Levels of Measurement and Choosing the Correct Statistical Test

Levels of Measurement
Most textbooks distinguish among nominal, ordinal, interval, and ratio scales based on a classification system developed by Stevens (1946).   Choice of the statistical analyses in the social sciences typically rests on a more general or cruder classification of measures into what I will call "continuous" and "discrete." Continuous refers to a variable with many possible values. By "discrete" I mean few categories. I, as well as others, often use the terms "dichotomous," "binary," "categorical," or "qualitative" synonymously with "discrete." These two general classes of measurement relate to two general classes of statistical tests—those based on *normal theory* and those based on *binomial theory.* Normal theory plays an important role in statistical tests with continuous dependent variables, such as t-tests, ANOVA, correlation, and regression, and binomial theory plays an important role in statistical tests with discrete dependent variables, such as chi-square and logistic regression.[1]

Ordinal scales with few categories (2,3, or possibly 4) and nominal measures are often classified as discrete and are analyzed using binomial class of statistical tests, whereas ordinal scales with many categories (5 or more), interval, and ratio, are usually analyzed with the normal theory class of statistical tests. Although the distinction is a somewhat fuzzy one, it is often a very useful distinction for choosing the preferred statistical test, especially when you are starting out.[2]

Type of Dependent Variable (or Scale)	Level of Measurement	General Class of Statistic	Examples of Statistical Procedures
Discrete	nominal, ordinal with 2, 3, or 4 levels	binomial	chi-square, logistic regression
Continuous	ordinal with more than 4 categories, interval, ratio	normal	ANOVA, regression, correlation, t-tests

Classifying the independent and the dependent variable as continuous or discrete will determine the type of analyses that are likely to be appropriate in a given situation.

		Dependent Variable	
		Discrete	**Continuous**
Independent variable	**Discrete** (binary and categorical)	Chi-square Logistic Regression Phi Cramer's V	t-test ANOVA Regression Point-biserial Correlation
	Continuous	Logistic Regression Point-biserial Correlation	Regression Correlation

---

[1] As we will discover later, the Pearson chi-square test really uses a normal distribution as an approximation, but the binomial (or multinomial) distribution is central to most statistics used with categorical dependent variables. I have placed chi-square with the binomial theory class of statistics, therefore, because the normal distribution is really just used as an efficient substitute for the binomial distribution.

[2] There is a longstanding debate about how to classify measurements and whether levels of measurement can be a successful guide to choice of data analysis type (e.g., Borgatta & Bohrnstedt, 1980; Townsend & Ashby, 1984). My intention is not to try to resolve the debate, but to offer a general simple heuristic as a starting place for deciding which type of analysis is used in common practice in the social sciences for general types of dependent variables. In reality, there are a number of other factors that must be considered in deciding on the most appropriate and statistically accurate analysis, including the distribution of the dependent variable, whether it is count data, and sample size among others. Think about the system I propose here as a kind of analysis triage or grand organizational scheme and trust that I will cover some of the caveats and other special considerations as we go along.

Common Practice

Although Likert-type scales are technically ordinal scales, most researchers treat them as continuous variables and use normal theory statistics with them. When there are 5 or more categories there is relatively little harm in doing this (Johnson & Creech, 1983; Zumbo & Zimmerman, 1993). Most researchers probably also use these statistics when there are 4 ordinal categories, although this may be problematic at times. Note that this distinction applies to the dependent variable used in the analysis, not necessarily the response categories used in a survey whenever multiple items are combined (e.g., by computing the mean or sum). Once two or more Likert or ordinal items are combined, the number of possible values for the composite variable begins to increase beyond 5 categories. Thus, it also is usual practice to treat these composite scores as continuous variables.

Ordinal Analyses

The contrast between discrete and continuous variables is an oversimplification. There really is a big gray area when there are 3 or 4 ordinal categories. Although in practice, most researchers only tend to use binomial and normal theory statistics, there is another class of statistical tests specifically designed for ordinal scales that are becoming increasingly available in software packages. There are several excellent references for ordinal statistical tests (Agresti, 1984, 2002; Cliff, 1996; Wickens, 1989). For regression models, Long's (1997) book is a very good, although technical, treatment. There is likely to be some statistical power advantage to using ordinal statistics over binomial statistics, and there is likely to be some accuracy gained in the statistical tests for using ordinal statistics over normal theory statistics when there are few categories or for certain other data conditions.

Problems with Crude Categorization and Artificial Dichotomization

One needs to be careful about converting continuous variables into dichotomous or categorical variables. One example is the practice of doing a "median split," which puts those with scores above and below the median into two categories, but other methods of artificial categorization can be just as problematic. Generally, a great deal of useful information is discarded, but other statistical issues arise. Although many papers have been published as far back as the 1940s on this topic, the practice of dichotomizing continuous variables is still quite prevalent. A paper by MacCullum, Zhang, Preacher, and Rucker (2002) is a superb overview of the problems and potentially serious consequences of this practice.

References and Further Readings

Agresti, A. (1984). *Analysis of ordinal categorical data.* NY: Wiley.

Agresti, A. (2002.) *Categorical Data Analysis, second edition.* NY: Wiley.

Borgatta, E.F., and Bohrnstedt, G.W. (1980). Level of measurement - Once over again. *Sociological Methods and Research, 9*, 147-160.

Cliff, N. (1996). Ordinal methods for behavioral data analysis. Mahwah, NJ: Erbaum.

Johnson, D.R., & Creech, J.C. (1983) Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review, 48*, 398-407.

Long, J.S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

MacCallum, R.X., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 1-40.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science,* 103, 677-680.

Townsend, J. T. and Ashby, F. G. (1984), Measurement Scales and Statistics: The Misconception Misconceived, *Psychological Bulletin, 96*, pp. 394-401.

Wickens, T.D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Erlbaum.

Zumbo, B.D., & Zimmerman, D.W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology, 34,* 390-400.