Parameters of continuous distributions

- **Location parameter (γ)**
  - x-axis location
  - usually the midpoint (mean for normal distribution) or lower endpoint
  - also called "shift"-parameter
  - changes in 'γ' shift the distribution left or right without changing it otherwise
- **Scale parameter (β)**
  - determines scale (unit) of measurement
  - standard deviation 'σ' for normal distribution
  - changes in 'β' compress or expand the associated distribution without altering its basic form
- **Shape parameter (α)**
  - determines basic form or shape of a distribution within the general family of distributions of interest
  - a change in 'α' generally alters a distribution's properties (skewness) more fundamentally than a change in location or scale

Expected value, formally

Discrete case:

$$E(X) = \sum_{\text{all } x} x_i p(x_i)$$

Continuous case:

$$E(X) = \int_{\text{all } x} x_i p(x_i) dx$$

Empirical Mean is a special case of Expected Value…

Sample mean, for a sample of n subjects: =

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)$$

The probability (frequency) of each person in the sample is 1/n.

Expected value isn’t everything though…

- Take the show “Deal or No Deal”
- Everyone know the rules?
- Let’s say you are down to two cases left. $1 and $400,000. The banker offers you $200,000.
- So, Deal or No Deal?

Deal or No Deal…

- This could really be represented as a probability distribution and a non-random variable:

x$	***p(x)***
+1	**.50**
+$400,000	**.50**

x$	***p(x)***
+$200,000	**1.0**

Expected value doesn't help…

x$	*p(x)*
+1	.50
+$400,000	.50

$$\mu = E(X) = \sum_{\text{all } x} x_i \, p(x_i) = +1(.50) + 400{,}000(.50) = 200{,}000$$

x$	*p(x)*
+$200,000	1.0

$$\mu = E(X) = 200{,}000$$

How to decide?

Variance!

- **If you take the deal, the variance/standard deviation is 0.**
- **If you don't take the deal, what is average deviation from the mean?**
- **What's your guess?**

Variance/standard deviation

"The average (expected) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

***We square because squaring has better properties than absolute value. Take square root to get back linear average distance from the mean (="standard deviation").*

Variance, formally

Discrete case:

$$Var(X) = \sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

Continuous case:

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 p(x_i) dx$$

Similarity to empirical variance

The variance of a sample: $s^2 =$

$$\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1} = \sum_{i=1}^{N}(x_i - \bar{x})^2 (\frac{1}{n-1})$$

Division by n-1 reflects the fact that we have lost a "degree of freedom" (piece of information) because we had to estimate the sample mean before we could estimate the sample variance.

Variance: Deal or No Deal

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i) =$$

$$= (1 - 200{,}000)^2 (.5) + (400{,}000 - 200{,}000)^2 (.5) = 200{,}000^2$$

$$\sigma = \sqrt{200{,}000^2} = 200{,}000$$

Now you examine your personal risk tolerance…

Handy calculation formula!

Handy calculation formula (if you ever need to calculate by hand!):

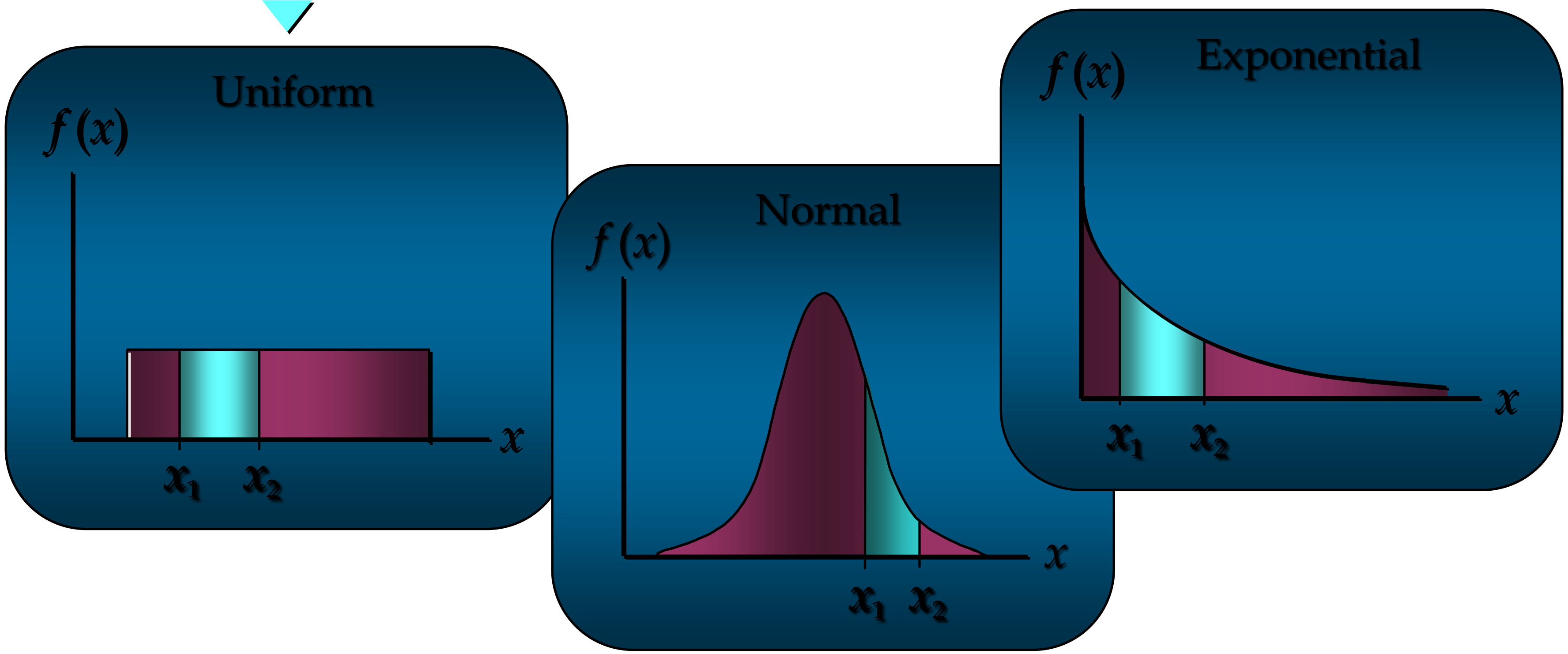
$$\boxed{Var(X) = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i) = \sum_{\text{all } x} x_i^{\,2} p(x_i) - (\mu)^2}$$

$$\boxed{= E(x^2) - [E(x)]^2}$$

Intervening algebra!

Continuous Probability Distributions

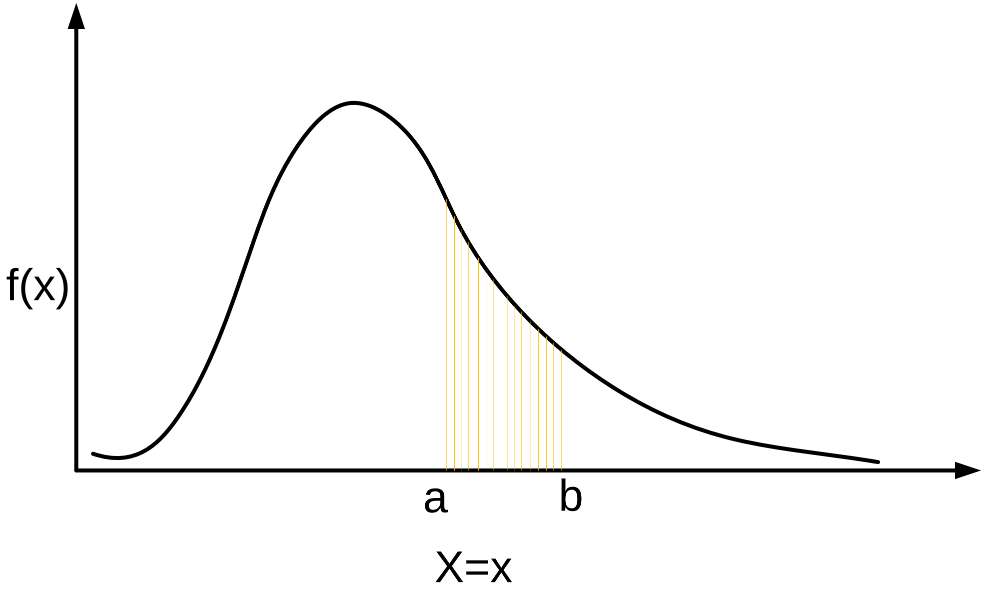
- The probability of the random variable assuming a value within some given interval from x_1 to x_2 is defined to be the area under the graph of the probability density function between x_1 and x_2.



Properties of Probability Density Function

The function $f(x)$ is a probability density function for the continuous random variable X, defined over the set of real numbers R, if

1. $f(x) \geq 0$, for all $x \in R$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$
4. $\mu = \int_{-\infty}^{\infty} xf(x)\,dx$
5. $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$



Uniform Probability Distribution

- A random variable is <u>uniformly distributed</u> whenever the probability is proportional to the interval's length.
- The <u>uniform probability density function</u> is:

$$f(x) = 1/(b - a) \quad \text{for } a \leq x \leq b$$
$$= 0 \quad \text{elsewhere}$$

where: a = smallest value the variable can assume

b = largest value the variable can assume

Uniform Probability Distribution

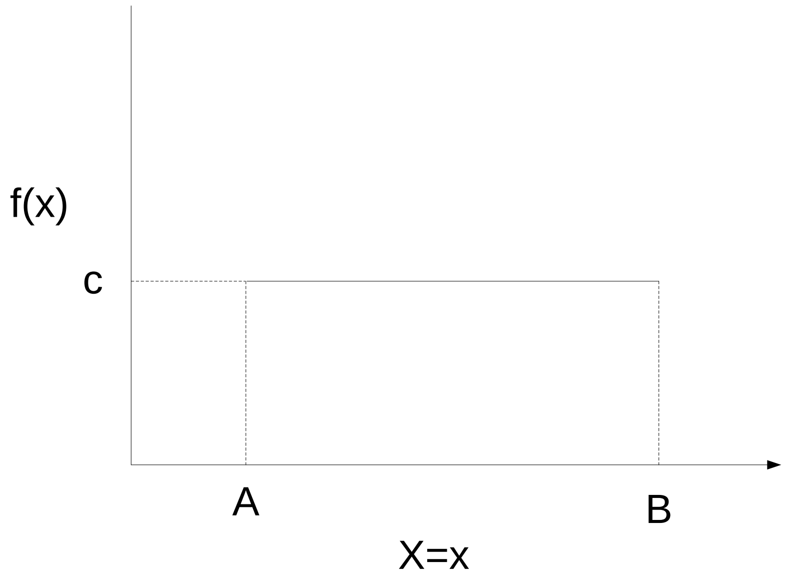
- Expected Value of x

$$E(x) = (a + b)/2$$

- Variance of x

$$\mathrm{Var}(x) = (b - a)^2/12$$

Continuous Uniform Distribution



Note:

a) $\int_{\infty}^{-\infty} f(x)dx = \frac{1}{B-A} \times (B - A) = 1$

b) $P(c < x < d) = \frac{d-c}{B-A}$ where both c and d are in the interval (A,B)

c) $\mu = \frac{A+B}{2}$

d) $\sigma^2 = \frac{(B-A)^2}{12}$

Uniform Probability Distribution

- Example: Slater's Buffet
  - Slater customers are charged for the amount of salad they take. Sampling suggests that the amount of salad taken is uniformly distributed between 5 ounces and 15 ounces.

Uniform Probability Distribution

- Uniform Probability Density Function

$$f(x) = 1/10 \quad \text{for } 5 \le x \le 15$$
$$= 0 \quad \text{elsewhere}$$

where:

x = salad plate filling weight

Uniform Probability Distribution

- Expected Value of x

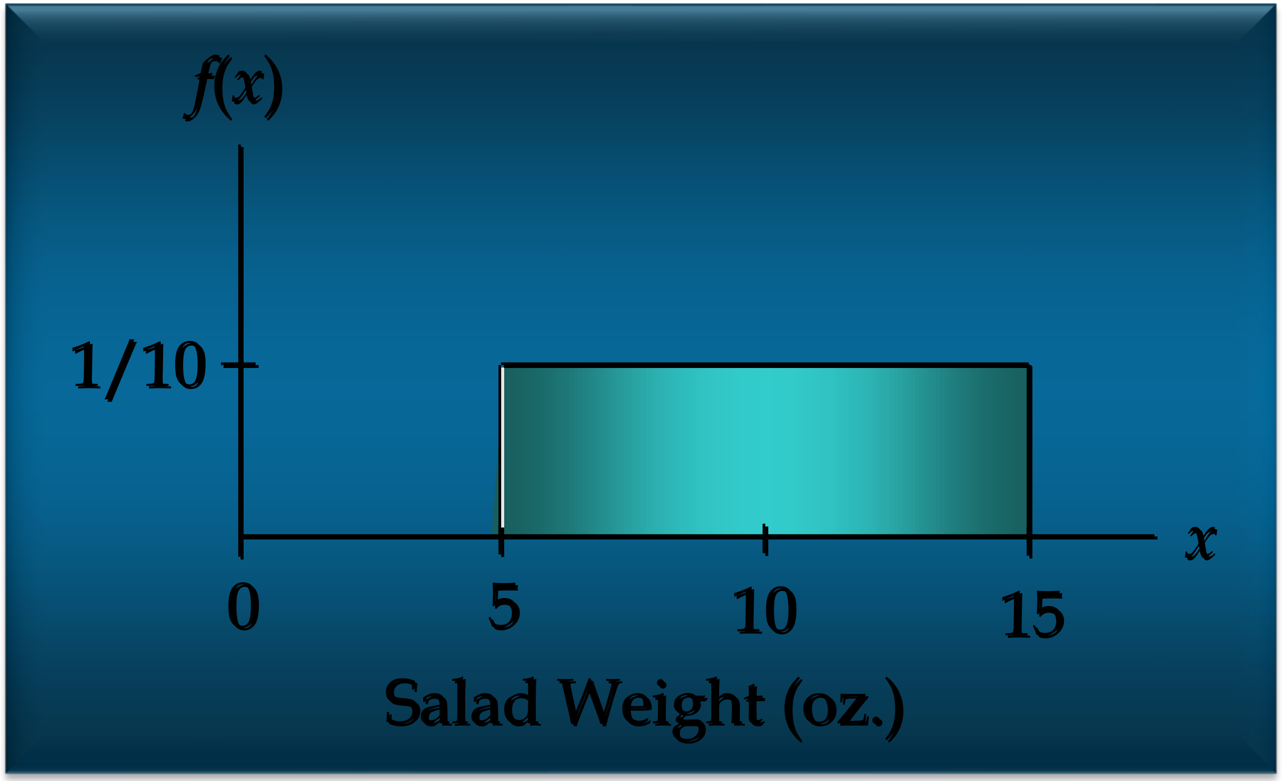
$$E(x) = (a + b)/2$$
$$= (5 + 15)/2$$
$$= 10$$

- Variance of x

$$Var(x) = (b - a)^2/12$$
$$= (15 - 5)^2/12$$
$$= 8.33$$

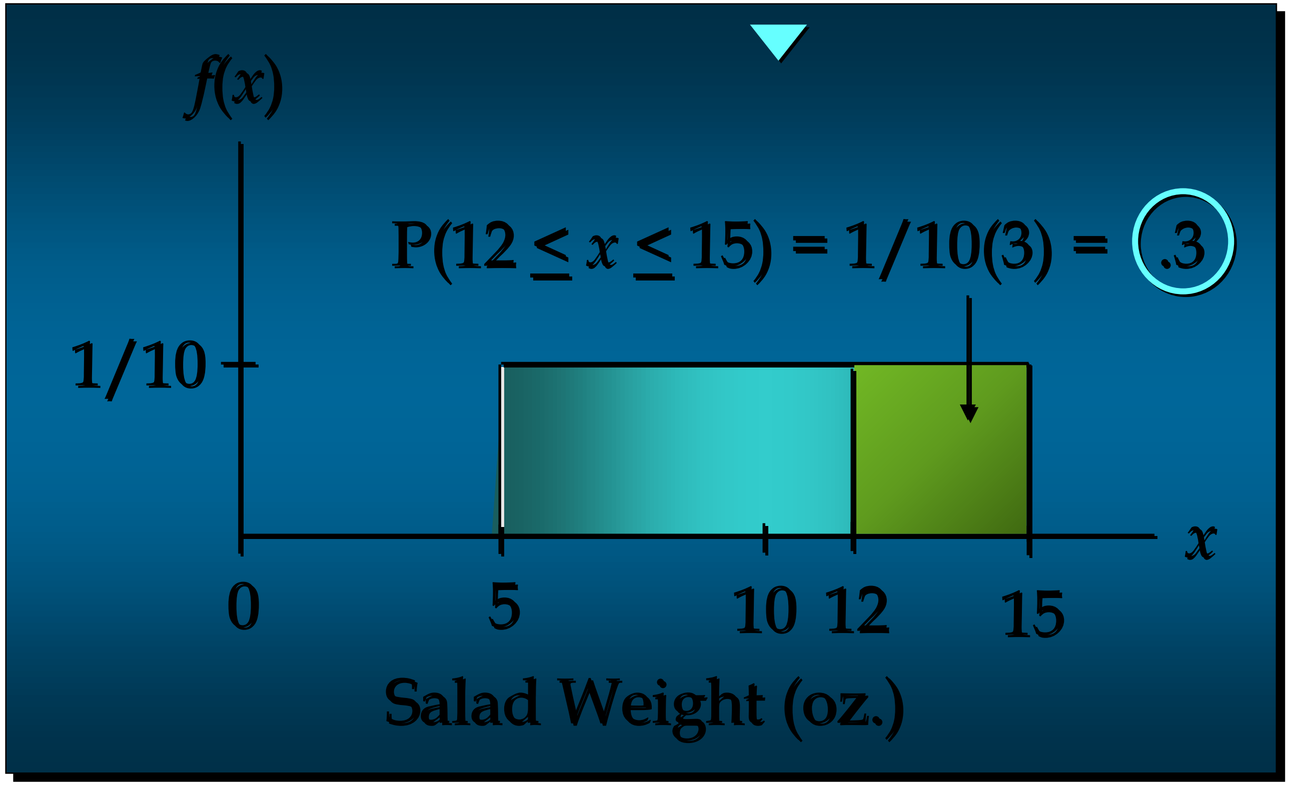
Uniform Probability Distribution

- Uniform Probability Distribution
  for Salad Plate Filling Weight



Uniform Probability Distribution

What is the probability that a customer will take between 12 and 15 ounces of salad?



Area as a Measure of Probability

- The area under the graph of $f(x)$ and probability are identical.
- This is valid for all continuous random variables.
- The probability that x takes on a value between some lower value x_1 and some higher value x_2 can be found by computing the area under the graph of $f(x)$ over the interval from x_1 to x_2.

Normal Distribution

- The most often used continuous probability distribution is the normal distribution; it is also known as **Gaussian** distribution.

- Its graph called the normal curve is the bell-shaped curve.

- Such a curve approximately describes many phenomenon occur in nature, industry and research.

    - Physical measurement in areas such as meteorological experiments, rainfall studies and measurement of manufacturing parts are often more than adequately explained with normal distribution.

- A continuous random variable X having the bell-shaped distribution is called a normal random variable.

Normal Probability Distribution

- The normal probability distribution is the most important distribution for describing a continuous random variable.
- It is widely used in statistical inference.
- It has been used in a wide variety of applications including:
  - Heights of people
  - Rainfall amounts
  - Test scores
  - Scientific measurements
- Abraham de Moivre, a French mathematician, published *The Doctrine of Chances* in 1733.
- He derived the normal distribution.

Normal Probability Distribution

- Normal Probability Density Function

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

where:

μ = mean

σ = standard deviation

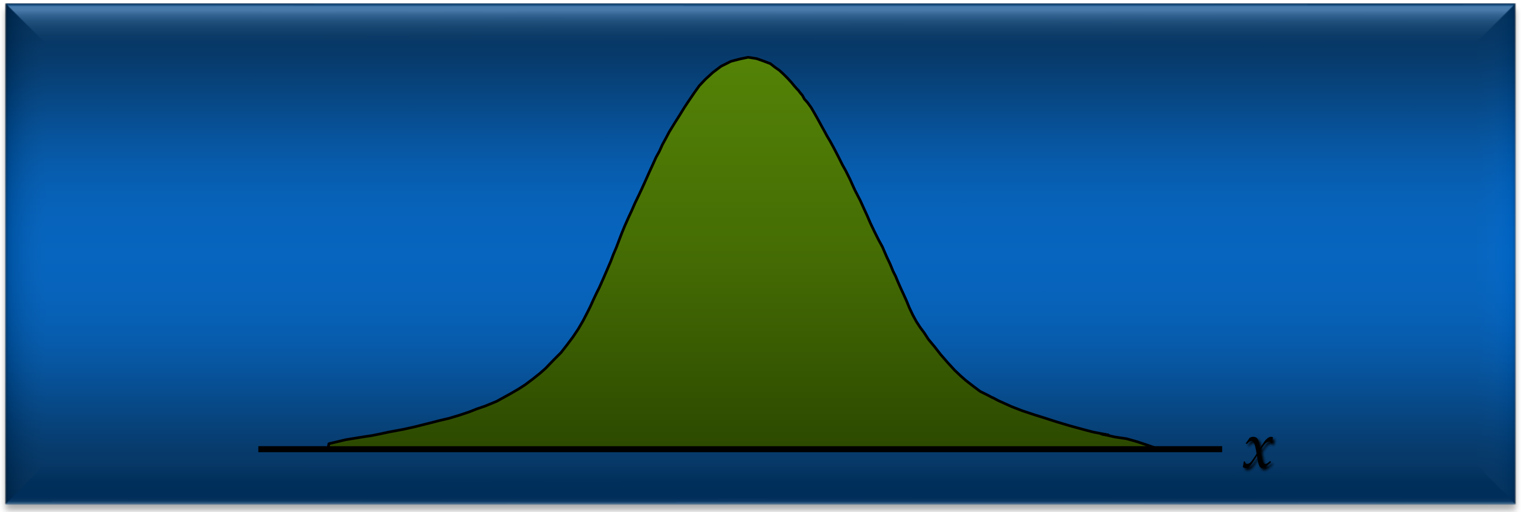
π = 3.14159

e = 2.71828

Normal Probability Distribution

- Characteristics

The distribution is symmetric; its skewness measure is zero.

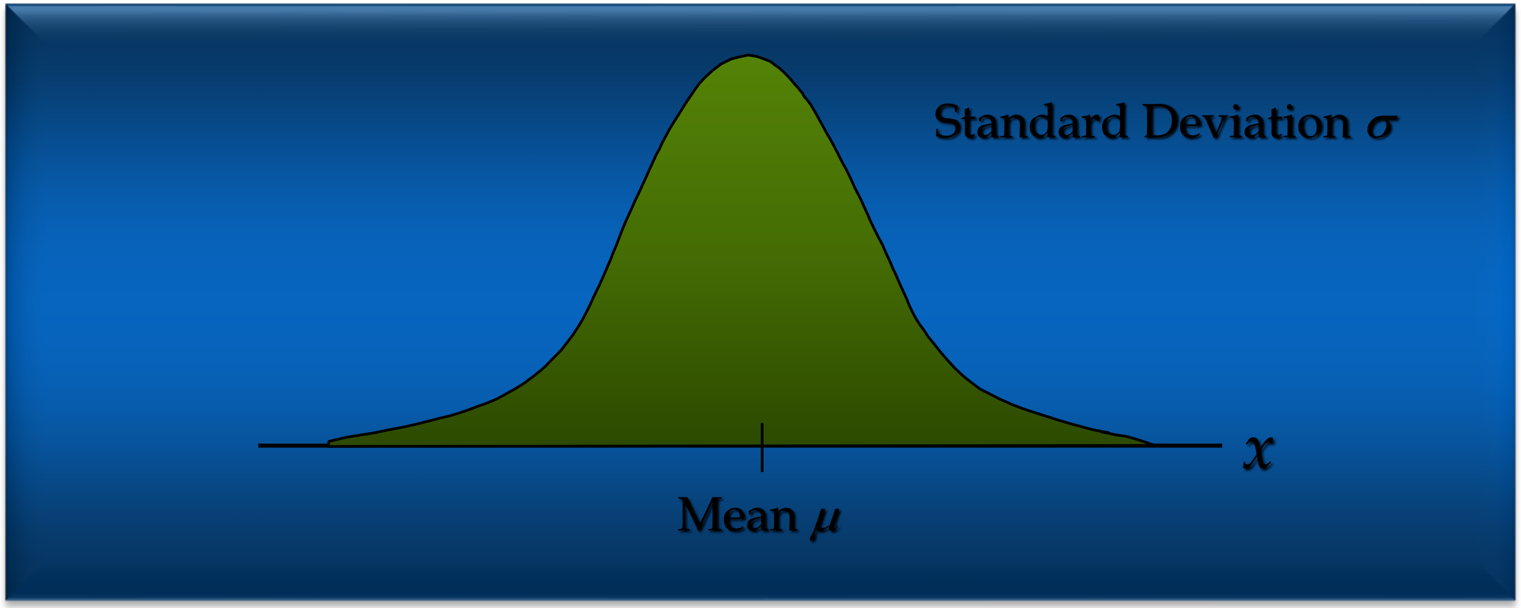


The random variable x can take any value from $-\infty$ to ∞.

Normal Probability Distribution

- Characteristics

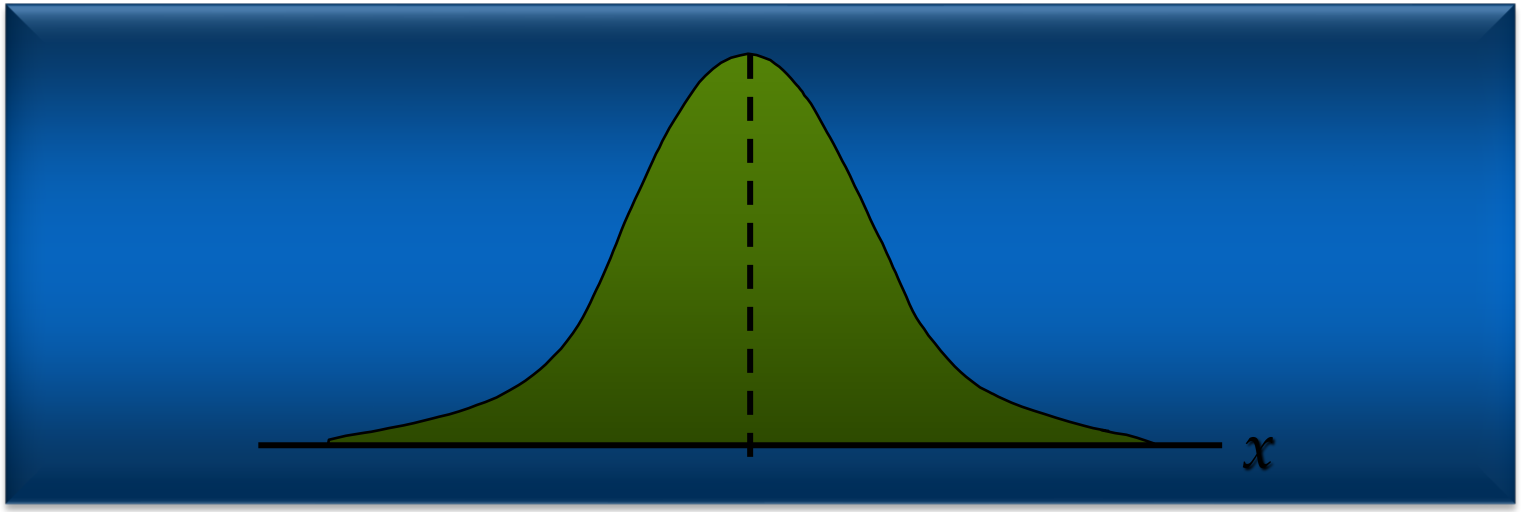
The entire family of normal probability distributions is defined by its mean μ and its standard deviation σ.



Normal Probability Distribution

- Characteristics

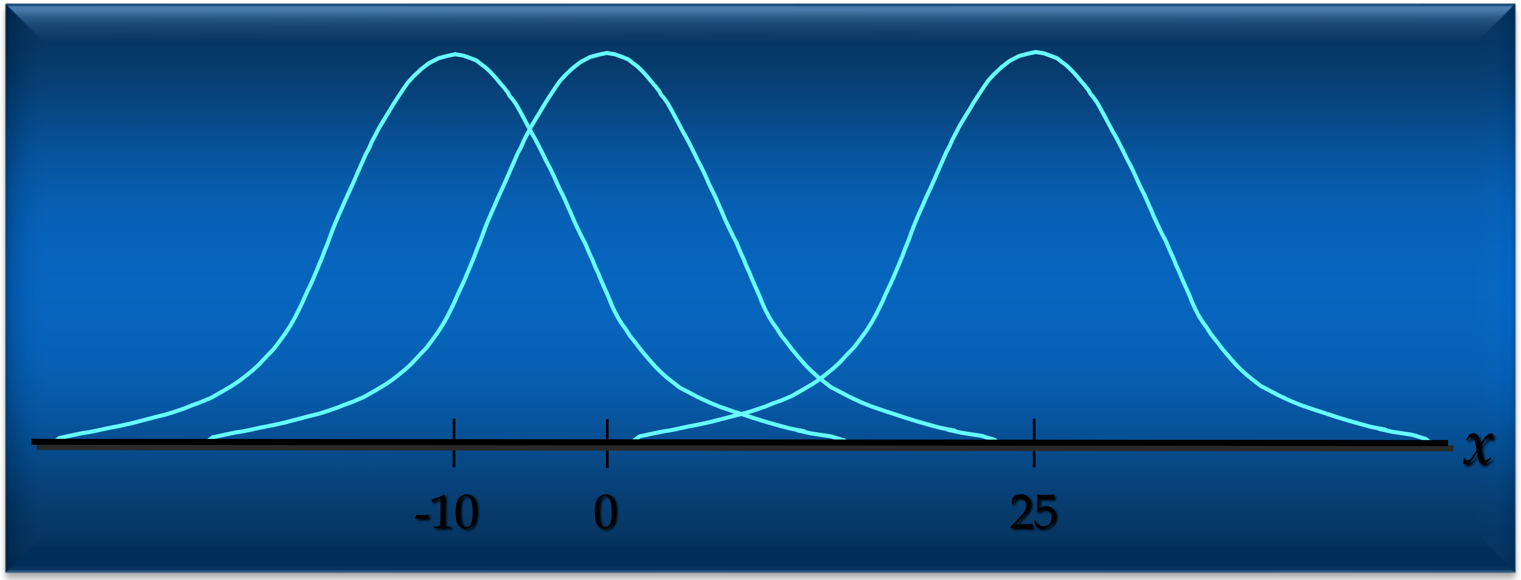
The highest point on the normal curve is at the mean, which is also the median and mode.



Normal Probability Distribution

- Characteristics

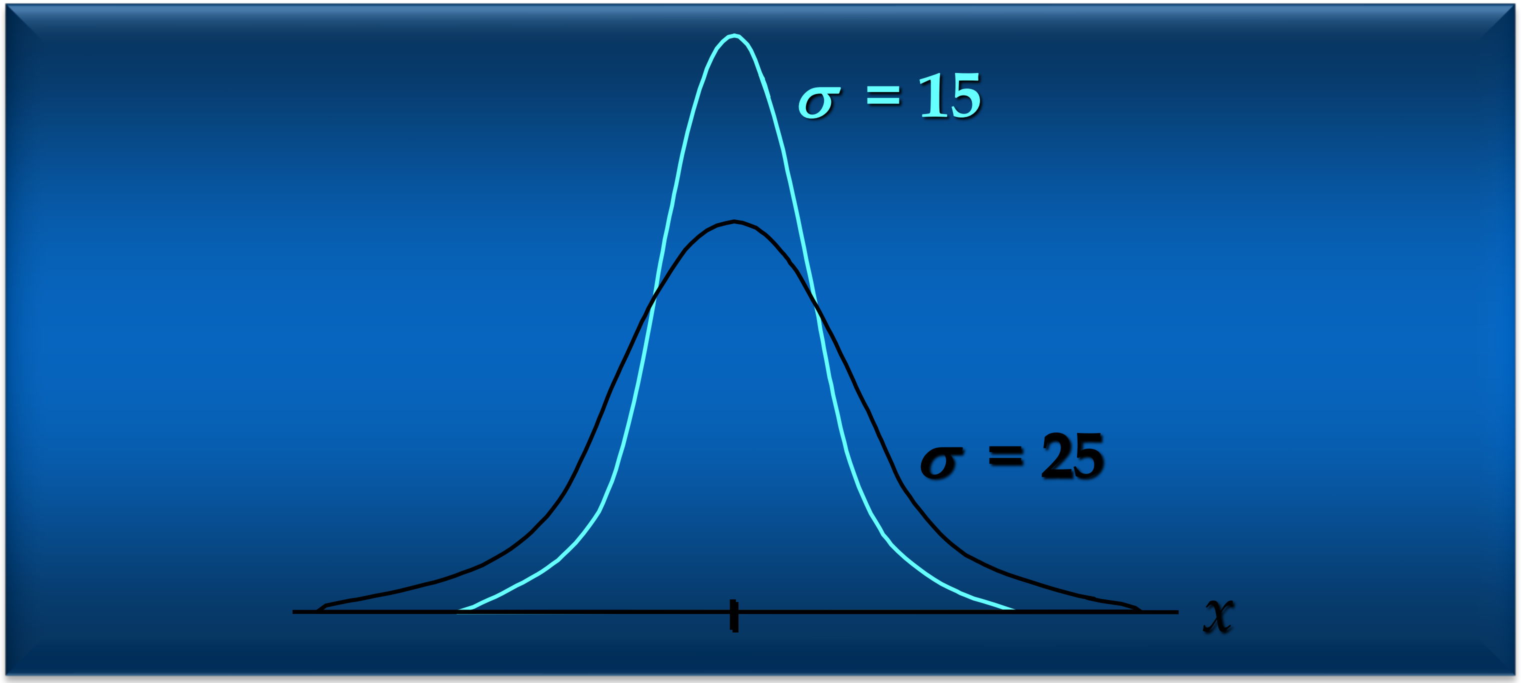
The mean can be any numerical value: negative, zero, or positive.



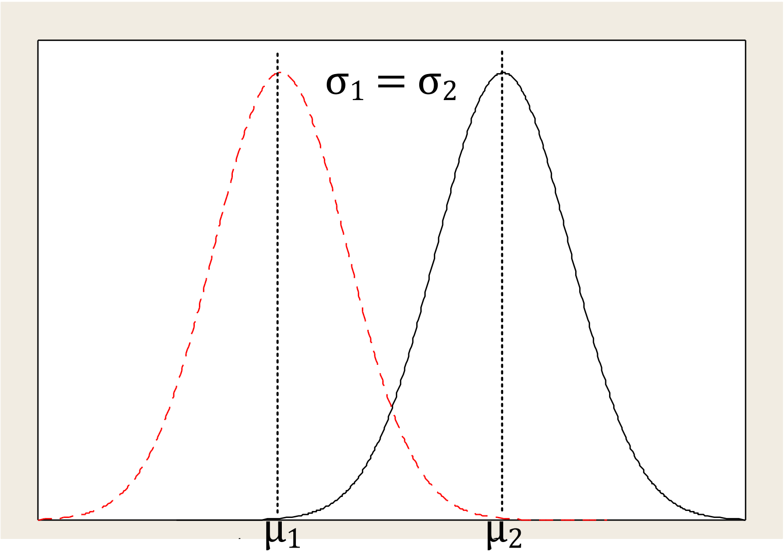
Normal Probability Distribution

- Characteristics

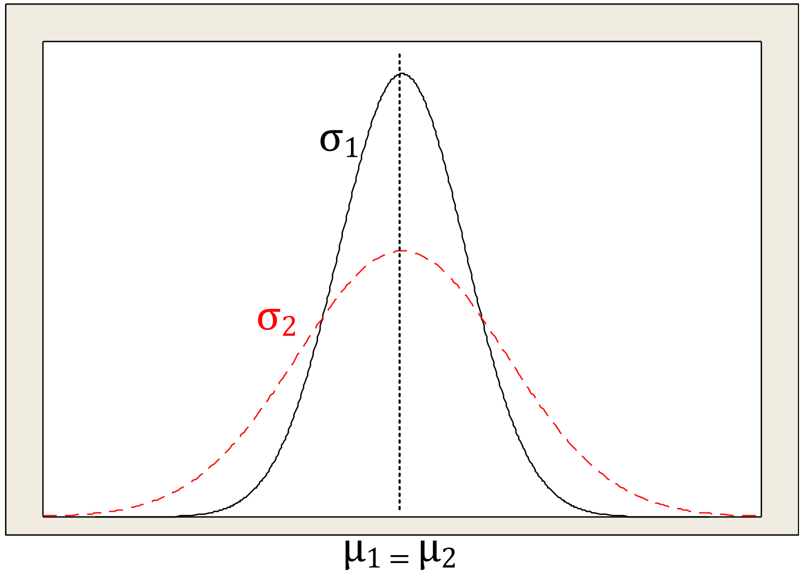
The standard deviation determines the width of the curve: larger values result in wider, flatter curves.



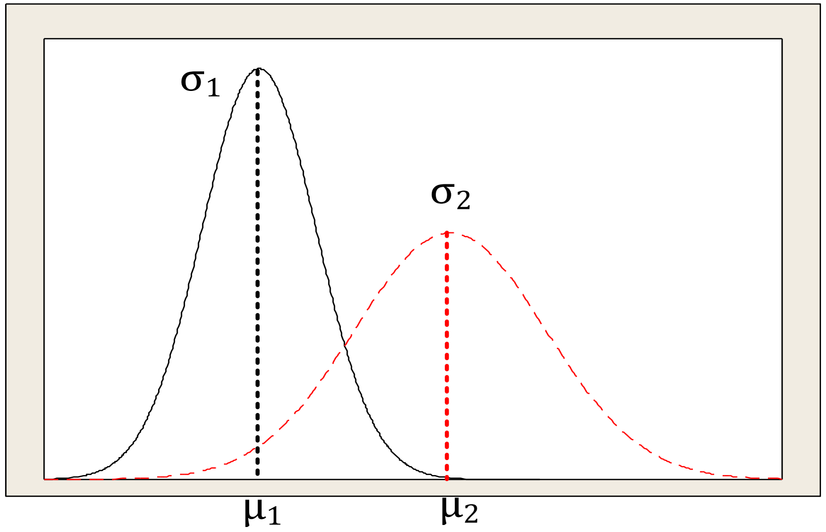
Normal Distribution



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$



Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$