VARIABLES

Variation is evident in almost every characteristic of patients, including their blood pressure and other physiologic measurements, diseases, environments, diets, and other aspects of their lifestyle.
A measure of a single characteristic that can vary is called a variable.

I. SOURCES OF VARIATION IN MEDICINE

Although variation in clinical medicine may be caused by biologic differences or the presence or absence of disease, it also may result from differences in the techniques and conditions of measurement, errors in measurement, and random variation. Some variation distorts data systematically in one direction, such as measuring and weighing patients while wearing shoes. This form of distortion is called systematic error and can introduce bias. Bias in turn may obscure or distort the truth being sought in a given study.

Other variation is random, such as slight, inevitable inaccuracies in obtaining any measure (e.g., blood pressure).

Biologic differences include factors such as differences in genes, nutrition, environmental exposures, age, gender, and race.

Blood pressure tends to be higher among individuals with high salt intake, in older persons, and in persons of African descent.

Tall parents usually have tall children.

Different conditions of measurement often account for the variations observed in medical data and include factors such as time of day, ambient temperature or noise, and the presence of fatigue or anxiety in the patient. Blood pressure is higher with anxiety or following exercise and lower after sleep. These differences in blood pressure are not errors of measurement, but of standardizing the conditions under which the data are obtained. Standardizing such conditions is important to avoid variation attributable to them and the introduction of bias.

Different techniques of measurement can produce different results.

A blood pressure (BP) measurement derived from the use of an intra-arterial catheter may differ from a measurement derived from the use of an arm cuff. This may result from differences in the measurement site (e.g., central or distal arterial site), thickness of the arm (which influences reading from BP cuff), rigidity of the artery (reflecting degree of atherosclerosis), and inter observer differences in the interpretation of BP sounds.

Some variation is caused by measurement error.

Two different BP cuffs of the same size may give different measurements in the same patient because of defective performance by one of the cuffs.

Different laboratory instruments or methods may produce different readings from the same sample.

Different x-ray machines may produce films of varying quality. When two clinicians examine the same patient or the same specimen (e.g., x-ray film), they may report different results

1. One radiologist may read a mammogram as abnormal and recommend further tests, such as a biopsy, whereas another radiologist may read the same mammogram as normal and not recommend further workup.

2 One clinician may detect a problem such as a retinal hemorrhage or a heart murmur, and another may fail to detect it.

Two clinicians may detect a heart murmur in the same patient but disagree on its characteristics. If two clinicians are asked to characterize a dark skin lesion, one may call it a “nevus,” whereas the other may say it is “suspicious for malignant melanoma.” A pathologic specimen would be used to resolve the difference, but that, too, is subject to interpretation, and two pathologists might differ.

II. STATISTICS AND VARIABLES

Statistical methods help clinicians and investigators understand and explain the variation in medical data.
The first step in understanding variation is to describe it.
How to describe variation in medical observations. Statistics can be viewed as a set of tools for working with data, just as brushes are tools used by an artist for painting. One reason for the choice of a specific tool over another is the type of material with which the tool would be used. One type of brush is needed for oil paints, another for tempera paints, and another type for watercolors. The artist must know the materials to be used to choose the correct tools. Similarly, a person who works with data must understand the different types of variables that exist in medicine.

A. Quantitative and Qualitative Data

The first question to answer before analyzing data is whether the data describe a quantitative or a qualitative characteristic.
A quantitative characteristic, such as a systolic blood pressure or serum sodium level, is characterized using a rigid, continuous measurement scale.
A qualitative characteristic, such as coloration of the skin, is described by its features, generally in words rather than numbers.
Normal skin can vary in color from pinkish white through tan to dark brown or black.
Medical problems can cause changes in skin color, with white denoting pallor, as in anemia;
red suggesting inflammation, as in a rash or a sunburn;
blue denoting cyanosis, as in cardiac or lung failure;
bluish purple occurring when blood has been released subcutaneously, as in a bruise; and
yellow suggesting the presence of jaundice, as in common bile duct obstruction or liver disease.

B. Types of Variables

Variables can be classified as nominal variables, dichotomous (binary) variables, ordinal (ranked) variables, continuous (dimensional) variables, ratio variables, and risks and proportions.

1. Nominal Variables

Nominal variables are “naming” or categorical variables that have no measurement scales and no rank order.
Examples are blood groups (O, A, B, and AB),
occupations,
food groups, and skin color.
If skin color is the variable being examined, a different number can be assigned to each color (e.g., 1 is bluish purple, 2 is black, 3 is white, 4 is blue, 5 is tan) before the information is entered into a computer data system. Any number could be assigned to any color, and that would make no difference to the statistical analysis. This is because the number is merely a numerical name for a color, and size of the number used has no inherent meaning; the number given to a particular color has nothing to do with the quality, value, or importance of the color.

2. Dichotomous (Binary) Variables

If all skin colors were included in one nominal variable, there is a problem: the variable does not distinguish between normal and abnormal skin color, which is usually the most important aspect of skin color for clinical and research purposes.

As discussed, abnormal skin color (e.g., pallor, jaundice, and cyanosis) may be a sign of numerous health problems (e.g., anemia, liver disease, cardiac failure).

Researchers might choose to create a variable with only two levels: normal skin color (coded as a 1) and abnormal skin color (coded as a 2). This new variable, which has only two levels, is said to be dichotomous (Greek, "cut into two").

Many dichotomous variables, such as well/sick, living/ dead, and normal/abnormal, have an implied direction that is favorable.

Other dichotomous variables, such as female/male and treatment/placebo, have no a priori qualitative direction. Dichotomous variables, although common and important, often are inadequate by themselves to describe something fully.

When analyzing cancer therapy, it is important to know not only whether the patient survives or dies (a dichotomous variable), but also how long the patient survives (time forms a continuous variable). A survival analysis or life table analysis, as described may be done. It is important to know the quality of patients' lives while they are receiving the therapy; this might be measured with an ordinal variable, discussed next. Similarly, for a study of heart murmurs, various types of data may be needed, such as dichotomous data concerning a murmur's timing (e.g., systolic or diastolic), nominal data on its location (e.g., aortic valve area) and character (e.g., rough), and ordinal data on its loudness (e.g., grade III).

Dichotomous variables and nominal variables sometimes are called discrete variables because the different categories are completely separate from each other.

3. Ordinal (Ranked) Variables

Many types of medical data can be characterized in terms of three or more qualitative values that have a clearly implied direction from better to worse.

An example might be "satisfaction with care" that could take on the values of "very satisfied," "fairly satisfied," or "not satisfied."

These data are not measured on a measurement scale. They form an ordinal (i.e., ordered or ranked) variable.

There are many clinical examples of ordinal variables. The amount of swelling in a patient's legs is estimated by the clinician and is usually reported as "none" or 1+, 2+, 3+, or 4+ pitting edema (puffiness).

A patient may have a systolic murmur ranging from 1+ to 6+.

Respiratory distress is reported as being absent, mild, moderate, or severe.

Although pain also may be reported as being absent, mild, moderate, or severe, in most cases, patients are asked to describe their pain on a scale from 0 to 10, with 0 being no pain and 10 the worst imaginable pain.

Ordinal variables are not measured on an exact measurement scale, but more information is contained in them than in nominal variables. It is possible to see the relationship between two ordinal categories and know whether one category is more desirable than another. Because they contain more information than nominal variables, ordinal variables enable more informative conclusions to be drawn.

4. Continuous (Dimensional) Variables

Many types of medically important data are measured on continuous (dimensional) measurement scales.

Patients' heights, weights, systolic and diastolic blood pressures, and serum glucose levels all are examples of data measured on continuous scales.

Continuous data often enable investigators to make more detailed inferences than do ordinal or nominal data. Relationships between continuous variables are not always linear (in a straight line). The relationship between the birth weight and the probability of survival of newborns is not linear.

Infants weighing less than 3000 g and infants weighing more than 4500 g are historically at greater risk for neonatal death than are infants weighing 3000 to 4500 g (~6.6-9.9 lb).

5. Ratio Variables

If a continuous scale has a true 0 point, the variables derived from it can be called ratio variables.

The Kelvin temperature scale is a ratio scale because 0 degrees on this scale is absolute 0. The centigrade temperature scale is a continuous scale, but not a ratio scale because 0 degrees on this scale does not mean the absence of heat.

For most statistical analyses, however, including significance testing, the distinction between continuous and ratio variables is not important.

C. Counts and Units of Observation

The unit of observation is the person or thing from which the data originated.

Common examples of units of observation in medical research are persons, animals, and cells.

Units of observation may be arranged in a frequency table, with one characteristic on the x-axis, another characteristic on the y-axis, and the appropriate counts in the cells of the table. Table which provides an example of this type of 2 × 2 table, shows that among 71 young professional persons studied, 63% of women and 57% of men previously had their cholesterol levels checked. Using these data and the chi-square test, one can determine whether or not the difference in the percentage of women and men with cholesterol checks was likely a result of chance variation (in this case the answer is "yes").

D. Combining Data

A continuous variable may be converted to an ordinal variable by grouping units with similar values together.

For example, the individual birth weights of infants (a continuous variable) can be converted to ranges of birth weights (an ordinal variable).

When the data are presented as categories or ranges (e.g., <500, 500-999, 1000-1499 g), information is lost because the individual weights of infants are no longer apparent.

An infant weighing 501 g is in the same category as an infant weighing 999 g, but the infant weighing 999 g is in a different category from an infant weighing 1000 g, just 1 g more.

The advantage is that now percentages can be created, and the relationship of birth weight to mortality is easier to show.

Three or more groups must be formed when converting a continuous variable to an ordinal variable. In the example of birth weight, the result of forming several groups is that it creates an ordinal variable that progresses from the heaviest to the lightest birth weight (or vice versa).

If a continuous variable such as birth weight is divided into only two groups, however, a dichotomous variable is created. Infant birth weight often is divided into two groups, creating a dichotomous variable of infants weighing less than 2500 g (low birth weight) and infants weighing 2500 g or more (normal birth weight). The fewer the number of groups formed from a continuous variable, however, the greater is the amount of information that is lost.