Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

Many experiments record frequencies with which events occur, and then these frequencies are used to calculate proportions to be used as estimates for population probabilities.

As a simple example, consider tossing a coin N times resulting in h heads and $N - h$ tails. Then the frequency of heads would be the integer h, while the proportion of heads would be the floating point number h/N, which would converge to the true probability of a head occurring for large values of N. In this case of dichotomous data we could define a variable x_i to have a value of 1 for a success (e.g. heads) and 0 for failure (e.g. tails) at the i'th trial, leading to a random variable X as the sum of the x_i values as follows

$$X = x_1 + x_2 + \cdots + x_N.$$

Then the appropriate statistical model would be the binomial distribution with parameters N and p, so that the probability of observing h successes in N independent trials would be

$$P(X = h) = \binom{N}{h} p^h (1-p)^{N-h}$$

and

$$\lim_{N\to\infty} \frac{h}{N} = p.$$

More generally, suppose that a total of N observations can be classified into k categories with frequencies consisting of y_i observations in category i, so that $0 \le y_i \le N$ and $\sum_{i=1}^{k} y_i = N$, then there are k proportions, that is ratios r_i of frequencies to sample size, defined as

$$r_i = y_i/N,$$

of which only $k - 1$ are independent due to the fact that

$$r_1 + r_2 + \cdots + r_k = 1.$$

If these proportions are then interpreted as estimates of the multinomial probabilities and it is wished to make inferences about these probabilities, then we are in a situation that can be described as the analysis of proportions, or the analysis of categorical data.

Since the observations are integer frequencies and not measurements, they are not normally distributed, so techniques like ANOVA should not be used, instead specialized methods to analyze frequencies must be employed. In particular, exact estimates for variances and confidence limits are not always available, and approximate confidence range estimates often exceed the theoretically possible limits since, for an estimate, say $\hat{p}$, with lower 95% confidence limit C_L, and upper 95% confidence limit C_U we must have

$$0 \le C_L \le \hat{p} \le C_U \le 1.$$

Furthermore, although exact confidence limits will not be symmetrical, approximate confidence limits will be. For example, with 2 successes in 10 trials the estimate for the binomial parameter would be

$$\hat{p} = 0.2,$$

but the exact 95% confidence range calculated by SIMFIT was found to be

$$0.0252 \le 0.2 \le 0.5561$$

so that $C_L = 0.2 - 0.1748$ while $C_U = 0.2 + 0.3561$. This illustrates a typical result that, for probability estimates less than 0.5 confidence ranges are skewed to the right, while for estimates greater than 0.5 confidence ranges are skewed to the left. So it is not accurate to report estimates as, e.g. $\hat{p} = (h/N) \pm \alpha$ for some α.