Research Minute

Statistics 101— One-Variable Statistics

Issue 15   **Sandra Burge, Ph.D.**

$$N \quad x^2$$
$$\overline{x} \quad \# \quad$$
$$p<.05 \quad r$$
$$\infty \quad \mu \quad t$$
$$\partial \quad \Sigma$$
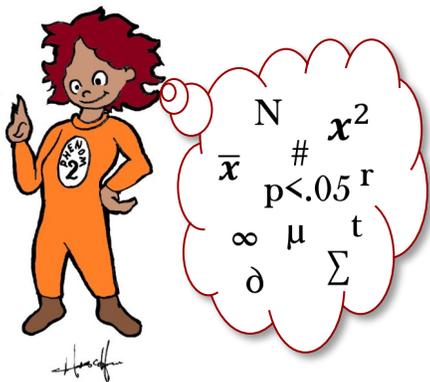
My experience with mathematical concepts is that they don't stay in my head unless I use them regularly. For example, I was a math major as an undergraduate, but I don't remember doodly-squat about calculus. However, I am adept at statistics because I use them often.

Let's start with the assumption that you and statistics have never been introduced. Or, if you have, it was a long time ago and you forgot her name. And face. And color of her hair. This *Research Minute* starts at the very beginning with one-variable statistics.

Your data can be characterized by three different types of description: (1) frequencies or counts, (2) measures of central tendency, and (3) measures of variation.

Frequencies

Researchers love to count things! Like...

GENDER	N	Percent
Men	30	15%
Women	170	85%
INSURANCE		
Private	60	30%
Government	100	50%
None	40	20%
BMI		
Normal	120	60%
Overweight	30	15%
Obese	50	25%

However, some things are cumbersome to count — primarily information that is measured with continuous numbers, like how many people have an A1c of 6.0? 6.1? 6.2? 6.3? 6.4? 6.5? 6.6? etc. Here, measures of central tendency and variation become more practical.

Central Tendency

Measures of Central Tendency refer to means, medians and modes.

Means are averages. You have 217 patients with diabetes (Figures A, B) and you want to know their average A1c. You add up individuals' A1cs and divide by 217.

Medians are the middle value. Line all your patients up by A1c value and pick the middle one. In both Figures the median is 8. Medians are useful when you have some outliers, like people with super-high labs—they'll throw your averages off. Economists use terms like "median household income," because the super-wealthy in a population will skew the group mean upward in a way that is not representative of the population.

Modes are the most common value. When we say "The FHC patient population is primarily Hispanic," we are describing the modal ethnic background of our patients.

Variation

While middle values give you information about a sample, something is missing. How high do the numbers go? Low? Widespread?

Variation is measured with range, variance, and standard deviation.

Range. Note the spread of A1c values in Figures A and B. In Figure A, A1c ranges from 4.3 to 13.0. In Figure B, the range is wider, 4.3 up to 20.0.

Standard Deviation measures the extent to which data points differ from the mean. It is *kind of* an average distance away from the mean, using the mathemagic of squares and square roots. To calculate:

1. find the group mean, then measure the distance of each person's A1c from the mean and square it;

2. add the squared distances together and divide by N, the number of subjects. This number is called the sample Variance.

3. Take the square root of the variance. This is the standard deviation.

Standard Deviations have "magical properties"! One standard deviation on either side of the mean will include 2/3 of your sample; two standard deviations on either side will include 95% of your sample.


Figure A. 217 Patients with Diabetes

Mean = 8.5
Median = 8.0
Std Dev = 2.1
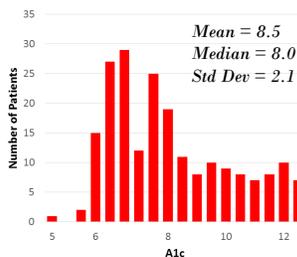
Figure B. 217 Patients with Diabetes

Mean = 8.6
Median = 8.0
Std Dev = 2.5