How To Know What To Use

Selection of statistical techniques depends on what we ask of the data and how we ask it. It also depends on the study design (esp. sampling method) and the measurement scales used to record variables. Other important considerations include whether the study is primarily descriptive or inferential, and whether the inferential focus is primarily one of estimation or hypothesis testing. Finally, the meeting or failure to meet underlying statistical assumptions (such as linearity, independence, normality, equal variance) must be considered.

Over the years, statistical texts (and now web sites) have come up with fairly rigorous and tight algorithms for selecting an appropriate statistical technique or test based on the criteria listed above (e.g., see http://members.aol.com/johnp71/javastat.html#WhichAnalysis). With this said, I think it is a mistake to get locked into any particular algorithm, primarily because no algorithm can address the most important aspect of selection: the judgement needed to know what you want to learn from the data. Therefore, only rough guides are given here for selecting the right *direction* for analysis.

Two key factors, of course, are

- The measurement scale of the outcome (dependent, disease) variable
- The sampling method used to collect data

In the HS167/HS267 sequence, we consider three primary measurement scales. These are continuous (quantitative, scale), ordinal (semi-quantitative, ranked), and categorical (qualitative, nominal). However, to simplify matters, we may treat ordinal outcomes as either continuous or categorical, based on judgement and depending on whether distributional assumptions can be met.[1] Therefore, we currently consider only two types of outcome measures. These are:

1. "Continuous" (quantitative, scale)
2. Categorical (qualitative, nominal)

We've also considered three general sampling methods. These were:

a. Single sample
b. Paired (matched) samples
c. Independent samples involving discrete groups
d. Independent samples involving a continuous predictor

Analyses can now be classified based on the nature of the outcome (dependent) variable and sample. For example, suppose you want to determine the relationship between fasting serum cholesterol (mg/dl) and behavior type, with behavior based on scores ranging from 0 to 100. The dependent variable (cholesterol, mg/dl) is continuous, and the independent variable (behavior scale) is also continuous. Assuming sampling independence, this would be classified as a 1.d. type of problem. On the other hand, if you we were investigating serum cholesterol level (mg/dl) and behavior type categorized as Type A or Type B, this would be a 1.c. problem. Finally, if you were look at cholesterol level classified as "hypercholesterolemic" or "normocholesterolemic (categorical dependent variable) in Type A and Type B people, you would have a 3.c. problem.

With this said, you should be able to find yourself to the correct *Chapter* in StatPrimer by which to select

[1] Often, methods based on normal sampling distributions should be avoided with ordinal outcomes, in which case they should be treated as a discrete categorical outcome or non-parametric methods (e.g., see Chapter 16) should be used.

appropriate techniques. To help in this regard, here is a table to help organize your thinking:

Inferential Techniques Covered in *StatPrimer* by Outcome Measure, Sampling Method, and Chapter

1. Continuous Outcomes
   a. Single Sample: Chapter 6 (Intro to Estimation); Chapter 7 (Intro to Null Hypothesis Testing); Chapter 15 (Inference About Variances)
   b. Paired Samples: Chapter 7
   c. Independent Samples, Discrete Groups: Chapter 8 (Two groups); Chapter 15 (Inference About Vriances); Chapter 16 (*k* groups)
   d. Independent Samples, Continuous Predictor: Chapter 17 (Correlation), Chapter 18 (Regression)

2. Categorical Outcomes
   a. Single Sample - Chapter 9
   b. Paired Samples: Chapter 14 (part of the Case-Control Chapter)
   c. Independent Samples, Discrete Groups: Chapter 10 (2x2 and R-by-C); Chapter 13 (Cohort); Chapter 14 (Case-control)
   d. Independent Samples, Continuous Predictor: not covered (take course in log-linear models or logistic regression).