Density Curves & Normal Distributions

Sections 4.1 & 4.2

Cathy Poliak, Ph.D.
cathy@math.uh.edu
Office in Fleming 11c

Department of Mathematics
University of Houston

Lecture 9 - 2311

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

1. Continuous Probability Distribution
2. Examples of Density Curves
3. Uniform Distribution
4. Shapes of Density Curves
5. The Normal Distributions
6. The Empirical Rule

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 05.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper 05 Questions

$n=10$; $p=0.3$; success = getting accepted

The probability that a student is accepted to a prestigious college is 0.3. Suppose 10 students apply to this college. What would be the best way to determine the following probabilities?

1. What is the probability that exactly 2 are accepted?

$P(X=2)$

a) pbinom(2,10,0.3) b) dbinom(2,10,0.3)
c) dbinom(1,10,0.3) d) pbinom(1,10,0.3)

2. What is the probability that at most 2 are accepted?

$P(X \leq 2)$

a) pbinom(2,10,0.3) b) dbinom(2,10,0.3)
c) 1 - dbinom(1,10,0.3) d) 1 - pbinom(1,10,0.3)

$P(X<2) = P(X \leq 1) = \text{pbinom}(1, 10, 0.3)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

3. What is the probability that at least 2 are accepted?

$P(X \geq 2)$

a) pbinom(2,10,0.3)
b) dbinom(2,10,0.3)
c) 1 - dbinom(1,10,0.3)
d) 1 - pbinom(1,10,0.3)

$1 - P(X \leq 1)$

4. What is the probability that between 2 and 6, inclusive, are accepted?

a) pbinom(6,10,0.3) - pbinom(2,10,0.3)
b) dbinom(6,10,0.3) - dbinom(2,10,0.3)
c) dbinom(6,10,0.3) - dbinom(1,10,0.3)
d) pbinom(6,10,0.3) - pbinom(1,10,0.3)

$P(2 \leq X \leq 6) = P(X \leq 6) - P(X \leq 1)$

0 1 2 3 4 5 6 7 8 9 10

$P(X \leq 6)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Probability Distribution

The **probability distribution** of a random variable X tells us what values X can take and how to assign probabilities to those values. Requirements for a probability distribution:

1. The sum of all the probabilities equal 1.

2. The probabilities are between 0 and 1, including 0 and 1.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Continuous Probability Distribution

- The probability distribution of a continuous random variable X is described by a density curve.
- The probability of any event is the area under the density curve and above the values of X that makes up the event.
- The mean is the center or expected value of that distribution.
- The standard deviation is the spread of that distribution.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

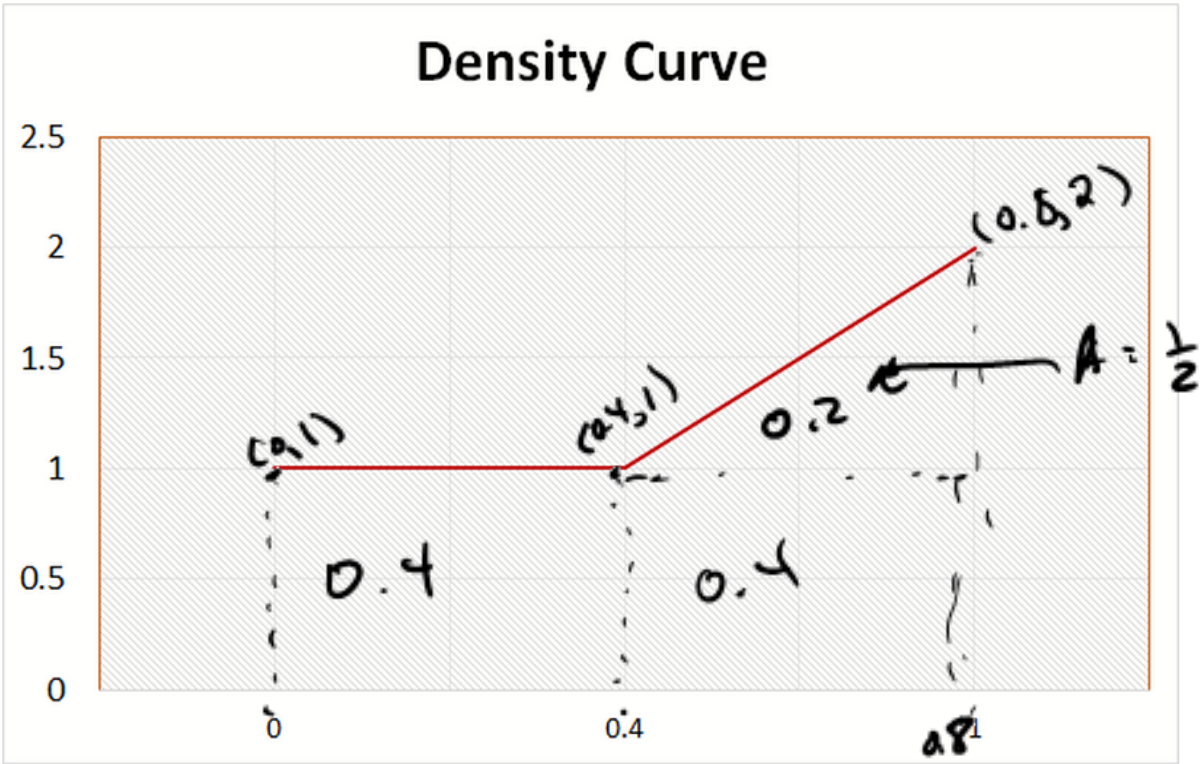
Density Curves

- A **mathematical model** for a probability distribution of a continuous random variable.
- This curve is always on or above the horizontal axis.
- The area under a density curve is exactly 1.
- The **area** under the curve and between any range of values is the probability that an observation falls in that range.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example of Density Curve

The following graph is an example of a density curve that consists of two line segments. The first goes from the point (0,1) to the point (0.4,1). The second goes from (0.4,1) to (0.8,2) in the xy-plane. Does this meet the requirements of a probability distribution?



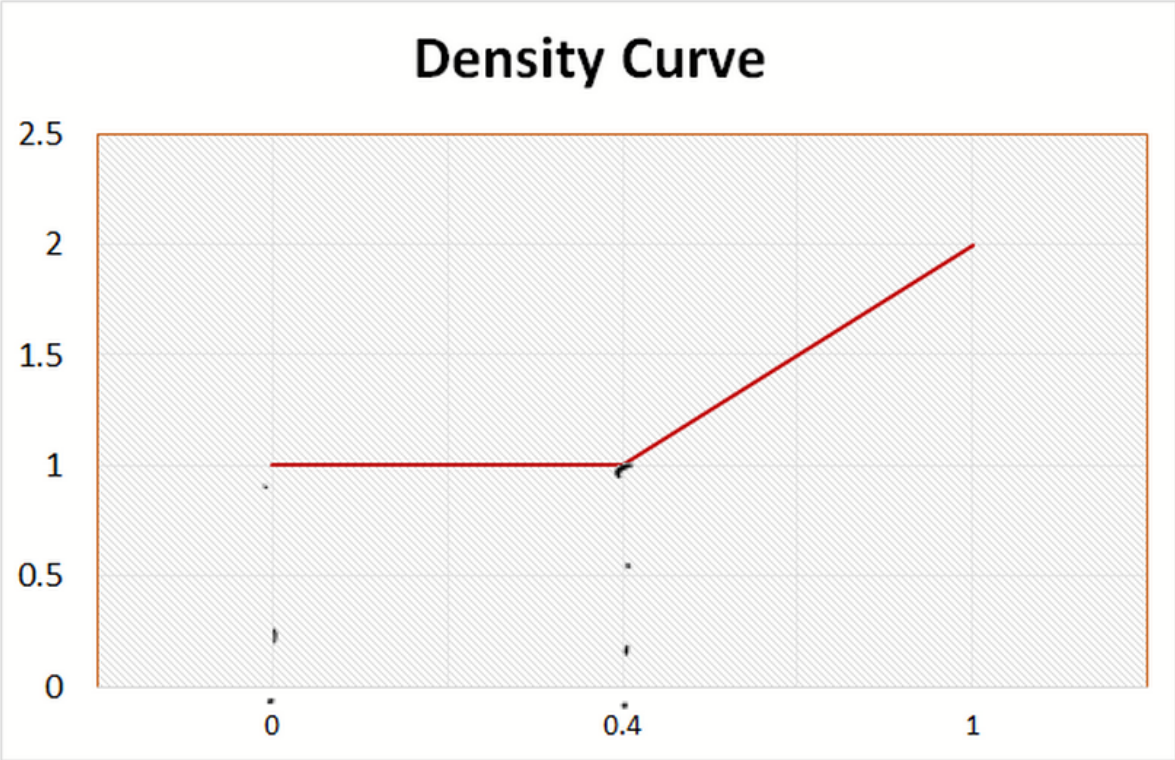
Area between $x=0$ and $x=0.4$

$A = w \cdot h = 0.4(1) = 0.4$

$A = \frac{1}{2}bh = \frac{1}{2}(0.4)(1) = 0.2$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

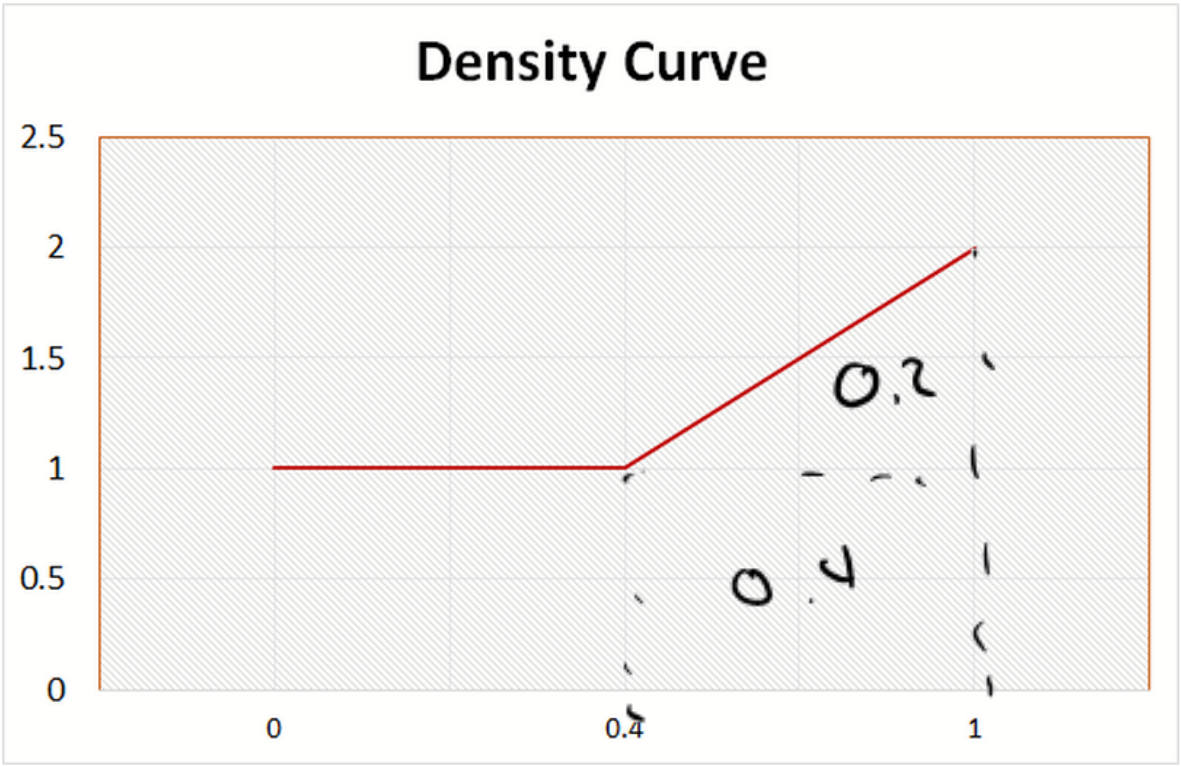
What percent of the observations fall below 0.4?



$P(X < 0.4) = 0.4(1) = 0.4$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What percent of the observations lie between 0.4 and 0.8?

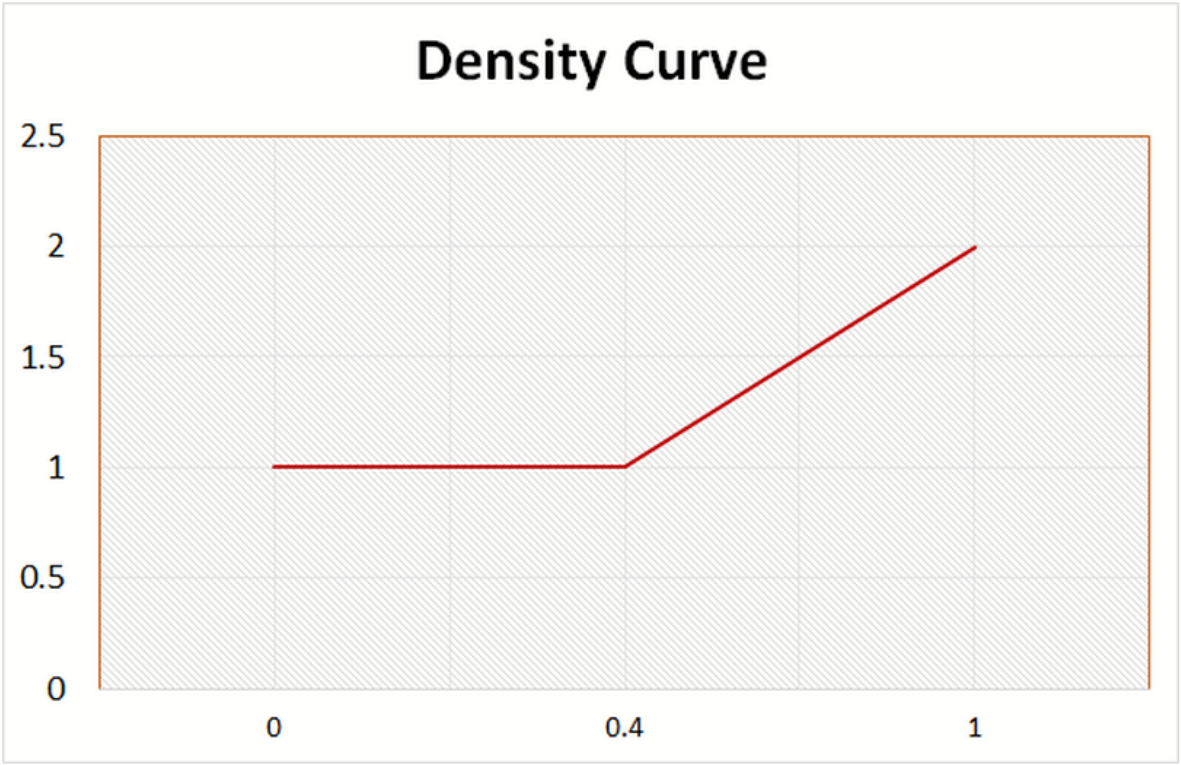


$P(0.4 < X < 0.8) = 0.4 + 0.2 = 0.6$

$P(0.4 \leq X \leq 0.8) = 0.6$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What percent of observations are equal to 0.4?



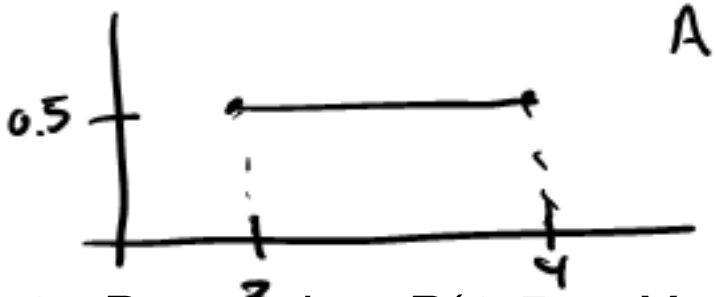
$P(X = 0.4) = 0$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

Suppose there is a line segment that goes from (2, 0.5) to (4, 0.5).

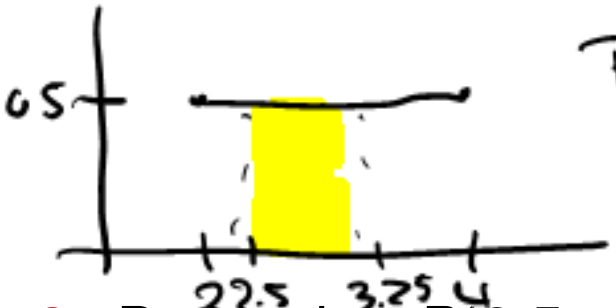
1. Does this meet the requirements of a probability distribution?



Area between 2 and 4

$A = bh = (4-2)(0.5) = 2(0.5) = 1$ ✓

2. Determine $P(2.5 \leq X \leq 3.25)$.



$P(2.5 \leq x \leq 3.25) = bh$

$= (3.25 - 2.5)(0.5)$

$= 0.75(0.5)$

$= 0.375$

3. Determine $P(2.5 < X < 3.25)$.

$P(2.5 < x < 3.25) = 0.375$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example of a density curve: Uniform distribution

- The **Uniform Distribution** describes a variable that takes values that are uniformly spread between a range of values.
- Thus it takes on a rectangular shape. The proportion (percent) of observations that lie within a range of values is equivalent to the area of the rectangle between the desired range of values.

$$Area = Height \times Width$$

The height of the rectangle is $\frac{1}{\text{highest value} - \text{lowest value}}$.

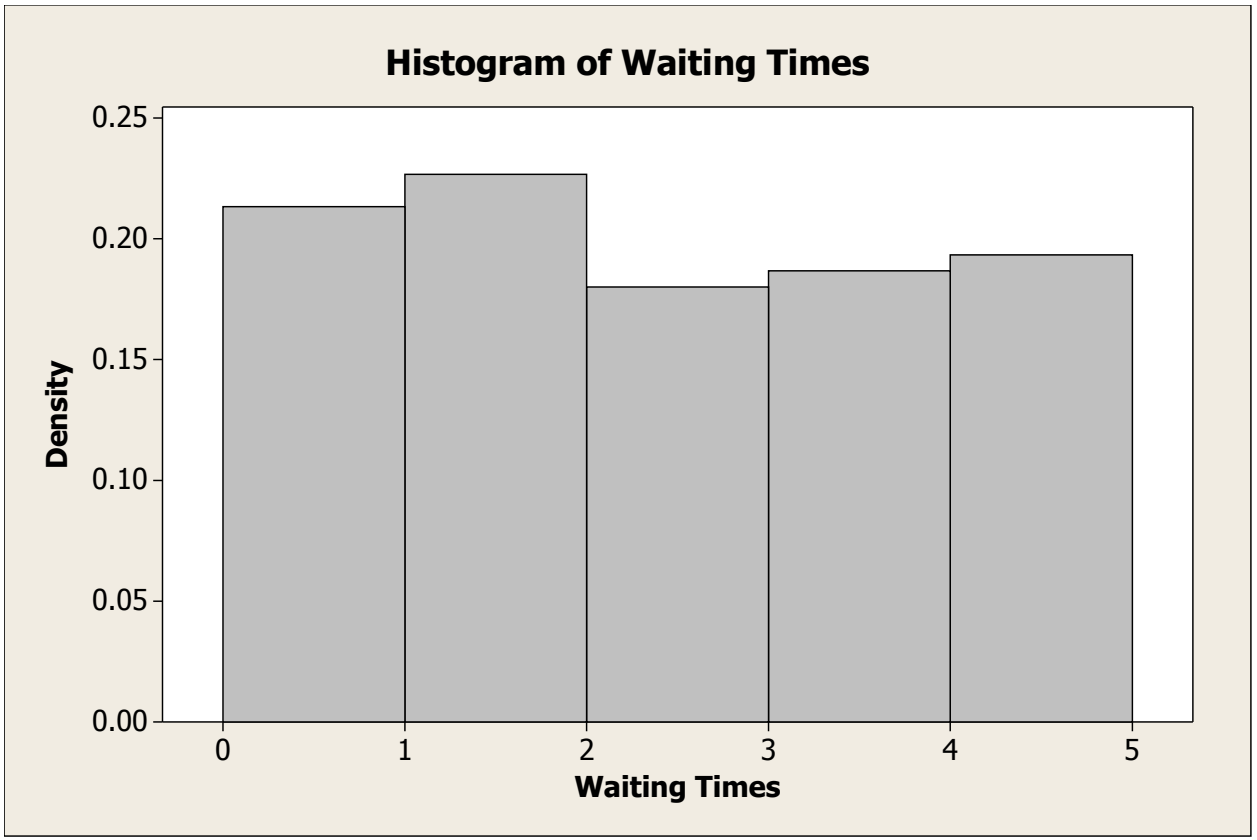
Previous example: max = 4 min = 2

$$\text{height} = \frac{1}{4-2} = \frac{1}{2} = 0.5$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Uniform distribution

The following histogram is of waiting times for an elevator where the longest waiting time is 5 minutes.



Uniform distribution example

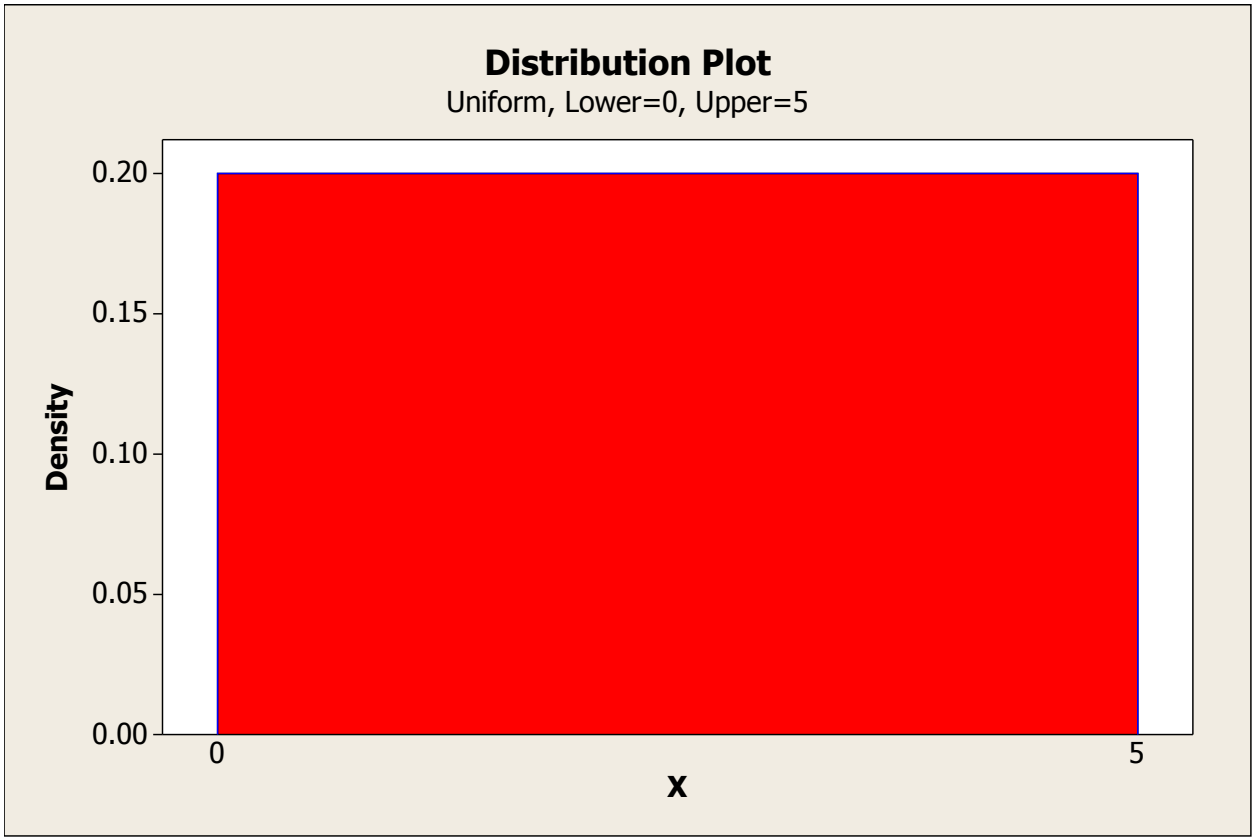
- The random variable X is the waiting time for the elevator.
- The possible values are $0 \leq X \leq 5$.
- Since we are looking at a random variable that assumes values corresponding to an interval this is a **continuous** random variable.
- The probabilities for this random variable is the same as the area under a density curve.
- In this example any one of the times has an equally likely chance of assuming a value between 0 and 5. Thus this curve is rectangular.

$$\text{height} = \frac{1}{\text{max} - \text{min}} = \frac{1}{5 - 0} = \frac{1}{5} = 0.2$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Density curve for waiting time

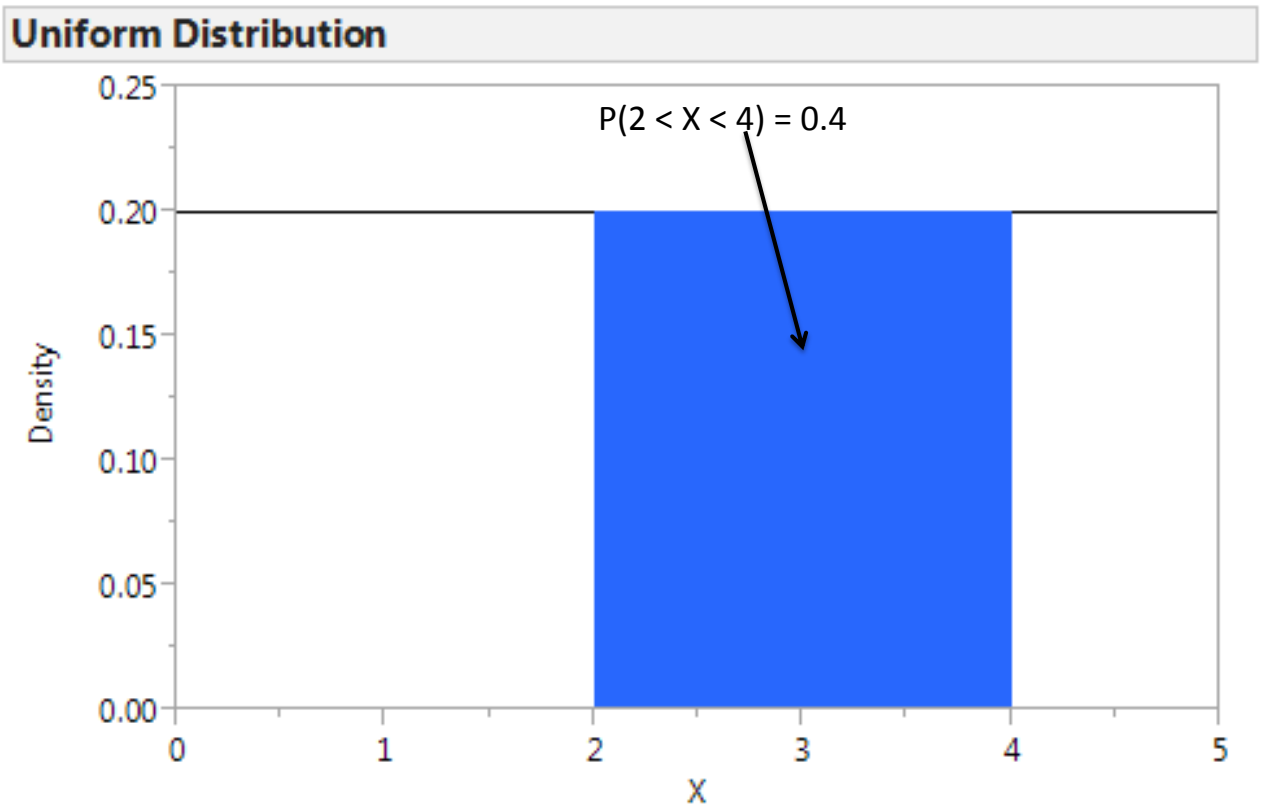
The rectangle ranges between 0 and 5. The height of the rectangle is:
$\frac{1}{\text{highest value} - \text{lowest value}} = \frac{1}{5-0} = 0.2.$



$P(2 < X < 4)$

The probability of any event between a range of values is the same as the area between the range under the density curve.

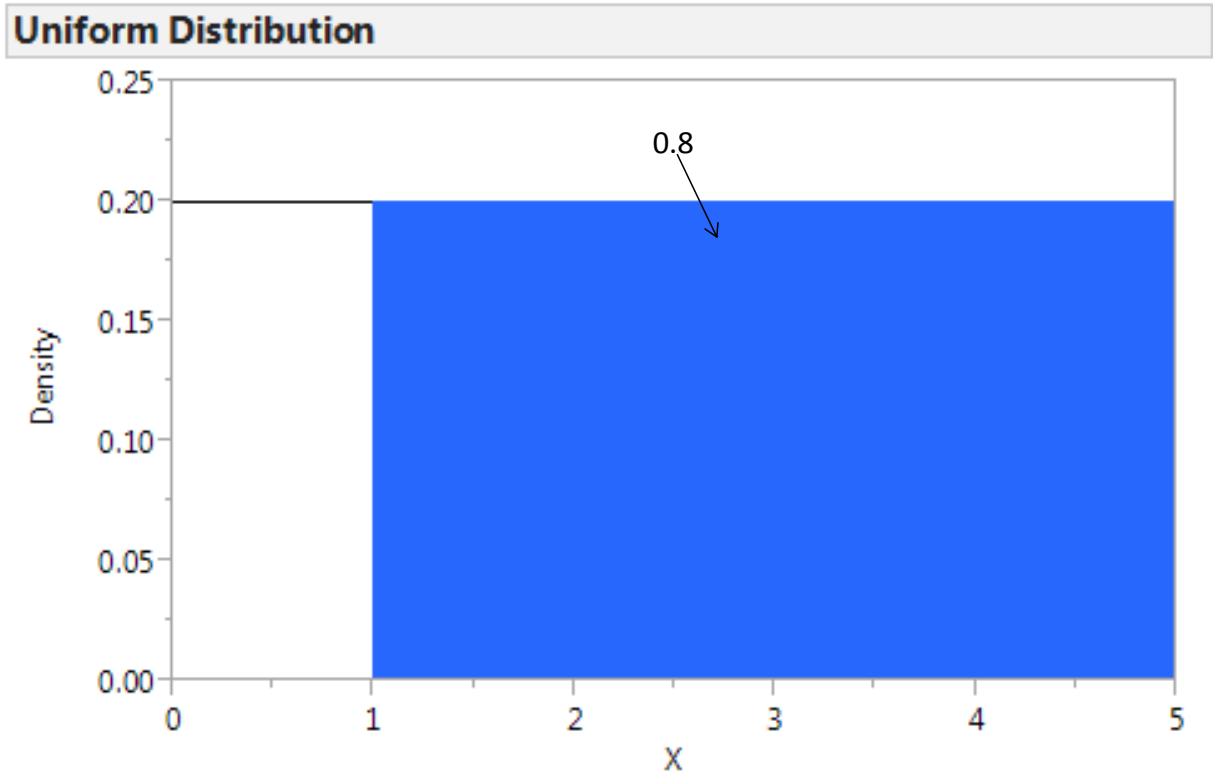
$$\text{Area of rectangle} = \text{height} \times \text{width} = 0.2 \times 2 = 0.4$$



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example continued

What is the probability that a person waits for at least one minute?



$P(X \geq 1) = 0.2\,(5-1) = 0.2(4) = 0.8$

$= 1 - P(0 \leq X \leq 1) = 1 - 0.2(1) = 0.8$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example continued

What is the probability that a person waits for at least one minute?

$$\begin{aligned} P(X \geq 1) &= \text{area above } 1 \\ &= \text{height} \times \text{width} \\ &= 0.2 \times 4 \\ &= 0.8 \end{aligned}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Mean and standard deviation

- The mean of a random variable that has a uniform distribution is:

$$\mu = \frac{\text{highest value} + \text{lowest value}}{2}$$

- The standard deviation of a random variable that has a uniform distribution is:

$$\sigma = \sqrt{\frac{(\text{highest value} - \text{lowest value})^2}{12}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Waiting time

- The expected waiting time for this elevator is:

$$\mu = \frac{5+0}{2} = 2.5$$

- The standard deviation for the waiting time for this elevator is:

$$\sigma = \sqrt{\frac{(5-0)^2}{12}} = \sqrt{\frac{25}{12}} = \sqrt{2.08333} = 1.4434$$

- We expect the waiting time to be 2.5 minutes give or take 1.4434 minutes or so.
- What is the median? 2.5

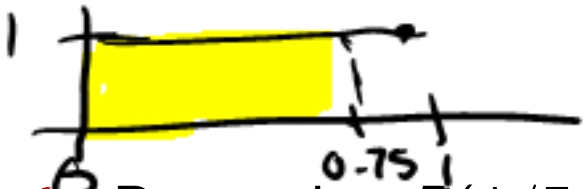
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

Consider a spinner that, after a spin, will point to a number between zero and 1 with “uniform probability."

1. What is the probability that the spinner will land on something less than 0.75?

$\text{height} = \frac{1}{\text{max-min}} = \frac{1}{1-0} = 1$



$P(X < 0.75) = 0.75(1) = 0.75$

2. Determine $P(1/5 \leq X \leq 3/8)$.



$= P(0.2 \leq X \leq 0.375)$

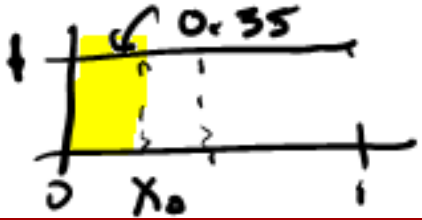
$= (0.375 - 0.2)(1)$

$= 0.175$

3. What is the value that the spinner is expected to land on?

$\text{mean} = E(X) = \frac{1+0}{2} = \frac{1}{2} = 0.5$

4. Determine the value of X_0 such that $P(X \leq X_0) < 0.35$.



$0.35 = X_0(1)$

$X_0 = 0.35$

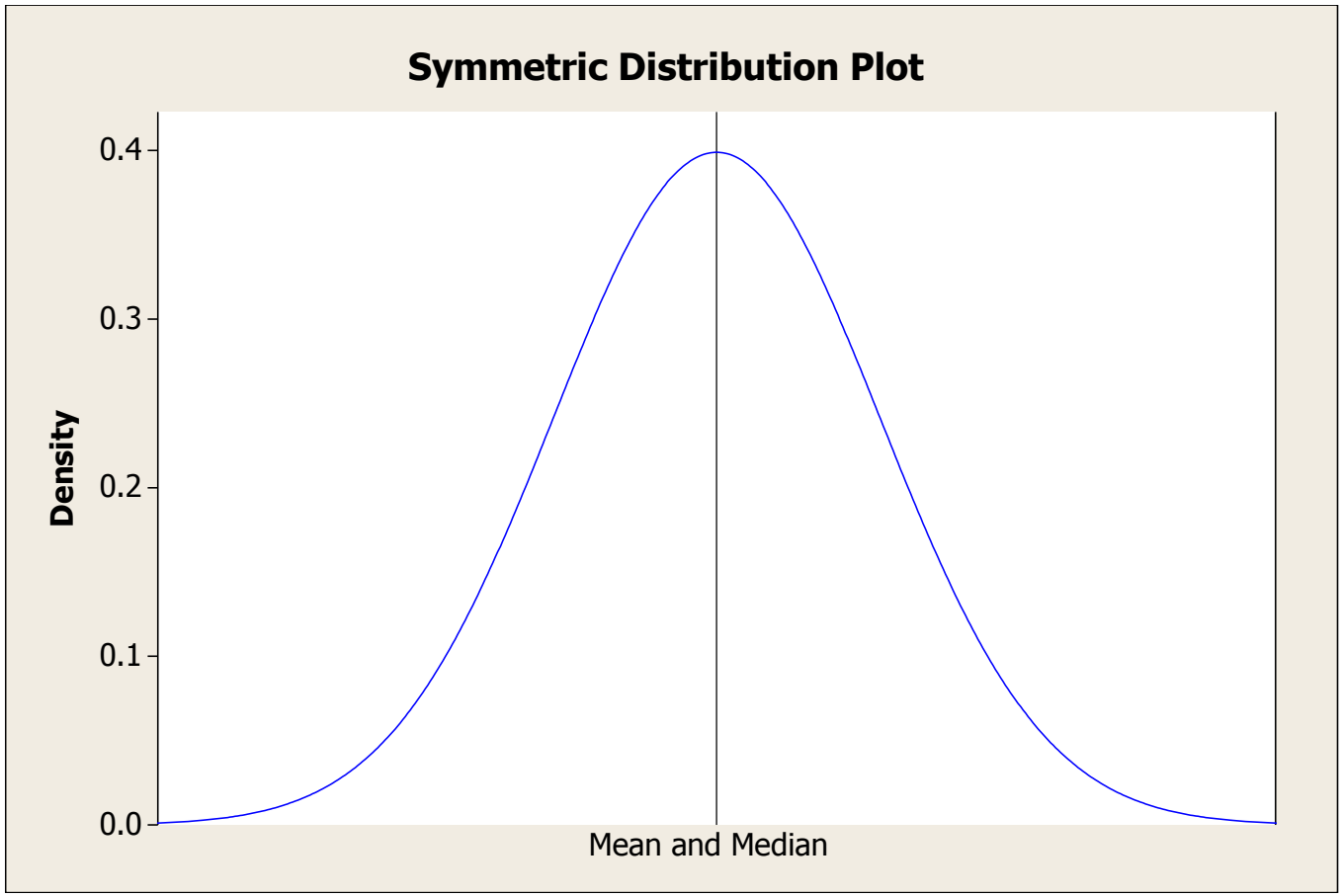
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Median and Mean of a Density Curve

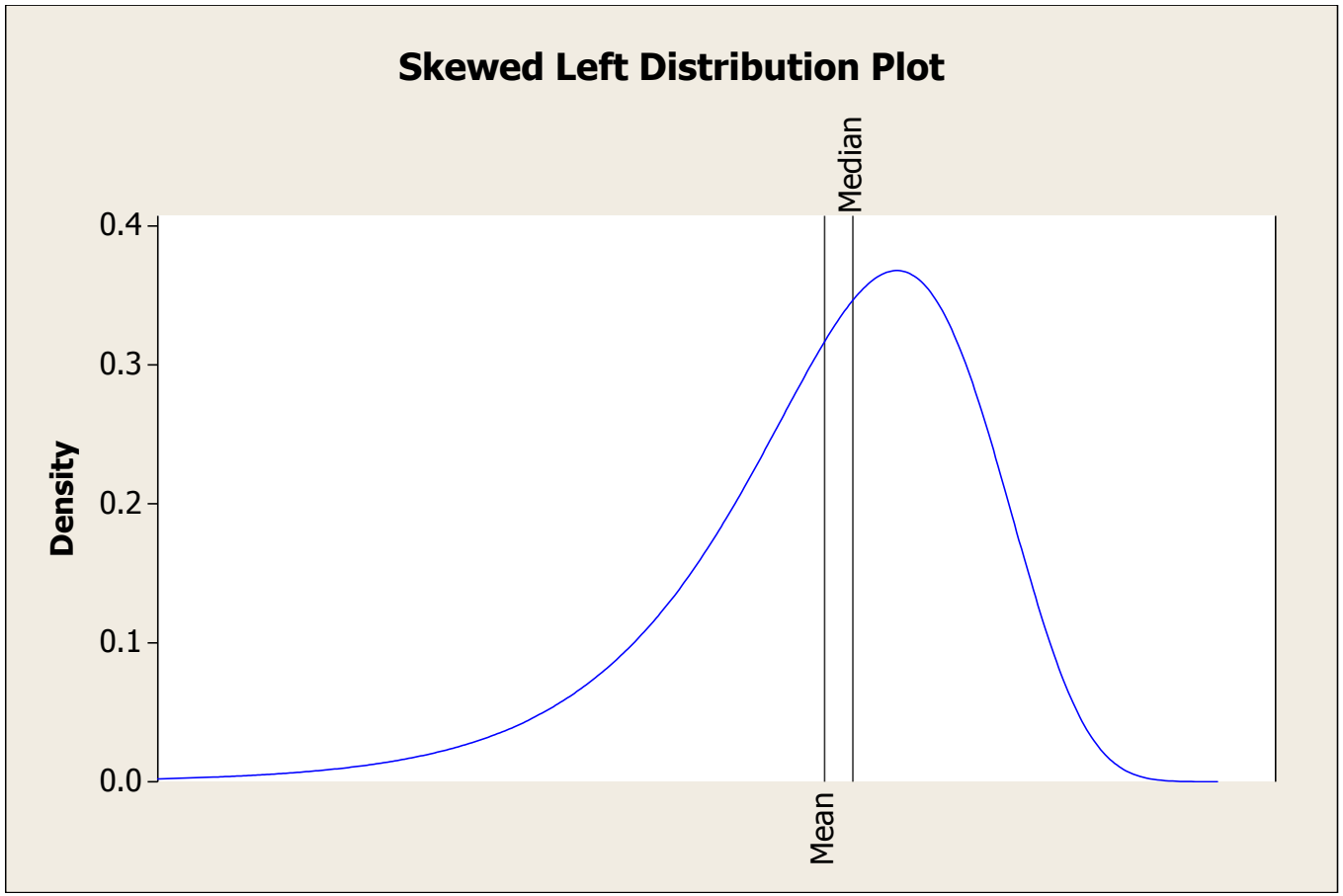
- The median of a density curve is the equal-areas point, the point that divides the area under the curve in half.
- The mean of a density curve is the balance point.
- The mean and the median is the same for symmetric density curve.
- The mean of a skewed curve is pulled away from the median in the direction of the long tail.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

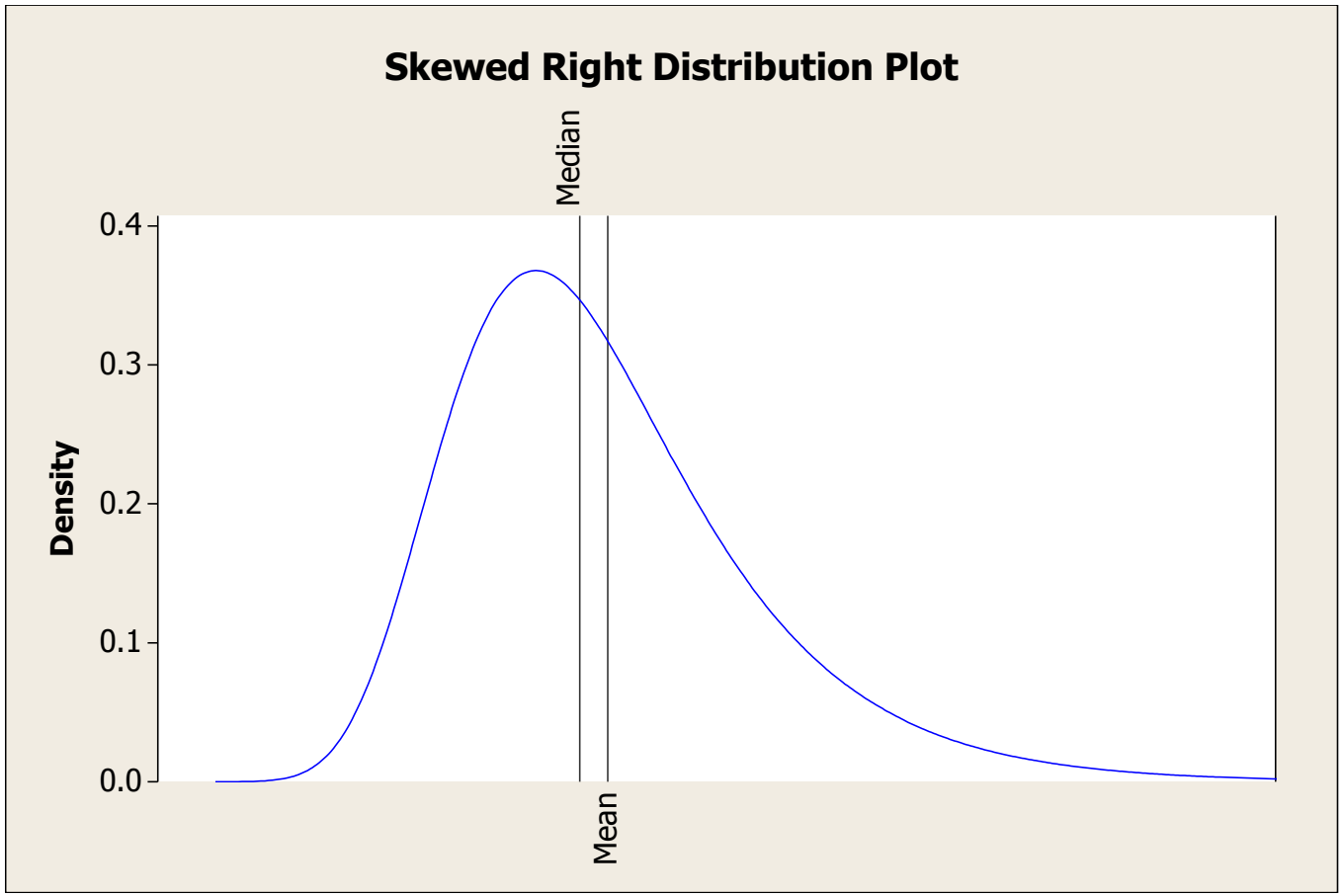
Symmetric Density Curve



Skewed Left Density Curve



Skewed Right Density Curve



Mean and Standard Deviation of a Density Curve

- Density curve is an idealized description of the distribution of data.
- $\bar{x}$ and s are computed from actual observations.
- Mean of an idealized distribution is μ called “mu”.
- Standard deviation of a density curve is σ called “sigma”.

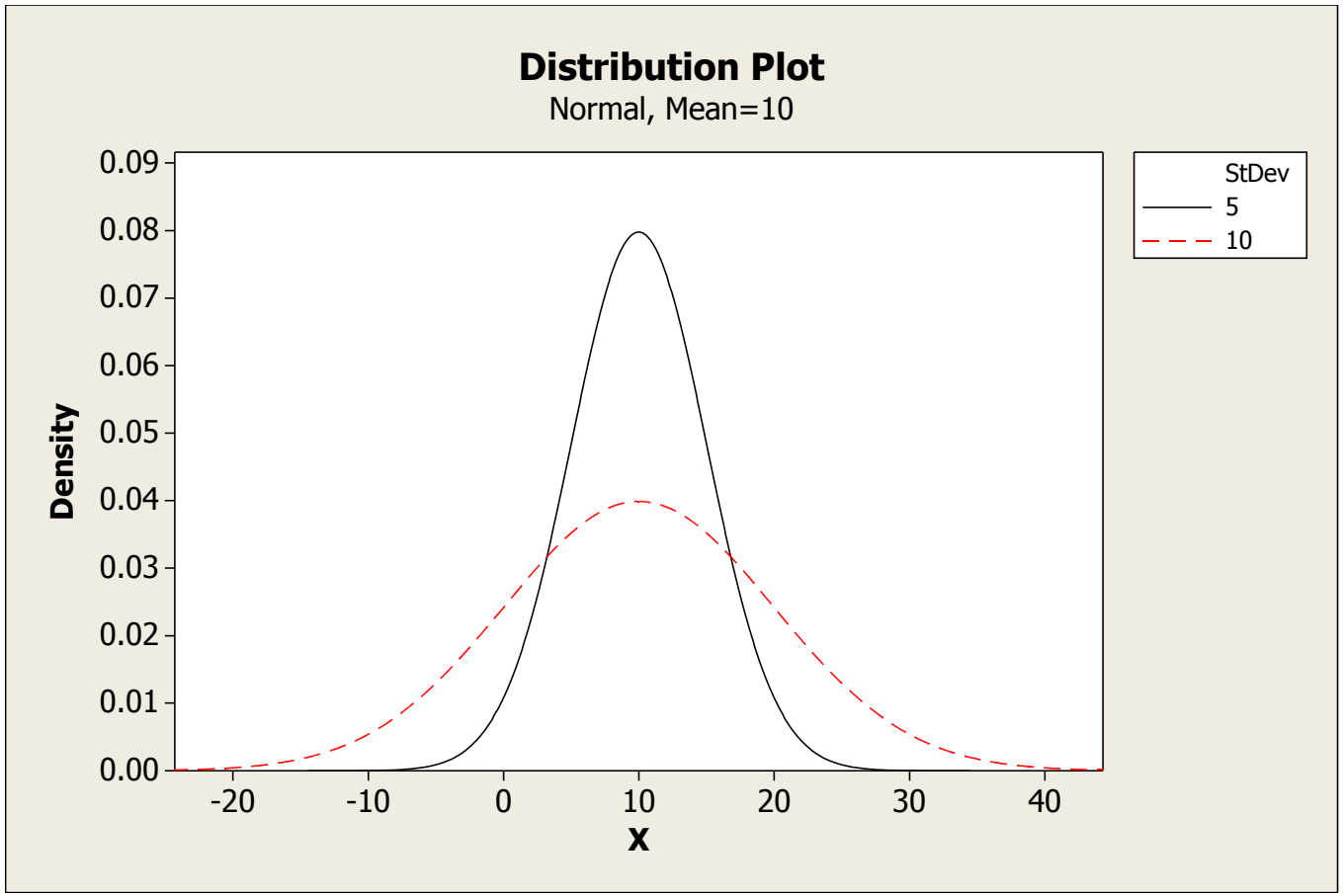
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Normal distributions

- Common type of probability distributions for continuous random variables.
- The highest probability is where the values are centered around the mean. Then the probability declines the further from the mean a value gets.
- These curves are symmetric, single-peaked, and bell-shaped.
- The mean μ is located at the center of the curve and is the same as the median.
- The standard deviation σ controls the spread of the curve.
- If σ is small then the curve is tall and slim.
- If σ is large then the curve is short and fat.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Normal Density Curves



Normal distributions important to statistics?

- Normal distributions are good descriptions for some distributions of real data.
- Normal distributions are good approximations to the results of many kinds of chance outcomes.
- Many statistical inference procedures based on Normal distributions work well for other roughly symmetric distributions.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Facts about the Normal distribution

- The curve is symmetric about the mean. That is, 50% of the area under the curve is below the mean. 50% of the area under the curve is above the mean.
- The spread of the curve is determined by the standard deviation.
- The area under the curve is with respect to the number of standard deviations a value is from the mean.
- Total area under the curve is 1.
- Area under the curve is the same a probability within a range of values.

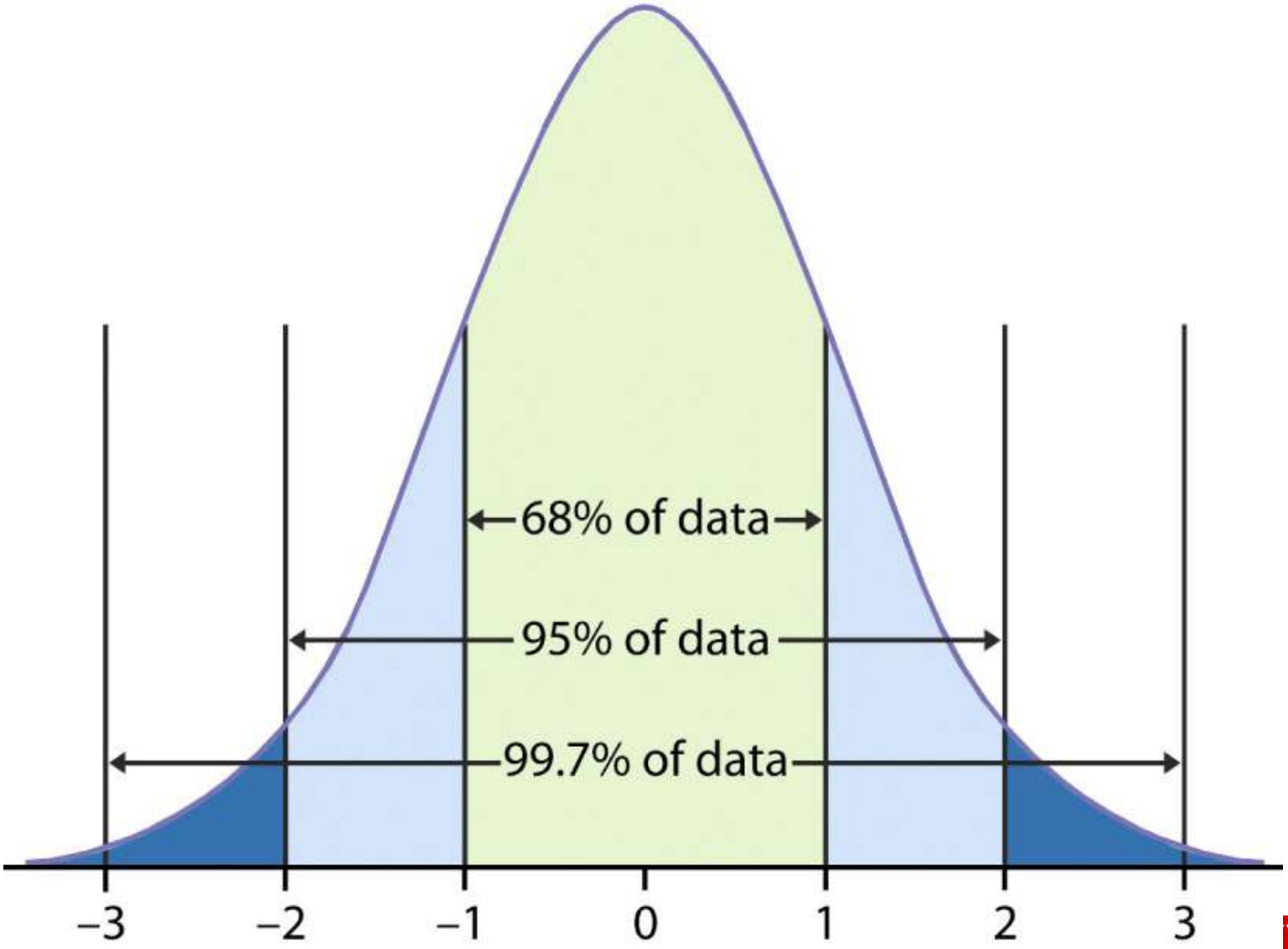
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Empirical Rule or 68-95-99.7 Rule

- Unfortunately to find the area under this density curve is not as easy to compute. Thus we can use the following approximate rule for the area under the **Normal** density curve.
- In the Normal Distribution with mean μ and standard deviation σ:
  - 68% of the observations fall within 1 standard deviation σ of the mean μ.
  - 95% of the observations fall within 2 standard deviations 2σ of the mean μ.
  - 99.7% of the observations fall within 3 standard deviations 3σ of the mean μ.

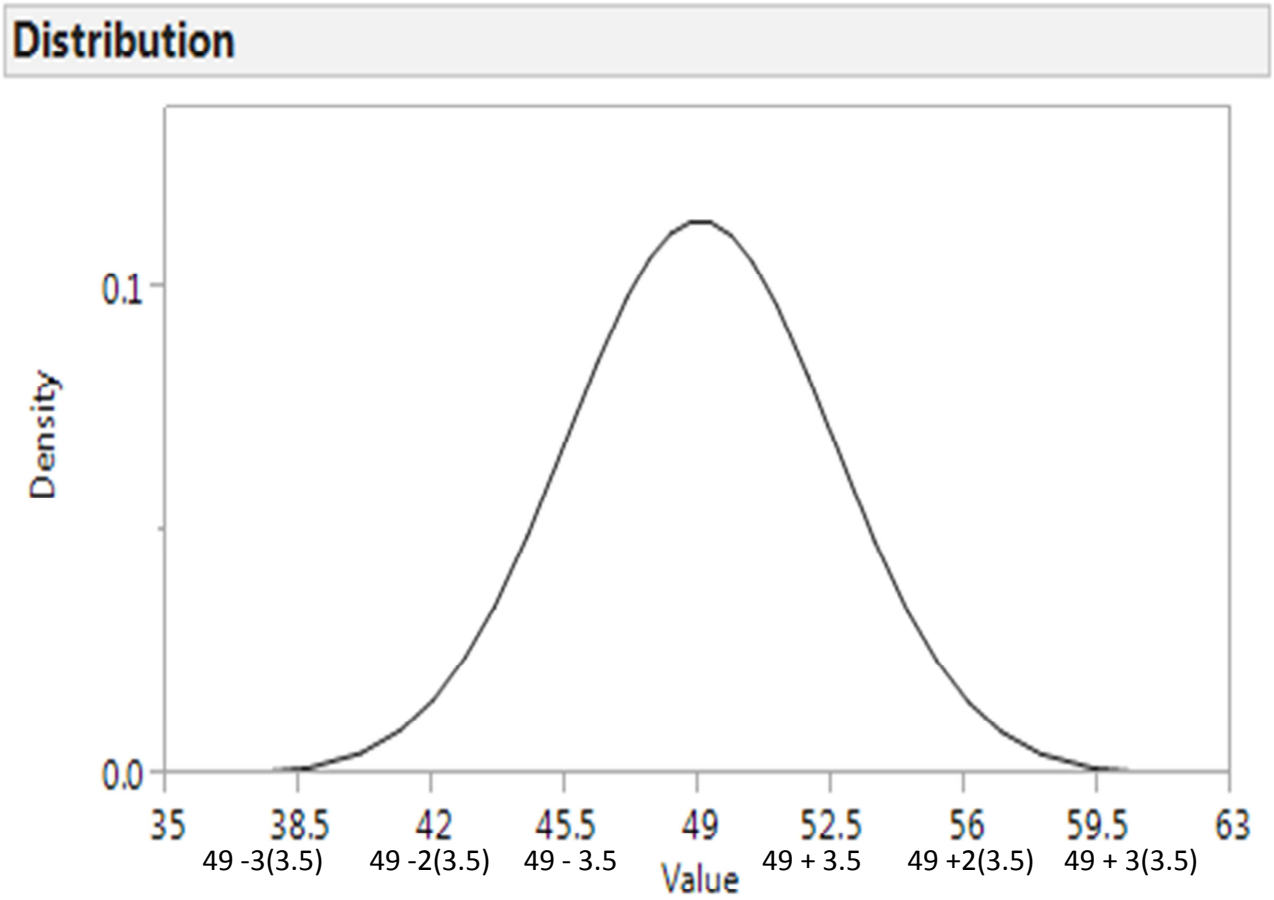
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The 68-95-99.7 rule for Normal distributions



MPG of Prius

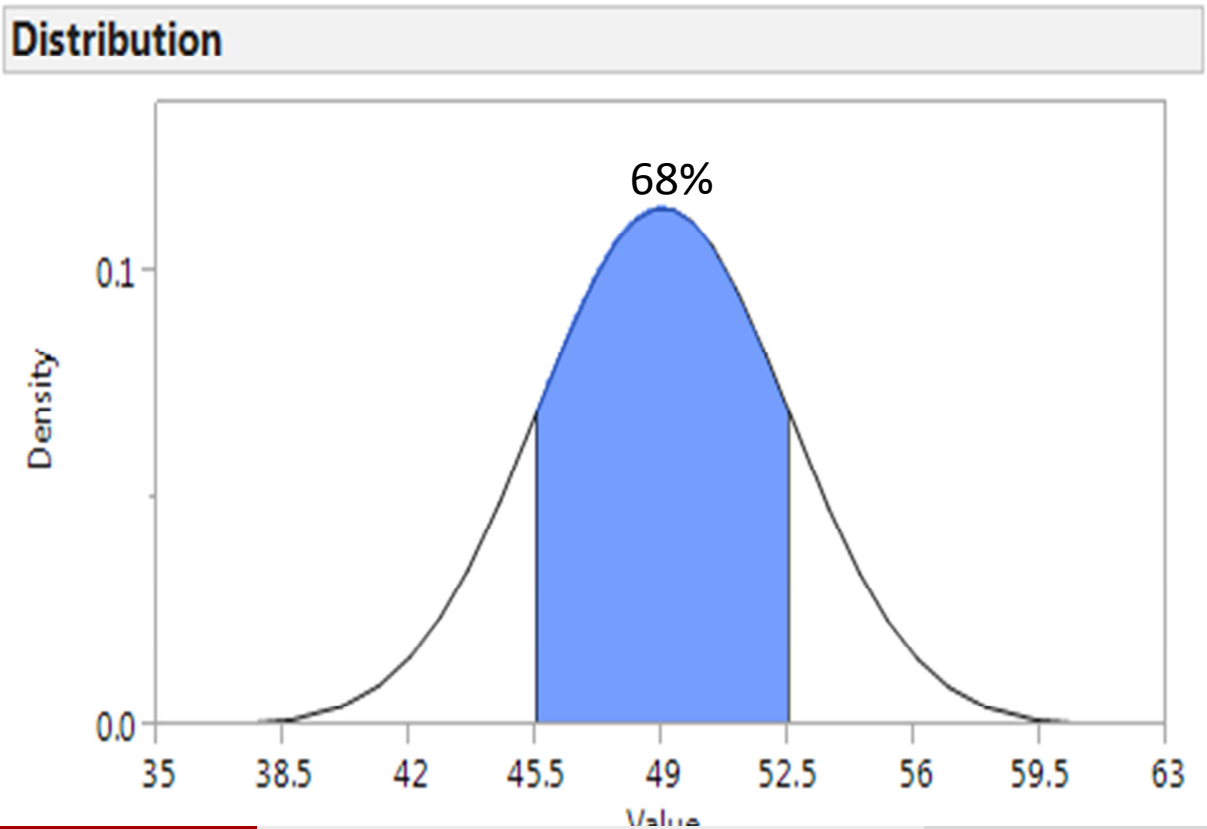
The MPG of Prius has a Normal distribution with mean $\mu = 49$ mpg and standard deviation $\sigma = 3.5$ mpg.



VERSITY of HOUSTON
EPARTMENT OF MATHEMATICS

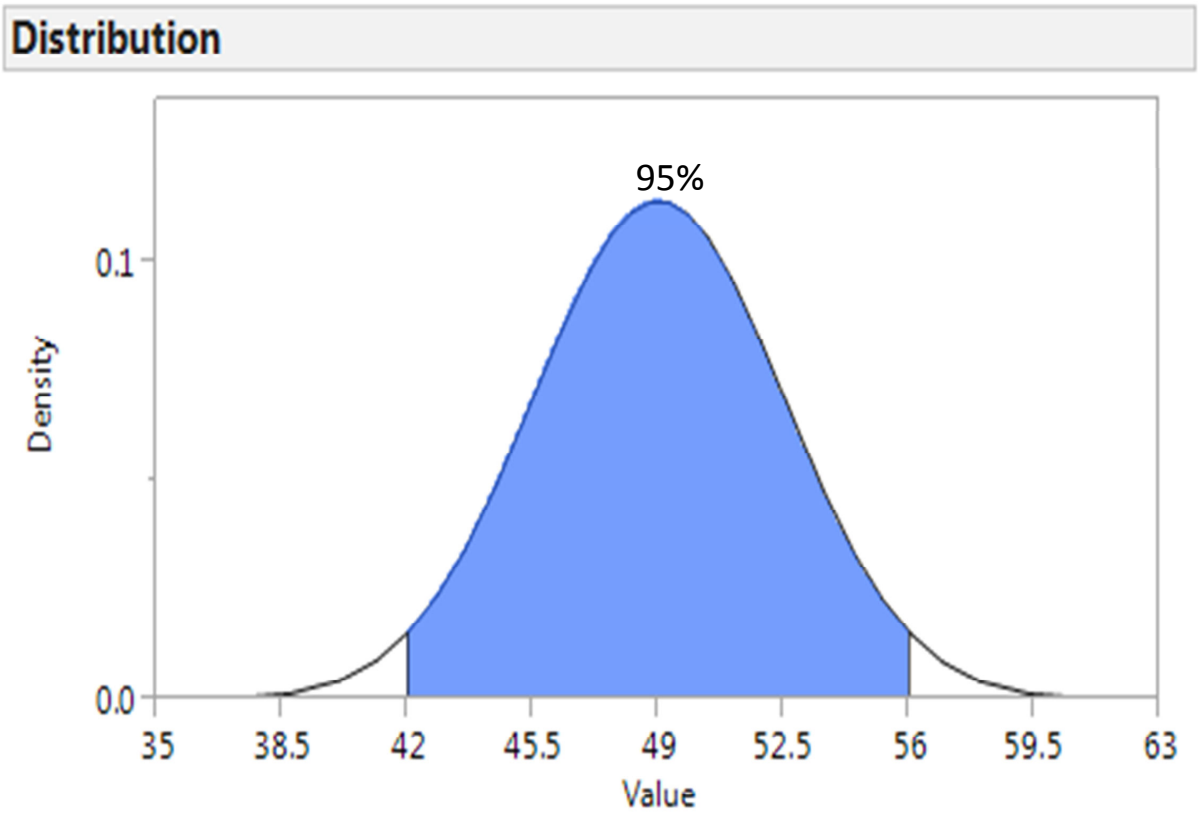
MPG of Prius

The percent within one standard deviation is 68%. That is about 68% of the Prius cars have a mpg between 45.5 and 52.5.



MPG of Prius

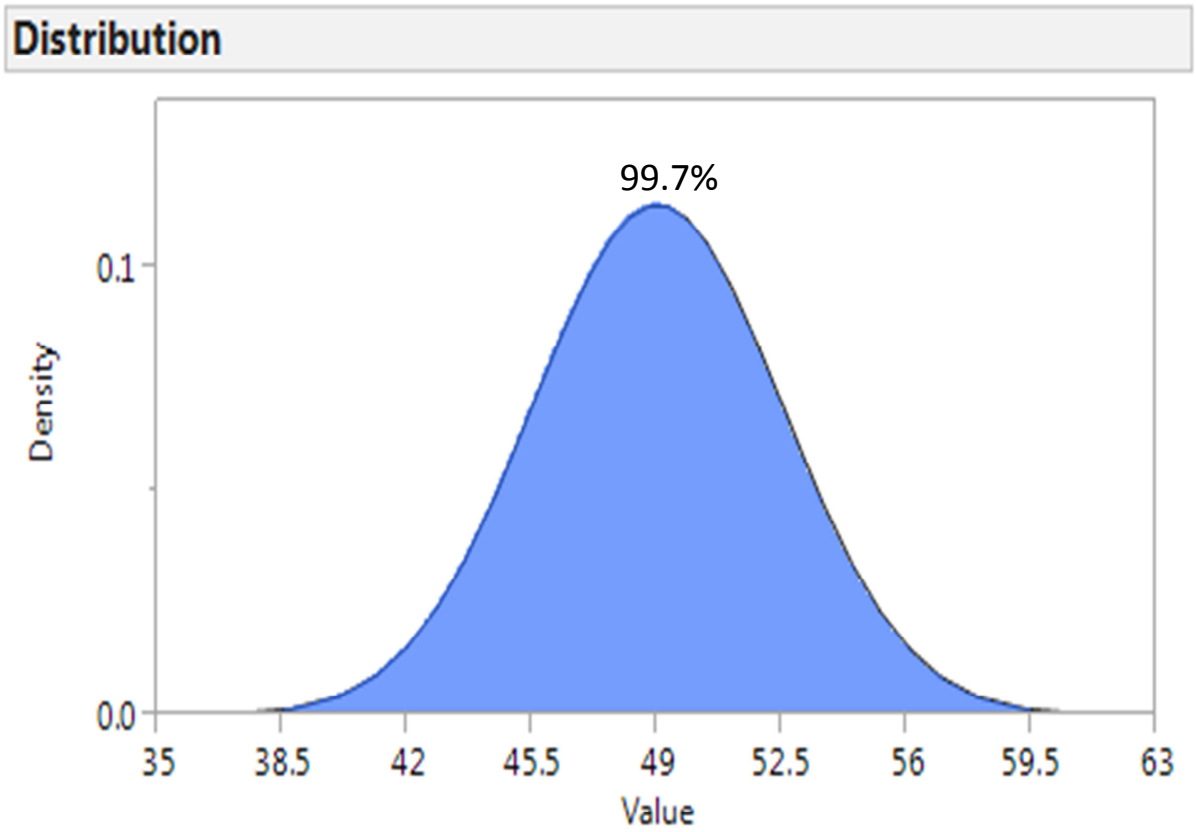
The percent within two standard deviations is 95%. That is about 95% of the Prius cars have a mpg between 42 and 56.



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

MPG of Prius

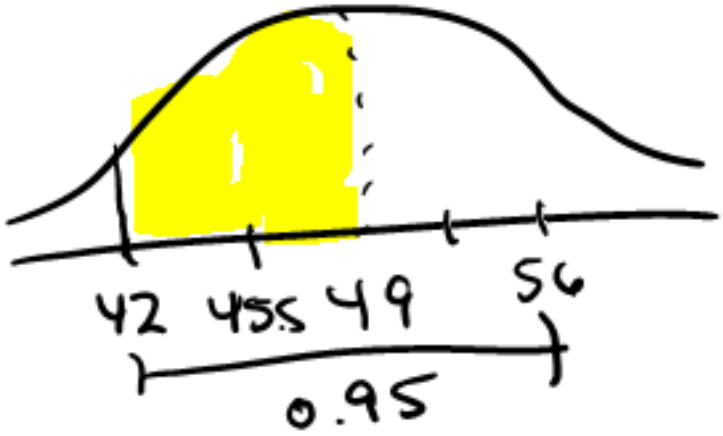
About 99.7% of the Prius cars have a mpg between 38.5 and 59.5. This is three standard deviations from the mean.



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

$\mu = 49 \quad \sigma = 3.5$

$P(42 < x < 49) = \frac{1}{2}(0.95) = 0.475$



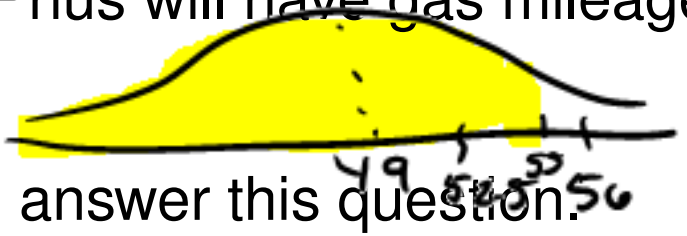
$P(X > 56) = \frac{1}{2}(1 - 0.95) = 0.025$



Not 1, 2 or 3 Standard Deviations

The MPG of Prius has a Normal distribution with mean $\mu = 49$ mpg and standard deviation $\sigma = 3.5$ mpg.

- What is the probability that a Prius will have gas mileage of less than 55 mpg? $P(X < 55)$.



- There are a couple of ways to answer this question.
  - R: pnorm(x,mean,sd), pnorm(55,49,3.5) = 0.9568
  - TI 83 or 84: normcdf(lowest limit, upper limit, mean, sd), normcdf(-1e99,55,49,3.5) = 0.9568
  - Table A: In the appendix of your book. This table is for Standard Normal Distribution.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Determine the probabilities?

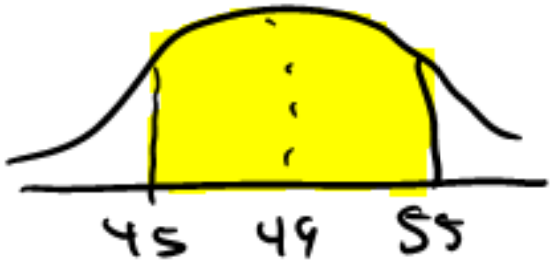
X = mpg of Prius with $N(49, 3.5)$. (mean = 49, SD = 3.5)

1. What is the probability that the gas mileage is more than 55 mpg?



$$P(X > 55) = 1 - P(X \le 55)$$
$$= 1 - \text{pnorm}(55, 49, 3.5)$$
$$= 1 - 0.9568 = 0.0432$$

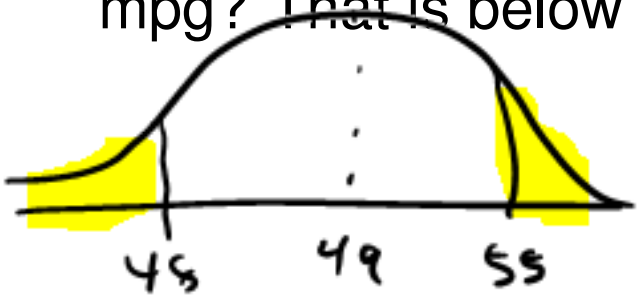
2. What is the probability that the gas mileage is between 45 and 55 mpg?



$P(45 < X < 55) =$ pnor

```
> pnorm(55,49,3.5)-pnorm(45,49,3.5)
[1] 0.8302129
```

3. What is the probabiity that the gas mileage is beyond 45 and 55 mpg? That is below 45 or above 55 mpg.



$$P(X < 45 \text{ or } X > 55)$$
$$= 1 - P(45 < X < 55)$$
$$= 1 - 0.8302 = 0.1698$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS