INTRODUCTION TO THE NORMAL DISTRIBUTION AND CURVE

A. Some basics

1. We will use the normal distribution and the normal curve to do a number of statistical tasks, especially to make increasingly complex statistical assertions.

2. This kind of argumentation depends upon the normal's ability to help us see the patterns in random variation.

3. also called Bell or Gaussian curve

Some basics (cont'd)

1. a family of **theoretical** distributions
   a. an infinite number of curves
   b. based on rough estimations from observations and mathematical reasoning – sprang from trying to find regularities in variation
   c. DeMoivre defined the normal mathematically
   d. x-axis – all possible values of the variable
   e. y-axis (rarely drawn) – probability of that value's occurrence

Some basics (cont'd)

a. Formulae

General formula for the normal

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Some basics (cont'd)

Formula for the unit or standard normal curve, where μ = 0 and σ = 1

$$y = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Where e is the base of the natural logarithms; for those of you curious about it, e is an irrational number approximately equal to 2.718281828 and is the sum of

$$1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots$$

B. Roscoe (1975) on the normal

1. (p. 80): the normal is a theoretical distribution, i.e., a "mathematical model for chance errors"

2. in this context, error means random variation – this "tension" about the meaning of the term "error" and its relationship with variation is a centuries-old characteristic of statistics as a body of thought

Roscoe (cont'd)

3. Roscoe summarizes:
   a. "small errors occur more than large errors"
   b."very large errors very seldom occur"
   c. "errors become progressively less numerous as they become larger"
   d. "negative errors (errors to the left)" are about as numerous as positive errors (errors to the right)"

C. Among the properties of the normal curve

1. Spatz (1997, p. 117) – generally speaking, a theoretical distribution "represents the 'best estimate' of how the events would actually occur."

2. Every normal curve is defined uniquely by its mean and standard deviation

3. Often used to describe the distribution of attributes that vary by chance in larger populations, e.g., IQ, height, life of batteries, number of chips in a bag, cereal boxes' weight

Properties of the normal curve (cont'd)

4. Bell-shaped and symmetrical – recall the definition of "symmetrical"

5. For the normal, mean = mode = median

6. Asymptotic to the x-axis, i.e., the curve comes infinitely close to the x-axis without touching it. Since the y-axis of the normal indicates probability, no value in the normal is presumed to have exactly 0 probability of occurring.

Properties of the normal curve (cont'd)

7. The curve is continuous – this is an important mathematical characteristic, but we will not explore it here.

8. See Roscoe (1975, pp. 83-84) for a very useful summary

9. Inflection points of curves indicate where the slope changes from positive to negative or vice versa. For the normal curve, the inflection points are $\pm 1\sigma$ from the mean, μ.