WHAT DO I DO WITH ALL THIS DATA—PART II

Dr. (Mrs) Marie Farrell and Dr. (Miss) Aparna Bhaduri

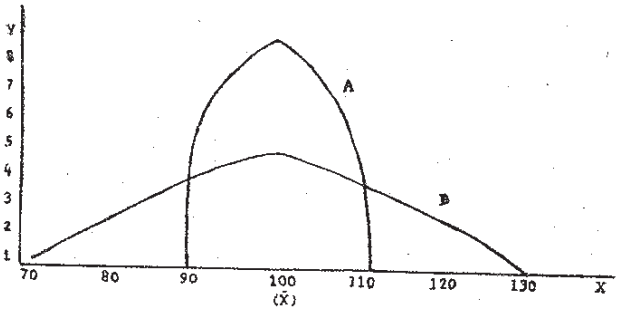
Dear Reader,

Measure of Variability

Earlier in our letter we introduced the range in order to discuss the steps in making a frequency distribution.

"Right—and from there we computed the mean."

Well, the range really belongs to a different kinds of measure group called the measures of variability. These measures reflect the spread-out-ness of the distribution, whereas the measures of central tendency reflect average-ness. Variability measures tell us more—they show us that some distributions **vary** more than those that seem "squashed together". The mean, median and mode are **points** on the baseline. Variability refers to **distances.** So that two separate groups may have the same mean, but be quite different because of the measure of variability as illustrated in Figure 1.



Symmetrical distributions with the *Same Mean*, having different degrees of dispersion.

Figure 1. Frequency distribution of knowledge of birth complications scores of trained birth attendants

Figure 1 shows two symmetrical distributions of two different groups of trained birth attendants. Both groups have the same mean score ($X = 100$) on a measure of knowledge of Birth Complications—the means are the same, but they show different pictures of dispersion. Group B is more spread out around the mean score of 100 very much more than Group A is. That is: Group A is less variable than Group B.

"Besides the range, what else comes under the measure of variability?"

Although other measures of variability are the average deviation and semi-interquartile range or quartile derivation, this discussion will be limited to the **standard deviation and variance.**

"I've heard people speak of the standard deviation—what does that term refer to?"

The standard deviation is based on the idea of individual deviations from the mean. The standard deviation takes a mean of the several deviation scores and then takes the square root of that mean. So if you have a mean (X) and you want to know how the scores are spread out on a baseline, you compute a deviation score. Thus the standard deviation is a kind of average of **individual deviations** from the mean of the distribution.[1]

The standard deviation is computed as follows:

X	X	X
5	+2	4
4	+1	1
3	0	0
2	−1	1
1	−2	4
$Ex=15$		$Ex^2=10$
$N=5$		

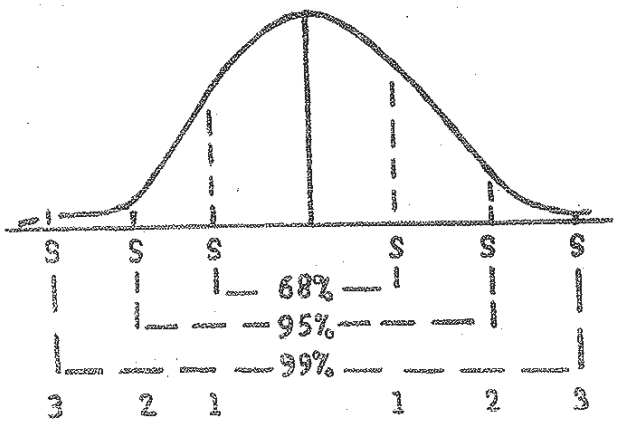
$$\bar{x}=\frac{Ex}{N} \qquad S.d.=\sqrt{\frac{Ex^2}{N}}$$

$$\bar{x}=\frac{15}{5} \qquad S.d.=\sqrt{\frac{10}{2}} \quad S.d.=1.4$$

$$\bar{x}=3 \qquad S.d.=\sqrt{2}$$

The standard deviation is one of the most commonly used statistic and is needed in computing other statistics. With standard deviation one can compute correlations (product moment) and some of the standard errors (these can be found in texts with more statistical emphases). Standard deviation relates to the normal distribution, as shown by the **bell shaped curve.** The

total curve has six standard deviations with one standard deviation equal to about 1/6 of the range. If you divide a normal curve, about 68% of the cases fall in the middle of the curve with half of these cases, or 34% to the left and 34% to the right. If you take in two standard deviations above and below the curve, you will have accounted for 95= of the cases. Three standard deviations above and below comprises 97% of the cases. See Figure 2.



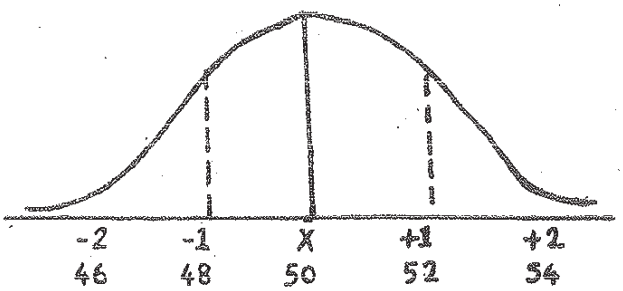
1. s.d. takes in 68% of cases
2. s.d. takes in 95% of cases
3. s.d. takes in 99% of cases

Figure 2. Normal curve and standard deviations

"Well, what does all this have to do with the trained birth attendants' score on the test of birth complications?"

Good question! Let us say that you scored the post-test of your sample, and obtained the raw scores for all subjects. You could create a frequency distribution and then compute the mean and standard deviation, which would tell you important information about your trained birth attendants. You could determine their average performance, and how the scores vary, by using the range and standard deviation. Then you could see exactly how one score compaires with all others in a given sample.

For example, look at Figure 3. If a person's score is 50, and the standard deviation is two, where will a score of 52 lie in a normally distributed sample?



$\bar{X}$=50
S.d.= 2

Figure 3. Normal distribution with a mean of 50 and standard deviation of two.

"A score of 52 will fall one standard deviation above the mean. Similarly, a score of 48 will fall one standard deviation below the mean."

That is correct. You remember, that the area enclosed within one standard deviation below and one standard above the mean represents 68% of the cases.

Measurements in health research can be of individual differences as well, but an individual's score can seldom be understood out of context. When we say that Ms X has achieved a 50 on a test for traditional birth attendants, we must ask, "in relation to what?" Her success cannot be understood except in relation to the successes of others who took the test.

Centiles or Percentiles

One very common measurement used in health research is the **centile**, or **percentile**. A centile shows where an individual stands in relation to a specified norm group. It does so by reporting first what percentage of that distribution falls below the selected individual's score.

"Then using percentiles would be useful for comparing our children to growth and development norms."

Yes. Height, weight and other anthropomorphic measures have been collected on a very large number of infants at various months of age. Measures of central tendency and measures of variability have been computed. This makes it possible to convert your infants score to other kinds of scores.

The norms that are published are usually revised to show the new norm tables in the form of percentiles. The raw scores are converted to a new form, which we will discuss here.

If you look at the curve below, you notice that it has a **baseline.** We can take this baseline and break it up into quarters of the distribution, or quartiles. If we divided the same baseline into tenths, each unit would be **decile.** See Figure 4.

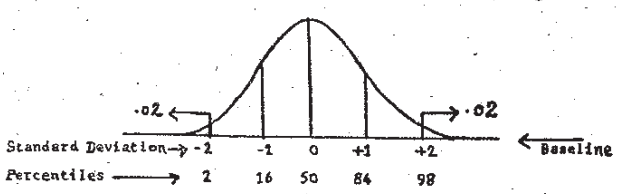


Figure 4. Normal distribution with baseline divided into standard deviation units showing proportion of total area subtended by each standard deviation and listing cumulative percentages (Percentiles)[2]

"And if we divide the baseline into one-hundredths, or 100 equal parts, we would have centiles."

That's exactly right. By definition, a centile is a "point on the baseline that lies above a specified number of hundredths of the distribution."[3] So when your infant measures a certain height and weight, that number is converted to a percentile. Now you have meaningful information. You can see how your infant compares with other infants in the group to which he is being compared.

"Alright. Let's say that the infant I just weighed scored at the 50th percentile. What would this mean, or how would I interpret that?"

You would say that his score is better than 50% of the infants of that sex and in that particular age group. It is very important to select appropriate norms for comparing **your** infants.

"Well, babies are babies. How can one group be so different from the next?"

There are very distinct differences in growth patterns of infants from different regions. In fact, even in one country, considerable variations may occur due to nutritional, environmental and other geographical factors. If there are no norms available, then you can begin to collect information to establish some norms for your area.

"How do I do that?"

Let's say you are the only health caretaker of a population of 10,000. By having the midwives and multipurpose workers keep simple charts on the infants, you, over time, can make your own reference group. Then you can compare an individual child to **that** group. The idea is to be able to compare **your** infant with some well-defined group. As you use the norms, you will be able to judge how well an individual child compares with other infants of the same sex, age in months and regional area.

In this letter, we have discussed the concept of variability and the measures of variability including the range and standard deviation, including calculation of both. Also discussed was the concept relating to centiles which were shown to be important in assessing one individual when compared to a reference group.

In our next letter, we will look at the other category of statistics and to do that we begin by looking at testing hypotheses—a most essential component in health research.

References

1. Phillips, John *Statistical Thanking* San Francisco: W.H. Freeman and Co., 1973, p. 25.
2. *Ibid.*, p. 59.
3. *Ibid.*, p. 62.