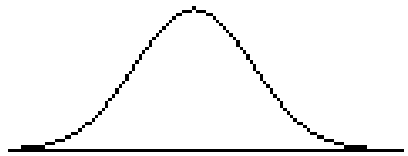
AnSc 5403
Biometry

Lecture Notes 9

I. Normal distribution

A. Many biologically related variables are normally distributed (follow a bell-shaped curve):



B. Properties of the normal distribution:

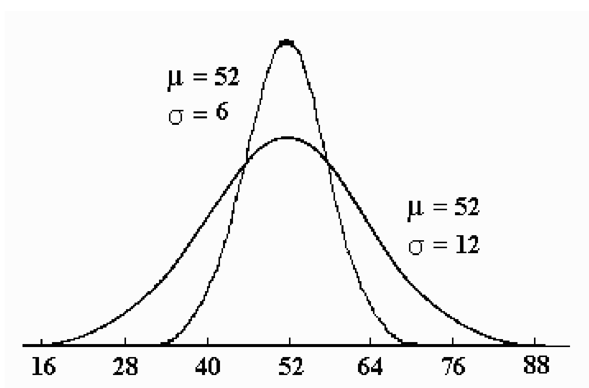
It is

It is

a. The curve has a single peak or mode at $x = \mu$

3. It is completely characterized by its mean and variance

a.



(a) Changing the variance (or standard deviation) has the effect of moving the inflection points closer or farther away from the mean

C. The probability density at any value of x for the normal distribution is given by the formula:

1. $1/\sqrt{2\pi\sigma^2} \cdot e^{-\frac{1}{2}([x-\mu]/\sigma)2}$

The importance of this probability distribution is to note that as x gets very far away from μ –

Example – values of x that are three standard deviations from the mean occur with a probability of –

b. This fact will be important later when we discuss hypothesis testing

II. The standard normal (Z) distribution

The standard normal distribution is defined as a normal distribution with –

1. It is useful for providing a simplification of the normal probability distribution
2. Because there are many possible normal distributions for a given variable, it would be convenient to transform the variables in such a way they might fit a standard distribution
3. The underlying distribution can be transformed to the standard normal distribution by converting it to a Z value:
   a. $Z = (x - \mu_x)/\sigma_x$
   b. Essentially Z converts our measurement into numbers of standard deviation units above and below the mean
      (a) Restated – Z is the number of standard deviations from μ at which the value of x is located
   c. If x is less than μ_x, values of Z will be negative, and if x is greater than μ_x, Z will be positive
   d. Thus, based on the symmetrical properties of the normal distribution, Z has $\mu = 0$ and $\sigma = 1$
4. A table of probabilities for the standard normal distribution (Microsoft Excel format) can be downloaded from www.afs.ttu.edu/ANSC5403/

B. Example of using the standard normal distribution

1. Assume that x is a normally distributed variable with $\mu = 70$ and $\sigma^2 = 25$
   a. This is typically written as x ~n(70, 25)
   b. What are the chances (what is the probability?) of selecting an individual in this population with a measurement of >75?
   c. First, convert to a Z value as follows: $Z = (75 - 70)/5 = 1$
   d. Look up a Z value of 1 in the Z table

      Value for Z = 1 =

      (b) This value is the area (probability) to the left of the given Z value, which would represent values < 75

So, the 1 – 0.8413 = =

III. The central limit theorem

A. The normal distribution is important for many reasons, but one significant reason is a mathematical result of the normal distribution that we have already discussed to some extent in Generalizations 1, 2, an 3

1. This result says – for random samples form an underlying distribution with a mean, μ, and variance, σ^2, regardless of how non-normal the distribution might be, the sampling distribution of $\bar{x}$, the sample mean of random samples of size n, is approximately normal

a. This sampling distribution of $\bar{x}$ will have a mean = μ of the underlying population and variance = σ^2/n, where σ^2 is the variance of the underlying population

This result is referred to as the –

2. The central limit theorem assumes only that:

a. The underlying population has a mean and a variance

The sampling from the population that results in a distribution of sample means and variances is –

When the underlying population is not normally distributed, the approximation of the sampling distribution to a normal distribution improves as –

a. For normally distributed populations, the sampling distribution does not simply approximate a normal distribution, it is a normal distribution

4. One note of caution is that the central limit theorem should not be used a reason to analyze data that are clearly non-normal using techniques designed for normal distributions

a. Many valuable "distribution-free" or non-parametric methods are available to analyze sets of data without the need for assumptions of normality

(a) It is probably advisable to use such methods when it is clear that the underlying distribution is not normal – e.g., binomial or clearly characterized by limited outcomes and non-uniform distribution