Classification of Statistical Problems by Type of Variable

Measurement Scales

A numerical measurement scale is the set of possible values that a variable can take on. We usually distinguish four levels of measurement scales.

Remark: For the purpose of illustration we will usually take people to be the objects that are being measured rather than say cars, houses, cats, dogs, fences, roads, or other such objects.

- Nominal Scales: On the nominal scale values of the variable are denoted by arbitrary numbers. For example suppose the variable is gender. Then, in measuring we might record the number –1 for a female and the number +1 for a male. In other words this level of measurement consists of just categories without order to these categories. Gender has no real numeric properties, and the numbers chosen here are arbitrary. Examples of nominal variables are: Region of the country, (say 1 for north, 2 for south, 3 for east, and 4 for west), Race, Color, Type of housing, *i.e.* any variable with values consisting of unordered categories.

- Ordinal Scales: Ordinal scales are similar to nominal scales in that they specify categories. However, they have the added feature that the categories can be ordered according to some property. For example, hospitals may describe the condition of patients as 1 (resting and comfortable), 2 (stable), 3 (guarded), and 4 (critical). These categories 1, 2, 3 and 4 are ordered but the numbers are arbitrary. Note that the interval between categories 4 and 3 may not be equal to the interval between categories 1 and 2. Examples of ordinal variables are Pain (none/moderate/severe), socio-economic status (low/middle/high) and class of university degree.

- Interval Scales: Interval scales categorize, order, and quantify comparisons between pairs of measurements. Variables measured on interval scales are numeric variables. Interval scales have a defined unit of measurement and the difference between two interval level numbers is a measure of the difference in the property being measured. However, on an interval scale the zero position on the scale is arbitrary. The Fahrenheit and Centigrade scales for temperature are examples of interval scales. For example, zero degrees Centigrade does not mean no temperature. Note that, because of a defined unit of measurement, interval scales have equal intervals between successive values.

- Ratio Scales: The ratio scale is similar to the interval scale, but the position of zero is unique and indicates absence of the property. A ratio scale categorizes, orders, and quantifies comparisons between variables. For example, the Kelvin temperature scale takes absolute zero as its unique zero (absence of heat), therefore the Kelvin temperature scale is a ratio scale. It is true that 40 degree K is twice as hot as 20 degree K. Length measured in feet, or miles, etc. is a ratio scale since it has a unit of measurement with an absolute origin, zero feet tall means absence of the property of height. In the ratio scale it is sensible to compare measurements, for example it is sensible to say that one person is twice as tall as another person. If it makes sense to say that one value is, say, k times another then the variable being measured is a ratio variable.

The statistical analysis of nominal variables is limited. Techniques such as frequency distributions, histograms, stem–and–leaf plots may be constructed. Contingency table analysis may be useful for examining any association between nominal variables. A useful measure of center for this data is the *mode*, since this tells us which value (or category) appears most frequently. However, the mean and median are not as useful since the numbers being used are arbitrary.

All the statistical techniques suitable for nominal data can be used on ordinal data plus ranking and non-parametric procedures. A more extensive contingency table analysis is possible for sets of ordinal variables. Since the median takes order into account it is a useful measure of centre. But the mean is not a sensible number to calculate since the numbers assigned to the categories of an ordinal variable are arbitrary.

For interval and ratio variables all the techniques useful on nominal and ordinal variables may be used plus many others. For example, the calculations can include calculation of the mean, the variance, and some percentiles, if desired.

Even though we've discussed four types of measurements, we will not be all that concerned with these types, in general. Instead, our focus will be on some broader categories: are the variables taken on meaningful numerical measures, and whether our measurements are restricted in the values they can assume.

Qualitative and Quantitative Variables

Quantitative variables are those for which their numerical value is meaningful, and for which operations such as differences or averages makes sense.

EX: Temperature, weekly salary, pulse rate.

From our above descriptions, the variables measured on either the interval or ratio scales would be quantitative.

Qualitative variables are those for which their numerical value is not meaningful. They simply record into which of several categories a subject belong. For this reason, these variables are also referred to as *categorical variables*.

EX: Hair colour, car model, gender.

From our above descriptions, the variables measured on either the nominal or ordinal scales would be qualitative.

Discrete and Continuous Variables

Variables are *discrete* if they can assume either a finite or countable number of values. Discrete variables can be either qualitative or quantitative.

EX: The number of broken eggs in a carton would be a discrete variable since it could only equal $0, 1, 2, \ldots, 12$. This is also a quantitative measurement.

EX: If our variable is a description of our level of pain as moderate or severe, this variable is discrete, and it is a qualitative measurement.

EX: The number of girls born at the Janeway in 2000 before the first boy was born would be discrete, even though we don't know what the largest outcome would be, *i.e.* there could be 0 girls born before the first boy, or 1, or 2, or 3 or ... This variable can take on a countable number of values.

Variables are *continuous* if they can assume any value in some interval. Continuous variables are qualitative.

EX: The speed of a car, in km/h, would be continuous, since it could be any speed from 0 up to the maximum value on the speedometer.

EX: The length of a randomly selected rattlesnake is continuous.

Why do we need to spend some time on the various types of measurements? It is because some statistical analysis procedures will only be suitable for certain types of data, *i.e.* some procedures will only be appropriate if our data is continuous. One example of this is mentioned above, which is the inappropriateness of using the mean to describe the centre, or average, in categorical data.