

Chapter 2

Characterizing data for analysis

2.1 Problem

Classify the following types of data by using Steven's measurement system: decibels of noise level, father's occupation, parts per million of an impurity in water, density of a piece of bone, rating of a wine by one judge, net profit of a firm, and score on an aptitude test.

Ans: ratio, nominal, ratio, ratio, ordinal, ratio, ratio

2.2 Problem

In a survey of users of a walk-in mental health clinic, data have been obtained on sex, age, household roster, race, education level (number of years in school), family income, reason for coming to the clinic, symptoms, and scores on screening examination. The investigator wishes to determine what variables affect whether or not coercion by the family, friends, or a governmental agency was used to get the patient to the clinic. Classify the data according to Stevens measurement system. What would you consider to be possible independent variables? Dependent Variables? Do you expect the dependent variables to be independent of each other?

Ans: sex - nominal; age - ratio; household roster - nominal; race - nominal; education level - ratio; family income - ratio; reason for coming to the clinic - nominal; symptoms - nominal; scores on screening exam - ordinal

Possible Independent Variables: sex, age, household roster, race, education level, family income, symptoms, scores on screening exam

*****WARNING: DOCUMENT HAS MANY PAGES, CONSIDER BEFORE PRINTING*****

Possible Dependent Variables: reason for coming to the clinic, scores on screening exam

The dependent variables are not necessarily independent. A patient's reason for coming to the clinic may indicate severity of illness, and if the illness is severe, the patient is likely to score poorly on the screening exam.

2.3 Problem

For the chronic respiratory study data described in Appendix A, classify each variable according to the Steven's scale and according to whether it is discrete or continuous. Pose two possible research questions and decide on the appropriate dependent and independent variables.

Ans: Data: age, height, weight, FVC, FEV1 - ratio, continuous; sex - nominal

Two possible questions: Is there a relationship between age (independent variable) and FVC (dependent variable)? Are FEV1 (dependent variable) levels different between men and women (sex: independent variable)?

2.4 Problem

Repeat problem 2.3 for the lung cancer data set described in Table 13.1.

Ans: ID hist treat painf smokfu smokbl death - nominal; staget stagen perflb - nominal or ordinal; days - ratio, continuous

Two possible questions: Is there a difference in time number of days (dependent) survived between patients with large and small tumors (staget - independent)? Is there a relationship between treatment (independent) and post-operative infection (dependent)?

2.5 Problem

From a field of statistical application (perhaps your own field of specialty), describe a data set and repeat the procedures described in Problem 2.3.

Ans: Blood Pressure Study: age, height, weight, systolic blood pressure, diastolic blood pressure - ratio, continuous; sex, race, marital status - nominal; education - ordinal

*****WARNING: DOCUMENT HAS MANY PAGES, CONSIDER BEFORE PRINTING*****

Two possible questions: Is there a difference in systolic blood pressure (dependent variable) between married and single (marital status - independent variable) people? Is there a difference in weight (dependent variable) among black, white, hispanic, and “other” (race - independent variable) patients?

2.6 Problem

If the RELIG variable described in Table 3.4 of this text was recoded, 1 = Catholic, 2 = Protestant, 3 = Jewish, 4 = none, and 5 = other, would this meet the basic empirical operation as defined by Stevens for an ordinal variable.

Ans: No, there is no hierarchy amongst religion, i.e. being one religion over another does not indicate an increase in status.

2.7 Problem

Give an example of nominal, ordinal, interval, and ratio variables from a field of application you are familiar with.

Ans: Patient Intake questionnaire: nominal - sex; ordinal - Feeling Today?(Likert Scale: 1 = poor to 5 = great); interval - (Patient’s body temperature (degrees F)); ratio - age

2.8 Problem

Data that are ordinal are often analyzed by methods that Stevens reserved for interval data. Give reasons why thoughtful investigators do this.

Ans: Thoughtful investigators may analyze do this because they are assuming that the ordinal variables have equally spaced levels.

2.9 Problem

The Parental HIV data set described in Appendix A includes the follow-

*****WARNING: DOCUMENT HAS MANY PAGES, CONSIDER BEFORE PRINTING*****

ing variables: job status of mother (JOBMO, 1=employed, 2=unemployed, 3=retired/disabled) and mother's education (EDUMO, 1=did not complete high school, 2=highschool diploma/GED, and 3=more than high school). Classify these two variables using Steven's measurement system.

Ans: JOBMO - nominal; EDUMO - ordinal

2.10 Problem

Give an example from a field that you are familiar with of an increased sophistication for measuring that has resulted in a measurement that used to be ordinal now being interval.

Ans: With medical advances in measurements of cancer cells, cancer stages can now be better defined and may be seen as interval rather than ordinal.