Mathematics 231

Lecture 4

Liam O'Brien

Announcements

- Reading
  - Today | M&M 1.2 | 45-47
  - | M&M 1.3 | 53-62
  - Next class | M&M 1.3 | 62-71

Linear Transformations, Standardizing, and the Normal Distribution

- Linear transformations: Impact on shape, center, and spread
- Standardizing
- Introduction to the normal (or Gaussian) distribution

Example: Linear Transformation

- Temperature: Celsius to Fahrenheit

$$^{0}F = 32 + \frac{9}{5}(^{0}C)$$

- Currency: Euro to U.S. dollar

1 USD = 0.814 Euros

Linear Transformation

- Let's pretend that we have nothing better to do but to imagine we have a set of *n* observations, $x_1, x_2, \ldots, x_n$.

- What we want is a set of variables, y_i, related to x_i by,

$$y_i = a + b(x_i)$$

Examples: Linear Transformation

- Temperature: Celsius to Fahrenheit

$$^{0}F = 32 + \frac{9}{5}(^{0}C)$$

$$y = a + bx$$

$$a = 32;\ \ b = \frac{9}{5}$$

Examples: Linear Transformation

- Euros to USD

$$USD = 0.814 \text{ Euros}$$

$$y = a + bx$$

$$a = 0; \; b = 0.814$$

Linear Transformations

$$y_i = a + bx_i$$

- A **linear** transformation is one that changes the data by adding a constant, multiplying by a constant, or both.
- Can we tell what will happen to the mean and standard deviation of the data is they undergo a linear transformation? Yes.

Effect of Linear Transformations on Measures of Location

$$y = a + bx$$

$$\text{mean of } y = a + b(\text{mean of } x)$$

$$\text{median of } y = a + b(\text{median of x})$$

Effect of Linear Transformations on Measures of Spread

$$\text{SD of } y = |b|(\text{SD of } x)$$

$$\text{variance of } y = b^2(\text{variance of } x)$$

$$\text{IQR of } y = |b|(\text{IQR of } x)$$

Note that measures of spread are not affected by the addition of a constant!

Standardizing

- Question: How far is an observation from the mean?
- Example: From a previous statistics class final exam, the mean grade was 66, with $s = 12$.
- Let's say that Tom scores 78 and Lisa scores 84.
- Tom's score was 12 points above the mean, while Lisa's was 18 above the mean.
- How similar/different are these two scores?

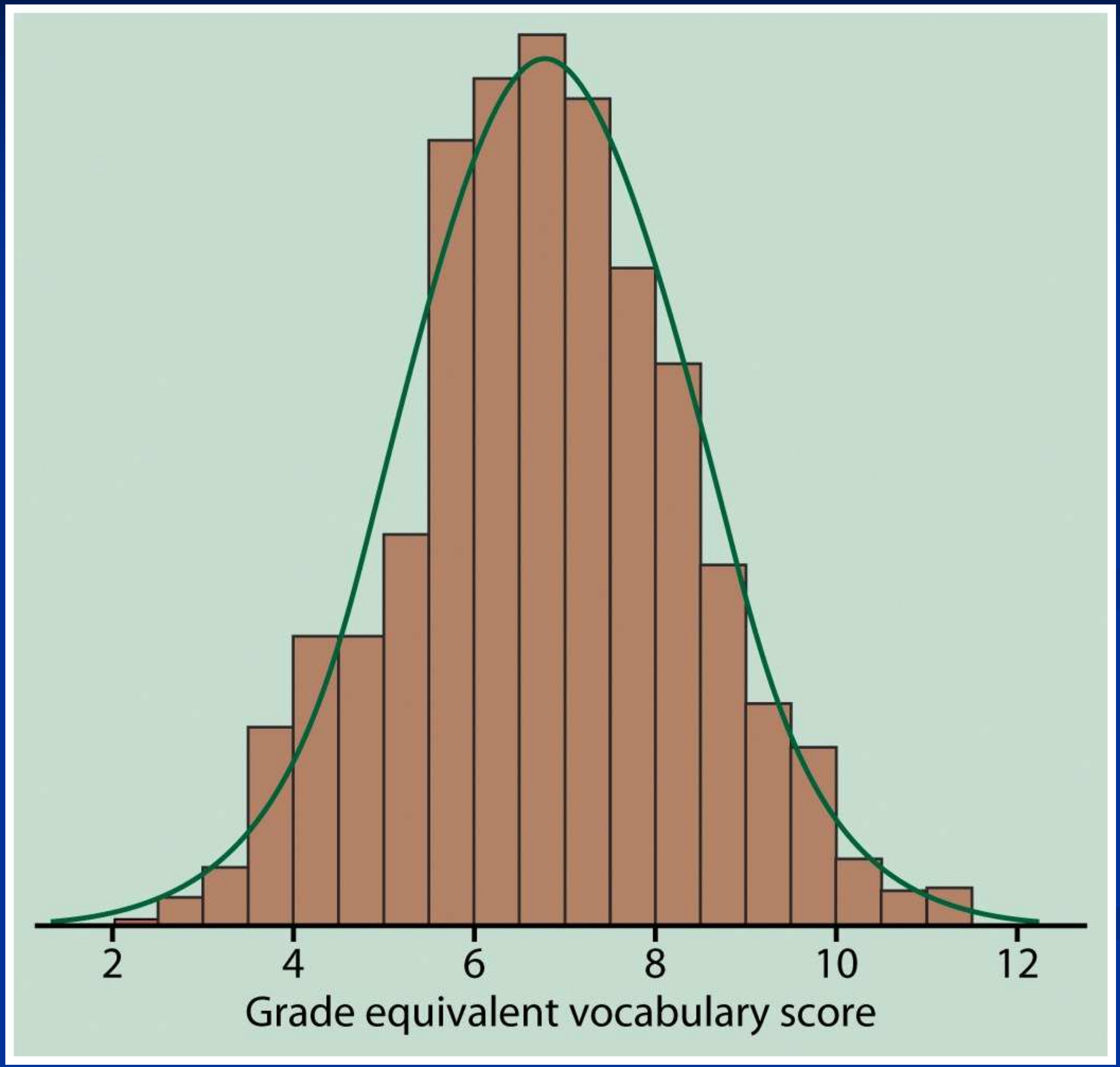
Standardizing

- We need to consider the spread of the data when answering this question.
- Consider how many SD's the scores are away from the mean.
- Since 1 SD = 12 points, Tom's score is 1 SD above the mean.
- Lisa's is 1.5 SD's above the mean (since 1.5 SD's = 1.5 * 12 points = 18 points).

Standardizing

- To compute how many standard deviations away from the mean a score is, we:
  1. Subtract the mean from the score.
  2. Divide the result in (1) by the standard deviation.
- This tells us how far a score is away from the mean score, in terms of the SD.
- This linear transformation is called **standardizing.**

Normal Distribution



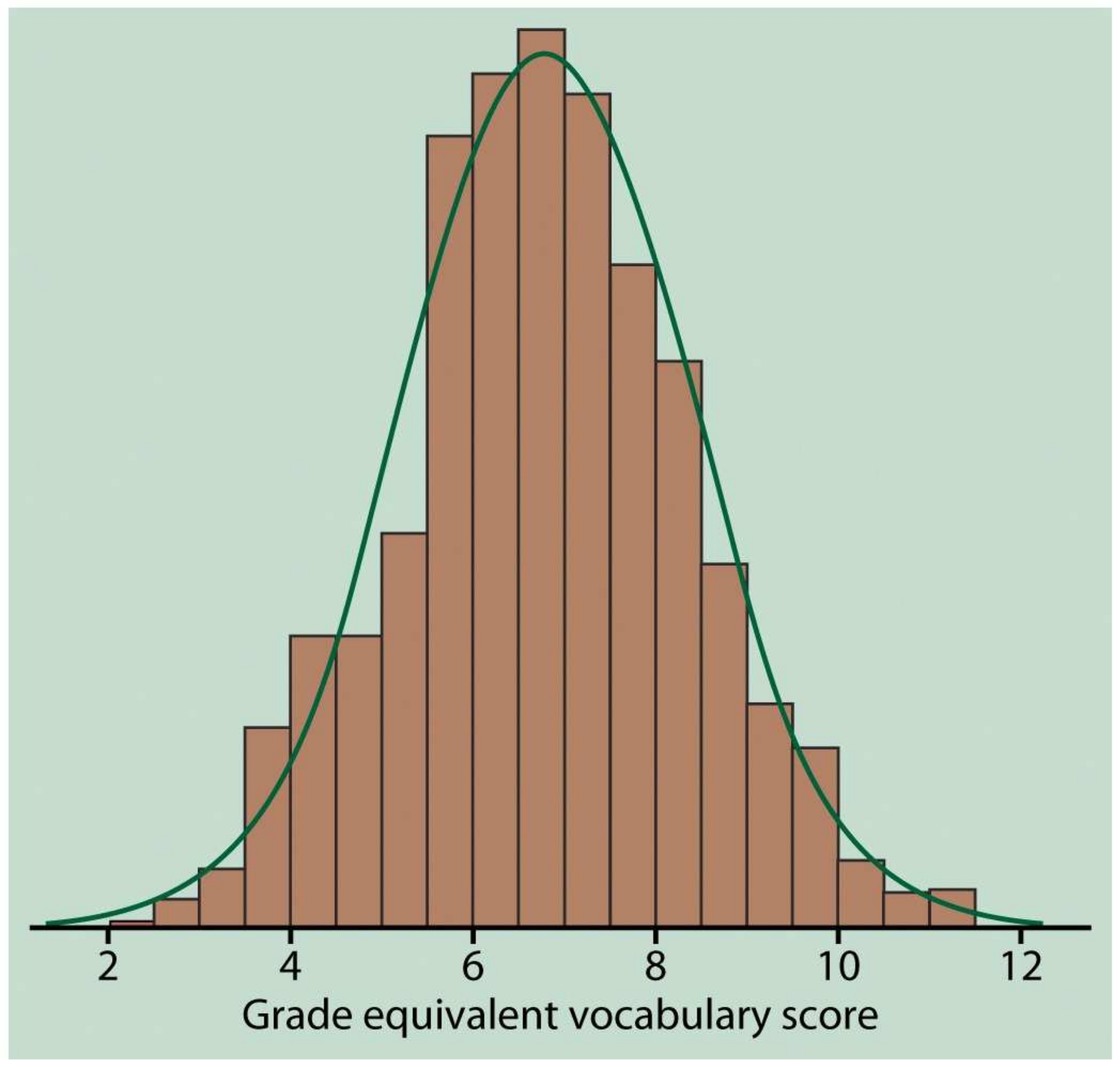
Normal Distribution

- Normal distribution is an idealized mathematical model for some distributions of data.
- Originally developed by A. DeMoivre in 1733, but named after C.F. Gauss.

Normal Distribution

- Properties:
  - All normal distributions have the same bell shape.
  - They differ in their center and spread, however.
  - Center: Mean (denoted by μ)
  - Spread: SD (denoted by σ)
- The different notation between what we used before for the mean and SD ($\bar{x}$ and s), comes from a subtlety that we'll cover soon. For now, you can consider the notation interchangeable.

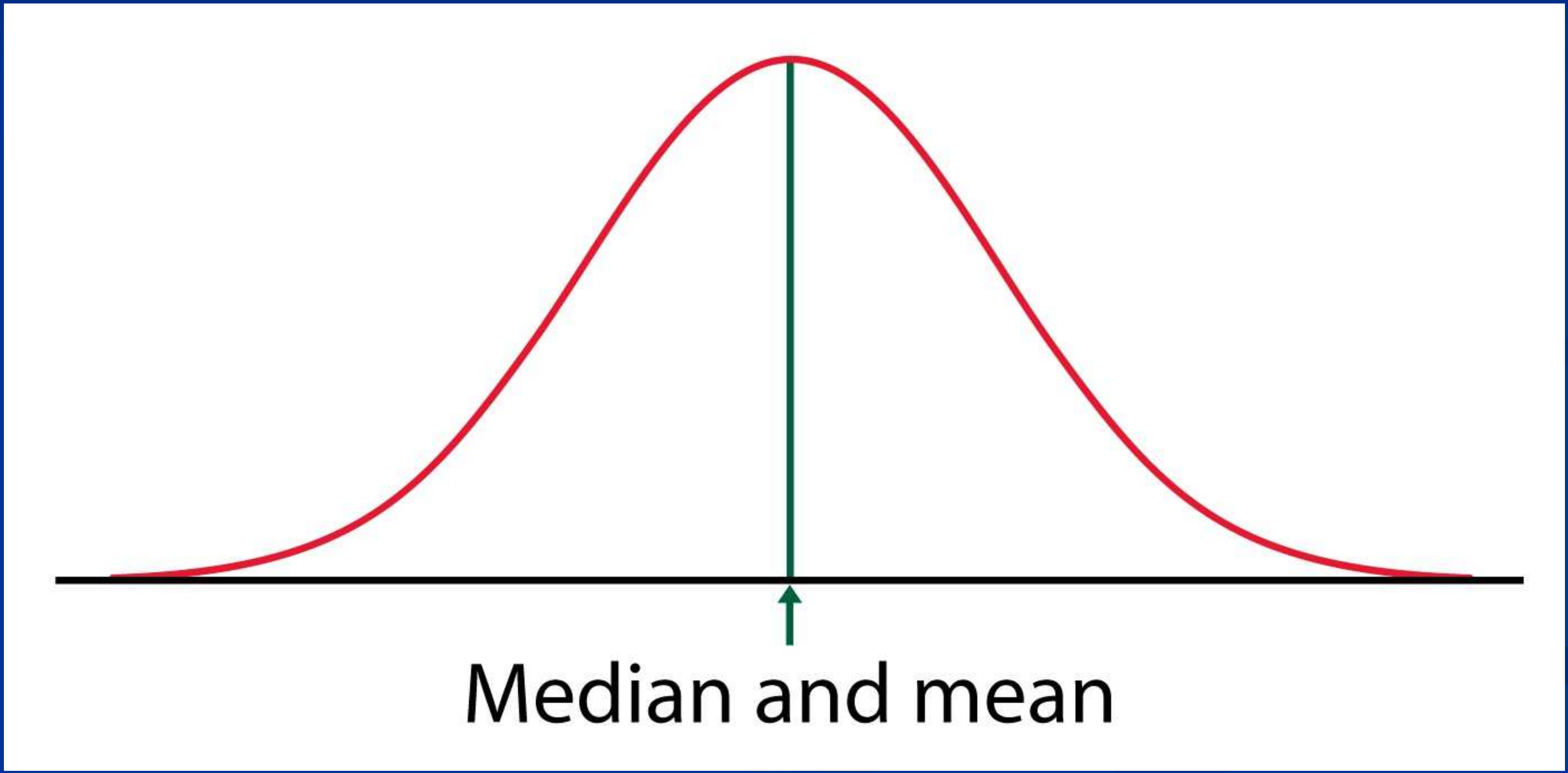
Example: Iowa Test Vocabulary Scores for 947 Gary, IN 7th Graders

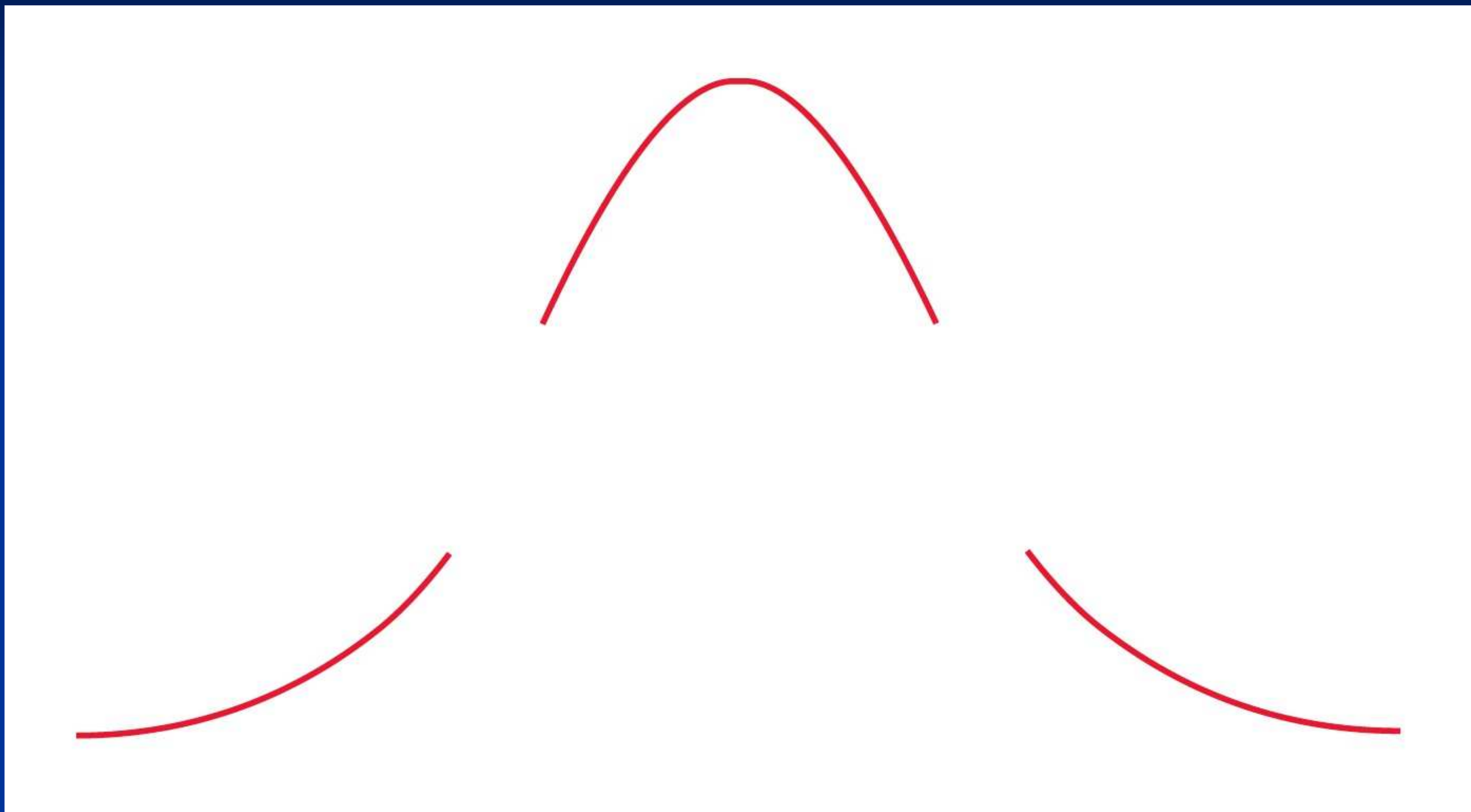


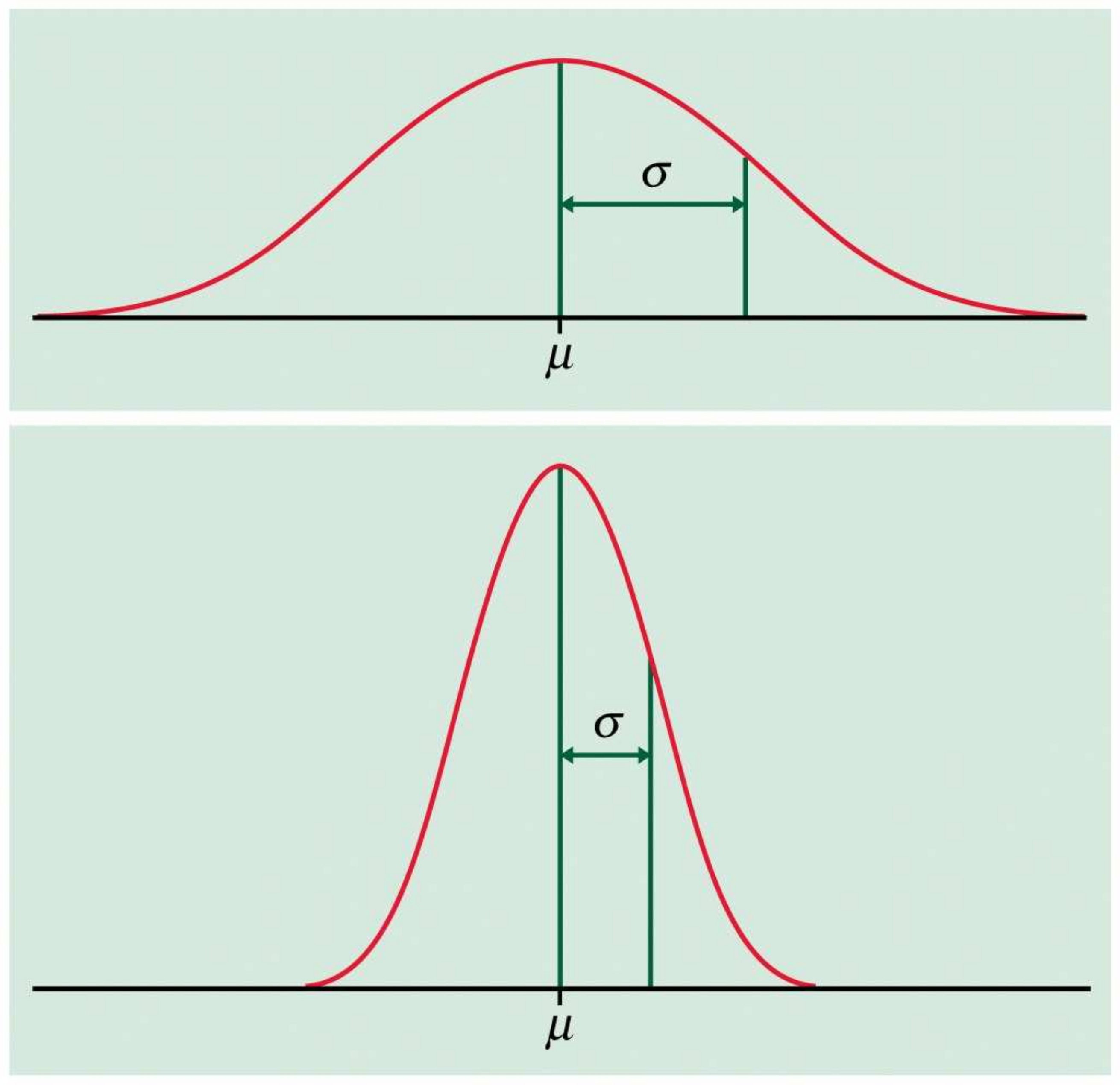
Additional Properties

- The curve is symmetric about its mean.
- The curve is unimodal (it has 1 peak only).
- Mean = Median = Mode
- The curve changes concavity at the points: $\mu \pm \sigma$
- Special Case: The **standard normal** distribution has mean = 0, and SD = 1.
- Total area under the curve is 1.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma} \right]$$






σ
μ
σ
μ

How "Normal" is Normal?

- Much effort has been poured into proving that all variables follow a normal distribution.

- "…the Law of Error upon which these Normal Values are based…finds a footing wherever the individual peculiarities are wholly due to the combined influence of a multitude of *accidents*…" Francis Galton, Natural Inheritance, 1889, pp. 54-55.

How "Normal" is Normal?

- Many variables do follow an approximately normal distribution. These include IQ scores, weight, height, blood pressure, cholesterol level.
- Many don't typically follow a normal distribution. Counts, income, expenditure are examples.
- So, if most people in the course are ECON majors, why are we talking about this?

Implications of Being Normal

- We will make the assumption that data are normal for the majority of things that we do.
- If it's not normal, we'll *make* it normal (well, sometimes).
- In fact, it can be shown that certain variables that are highly non-normal, can be analyzed using techniques that require a normality assumption, and that this is valid.
- We will return to this idea when we talk about sampling methods.