Categorizing statistics problems

Below are some questions you should ask yourself if you're not sure what kind of problem you're dealing with in statistics. Remember that not all of these questions apply to every problem, and also that this list is a work in progress and therefore may be incomplete. Suggestions are welcome (to Alan Weiss in the Math Lab).

Is this just a simple computation?
After you've been doing a lot of more complicated things, don't forget that you'll still get some problems that are simple things like computing the probability that a variable with a normal distribution is in a certain range of values, or computing a standard deviation. Especially, don't forget that probability computations for uniform distributions are very simple and don't require the use of functions like normalcdf.

Is this a hypothesis test or a confidence interval?
A hypothesis test means choosing which of two alternatives is a better description of the population. A confidence interval is a range of values within which a certain parameter value is likely to be. These two things are not entirely unrelated, but you are generally asked to do one or the other.

Are you dealing with a mean or a proportion?
That is, is the problem about the parameter μ or p? In a means problem, there is a measurement made for each subject or unit in the sample and the mean of those measurements (the sample mean, $\bar{x}$) is computed. In a proportion problem a yes-or-no question is answered for each subject or unit and the number of "yes" answers (the number of "successes") is divided by the sample size to get the sample proportion, $\hat{p}$.

For means, should you use z or t?
If you know σ (the population standard deviation) then always use z. If not, use t, unless you're using the Larson book and the sample size is 30 or more, in which case use z. Remember that sample proportions always have a normal (i.e., z) distribution, never t.

Are there one, two, or more populations, and, if two, are they dependent or independent?
Different procedures are used for a single population (where a mean or proportion is compared to a particular numeric value), two populations (where the means or proportions of two groups are compared with each other), or more than two populations (where means – never proportions – are compared using ANOVA). For means of two populations, you must decide whether the populations are dependent (i.e., paired), or independent.

Is this bivariate data?
That is, are there sets of numbers in pairs (e.g., age and forearm length) measured for each unit or subject? If so, you should probably be thinking about a least-squares regression and finding the correlation coefficient, r.

Are you seeing if categorical (qualitative) counts match expected values?
If so, you should be using a χ^2 (chi-square) test.

Last updated 2012-08-12 using LaTeX 2ε.