Test finder and help guide

Test finder

A Looking for a difference or a trend? (if unsure, go to *Help* 1)

- Trend	– go to B
- Difference	– go to C

B Have you fixed the X-values experimentally (looking to see whether changes in X cause changes in Y) or do you require the equation of the line itself (e.g. for predicting new values)? (if unsure, go to *Help* 2)

- No	– correlation
- Yes	– linear regression

C Does each data value belong to treatments within one level of grouping, or does it belong to more than one level of grouping? (if unsure, go to *Help* 3 and *Help* 6)

- One level of grouping (one-way designs)	– go to D
- More than one level of grouping (n-way designs)	– go to G

D Are data for each treatment replicated? (if unsure, go to *Help* 4)

- No	– go to E
- Yes	– go to F

E Data are in the form of counts

- No	– (not analyzable)
- Yes	– $1 \times n$ chi-squared

F Data in one treatment are independent of those in other treatments (if unsure, go to *Help* 5)

- No	– only two treatments	– matched-pairs tests (or one-way repeated-measures ANOVA)
	– two or more treatments	– one-way repeated-measures ANOVA
- Yes	– only two treatments	– *t*-test or Mann–Whitney test (or one-way ANOVA)
	– two or more treatments	– one-way ANOVA

G Are data for treatments within each level of grouping replicated? (if unsure, go to *Help* 6)

- No	– $n \times n$ chi-squared
- Yes	– *n*-way ANOVA

Help 1 Difference or trend?

- *Difference* predictions are concerned with some kind of difference between two or more groups of data. The groups could be based on any characteristic that can be used to make a clear-cut distinction, e.g. sex, drug treatment, habitat. Thus a difference might be predicted between the growth rates of men and women, or between the development of disease in rats given drug A versus those given drug B versus those given a placebo.
- *Trend* predictions are concerned not with differences between mutually exclusive groupings but with the relationship between two more or less continuously distributed measures, e.g. the relationship between the size of a shark and the size of prey it takes, or the relationship between the amount of rainfall in a growing season and the number of apples produced by an apple tree.

Help 2 What sort of trend?

The basic choice here is between fitting a line (regression) or not (correlation).

Correlation is used whenever we merely want to know whether there is an association between X and Y – there is no real dependent or independent variable, and when plotted as a scattergraph you could equally well plot Y against X as opposed to X against Y.

Regression was developed for situations where the experimenter manipulates the values of the *independent* variable (X) and measures the impact of these manipulations on another *dependent* variable (Y). You are therefore looking for a causal relationship between X and Y – changes in X cause changes in Y. However, because knowing the slope of the relationship between X and Y is useful in many other contexts, the use of regression has expanded to incorporate many cases where the X values are merely measured rather than manipulated. Technically this is wrong, but this usage is so firmly embedded in biological practice that the majority of scientific investigations do it. An example is when you want to use the relationship to predict the value of Y for a particular X value. Comparisons of slopes are also extremely informative, and used extensively in biological research.

Help 3 Levels of grouping

Many difference predictions are concerned with differences at just *one* level of grouping, e.g. differences in faecal egg counts following treatment of mice with one of four different anthelminthic drugs. Here drug treatment is the only level of grouping in which we are interested. However, if we wished, say, to distinguish between the effects of different drugs on male and female mice, we should be dealing with *two* levels of grouping: drug treatment and sex.

Help 4 Replication

Replication simply means that each treatment within a level of grouping has more than one data value in it. The table below shows replicated data as columns of values within each treatment in a one-way (i.e. one level of grouping) design:

	Group		
	Pesticide A	**Pesticide B**	**Control**
	10	30	0
	5	27	1
% mortality of pest	3	50	1
	0	6	0
	1	3	2
	20	3	5

Help 5 Independence

Unless specifically allowed for in the analysis, all statistical analyses assume that each data value is independent of all others. Each value in one group is *non-independent* of one from each of the other groups if they have something in common (e.g. it is measured on the same individual, or derives from animals from the same cage – i.e. a given individual is exposed to each treatment in turn, or the same cage provides animals for each treatment). The source of non-independence thus needs to be taken into account in any analysis, and so requires a so-called repeated-measures design.

Help 6 Rows and columns

If data have been collected at two levels of grouping, then each data value can be thought of as belonging to both a row and a column (i.e. to one row/column cell) in a table, where rows refer to one level of grouping (say sex – *see* Help 4) and columns to the other (drug treatment – *see* Help 4). If there are several values per row/column cell, as below for the number of individuals dying during a period of observation:

		Treatment	
		Experimental	**Control**
Sex	Male	3, 4, 8, 12	23, 24, 12, 32
	Female	1, 0, 2, 9	32, 45, 31, 21

then a two-way analysis of variance is appropriate. If there is just a single *count* in each cell, as in the number of male and female fish responding to an experimental or control odour stimulus:

		Treatment	
		Experimental	**Control**
Sex	Male	27	91
	Female	12	129

then an $n \times n$ chi-squared test is appropriate.