What is the difference between categorical, ordinal and interval variables?

In talking about variables, sometimes you hear variables being described as categorical (or sometimes nominal), or ordinal, or interval. Below we will define these terms and explain why they are important.

Categorical

A categorical variable (sometimes called a nominal variable) is one that has two or more categories, but there is no intrinsic ordering to the categories. For example, gender is a categorical variable having two categories (male and female) and there is no intrinsic ordering to the categories. Hair color is also a categorical variable having a number of categories (blonde, brown, brunette, red, etc.) and again, there is no agreed way to order these from highest to lowest. A purely categorical variable is one that simply allows you to assign categories but you cannot clearly order the variables. If the variable has a clear ordering, then that variable would be an ordinal variable, as described below.

Ordinal

An ordinal variable is similar to a categorical variable. The difference between the two is that there is a clear ordering of the variables. For example, suppose you have a variable, economic status, with three categories (low, medium and high). In addition to being able to classify people into these three categories, you can order the categories as low, medium and high. Now consider a variable like educational experience (with values such as elementary school graduate, high school graduate, some college and college graduate). These also can be ordered as elementary school, high school, some college, and college graduate. Even though we can order these from lowest to highest, the spacing between the values may not be the same across the levels of the variables. Say we assign scores 1, 2, 3 and 4 to these four levels of educational experience and we compare the difference in education between categories one and two with the difference in educational experience between categories two and three, or the difference between categories three and four. The difference between categories one and two (elementary and high school) is probably much bigger than the difference between categories two and three (high school and some college). In this example, we can order the people in level of educational experience but the size of the difference between categories is inconsistent (because the spacing between categories one and two is bigger than categories two and three). If these categories were equally spaced, then the variable would be an interval variable.

Interval

An interval variable is similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced. For example, suppose you have a variable such as annual income that is measured in dollars, and we have three people who make $10,000, $15,000 and $20,000. The second person makes $5,000 more than the first person and $5,000 less than the third person, and the size of these intervals is the same. If there were two other people who make $90,000 and $95,000, the size of that interval between these two people is also the same ($5,000).

Why does it matter whether a variable is categorical, ordinal or interval?

Statistical computations and analyses assume that the variables have a specific levels of measurement. For example, it would not make sense to compute an average hair color. An average of a categorical variable

does not make much sense because there is no intrinsic ordering of the levels of the categories. Moreover, if you tried to compute the average of educational experience as defined in the ordinal section above, you would also obtain a nonsensical result. Because the spacing between the four levels of educational experience is very uneven, the meaning of this average would be very questionable. In short, an average requires a variable to be interval. Sometimes you have variables that are "in between" ordinal and interval, for example, a five-point likert scale with values "strongly agree", "agree", "neutral", "disagree" and "strongly disagree". If we cannot be sure that the intervals between each of these five values are the same, then we would not be able to say that this is an interval variable, but we would say that it is an ordinal variable. However, in order to be able to use statistics that assume the variable is interval, we will assume that the intervals are equally spaced.

Does it matter if my dependent variable is normally distributed?

When you are doing a t-test or ANOVA, the assumption is that the distribution of the sample means are normally distributed. One way to guarantee this is for the distribution of the individual observations from the sample to be normal. However, even if the distribution of the individual observations is not normal, the distribution of the sample means will be normally distributed if your sample size is about 30 or larger. This is due to the "central limit theorem" that shows that even when a population is non-normally distributed, the distribution of the "sample means" will be normally distributed when the sample size is 30 or more, for example see Central limit theorem demonstration .

If you are doing a regression analysis, then the assumption is that your residuals are normally distributed. One way to make it very likely to have normal residuals is to have a dependent variable that is normally distributed and predictors that are all normally distributed, however this is not necessary for your residuals to be normally distributed. You can see Regression with SAS: Chapter 2 - Regression Diagnostics, Regression with SAS: Chapter 2 - Regression Diagnostics, or Regression with SAS: Chapter 2 - Regression Diagnostics