Chapter 6: Random Variables

Section 6.1

Discrete and Continuous Random Variables

The Practice of Statistics, 4th edition – For AP*

STARNES, YATES, MOORE

Chapter 6
Random Variables

- 6.1 **Discrete and Continuous Random Variables**
- 6.2 Transforming and Combining Random Variables
- 6.3 Binomial and Geometric Random Variables

Section 6.1
Discrete and Continuous Random Variables

Learning Objectives

After this section, you should be able to…

- ✓ APPLY the concept of discrete random variables to a variety of statistical settings
- ✓ CALCULATE and INTERPRET the mean (expected value) of a discrete random variable
- ✓ CALCULATE and INTERPRET the standard deviation (and variance) of a discrete random variable
- ✓ DESCRIBE continuous random variables

Random Variable and Probability Distribution

A **probability model** describes the possible outcomes of a chance process and the likelihood that those outcomes will occur.

A numerical variable that describes the outcomes of a chance process is called a **random variable**. The probability model for a random variable is its probability distribution

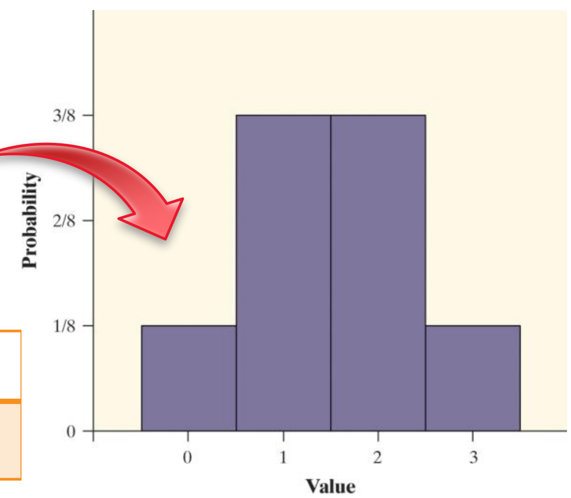
Definition:

A **random variable** takes numerical values that describe the outcomes of some chance process. The **probability distribution** of a random variable gives its possible values and their probabilities.

Example: Consider tossing a fair coin 3 times. Define X = the number of heads obtained

X = 0: TTT
X = 1: HTT THT TTH
X = 2: HHT HTH THH
X = 3: HHH

Value	0	1	2	3
Probability	1/8	3/8	3/8	1/8



Discrete Random Variables

There are two main types of random variables: *discrete* and *continuous*. If we can find a way to list all possible outcomes for a random variable and assign probabilities to each one, we have a **discrete random variable**.

Discrete Random Variables and Their Probability Distributions

A **discrete random variable** X takes a fixed set of possible values with gaps between. The probability distribution of a discrete random variable X lists the values x_i and their probabilities p_i:

Value:	x_1	x_2	x_3	…
Probability:	p_1	p_2	p_3	…

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. The sum of the probabilities is 1.

To find the probability of any event, add the probabilities p_i of the particular values x_i that make up the event.

Example: Babies' Health at Birth

Read the example on page 343.

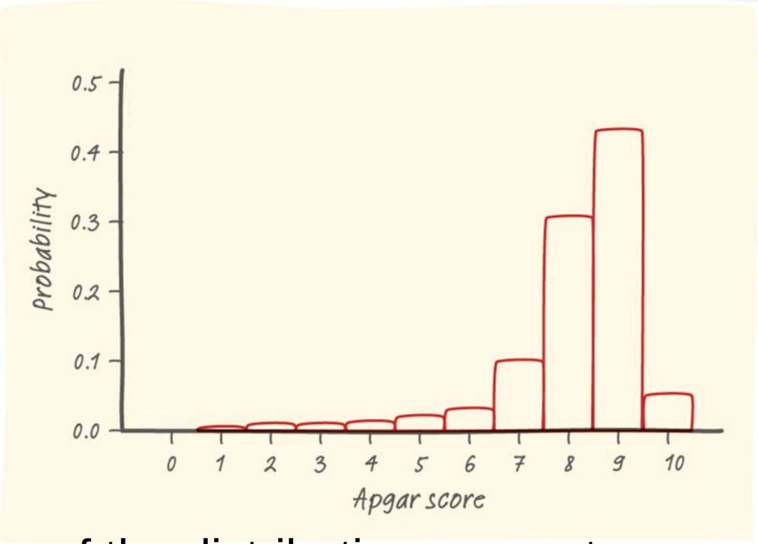
(a) **Show that the probability distribution for X is legitimate.**

(b) **Make a histogram of the probability distribution. Describe what you see.**

(c) **Apgar scores of 7 or higher indicate a healthy baby. What is $P(X \geq 7)$?**

Value:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

(a) All probabilities are between 0 and 1 and they add up to 1. This is a legitimate probability distribution.



(c) $P(X \geq 7) = .908$
We'd have a 91 % chance of randomly choosing a healthy baby.

(b) The left-skewed shape of the distribution suggests a randomly selected newborn will have an Apgar score at the high end of the scale. There is a small chance of getting a baby with a score of 5 or lower.

Alternate Example: NHL goals

Read the example on page 343.

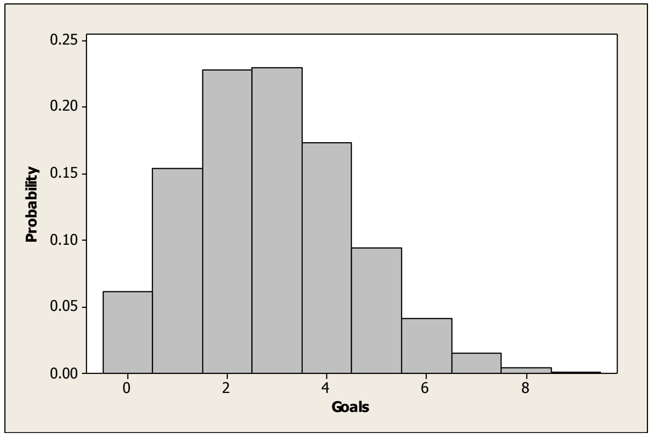
(a) **Show that the probability distribution for X is legitimate.**

(b) **Make a histogram of the probability distribution. Describe what you see.**

(c) **What is the probability that the number of goals scored by a randomly selected team in a randomly selected game is at least 6?**

Goals:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Probability:	0.061	0.154	0.007	0.228	0.229	0.173	0.094	0.041	0.015	0.001

(a) All probabilities are between 0 and 1 and they add up to 1. This is a legitimate probability distribution.



(c) $P(X \geq 6) = .061$
A randomly selected team in a randomly selected game has a 6.1% chance of scoring at least 6 goals.

(b) The histogram is skewed to the right, which means that the majority of games are relatively low scoring. It is pretty unusual for a team to score 6 or more goals.

Mean of a Discrete Random Variable

When analyzing discrete random variables, we'll follow the same strategy we used with quantitative data – describe the shape, center, and spread, and identify any outliers.

The mean of any discrete random variable is an average of the possible outcomes, with each outcome weighted by its probability.

Definition:

Suppose that *X* is a discrete random variable whose probability distribution is

Value:	x_1	x_2	x_3	…
Probability:	p_1	p_2	p_3	…

To find the **mean (expected value)** of *X*, multiply each possible value by its probability, then add all the products:

$$\mu_x = E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots$$

$$= \sum x_i p_i$$

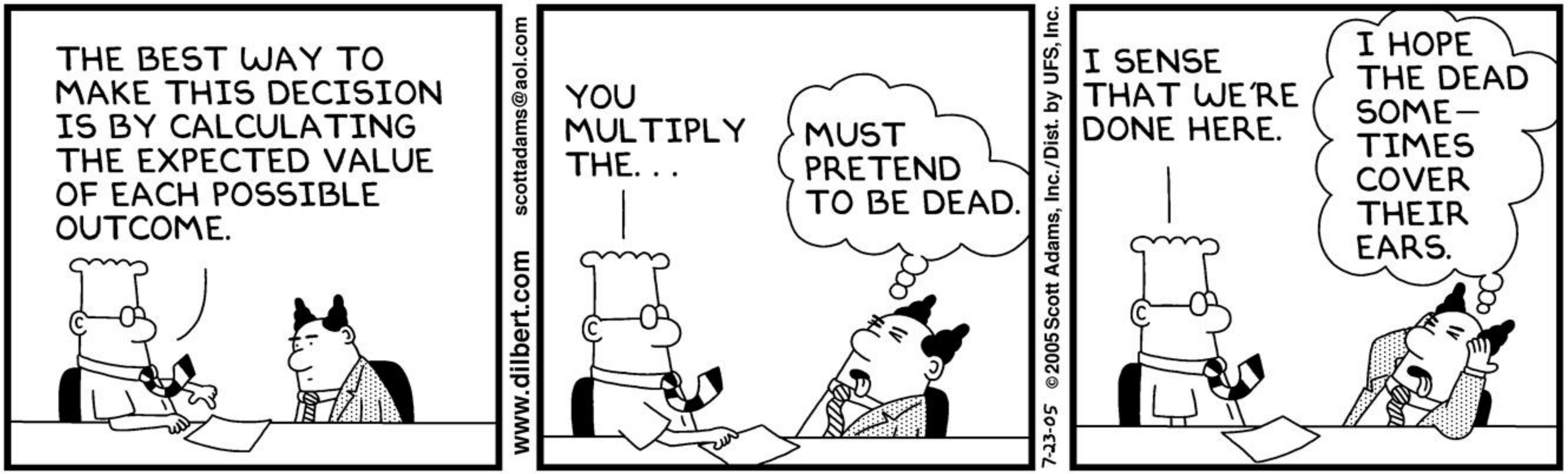
Example: Apgar Scores – What's Typical?

Consider the random variable X = Apgar Score

Compute the mean of the random variable X and interpret it in context.

Value:	0	1	2	3	4	5	6	7	8	9	10
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

$$\mu_x = E(X) = \sum x_i p_i$$



Alternate Example: NHL goals

Consider the random variable X = The number of goals scored by a randomly selected team in a randomly selected game.

Compute the mean of the random variable *X* and interpret it in context.

Goals:	0	1	2	3	4	5	6	7	8	9
Probability:	0.061	0.154	0.007	0.228	0.229	0.173	0.094	0.041	0.015	0.001

$$\mu_x = E(X) = \sum x_i p_i$$

$$= (0)(0.061) + (1)(0.154) + (2)(0.007) + \ldots + (9)(0.001)$$

$$= 2.851$$

The mean number of goals for a randomly selected team in a randomly selected game is 2.851. If you were to repeat the random sampling process over and over again, the mean number of goals scored would be about 2.851

Standard Deviation of a Discrete Random Variable

Since we use the mean as the measure of center for a discrete random variable, we'll use the standard deviation as our measure of spread. The definition of the **variance of a random variable** is similar to the definition of the variance for a set of quantitative data.

Definition:

Suppose that X is a discrete random variable whose probability distribution is

Value:	x_1	x_2	x_3	…
Probability:	p_1	p_2	p_3	…

and that μ_X is the mean of X. The **variance** of X is

$$Var(X) = \sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + (x_3 - \mu_X)^2 p_3 + \ldots$$
$$= \sum (x_i - \mu_X)^2 p_i$$

To get the **standard deviation of a random variable**, take the square root of the variance.

Example: Apgar Scores – How Variable Are They?

Consider the random variable X = Apgar Score

Compute the standard deviation of the random variable X and interpret it in context.

Value:	0	1	2	3	4	5	6	7	8	9	10
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$

$$= (0-8.128)^2(0.001) + (1-8.128)^2(0.006) + \ldots + (10-8.128)^2(0.053)$$

$$= 2.066$$

Variance

$$\sigma_X = \sqrt{2.066} = 1.437$$

The standard deviation of X is 1.437. On average, a randomly selected baby's Apgar score will differ from the mean 8.128 by about 1.4 units.

Alternate Example: NHL goals

Consider the random variable X = The number of goals scored by a randomly selected team in a randomly selected game.

Compute the standard deviation of the random variable *X* and interpret it in context.

Goals:	0	1	2	3	4	5	6	7	8	9
Probability:	0.061	0.154	0.007	0.228	0.229	0.173	0.094	0.041	0.015	0.001

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$

$$= (0-2.851)^2(0.061) + (1-2.851)^2(0.154) + \ldots + (9-2.851)^2(0.001)$$

$$= 2.66$$

$$\sigma_X = \sqrt{2.66} = 1.63$$

The standard deviation of X is 1.63. On average, a randomly selected team's number of goals in a randomly selected game will differ from the mean by about 1.63 goals.

Continuous Random Variables

Discrete random variables commonly arise from situations that involve counting something. Situations that involve measuring something often result in a **continuous random variable**.

Definition:

A **continuous random variable** *X* takes on all values in an interval of numbers. The probability distribution of *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 make up the event.

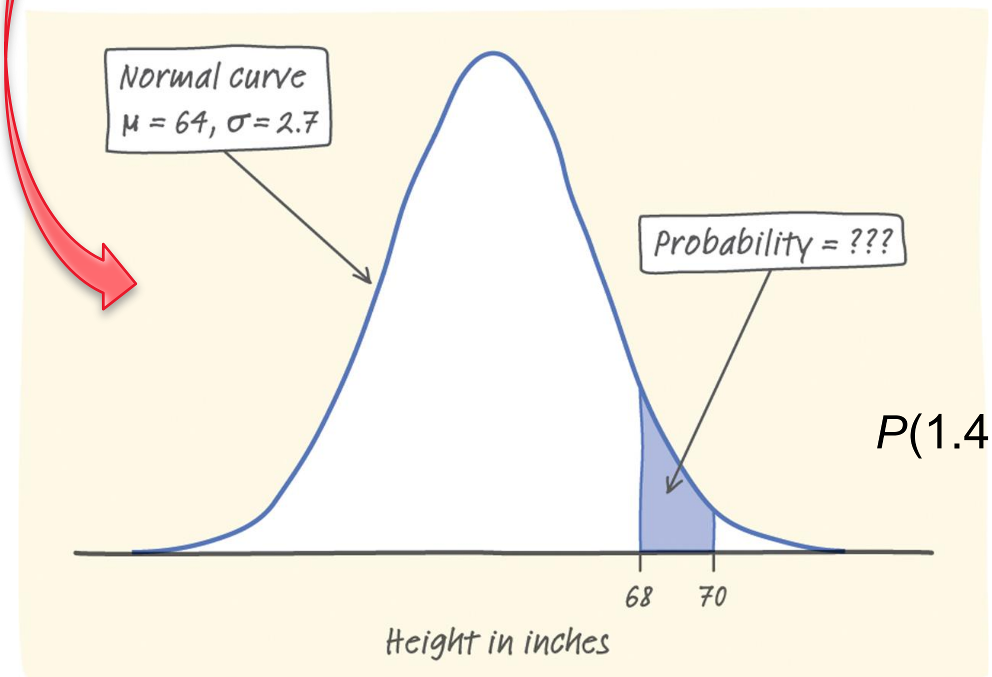
The probability model of a discrete random variable *X* assigns a probability between 0 and 1 to each possible value of *X*.

A continuous random variable *Y* has *infinitely many* possible values. All continuous probability models assign probability 0 to every individual outcome. Only *intervals* of values have positive probability.

Example: Young Women's Heights

Read the example on page 351. Define *Y* as the height of a randomly chosen young woman. *Y* is a continuous random variable whose probability distribution is *N*(64, 2.7).

What is the probability that a randomly chosen young woman has height between 68 and 70 inches?



$$P(68 \le Y \le 70) = ???$$

$$z = \frac{68-64}{2.7} = 1.48 \qquad z = \frac{70-64}{2.7} = 2.22$$

$$P(1.48 \le Z \le 2.22) = P(Z \le 2.22) - P(Z \le 1.48)$$

$$= 0.9868 - 0.9306$$

$$= 0.0562$$

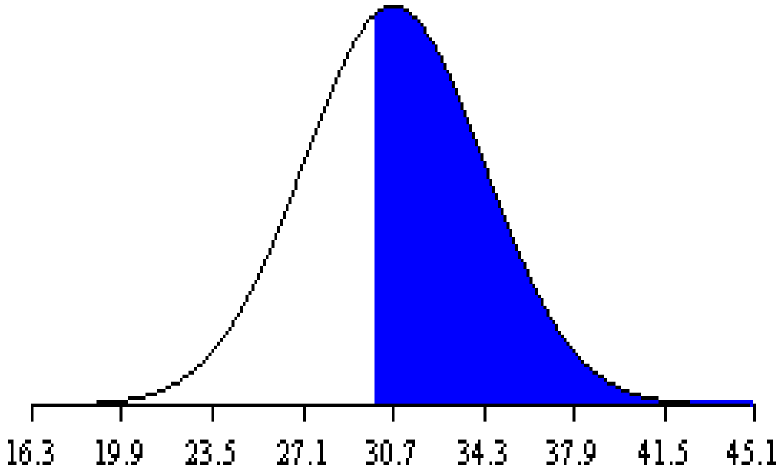
There is about a 5.6% chance that a randomly chosen young woman has a height between 68 and 70 inches.

Alternate Example: Weights of 3-year-old females

Read the example on page 351. Define *X* as the weight of a randomly chosen 3-year-old female. *X* is a continuous random variable whose probability distribution is *N*(30.7, 3.6).

State: What is the probability that a randomly chosen 3-year-old female weighs at least 30 lbs?

Plan: The weight X of the 3-year old female we chose has the *N*(30.7, 3.6) distribution. We want to find P(X≥30). We will standardize this weight and use Table A to find the shaded area.



Do:

$$z = \frac{30 - 30.7}{3.6}$$

$$= -.19$$

P(*Z* < -0.19) = 0.4247, so

P(Z>-0.19) = 1-0.427

= 0.5753

Conclude: There is about a 58% chance that a randomly chosen 3-year-old female will weigh at least 30 lbs.

+

Section 6.1
Discrete and Continuous Random Variables

Summary

In this section, we learned that…

- ✓ A **random variable** is a variable taking numerical values determined by the outcome of a chance process. The **probability distribution** of a random variable *X* tells us what the possible values of *X* are and how probabilities are assigned to those values.
- ✓ A **discrete random variable** has a fixed set of possible values with gaps between them. The probability distribution assigns each of these values a probability between 0 and 1 such that the sum of all the probabilities is exactly 1.
- ✓ A **continuous random variable** takes all values in some interval of numbers. A density curve describes the probability distribution of a continuous random variable.

+

Section 6.1
Discrete and Continuous Random Variables

Summary

In this section, we learned that…

- ✓ The **mean of a random variable** is the long-run average value of the variable after many repetitions of the chance process. It is also known as the **expected value** of the random variable.

- ✓ The expected value of a discrete random variable *X* is

$$\mu_x = \sum x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots$$

- ✓ The **variance of a random variable** is the average squared deviation of the values of the variable from their mean. The **standard deviation** is the square root of the variance. For a discrete random variable *X*,

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + (x_3 - \mu_X)^2 p_3 + \ldots$$

Looking Ahead…

<u>In the next Section…</u>

We'll learn how to determine the mean and standard deviation when we transform or combine random variables.

We'll learn about

- ✓ **Linear Transformations**
- ✓ **Combining Random Variables**
- ✓ **Combining Normal Random Variables**