Modeling Continuous Variables

Lecture 19

Sections 6.1 - 6.3.1

Robb T. Koether

Hampden-Sydney College

Tue, Sep 30, 2008

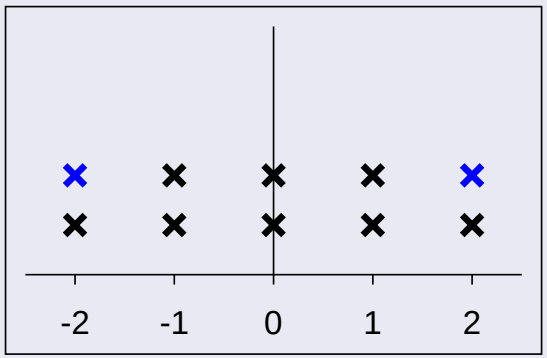
Outline

Homework Review

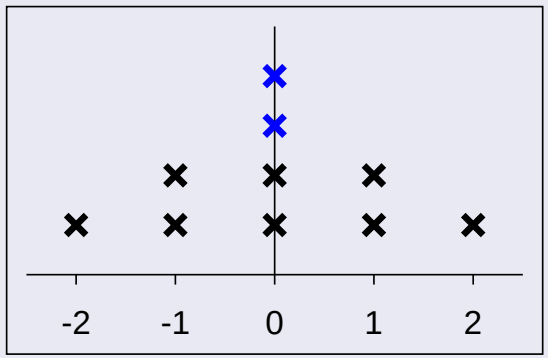
Exercise 5.20, page 336.

Consider the following two frequency plots based on two sets of data with 10 observations:

Frequency Plot 1



Frequency Plot 2



Note that Plot 1 can be converted to Plot 2 by moving the two blue X's.

Homework Review

Exercise 5.20, page 336.

(a) Do you think the range for Plot 1 is smaller than, equal to, or larger than the range for Plot 2?

(b) Do you think the mean for Plot 1 is smaller than, equal to, or larger than the mean for Plot 2?

(c) Do you think the standard deviation for Plot 1 is smaller than, equal to, or larger than the standard deviation for Plot 2? Explain.

(d) Check your answers to parts (a), (b), and (c) by computing the range, mean, and standard deviation for each set of data.

Homework Review

Solution

(a) The ranges are clearly equal.

(b) Both distributions are symmetric, so the means are in the center, at 0. Thus, they are equal.

(c) The two blue points in Plot 2 are closer to the mean than they are in Plot 1, so they have smaller deviations. All other points are the same between the two distributions. Thus, Plot 2 has a smaller standard deviation.

Homework Review

Solution

(d)

- For Plot 1,

$$\begin{aligned}
\text{range} &= 2-(-2)=4.\\
\bar{x} &= \frac{0}{10}=0.\\
\text{SSX} &= 4+4+1+1+0+0+1+1+4+4\\
&= 20.\\
s &= \sqrt{\frac{20}{9}}\\
&= 1.491.
\end{aligned}$$

Homework Review

Solution

(d)

- For Plot 2,

$$\begin{aligned}
\text{range} &= 2-(-2)=4. \\
\bar{x} &= \frac{0}{10}=0. \\
\text{SSX} &= 4+1+1+0+0+0+0+1+1+4 \\
&= 12. \\
s &= \sqrt{\frac{12}{9}} \\
&= 1.155.
\end{aligned}$$

Models

Definition (Mathematical model)

A mathematical model is a mathematical abstraction and, therefore, a simplification of a real situation, one that retains the essential features.

- Real situations are usually much too complicated to deal with in all their details.

Example

- The “bell curve” is a model (an abstraction) of many populations.
- Real populations have all sorts of bumps and twists and irregularities.
- The bell curve is smooth and perfectly symmetric.
- In statistics, the bell curve is called the normal curve, or normal distribution.

Models

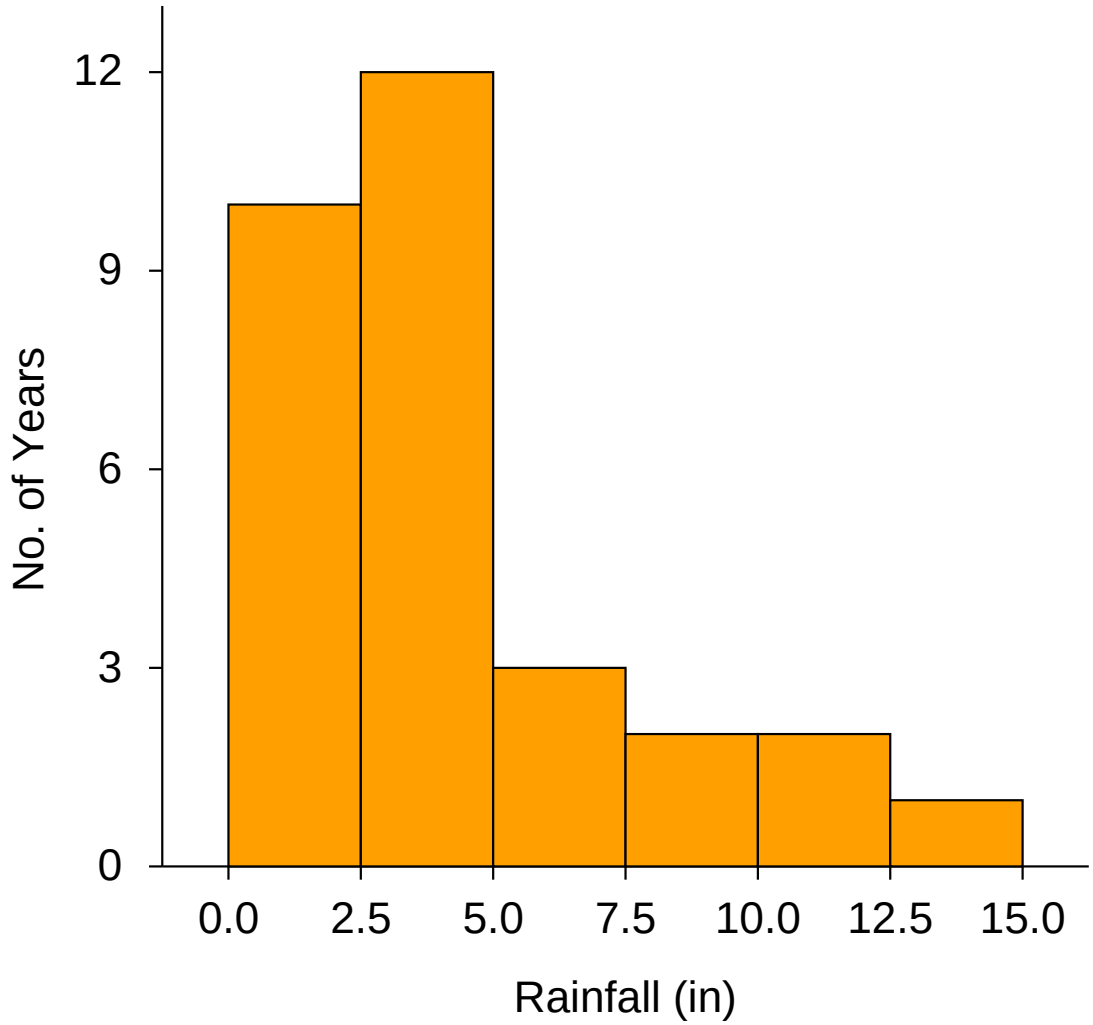
- Our models will be models of distributions, presented either as histograms or as continuous distributions.

Histograms and Area

- In a histogram, frequency is represented by area.
- Consider the following distribution of rainfall data.

Rainfall (in)	Frequency
0.00 - 2.49	10
2.50 - 4.99	12
5.00 - 7.49	3
7.50 - 9.99	2
10.00 - 12.49	2
12.50 - 14.99	1

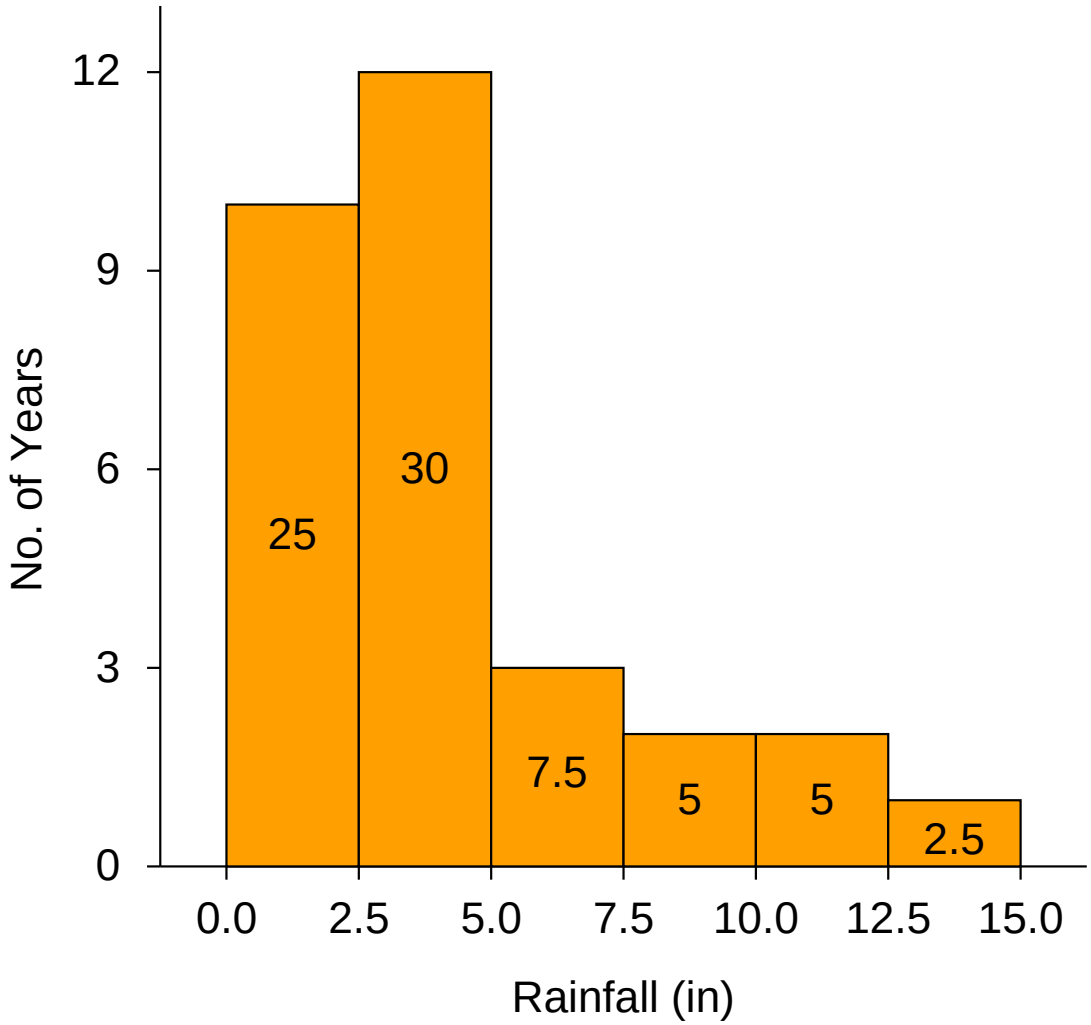
Histograms and Area



Histograms and Area

- What is the total area of this histogram?

Histograms and Area



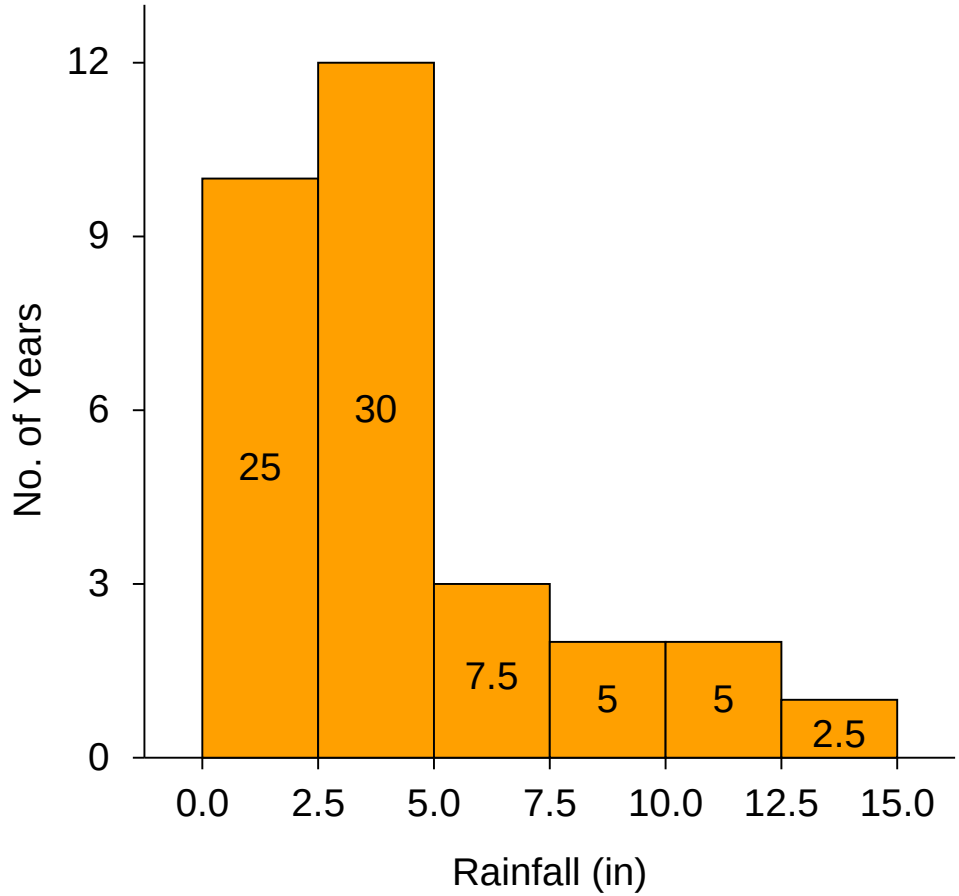
Histograms and Area

- The total area is

$$25 + 30 + 7.5 + 5 + 5 + 2.5 = 75.$$

Histograms and Probability

- If we select one year at random, what is the probability that September rainfall was between 5 inches and 10 inches?



Histograms and Probability

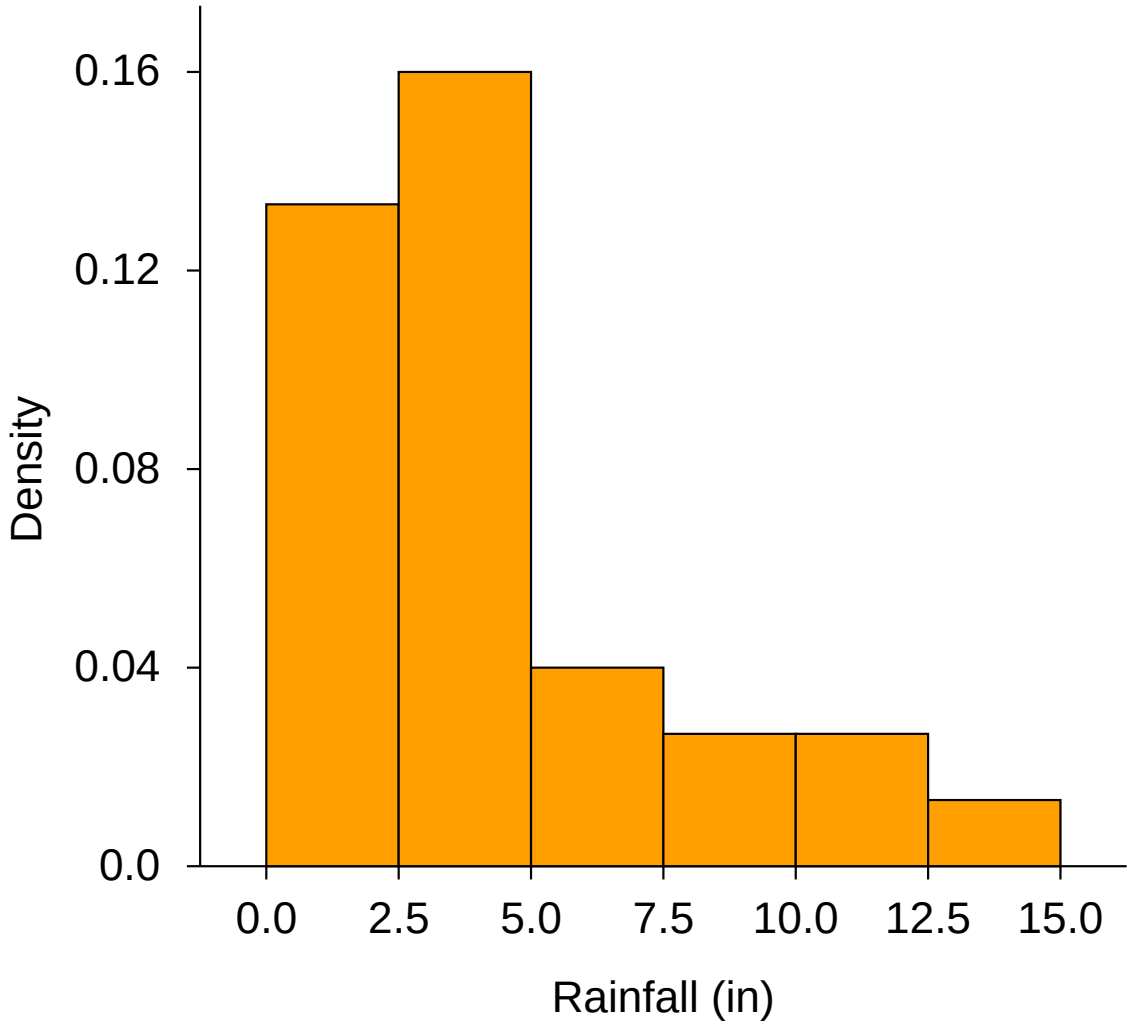
- We will rescale the vertical axis so that the total area equals 1, representing 100%.
- This new scale is called the density.

Histograms and Area

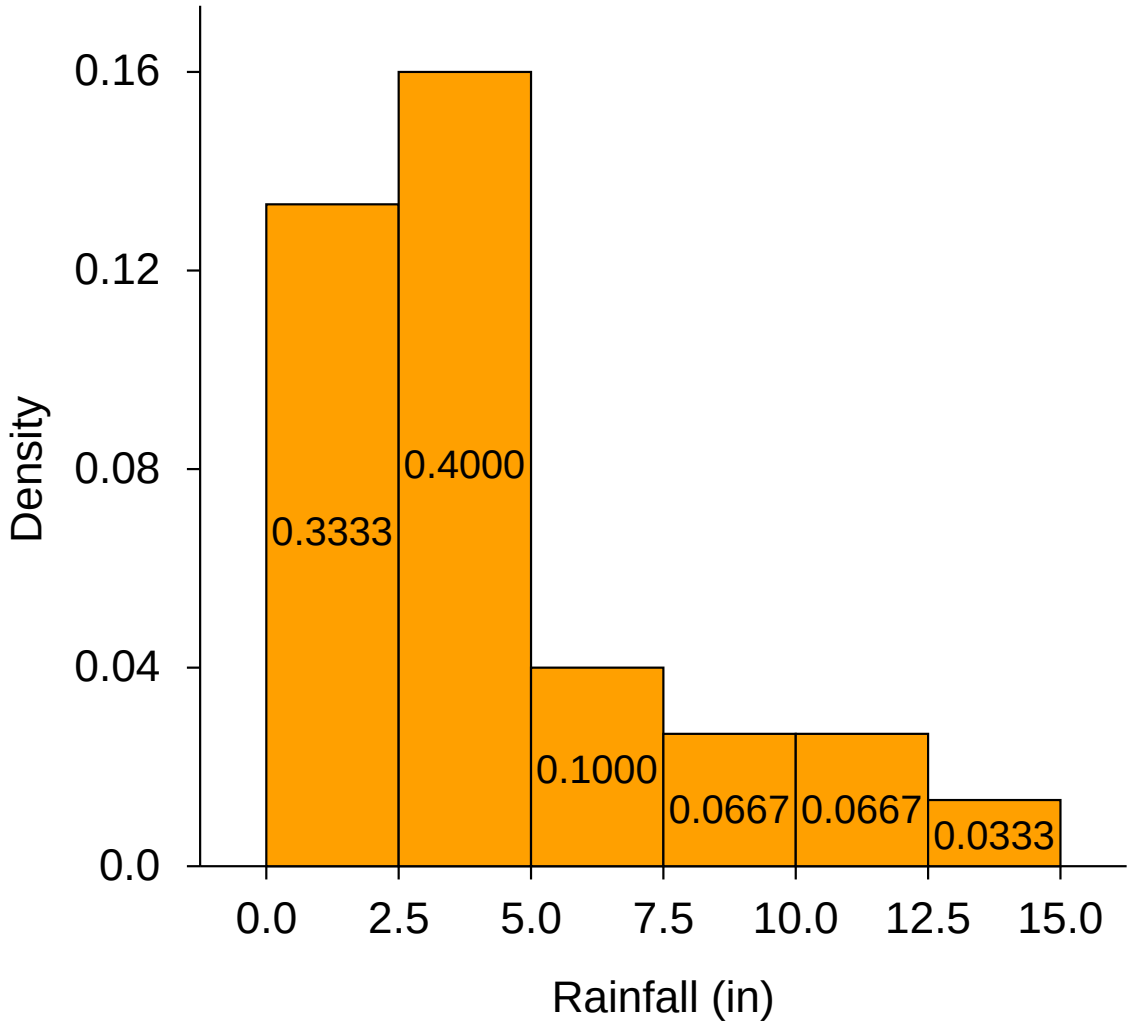
- Divide the frequencies by the original area to get the density.

Rainfall (in)	Frequency	Density
0.00 - 2.49	10	0.1333
2.50 - 4.99	12	0.1600
5.00 - 7.49	3	0.0400
7.50 - 9.99	2	0.0267
10.00 - 12.49	2	0.0267
12.50 - 14.99	1	0.0133

Histograms and Area

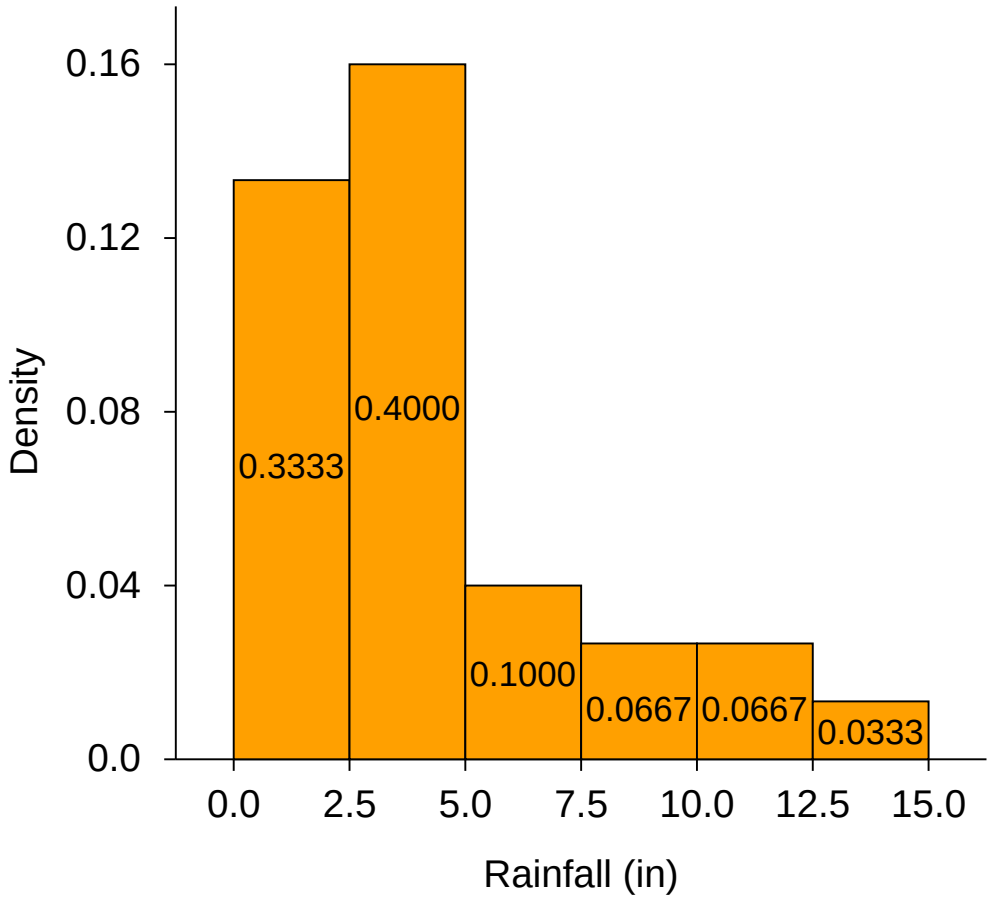


Histograms and Area



Histograms and Area

- If we select one year at random, what is the probability that September rainfall was between 5 inches and 10 inches?



Density Functions

AREA = PROPORTION = PROBABILITY

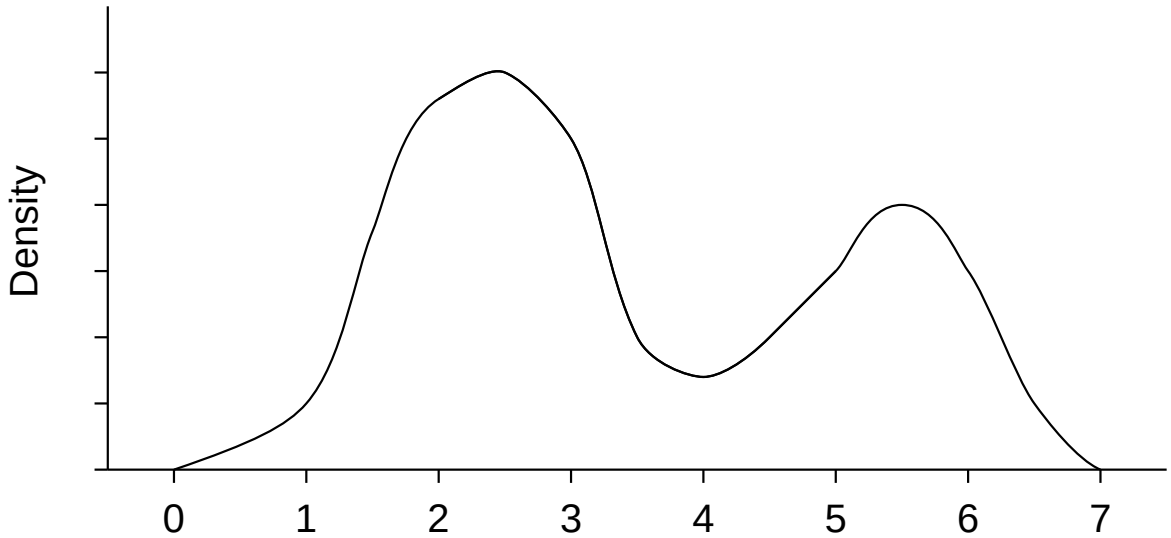
- This is the fundamental property that connects the graph of a continuous model to the population that it represents, namely:
  - The area under the graph between two points on the x-axis represents the proportion of the population that lies between those two points.

Density Functions

- Consider an arbitrary continuous distribution.

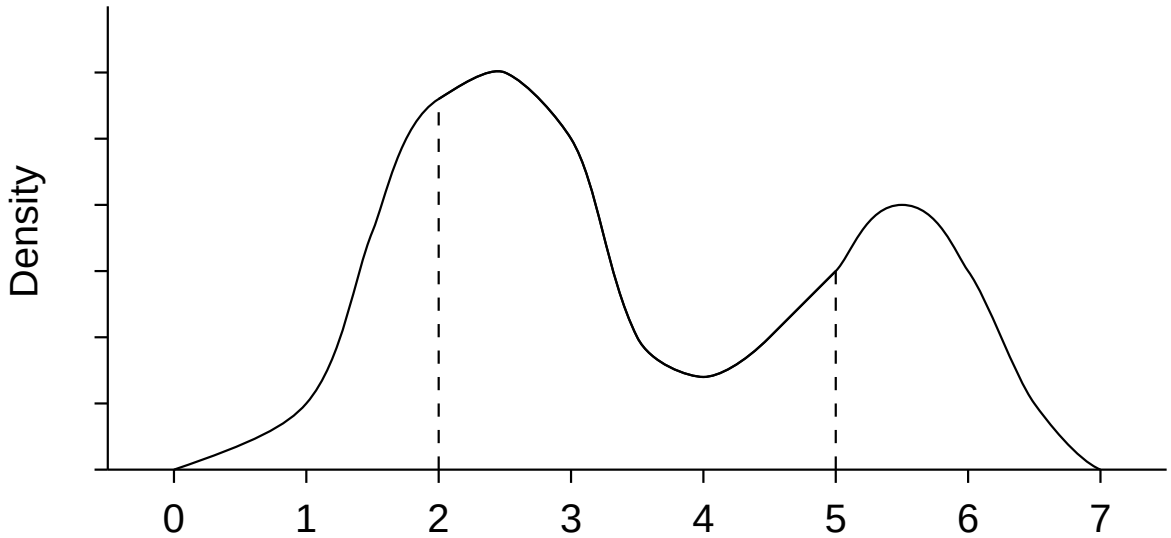
Density Functions

- The area under the curve between 2 and 5 is the proportion of the values of x that lie between 2 and 5.



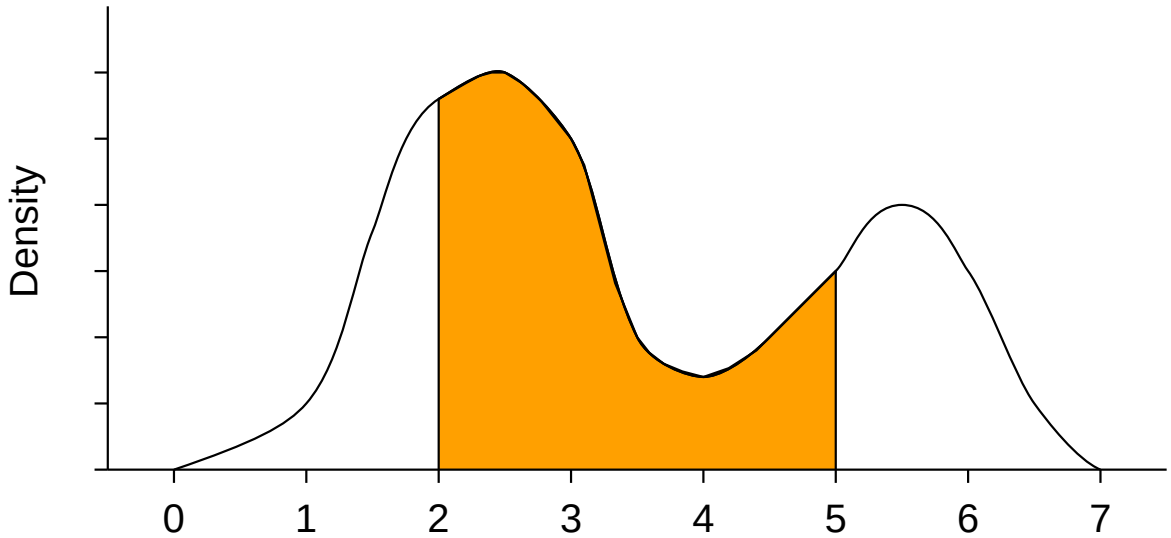
Density Functions

- The area under the curve between 2 and 5 is the proportion of the values of x that lie between 2 and 5.



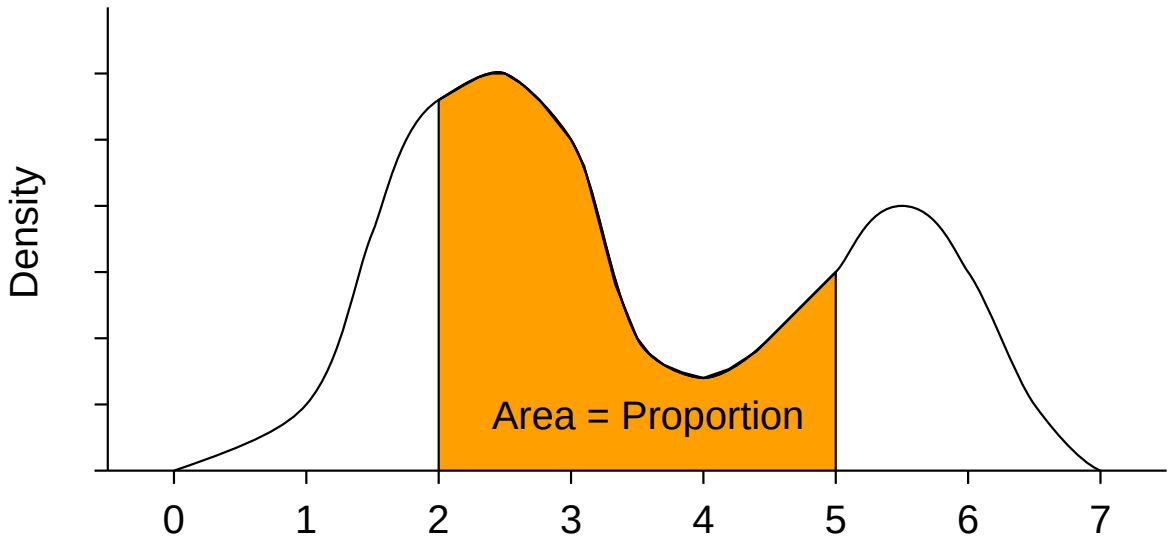
Density Functions

- The area under the curve between 2 and 5 is the proportion of the values of x that lie between 2 and 5.



Density Functions

- The area under the curve between 2 and 5 is the proportion of the values of x that lie between 2 and 5.



The Normal Distribution

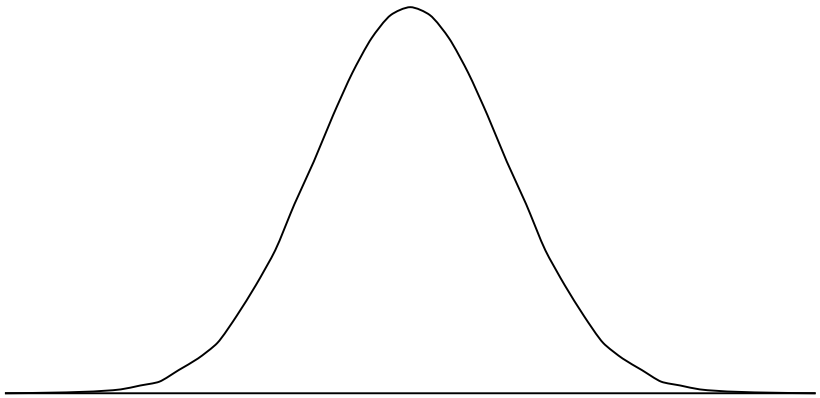
Definition (Normal distribution)

The normal distribution is the statistician's name for the bell curve.

- It is a density function in the shape of a "bell."
  - Symmetric.
  - Unimodal.
  - Extends over the entire real line (no endpoints).
  - "Main part" lies within $\pm 3\sigma$ of the mean.

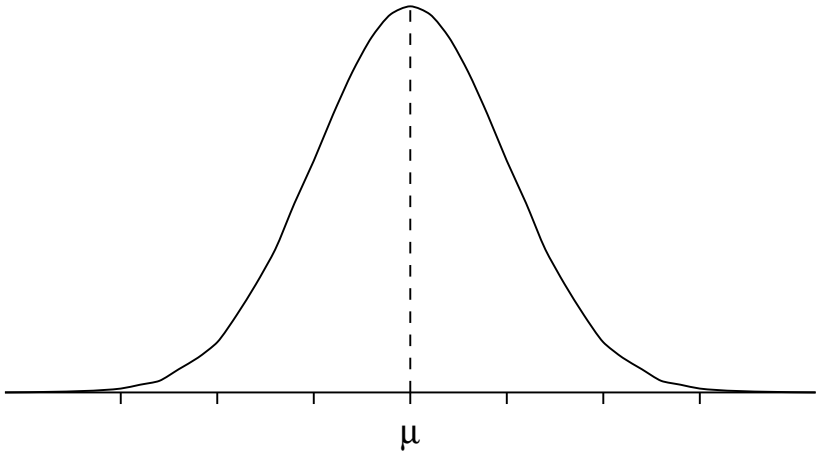
The Normal Distribution

- The curve has a bell shape, with infinitely long tails in both directions.



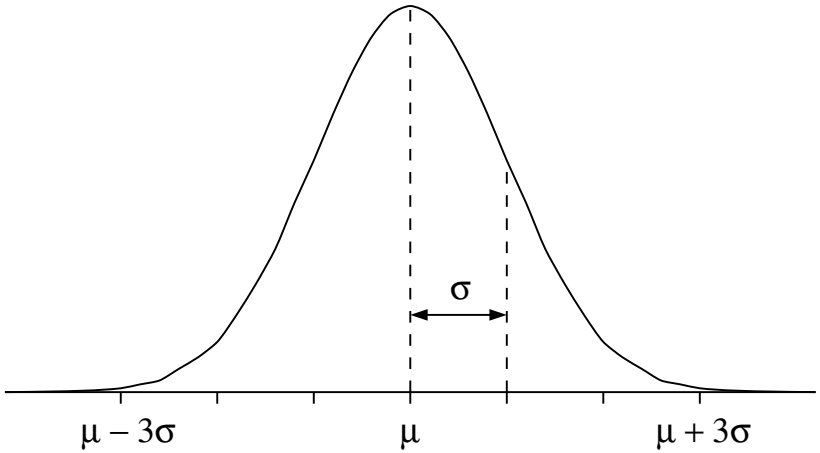
The Normal Distribution

- The mean μ is located in the center, at the peak.



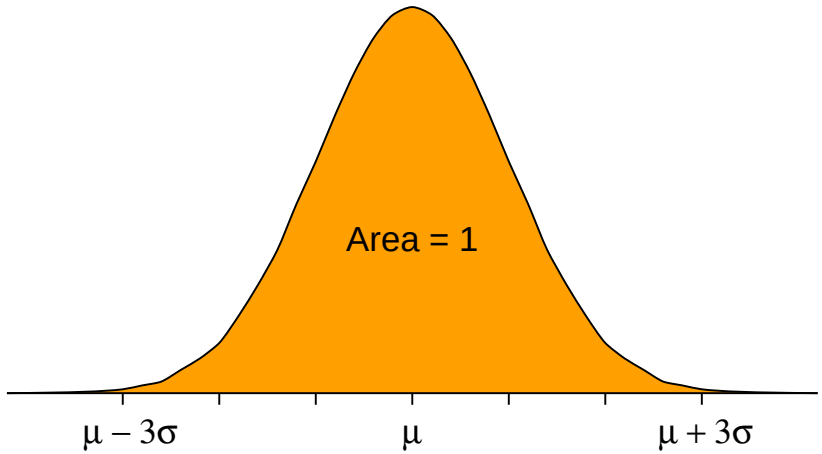
The Normal Distribution

- The width of the “main” part of the curve is 6 standard deviations wide (3 standard deviations each way from the mean).



The Normal Distribution

- The area under the entire curve is 1.
- (The area outside of $\mu \pm 3\sigma$ is approx. 0.0027.)

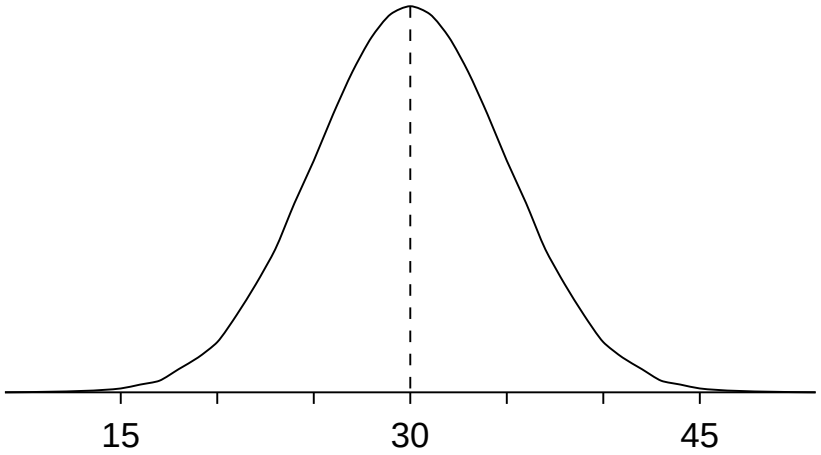


The Normal Distribution

- The normal distribution with mean μ and standard deviation σ is denoted $N(\mu, \sigma)$.
- For example, if X is a variable whose distribution is normal with mean 30 and standard deviation 5, then we say that "X is $N(30, 5)$."

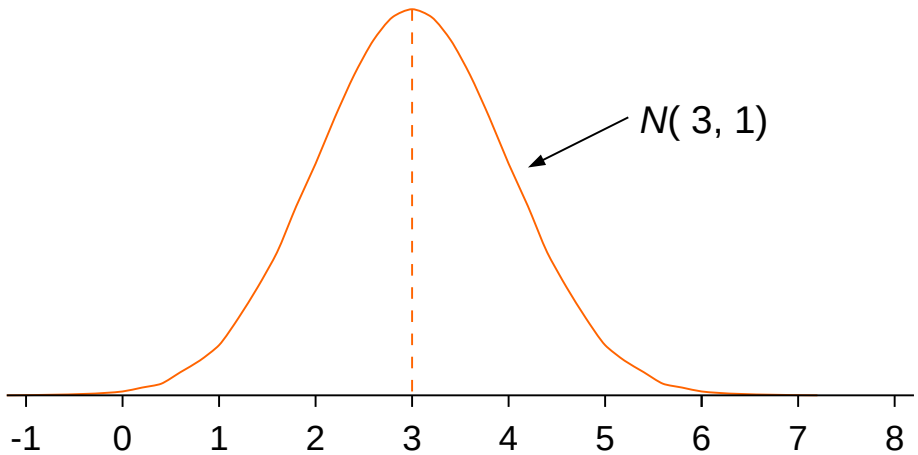
The Normal Distribution

- If X is $N(30, 5)$, then the distribution of X looks like this:



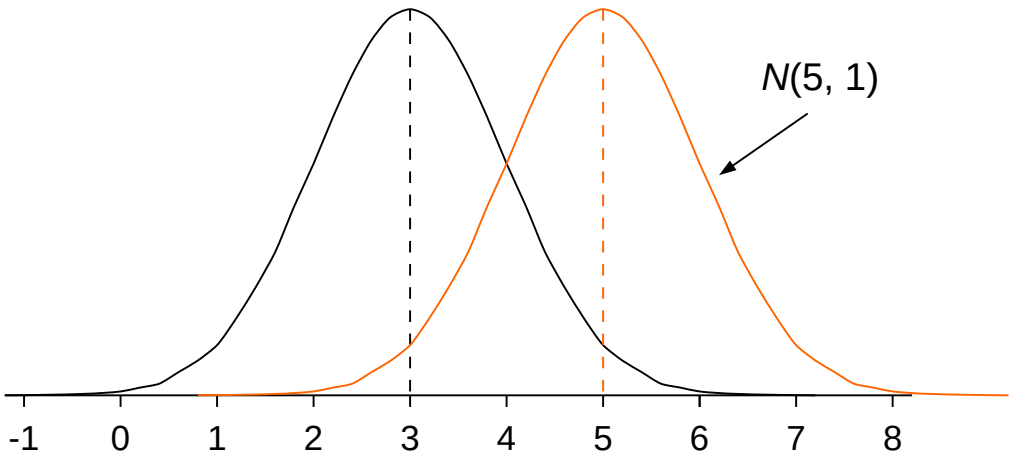
Some Normal Distributions

- Some other examples.



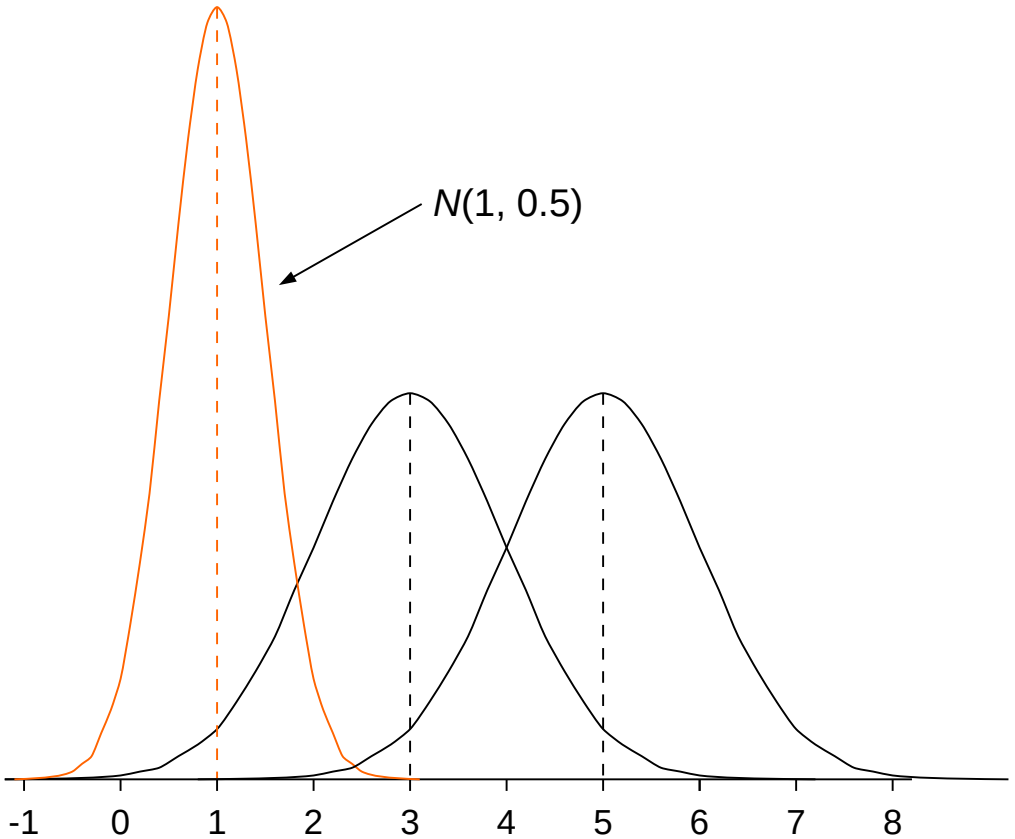
Some Normal Distributions

- Some other examples.



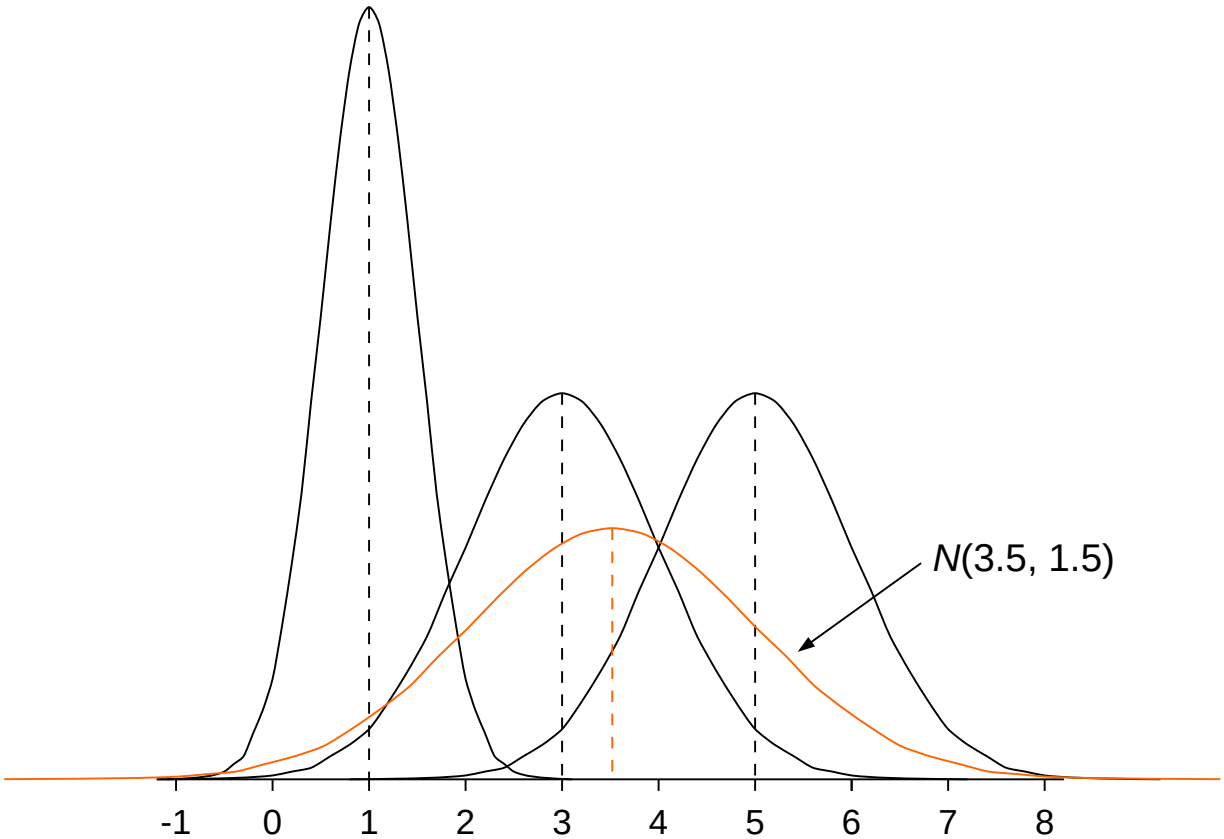
Some Normal Distributions

- Some other examples.



Some Normal Distributions

- Some other examples.



Assignment

Homework

- Read Section 6.1 - 6.3.1, pages 357 - 362.
- Let's Do It! 6.1.
- Exercises 1, 2, 3, 32, page 376.