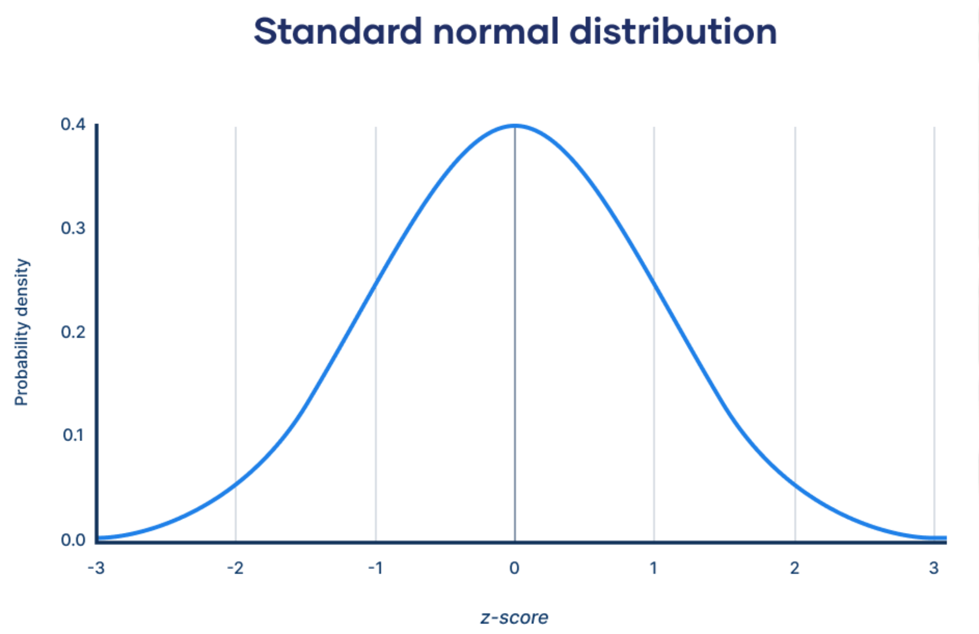
The Normal Curve

Unit 7

The Normal Curve

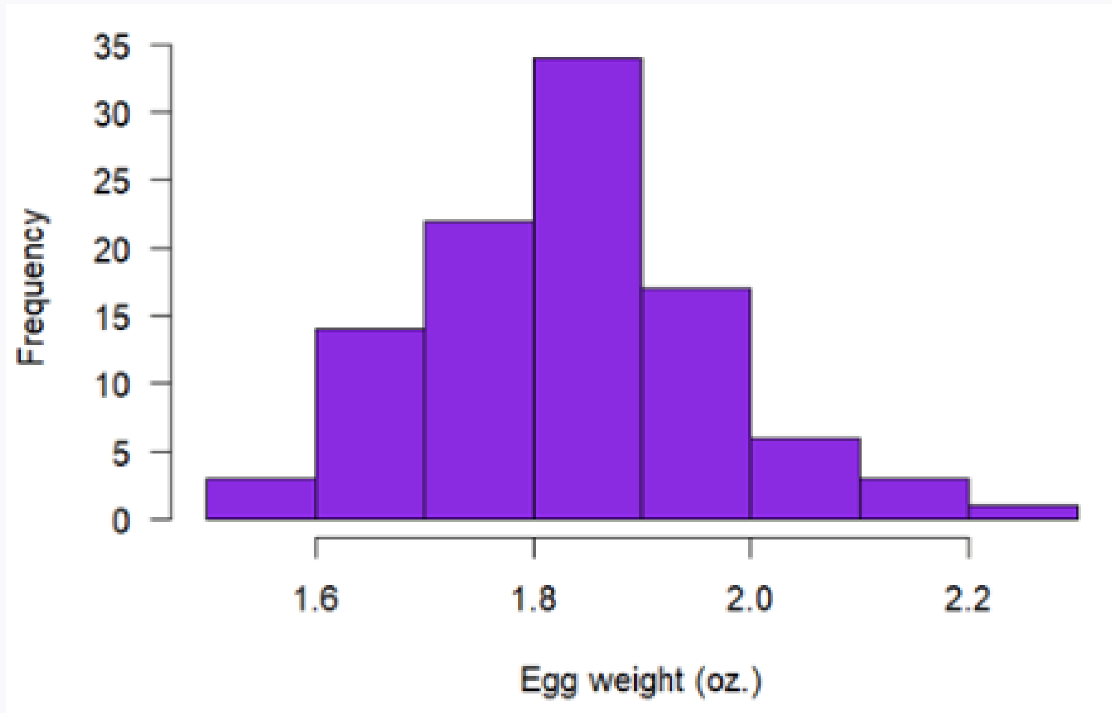
- Normal distribution, also known as the Gaussian distribution, is a probability distribution that appears as a "bell curve when graphed.



Probability Distributions

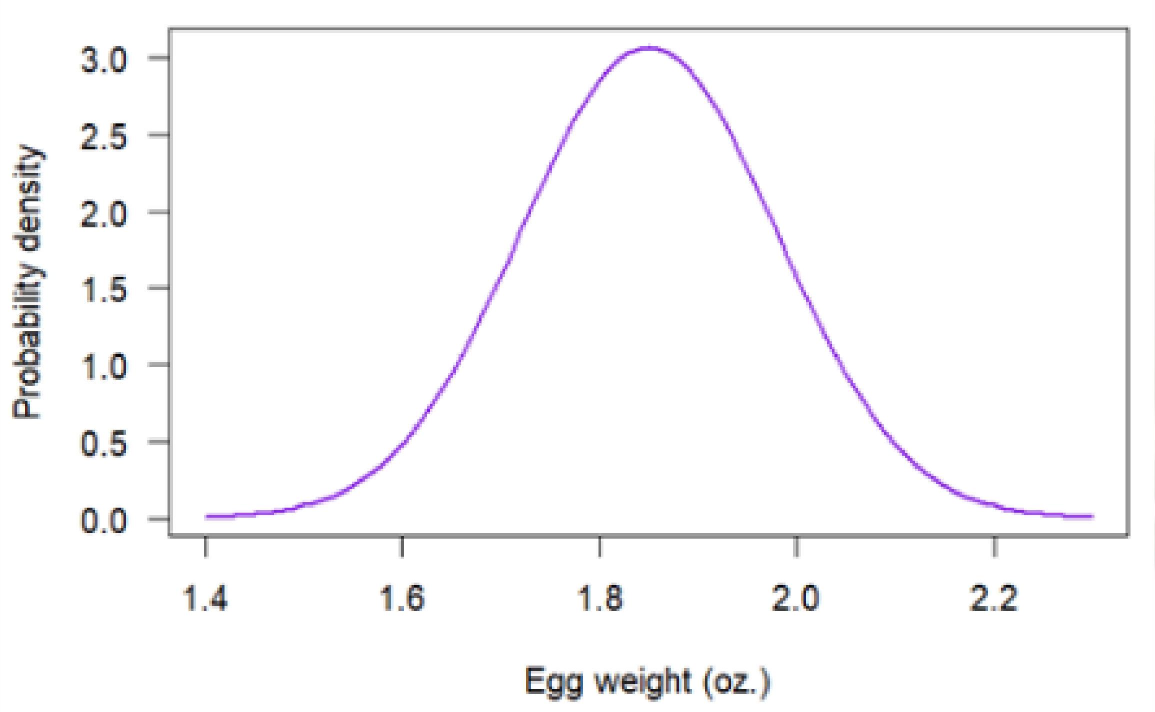
- A **frequency distribution** describes the number of times each possible value of a variable occurs in the dataset.

The farmer weighs 100 random eggs and describes their frequency distribution using a histogram:



Probability Distributions

- A **probability distribution** is an idealized frequency distribution.
  - describe the population the sample was drawn from.



Probability Distributions

- Probability is a number between 0 and 1 that says how likely something is to occur
  - 0 means it's impossible.
  - 1 means it's certain.
- The higher the probability of a value, the higher its frequency in a sample.
- More specifically, the probability of a value is its relative frequency in an **infinitely large sample.**

Random Variables

- **Random variables** are variables that follow a probability distribution
- **Notations**
  - Random variables are usually denoted by X
  - The ~ (tilde) symbol means "follows the distribution."
  - The distribution is denoted by a capital letter (usually the first letter of the distribution's name), followed by brackets that contain the distribution's parameters.
    - $X \sim N(\mu, \sigma^2)$
    - The random variable X follows a normal distribution with a mean of μ and a variance of σ^2

Types of Probability Distributions

- Discrete Probability Distributions
  - Categorical or discrete variable
    - Doesn't include any values with a probability of zero
      - Binomial
      - Poisson
- Continuous Probability Distributions
  - Continuous variables
    - Probability of occurrence is infinitely small to have a probability of zero

Continuous Probability Distributions

- A **continuous variable** is a variable that can assume any value on a continuum (can assume an uncountable number of values)
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height, in inches
- These can potentially take on any value depending only on the ability to precisely and accurately measure

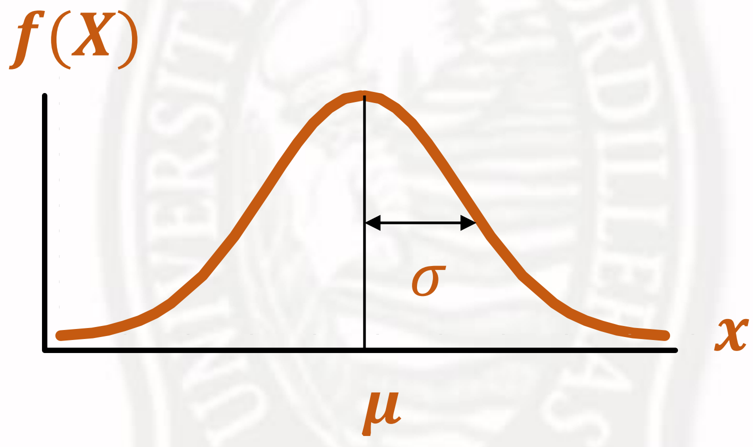
The Normal Distribution

- Bell Shaped
- Symmetrical
- Mean, Median and Mode are Equal

Location is determined by the mean, μ

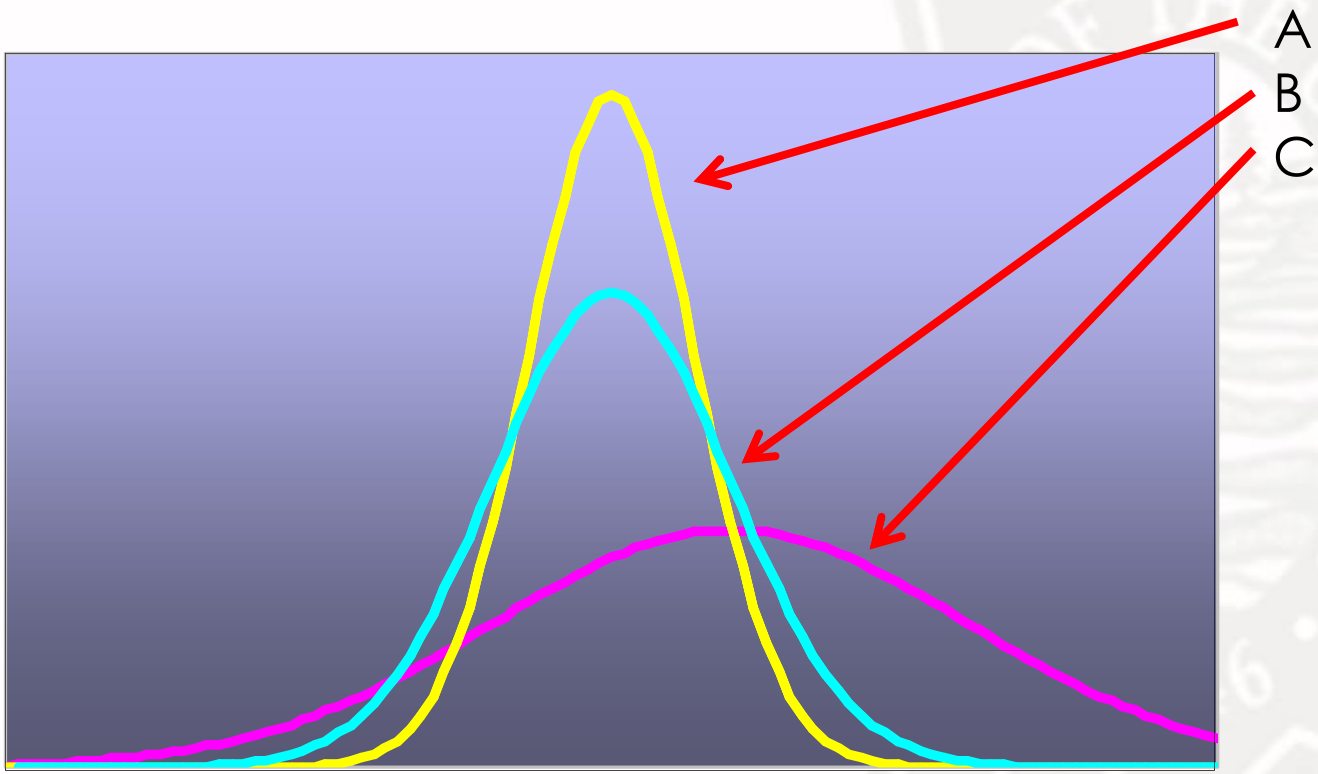
Spread is determined by the standard deviation, σ

The random variable has an infinite theoretical range: $+\infty$ to $-\infty$



Mean
= Median
= Mode

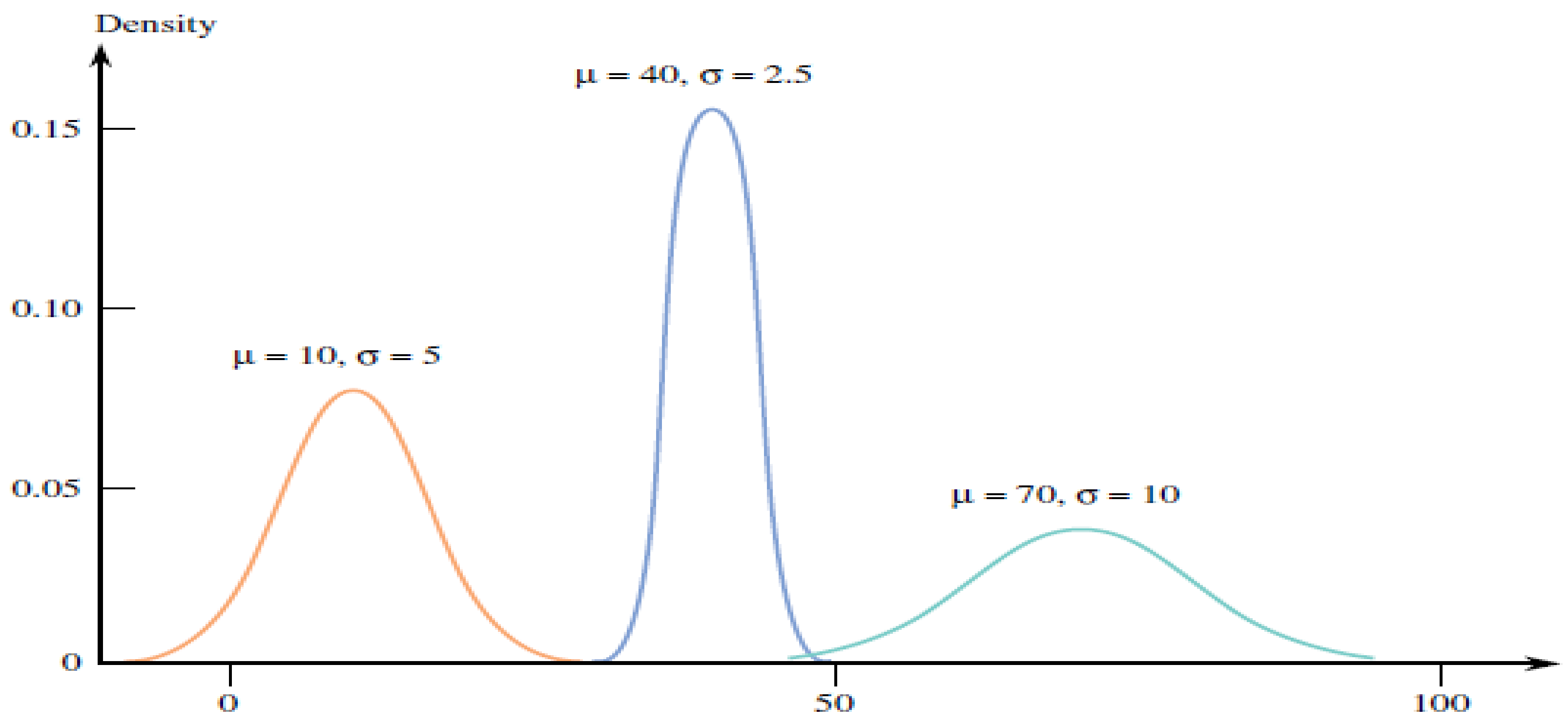
By varying the parameters μ and σ, we obtain different normal distributions



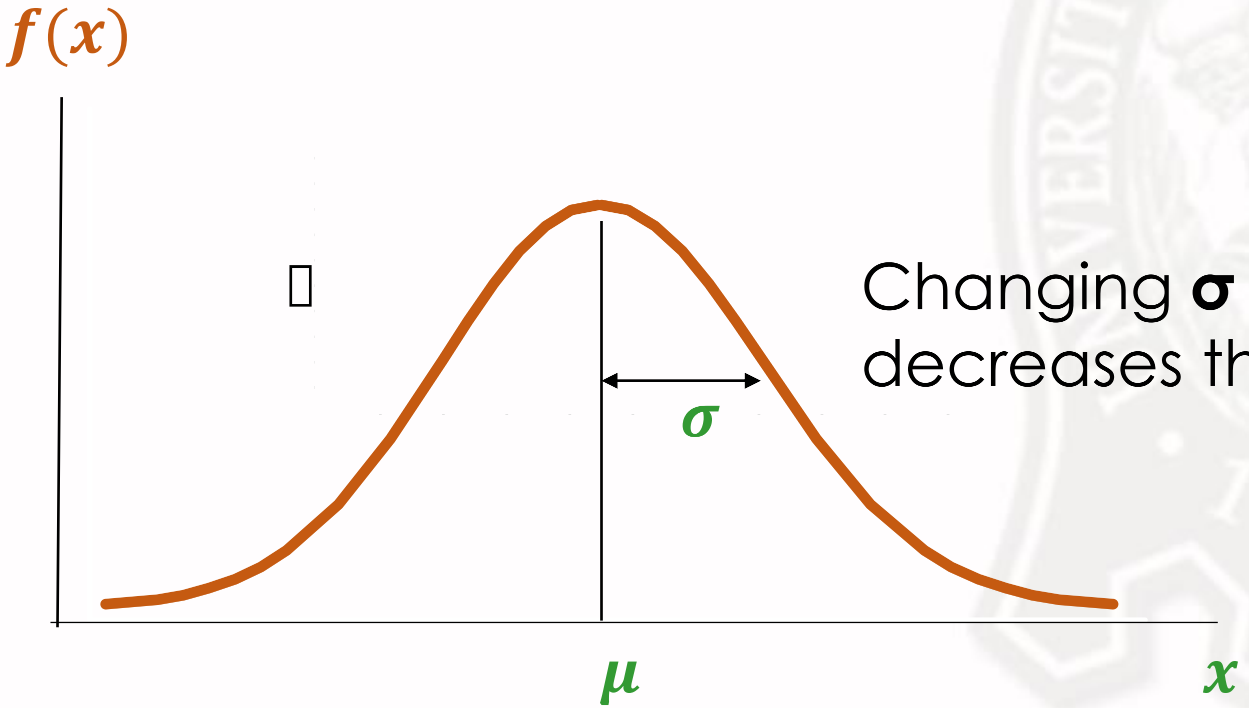
A and B have the same mean but different standard deviations.

B and C have different means and different standard deviations.

By varying the parameters ***μ*** and ***σ***, we obtain different normal distributions



The Normal Distribution Shape



Changing **σ** increases or decreases the spread.

Changing **μ** shifts the distribution left or right.

The Standardized Normal

- Any normal distribution (with any mean and standard deviation combination) can be transformed into the standardized normal distribution (z)

- To compute normal probabilities need to transform x units into z units

- The standardized normal distribution (z) has a mean, $\mu = 0$ and a standard deviation, $\sigma = 1$

Translation to the Standardized Normal Distribution

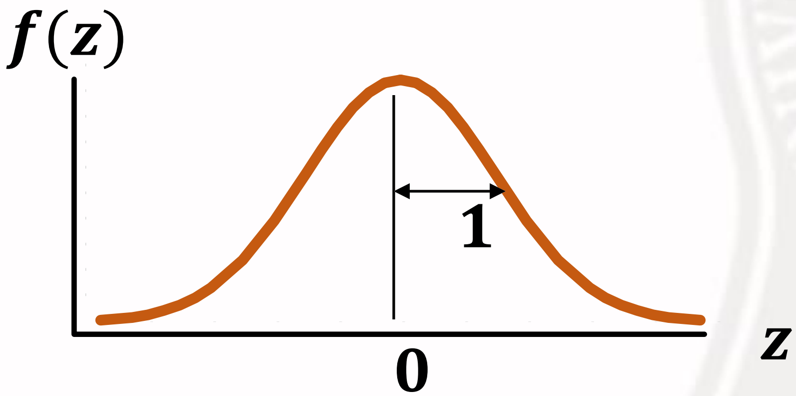
- Translate from x to the standardized normal (the "z" distribution) by subtracting the mean of x and dividing by its standard deviation:

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

The Z distribution always has $\mu = 0$ and $\sigma = 1$

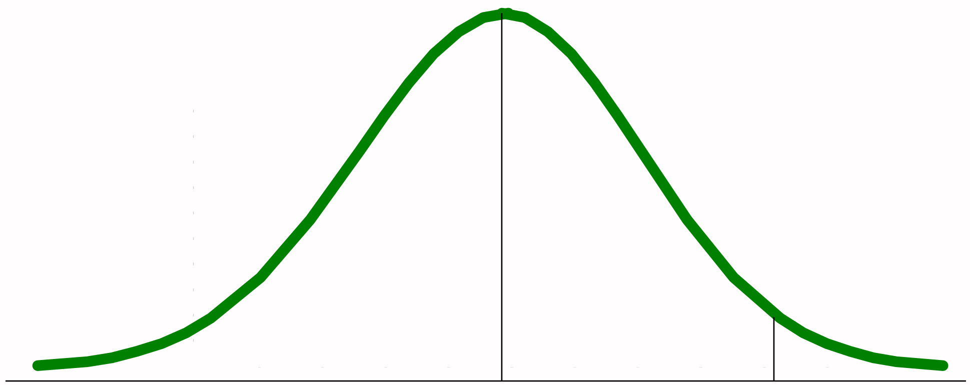
The Standardized Normal Distribution

- Also known as the "z" distribution
- Mean is 0
- Standard Deviation is 1



Values above the mean have positive z-values.
Values below the mean have negative z-values

Comparing X and Z units



\$100 **\$200** $\$X \; (\mu = \$100, \sigma = \$50)$

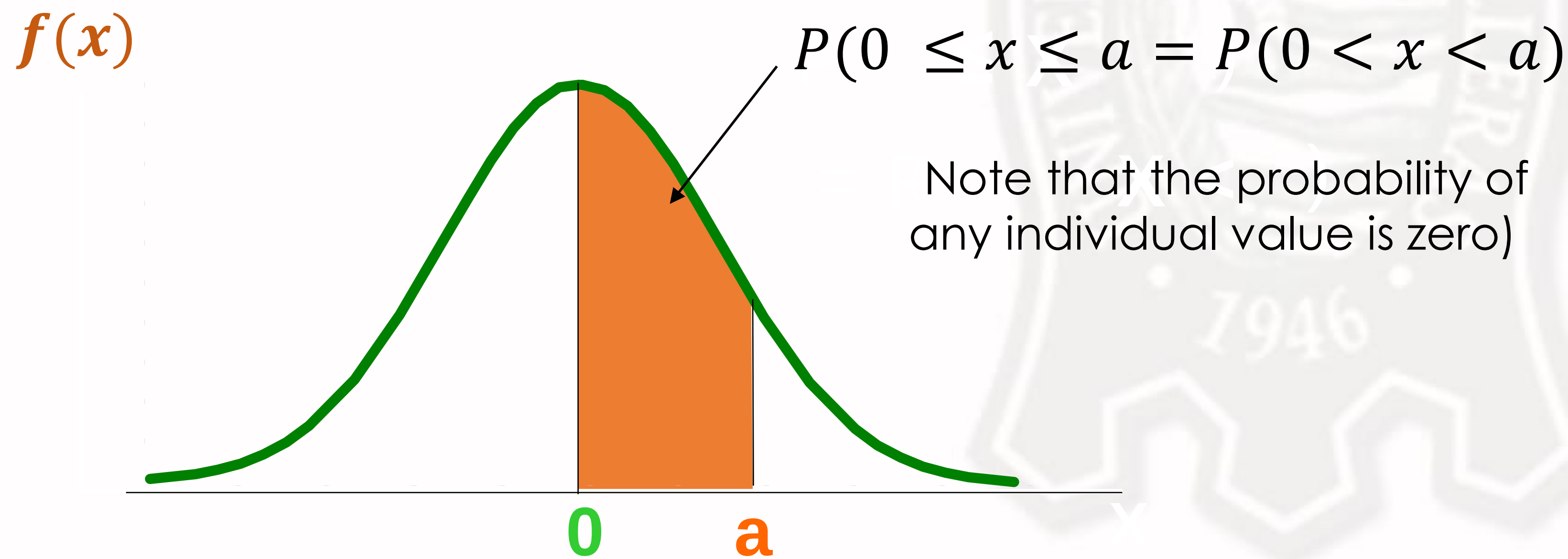
0 **2.0** $z \; (\mu = 0, \sigma = 1)$

Note that the shape of the distribution is the same, only the scale has changed.

We can express the problem in the original units (X in dollars) or in standardized units (z)

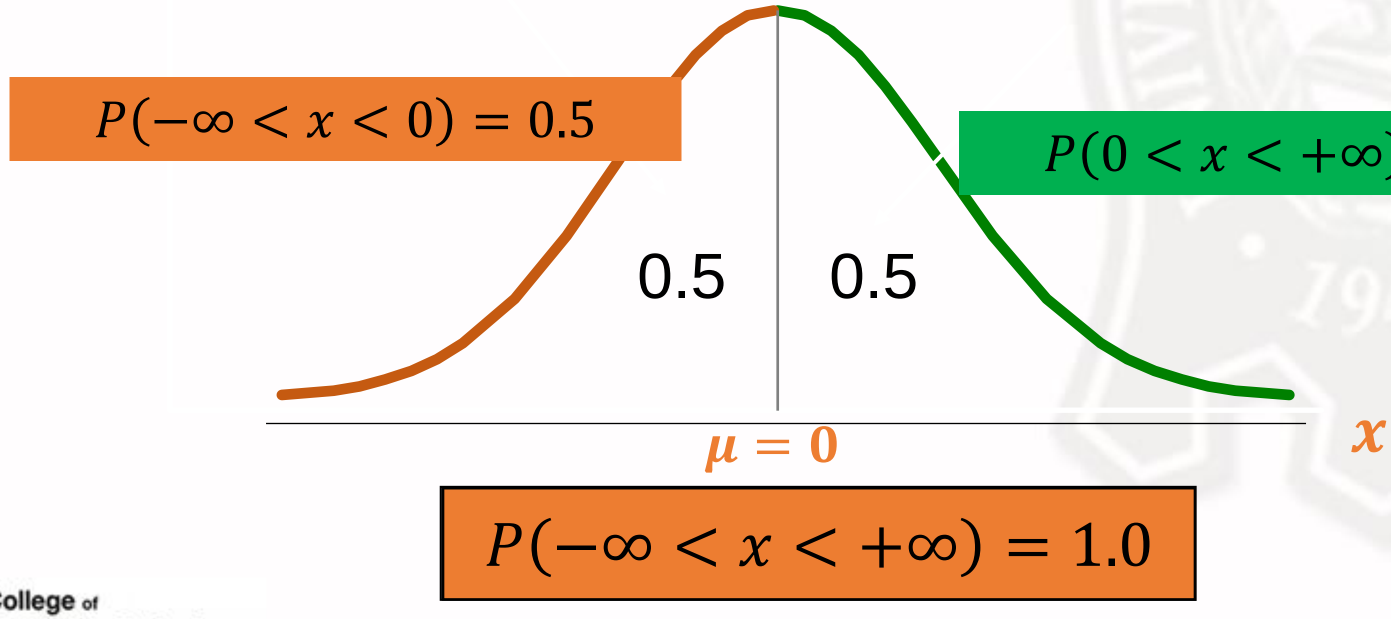
Finding Normal Probabilities

Probability is measured by the area under the curve

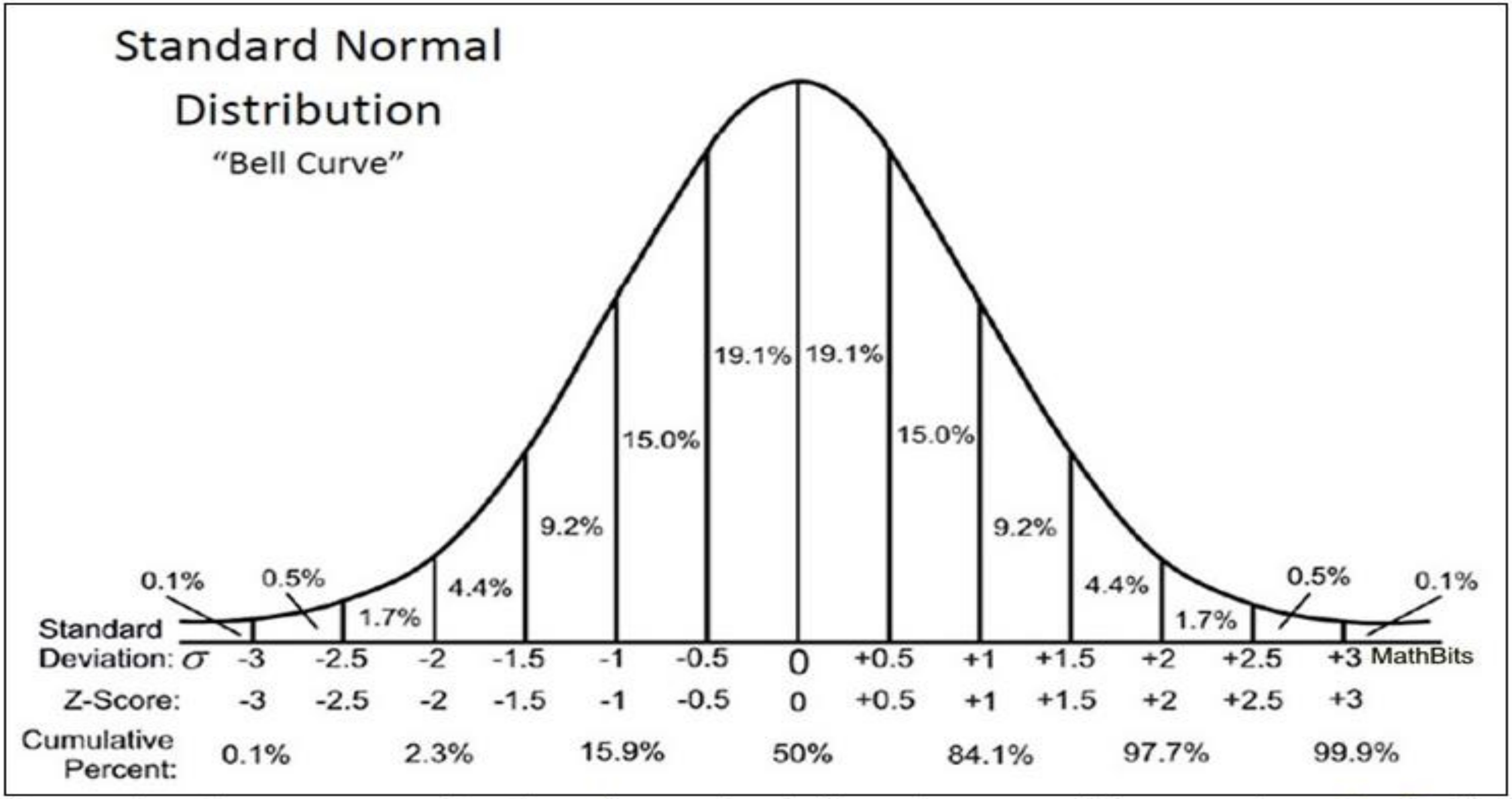


Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below



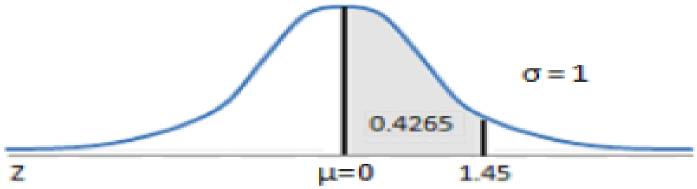
Probability as Area Under the Curve



Remember that z-scores tell us how far a value is from the mean. When you "standardize" a variable, its mean becomes zero and its standard deviation becomes one.

The Standardized Normal Table (Cumulative from the Mean)

This table provides the area between the mean and some Z score.
For example, when Z score = 1.45 the area = 0.4265.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890

Applications: Finding Normal Probabilities

- Finding the area under the normal curve
  a. Area from the mean to any $z - value$
     - $P(0 < z < a)$ or $P(a < z < 0)$
  b. Area between two $z - values$
     - $P(a < z < b)$
  c. Area to the right or left of a $z - value$
     - $P(z < a)$
- Finding x, given a normal probability

General Procedure for Finding Normal Probabilities

1. Translate x-values to z-values
$$z = \frac{x - \mu}{\sigma}$$
2. Draw the standard normal curve and shade the area required.
3. Use the Standardized Normal Table

Let's try:

- Suppose that the variable x is normally distributed with $\mu = 100$ and $\sigma = 15$. Find the following probabilities.
    - $P(100 < x < 126)$
    - $P(80 < x < 100)$
    - $P(105 < x < 112)$
    - $P(87 < x < 98)$
    - $P(x > 126)$
    - $P(x < 115)$
    - $P(x > 87)$
    - $P(x < 115)$

Let’s try:

- Data from the article “The Osteological Paradox: Problems in Inferring Prehistoric Health from Skeletal Samples” (*Current Anthropology* [1992]: 343–370) suggest that a reasonable model for the probability distribution of the continuous numerical variable x = height of a randomly selected 5-year-old child is a normal distribution with a mean, μ of 39.4” and standard deviation, σ of 2.4”. What proportion of the heights is
- Taller than 39.4” but smaller than 42”?
- Smaller than 39.4” but taller than 36”?
- between 36” and 42”?
- Between 42” and 46”?
- Smaller than 35”
- Taller than 41”

Given a Normal Probability Find the X Value

- Steps to find the X value for a known probability:
  1. Find the z value for the known probability
  2. Convert to x units using the formula:

$$x = z\sigma + \mu$$

Let's try: Given a Normal Probability Find the x Value

- Data from the article "Determining Statistical Characteristics of a Vehicle Emissions Audit Procedure" (*Technometrics* [1980]: 483–493) suggest that the emissions of nitrogen oxides, which are major constituents of smog, can be plausibly modeled using a normal distribution. Let x denote the amount of this pollutant emitted by a randomly selected vehicle. The distribution of x can be described by a normal distribution with $\mu = 1.6$ and $\sigma = 0.4$. Suppose that the EPA wants to offer some sort of incentive to get the worst polluters off the road. What emission levels constitute the worst 10% of the vehicles?