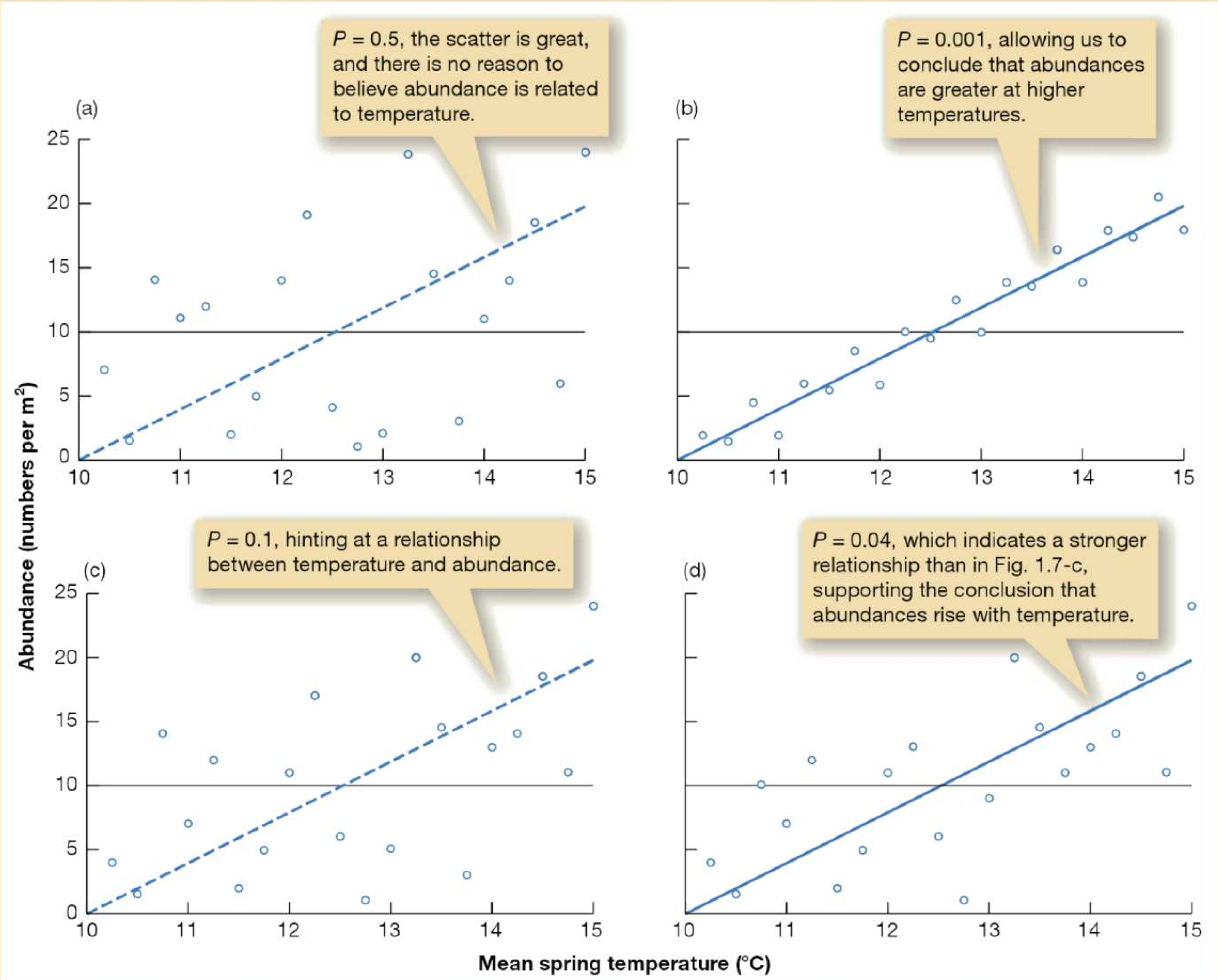

FIGURE 1.7 The results from four hypothetical studies of the relationship between insect pest abundance in summer and mean temperature the previous spring. In each case, the points are the data actually collected. Horizontal lines represent the *null hypothesis* – that there is no association between abundance and temperature, and thus the best estimate of expected insect abundance, irrespective of spring temperature, is the mean insect abundance overall. The second line is the *line of best fit* to the data, which in each case offers some suggestion that abundance rises as temperature rises. However, whether we can be confident in concluding that abundance does rise with temperature depends, as explained in the text, on statistical tests applied to the data sets. (a) The suggestion of a relationship is weak ($P = 0.5$). There are no good grounds for concluding that the true relationship differs from that supposed by the null hypothesis and no grounds for concluding that abundance is related to temperature. (b) The relationship is strong ($P = 0.001$) and we can be confident in concluding that abundance increases with temperature. (c) The results are suggestive ($P = 0.1$) but it would not be safe to conclude from them that abundance rises with temperature. (d) The results are not vastly different from those in (c) but are powerful enough ($P = 0.04$, i.e., $P < 0.05$) for the conclusion that abundance rises with temperature to be considered safe.

Standard errors and confidence intervals

Another way in which our confidence in results is assessed is through reference to 'standard errors,' which statistical tests often allow to be attached either to mean values calculated from a set of observations or to slopes of lines like those in Figure 1.7. These mean values and slopes can only ever be estimates of the 'true' mean value or slope, because they are calculated from data that are only a sample of all the imaginable items of data that could be collected. The standard error, then, sets a band around the estimated value within which the true value can be expected to lie, with a given, stated probability. In particular, there is a 95% probability that the true mean lies within roughly two standard errors (2 SE) of the estimated mean; we call this the *95% confidence interval*.

Large standard errors (little confidence in the estimated value) can arise when data are, for whatever reason, highly variable; but they may also be due to only a small data set having been collected. Standard errors are smaller, and confidence in estimates greater, *both* when data are more consistent (less variable) *and* when there are more data.