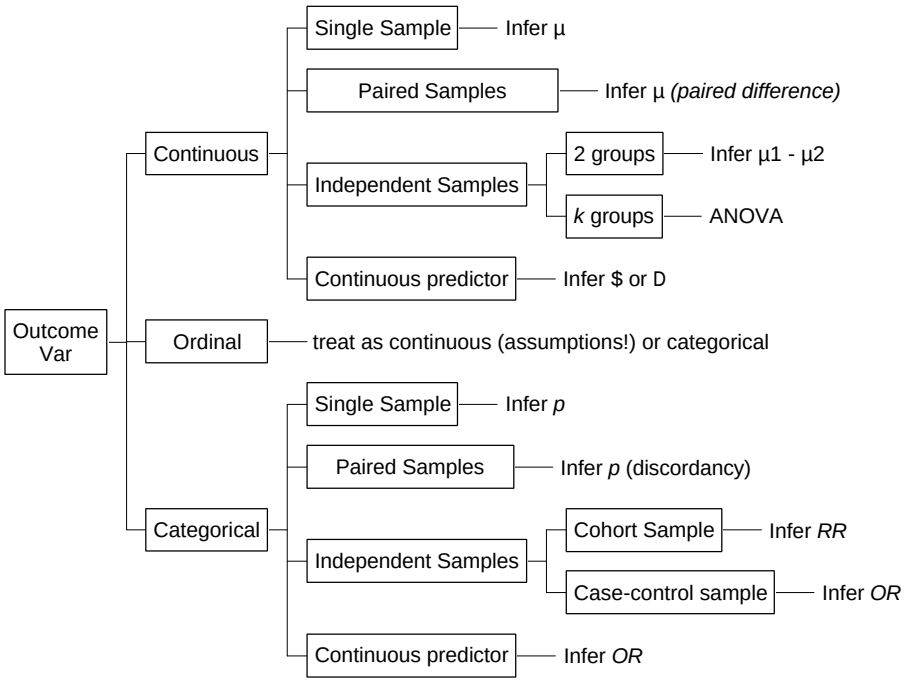
How To Know What To Use

Selection of valid statistical techniques depends on an understanding of the research question, the variables being measured, and the way data were collected. The *research question* determines conceptualization of the variables and relationship among them. For example, we may wish to study the relationship between aluminum in the diet (the exposure) and the risk of Alzheimer's disease (the outcome). All other variables (e.g., sex, age, ethnicity, social class) then become extraneous or, as we like to say, potential confounders. *Variables* may be classified as either continuous (quantitatively, scale), ordinal (ranked), or categorical (qualitative, nominal). The *sample* may be classified as a single, paired (blocked, matched), or independent. We will also consider the special case of independent samples, in which the predictor (independent variable is continuous). This, then, provides a framework for selecting appropriate statistics:



For example, if you want to study the relationship between cholesterol (continuous outcome) and type A and B behavior (independent groups), you'd want to estimate the independent mean difference and test it for significance. If you want to study heart attacks (categorical outcome) and type A and B behavior (categorical predictor), you'd want to compare the incidence of heart attacks in the form of a relative risk.

In using inferential methods, its important to remember that all inferential methods entail underlying assumptions. Assumptions may address the sampled method (e.g., independence), nature of the statistic's sampling distribution (e.g., normal), the functional relationship between factors (e.g., linear), and so on. We must always assume that the measurements are validity and sample is representative, for bias tends to

dominate variability.

Finally, never overlook the importance of judgement during statistical analyses. Always start with data of good quality and know the strength and weakness of the data set in detail. Then make careful description the first step, and "define" the population being studied as precisely as possible. Reduce the data to simple descriptions and graphs, comparing groups in intuitive ways. The strongest case for a result is one that meets causal criteria. The weakest case is simple statistical significance; always try to determine the role that bias may play in explaining your results. Finally, in assessing an association, do not rely on hypothesis tests of significance. "There are innumerable situations in which tests of statistical significance are totally unnecessary because the difference is grotesquely obvious, because it is negligible, or because, whether it be formally significant or not, it is too small to be of any practical importance" (Bradford Hill).

Exercises

Identify the outcome variable and its type for each of the study summaries described below. Then, place it into it's proper rubric, using the scheme described in this article. Identify appropriate descriptive and inferential techniques that may be used during analysis.

1. **HDL**: An investigator wishes to determine whether high density lipoprotein levels (mg/dl) differ in men and women.

2. **GLAUCOMA**: An investigator treats one eye of bilateral glaucoma with a new drug intended to lower intra-ocular pressure (mm Hg) and the other eye with a placebo.

3. **BIRTHWT**: An investigator studies the functional relationship between gestational age (weeks) and birth weight (grams).

4. **CARDIAC**: An investigator studies the relationship between cardiac output (liters/minute) and body weight (kilograms).

5. **HEADTRAU**: An investigator hypothesizes that head trauma during childhood is associated with the development of seizures. A series of head trauma cases and controls is compared with respect to their history of head trauma (yes or no).

6. **ROGAINE**: An investigator treats 50 men with hair loss pattern with either treatment I, II, or III and then studies the hair growth (new follicles per centimeter) associated with each treatment.

7. **WT&BP**: An investigator wants to determine whether systolic blood pressure (mm Hg) is related to body weight (kilograms).

8. **CLIN-TRI**: An investigator randomizing 100 women with breast cancer to a treatment and 100 women a standard therapy. Subjects are followed for 5 years to determine whether or not they survived.

9. **ANALGESIC**: We want to investigate the relationship between analgesic abuse and kidney disease by studying creatinine levels (mg/dl) in analgesic abusers and a control group.