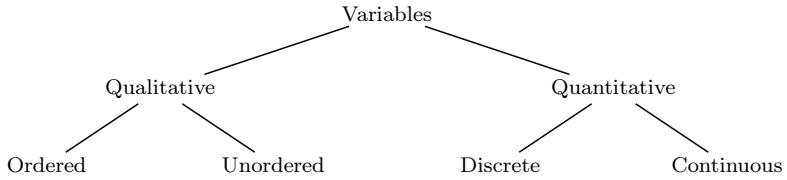
Variables

A (random) **variable** is a measurement that depends on the outcome of a (random) event.

- **Quantitative** variables have numeric values.
  - **Continuous** variables can assume a continuum of values: Examples include income, temperature, pressure, mass, and speed.
  - A **discrete** variable can assume only finitely (or countably) many values. Examples include: family size, and number of engine cylinders.
- **Qualitative** variables are non-numeric.
  - A qualitative variable can be **ordered** such as customer satisfaction (good, better or best)
  - or **unordered** such as eye color or automobile transmission type.

When working with a variable, it is important to understand what kind of variable it is. The kind of variable we are working with influences the statistical things that we can do with it. We can compute averages, standard deviations and other numeric statistics for quantitative variables. We can't do this for qualitative variables. We can make histograms of any kind of variable. If the variable is qualitative but unordered, then the histogram boxes can be placed in any order. Variable type also determines the kinds of computer data structures used to store its values.



1. Classify each of the following variables as qualitative or quantitative; if quantitative, as discrete or continuous; if qualitative, as ordered or unordered.

(a) occupation (b) region of residence (c) weight
(d) height (e) number of automobiles owned

2. The following variables are in the `adult.csv` file of the NHANESIII data set. Classify each as qualitative or quantitative; if quantitative, as discrete or continuous; if qualitative, as ordered or unordered.

(a) `DMARACER` (b) `HSAGEIR` (c) `MXPLANG` (d) `HFA8R`
(e) `HFA13` (f) `HFB1` (g) `HAB6S` (h) `SEQN`

3. The following variables are in the `exam.csv` file of the NHANESIII data set. Classify each as qualitative or quantitative; if quantitative, as discrete or continuous; if qualitative, as ordered or unordered.

(a) `BMPWT` (b) `BMPHT` (c) `PEP5R` (d) `PEPPREG`
(e) `DMPCREGN` (f) `DMPPIR` (g) `DEPCS2` (h) `SEQN`

4. The following variables are in the `lab.csv` file of the NHANESIII data set. Classify each as qualitative or quantitative; if quantitative, as discrete or continuous; if qualitative, as ordered or unordered.

(a) `PHPCHM2` (b) `PHPSNDA` (c) `WCP` (d) `RCP`
(e) `PBP` (f) `LCP` (g) `UIP` (h) `SEQN`

Solutions

1. (a) occupation is a qualitative variable. It's not a number. (b) region of residence is also qualitative. It's not a number either. (c) weight is quantitative. It's a continuous value (can be measured to many decimal places if you have the right equipment). (d) height is quantitative and continuous just like weight is. (e) number of automobiles owned is quantitative and discrete (assume you can't own fractions of a car!).

2.

(a) `DMARACER` is an unordered, qualitative variable that gives information about the subject's race.

(b) `HSAGEIR` is a discrete, quantitative variable that gives the age of the subject in years.

(c) `MXPLANG` is an unordered, qualitative variable that gives information about the language used in the interview.

(d) `HFA8R` is a discrete, quantitative variable that gives information about the number of years of school completed by the subject. Blank and Don't know are two possible responses that need to be omitted to compute statistics.

(e) `HFA13` is an unordered, qualitative variable that gives information about whether or not the subject ever served in the U.S. armed forces.

(f) `HFB1` is an unordered, qualitative variable that gives information about whether or not the subject is covered by Medicare.

(g) `HAB6S` is a discrete, quantitative variable that gives the number of months since the subject last saw or talked to a medical doctor or other health professional.

(h) `SEQN` is the identification number of the subject. Typically this would be viewed as an ordered, qualitative variable.

3. (a) `BMPWT` is weight in kg. If we eliminate the blank responses, this is a continuous, quantitative variable.

(b) `BMPHT` is standing height in cm. It is a continuous, quantitative variable.

(c) `PEP5R` is central pulse rate in beats per minute. It is a discrete, quantitative variable.

(d) `PEPPREG` is an unordered, qualitative variable indicating whether or not the subject is pregnant. Blank and Don't know are also possible responses.

(e) `DMPCREGN` is the census region (NW, MW, S, W) of the subject. It is an unordered, qualitative variable.

(f) `DMPPIR` is the subject's poverty income ratio. It is a continuous, quantitative variable.

(g) `DEPCS2` gives unordered, qualitative information about a the upper left central incisor tooth.

(h) `SEQN` is the identification number of the subject. Typically this would be viewed as an ordered, qualitative variable.

4.

(a) `PHPCHM2` is an unordered, qualitative variable indicating whether or not the subject has received chemotherapy treatment within the past four weeks.

(b) `PHPSNDA` is an ordered, qualitative variable indicating the day that the participant last ate.

(c) `WCP` is white blood cell count. It is a continuous, quantitative variable.

(d) `RCP` is red blood cell count. It is a continuous, quantitative variable.

(e) `PBP` is lead level in the blood sample measured in ug/dL. It is a continuous, quantitative variable.

(f) `LCP` is serum LDL cholesterol measured in mg/dL. It is a continuous, quantitative variable.

(g) `UIP` is urinary iodine measured in ug/dL. It is a continuous, quantitative variable.

(h) `SEQN` is the identification number of the subject. Typically this would be viewed as an ordered, qualitative variable.