What is Normal, Anyway?

The **normal distribution** is suitable for many naturally occurring variables and one of the most important distributions in statistics!
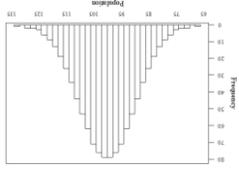

Basic Stat Vocabulary

Define the following terms, in your own words:

1. Population
   All members of a defined group.

2. Sample
   A subset of the population.

3. Random Sample
   A subset of the population where members must have:
   1) An equal opportunity to be selected and
   2) The same characteristics of the population.

4. Mean (μ)
   The arithmetic average; sum of the data values divided by the total number of data values.

5. Variance (σ^2)
   The arithmetic mean of the squared differences between each data value and the mean; Combines all of the values in a data set to produce a measure of spread.

6. Standard Deviation (σ)
   The square root of variance; the measure of spread of a distribution with the same units as the data.

Example

The lengths of adult Loggerhead sea turtle shells follow a normal distribution. It is known that 20% of these turtle shells have a length less than 85 cm and 10% have a length greater than 103 cm. Find the value of the mean μ and the standard deviation σ.

$$z = \frac{x - \mu}{\sigma}$$

$z = InverseNorm(0.2, 0, 1) = -0.84$

$z = InverseNorm(0.9, 0, 1) = 1.28$

Set up a system of equations to solve for μ and σ:
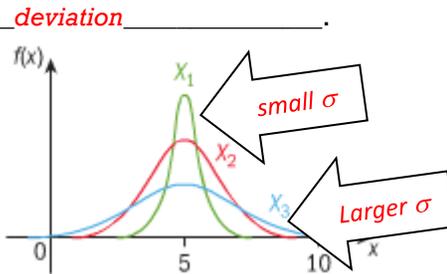$$\frac{103-\mu}{\sigma} = 1.28 \quad \text{and} \quad \frac{85-\mu}{\sigma} = -0.84$$

$\mu = 92.1\ cm$ and $\sigma = 8.5\ cm$

The Normal Curve

There is no one normal curve; it will ___vary___ depending on the spread of the data, or _standard_ ___deviation___.


Z-Scores

Using the mean and standard deviation of the data, we can standardize the normal distribution by finding the normal deviate, or ___z-score___, in each situation.

This is done through the formula:
$$z = \frac{x-\mu}{\sigma}.$$

Set the mean to ___0___ and each 1 unit of the x-axis is ___1___ standard deviation.
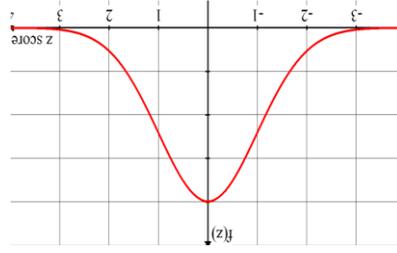
Properties of the Normal Curve

- The curve is _bell-shaped_.

- It is symmetrical about the __ _mean_.

- The mean, median, and mode are the ___ _same_.

- The area under the curve represents __100%___ of the data.

- The Empirical Rule states _68%_ of the data lies within one, _95%_ within two, and _99.7%_ within three standard deviations from the mean.

Example

The weights of Kemp's ridley (*Lepidochelys kempii*) sea turtles observed off the coast of Florida are normally distributed with a mean of 45 kg and a standard deviation of 4 kg.
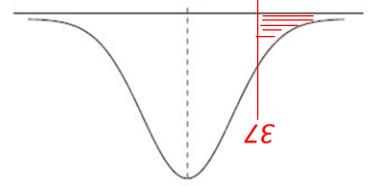
A turtle is chosen at random. Find the probability that it weighs less than 37 kg.

Using Empirical Rule: $P(x < 37) = 2.5\%$

Using Calculator: $P(x < 37) =$ $normalCDF(0, 37, 45, 4) = 2.3\%$

Shade the region of the curve that represents this probability.