USC School of Business

Data Driven Decision Making: Uniform & Normal Distributions

GSBA 545, Fall 2021

Professor Dawn Porter

Uniform & Normal Distributions

- Uniform Random Variables
  - Probability Calculations
  - Summary Statistics for Continuous Random Variables

- Normal Distribution Basics
  - Standard Normal
  - *Z*-scores
  - Percentiles
  - Using the Normal Distribution Table

Continuous Distributions

Consider historic stock market returns.

- Could the 2020 S&P annual return be 5%?
- How about 55%?
- What about -50%?
- What should the probability be that the return is *exactly* 4.83%?
- What should the probability be that the return is between 4% and 5%?

- We've seen values below and above 5%, but not *exactly* 5%.
- We have not yet seen a value higher than 52.56%.
- We have not yet seen a value lower than -43.84%.
- This is a value we have seen before.
- Assigning a probability to an interval is more sensible.

A **continuous probability distribution** assigns probability to intervals (not point values) and can have an infinite upper bound and/or an infinite lower bound.

Continuous Distributions

A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.

We cannot find the probability of a random variable being a particular value. Instead, we find the probability of the random variable being within a given interval.

- Probability of a random variable having a value in an interval $[x_1, x_2]$ is the same as the area under the curve between x_1 and x_2.
- So probability and area are interchangeable for continuous functions.

Uniform Distribution

So far, we have analyzed situations where an underlying random variable takes on *discrete* values. That may not always be the case.

- Consider a random number generator, such as the RAND () function in Excel. Each observation from RAND () can be viewed as a sample from a distribution.

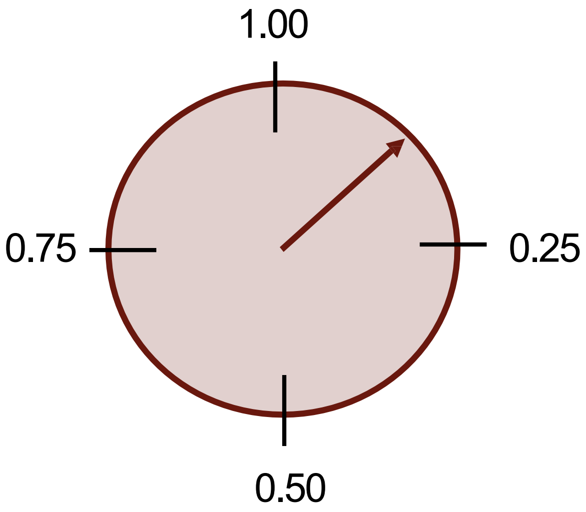
Observation Number	Observation Value
1	0.5567
2	0.1732
And so on…	

What shape will represent the *distribution* of this function?

- Let’s take 1000 observations and produce a histogram.

Uniform Random Variables

Let's start with a very simple example… a game *spinner.*



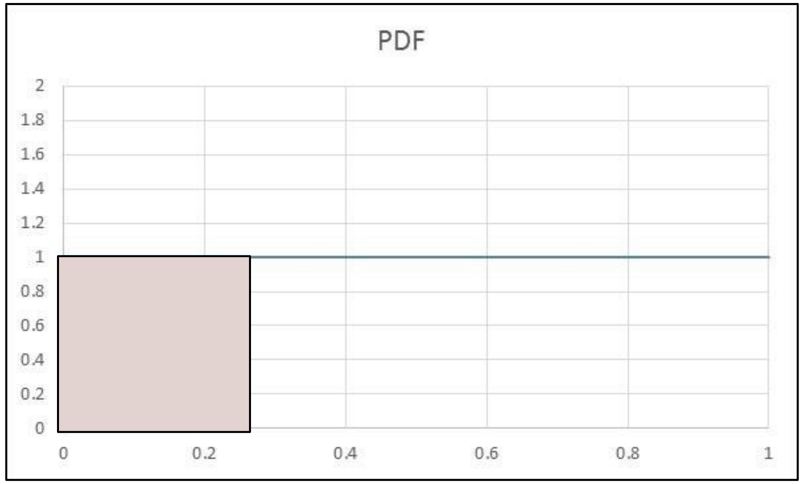
- X represents where the spinner stops, equally likely to be anywhere between 0 and 1.
- This is represented by a *uniform distribution* between a = 0 and b = 1, or $U(0,1)$.
- This is a continuous random variable. How do we calculate probabilities?

Spinner Probabilities

If E is the event the spinner stops below k = 0.25, what is the *probability* of E happening (or $P(E)$)?

$$\boxed{P(X < 0.25) = (k - 0)\times\frac{1}{b-a} = 0.25\times\frac{1}{1-0} = 0.25.}$$

Also, we can think of this as being 0.25 of the *total* area of 1.



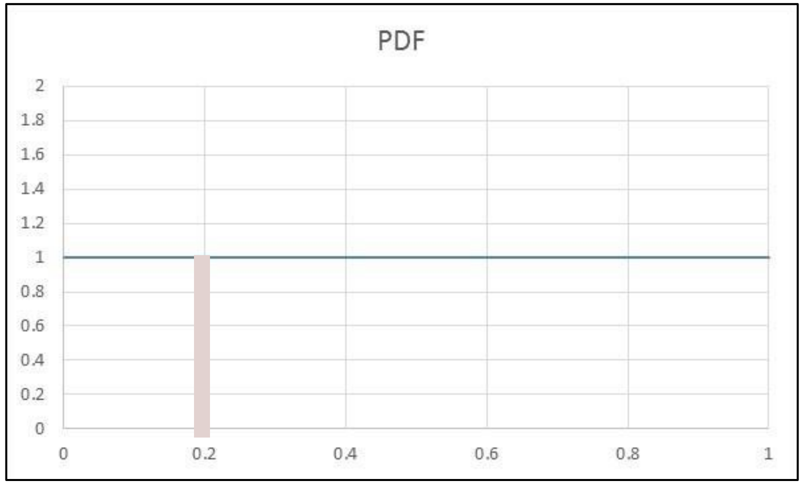
The probability is the area under the curve.

Spinner Probabilities

If E is the event that the spinner stops between 0.195 and 0.205, what is the *probability* of E happening (or $P(E)$)?

$$\boxed{P(0.195 < X < 0.205) = (0.205 - 0.195)\times\frac{1}{b-a} = 0.01\times\frac{1}{1-0} = 0.01.}$$

Also, we can think of this as being 0.01 of the *total* area of 1.



The probability is the area under the curve.

Uniform RV Summary Statistics

For a $U(a,b)$ random variable:

- Mean. $\mu = \frac{a+b}{2}$
- Median. $M_d = \frac{a+b}{2}$
- Variance. $\sigma^2 = \frac{(b-a)^2}{12}$
- p^{th} Percentile. $x_p = a + p \times (b - a)$

  This is where x is such that $P(X \leq x_p) = p$.

Example: Bus Waiting Times

Assume the amount of time, in minutes, that a person must wait for a bus is *uniformly distributed* between 0 and 15 minutes, so

$$X \sim U(0,15).$$

a) What is the probability a person waits fewer than 12.5 minutes?

b) On *average*, how long does a person have to wait for the bus?

c) What is the *standard deviation* of the waiting times?

d) Ninety percent of the time, the minutes a person must wait falls *below* what value?

Solutions: Bus Waiting Times

Assume the amount of time, in minutes, that a person must wait for a bus is *uniformly distributed* between 0 and 15 minutes, so $X \sim U(0,15)$.

a) What is the probability a person waits fewer than 12.5 minutes?

$$P(X < 12.5) = (12.5 - 0) \times \frac{1}{15-0} = 12.5 \times \frac{1}{15} = \boxed{0.8333}$$

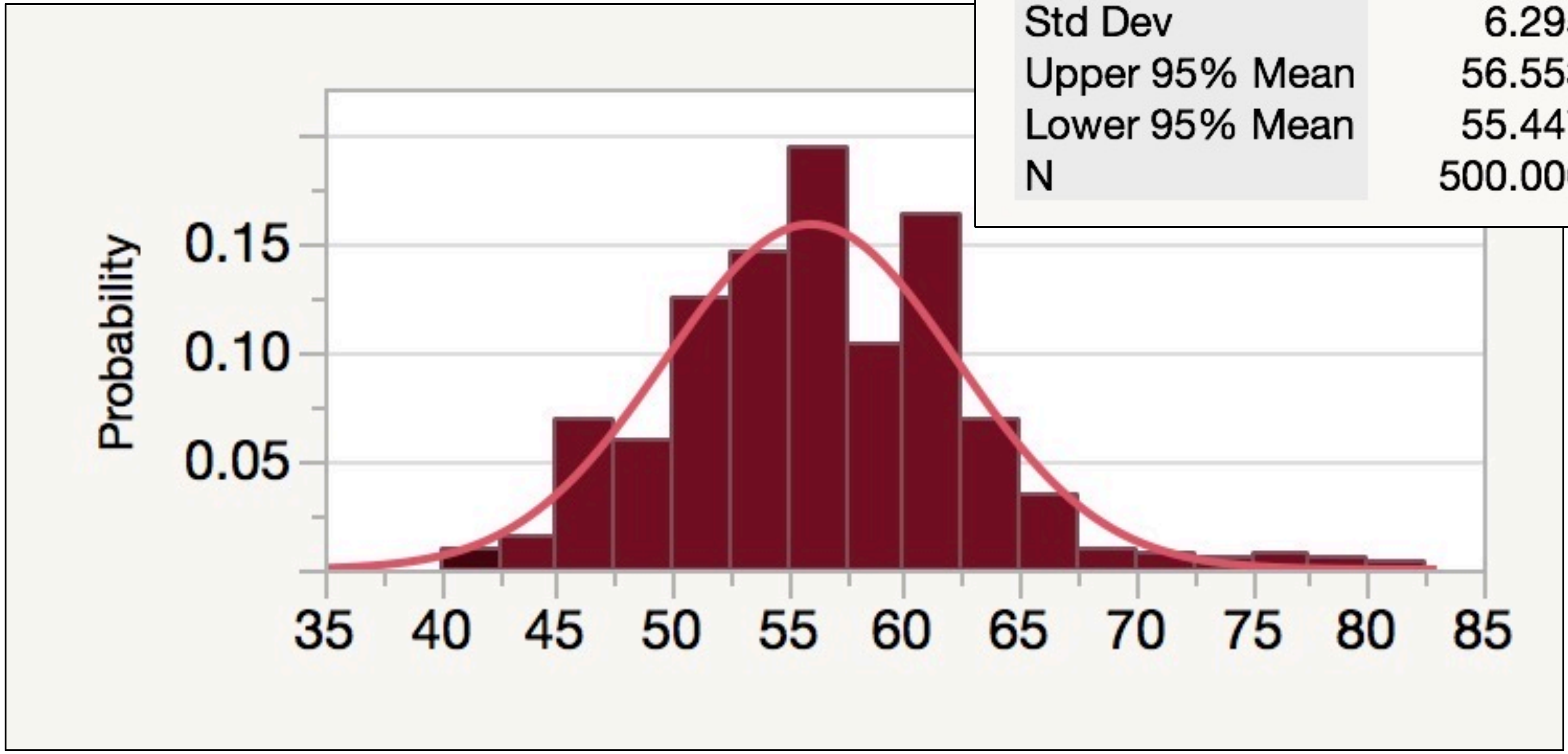
b) On *average*, how long does a person have to wait for the bus? $\mu = \frac{a+b}{2} = \frac{0+15}{2} = \boxed{7.5 \text{ mins}}$

c) What is the *standard deviation* of the waiting times? $\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(15-0)^2}{12}} = \boxed{4.3 \text{ mins}}$

d) Ninety percent of the time, the minutes a person must wait falls *below* what value?

$$P(X \leq x_p) = p \rightarrow x_{0.9} = 0.9 \times (15 - 0) = \boxed{13.5 \text{ mins}}$$

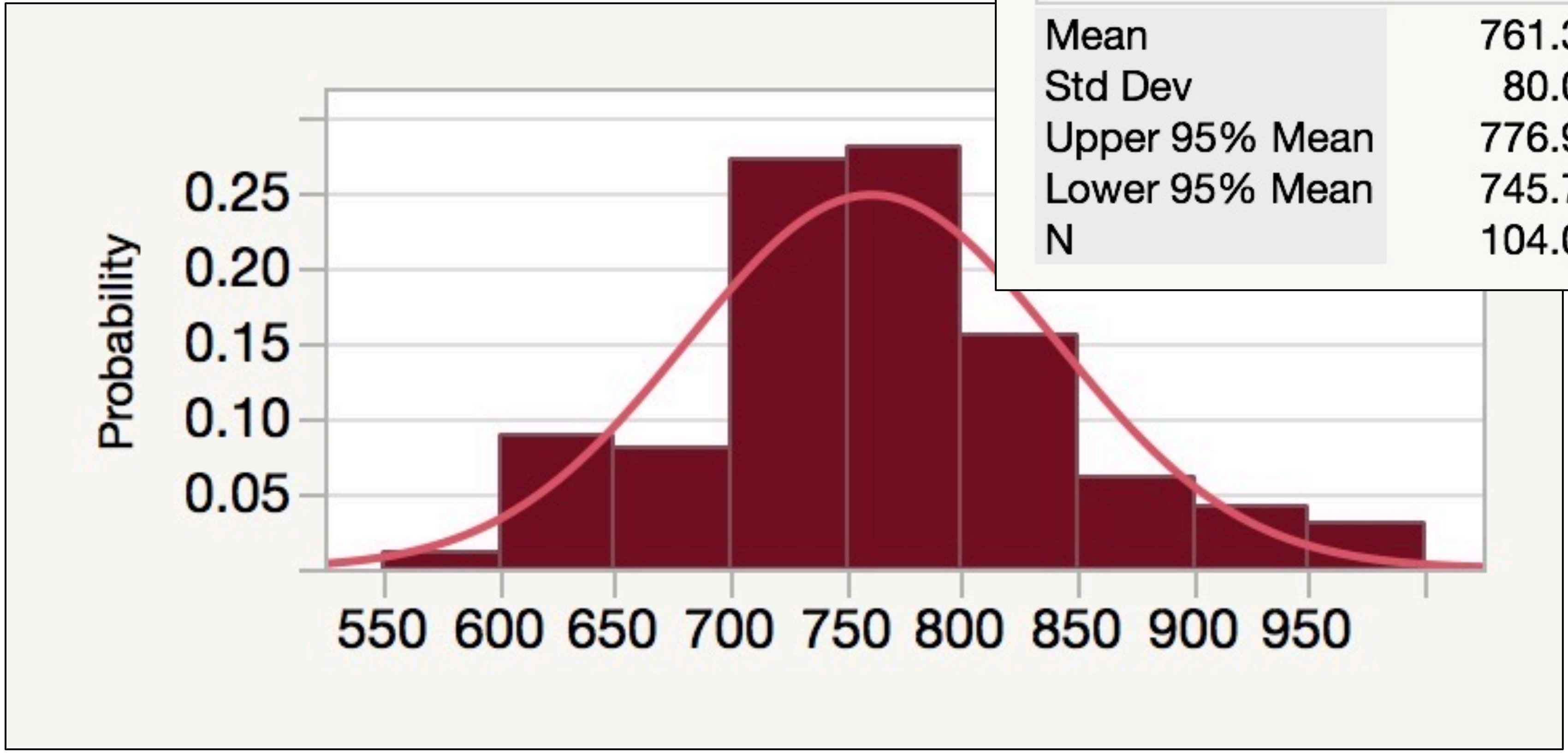
Top CEO Ages



Summary Statistics

Mean	56.000
Std Dev	6.295
Upper 95% Mean	56.553
Lower 95% Mean	55.447
N	500.000

Google Weekly Stock Prices



Summary Statistics

Mean	761.35
Std Dev	80.04
Upper 95% Mean	776.91
Lower 95% Mean	745.78
N	104.00

*Data from Yahoo! Finance, 8/3/15 – 7/24/17

Google Weekly Stock Prices

Mean, or *expected value*, based on history of the stock is $761.35 and the standard deviation is $80.04.

You bought stock at $700 a share and will sell if it exceeds $850.

Assuming we can consider this to be a normal distribution,

1. How likely is it that the price goes above $850?
2. What is the probability of losing money (i.e. the price dips below $700)?
3. What is a reasonable range of prices you could expect to see?

Normal Distribution

The normal probability distribution is the most important distribution for describing a continuous random variable.

It is commonly used in statistical inference and is found in a wide variety of applications including:

- Heights of people
- Amounts of rainfall
- Test scores
- Scientific measurements

Even though not all things follow a *normal distribution*, this will still be important and useful as we move into statistical inference.

Normal Distribution Properties

1. The curve has a bell shape, is symmetrical about μ, and reaches its maximum at μ.

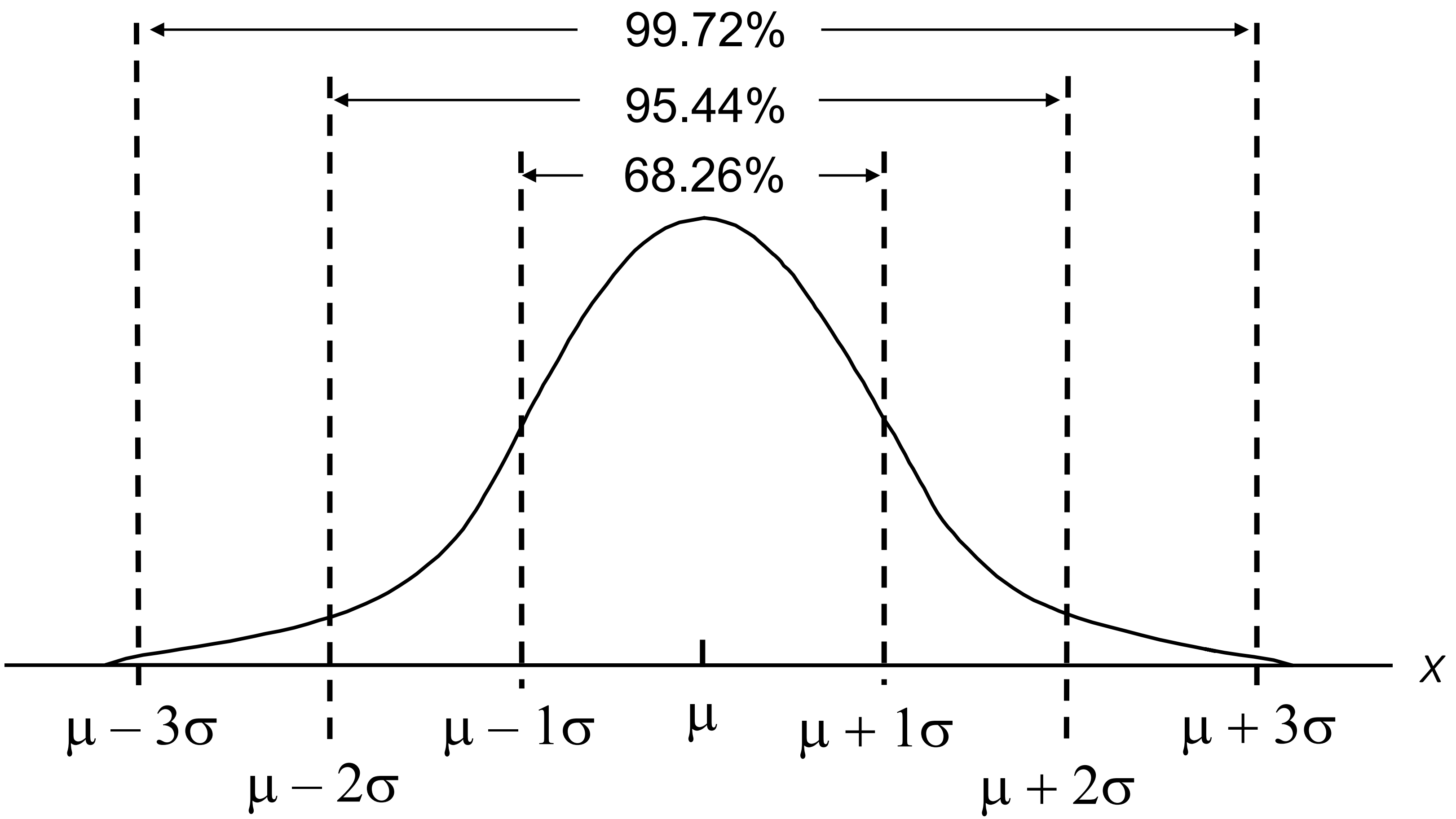
2. μ and σ determine center and spread of distribution.

3. Tails, or ends, of curve extend out to $\pm \infty$ and never actually touch zero.

4. **Empirical rule** holds for all normal distributions:

68% of area under curve lies between	$(\mu - \sigma, \mu + \sigma)$
95% of area under curve lies between	$(\mu - 2\sigma, \mu + 2\sigma)$
99.7% of area under curve lies between	$(\mu - 3\sigma, \mu + 3\sigma)$

Normal Distribution



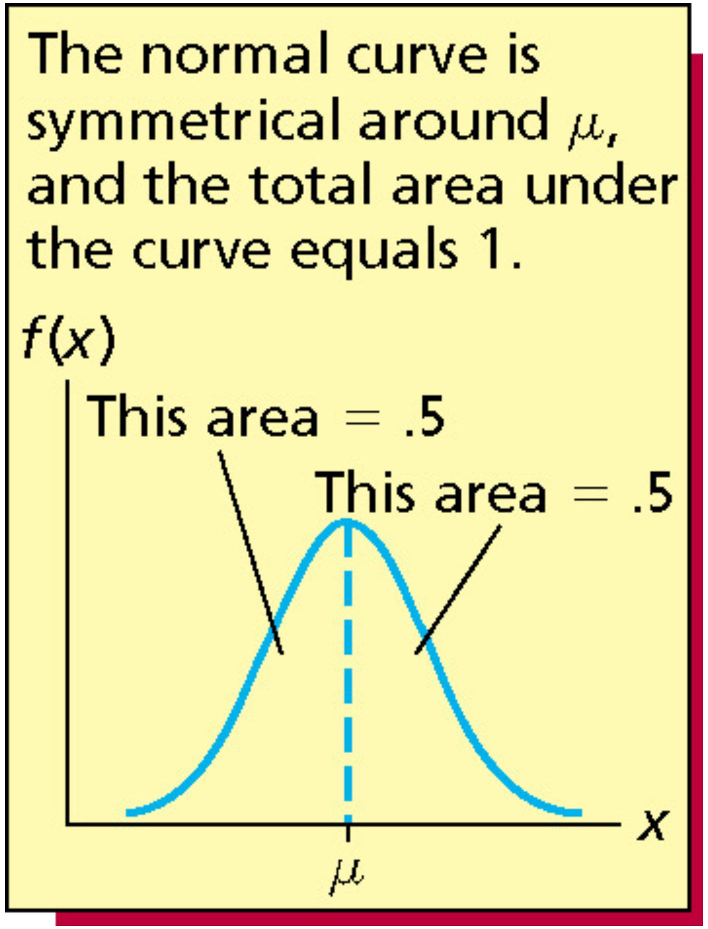
Normal Distribution

The **Normal Probability Distribution** is defined by the equation

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

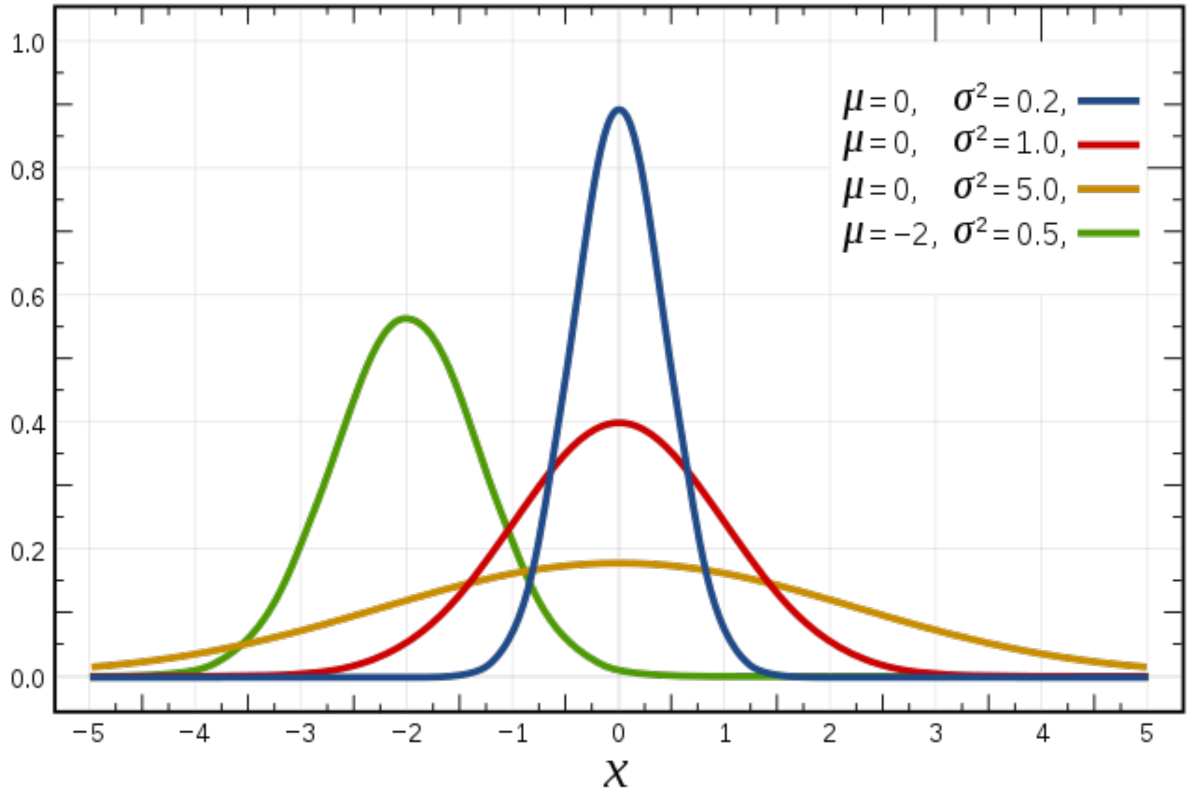
where

μ = mean
σ = standard deviation
π = 3.14159 …
e = 2.71828



Normal Distribution

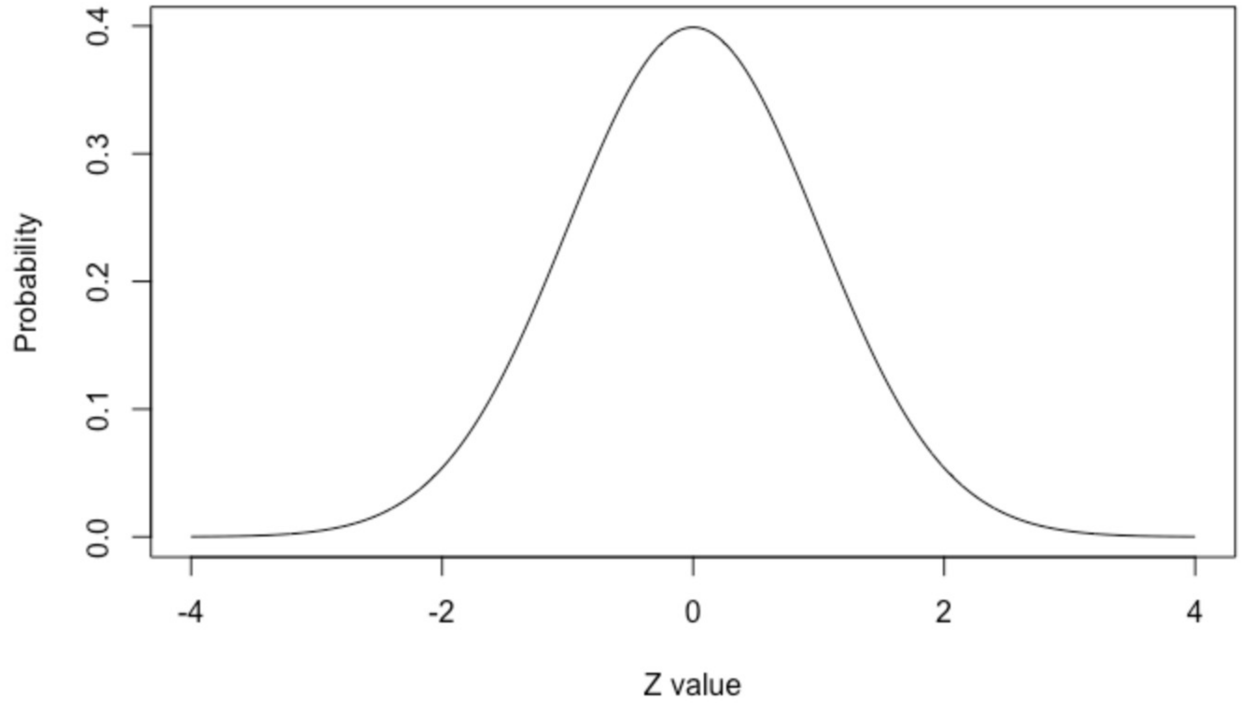
There are an infinite number of normal distributions, one for each choice of the parameters μ and σ.



Standard Normal Distribution

How do we compare normal random variables that don't have the same mean or std deviation?

- We can *standardize* them. A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is said to have a **standard normal distribution** and is denoted "Z."
- There are infinitely many normal distributions, but only one standard normal distribution.

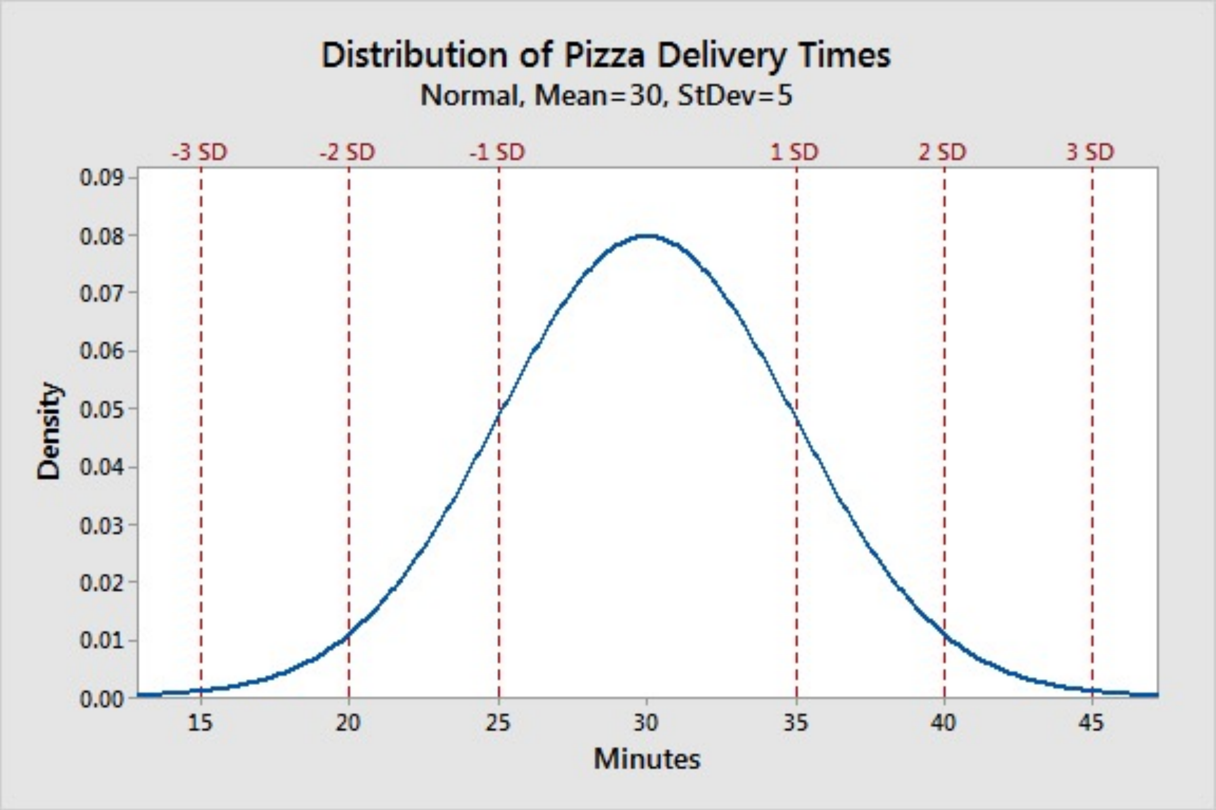


Standard Normal Distribution

If X is normally distributed with μ and σ, then $Z = \frac{x-\mu}{\sigma}$.

Z is now ALSO normally distributed, with μ = 0 and σ = 1, and Z follows a **standard normal distribution.**

*Z-scores measure how many standard deviations X is away from its mean.



Normal Distribution: Pep Zone

What is the probability of a stockout at Pep Zone?

- Pep Zone sells a popular multi-grade motor oil.
- When the stock of this oil drops to 20 gallons, a replenishment is ordered.
- The store manager is concerned that sales are being lost due to stockouts while waiting for a replenishment order.
- Demand during this time is normally distributed with $\mu = 15$ gallons and $\sigma = 6$ gallons.

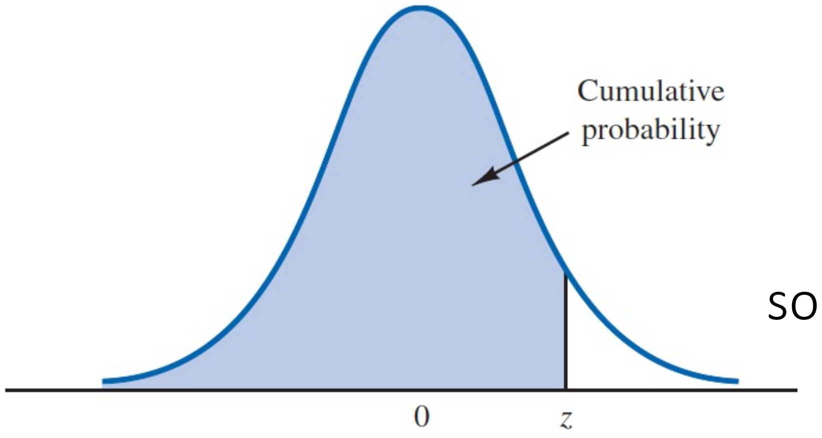
$$\mathrm{P}(X > 20) = ?$$

$$\mathrm{P}(X > 20) = \mathrm{P}\left(Z > \frac{20 - 15}{6}\right) = \boxed{\mathrm{P}(Z > 0.83)}$$

So we need to find the area to the right of *z* = 0.83 in a Normal Table.

Normal Distribution: Pep Zone

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389



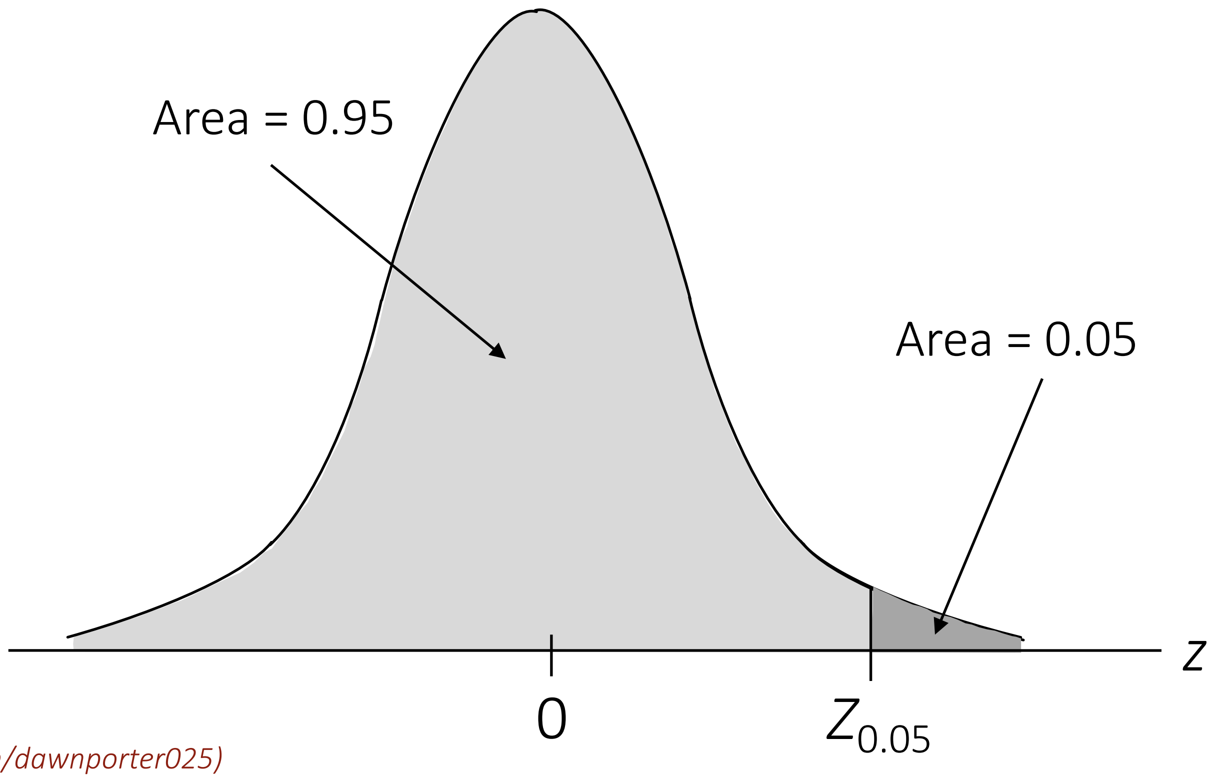
We know $P(Z \leq 0.83) = 0.7967$,

so $P(Z > 0.83) = 1 - 0.7967 = \boxed{0.2033}$

Normal Distribution: Pep Zone

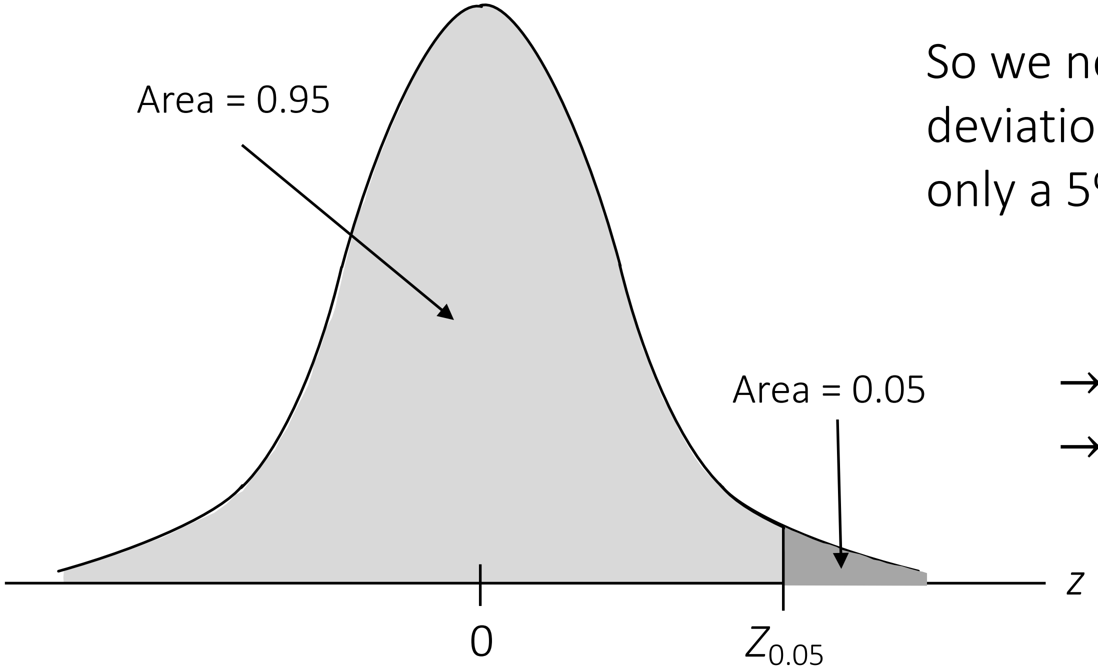
The store manager wants the probability of a **stockout** to be no more than 0.05. What should the reorder point be?

- Given a probability of 0.05, we can use the standard normal table backwards to find the corresponding *z* value.



Normal Distribution: Pep Zone

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767



So we need to be 1.645 standard deviations above the mean to ensure only a 5% chance of **stockout.**

$X = \mu + 1.645\,\sigma$

$\rightarrow X = 15 + 1.645\,(6)$

$\rightarrow X = 24.87$, or $\boxed{\textbf{25 gallons}}$

Normal Distribution: Pregnancy Days

The length (in days) of a randomly chosen human pregnancy is normally distributed with a mean of 266 days and standard deviation of 16 days.

a) What is the probability a randomly chosen pregnancy will last less than 246 days?
b) What is the probability a randomly chosen pregnancy will last longer than 240 days?
c) Suppose a pregnant woman's partner has scheduled business trips so as to be in town between the 235th and 295th days. What is the probability that the birth will take place during that time?

Normal Distn Solns: Pregnancy Days

The length (in days) of a randomly chosen human pregnancy is normally distributed with a mean of 266 days and standard deviation of 16 days, so $X \sim N(266, 16)$.

a) What is the probability a randomly chosen pregnancy will last less than 246 days?

$$P(X < 246) = P\left(Z < \frac{246 - 266}{16}\right) = P(Z < -1.25) = \boxed{0.1056}$$

b) What is the probability a randomly chosen pregnancy will last longer than 240 days?

$$P(X > 240) = P\left(Z > \frac{240-266}{16}\right) = P(Z > -1.63) = 1 - 0.0516 = \boxed{0.9484}$$

c) Suppose a pregnant woman's partner has scheduled business trips so as to be in town between the 235th and 295th days. What is the probability that the birth will take place during that time? $P(235 < X < 295) = P\left(\frac{235-266}{16} < Z < \frac{295-266}{16}\right)$

$$= P(-1.94 < Z < 1.81) = 0.9649 - 0.0262 = \boxed{0.9387}$$