Section 7.2-7.3
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

Source : TPS, 4th edition (Chapter 6)

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 and Continuous Random Variables

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	$\dots$
Probability:	p_1	p_2	p_3	$\dots$

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: Apgar Scores: Babies' Health at Birth

- In 1952, Dr Virginia Apgar suggested 5 criteria for measuring a baby's health at birth: skin color, heart rate, muscle tone, breathing and response when stimulated. She developed a 0-1-2 scale to rate a new born on each of these 5 criteria. A baby's Apgar score is a whole number value 0-10. Apgar scores are still used today to evaluate the health on a newborns.
- What Apgar scores are typical? Research was done on over 2 million newborns in a single year.
- Let's **define a random variable**:

  X= Apgar score of a randomly selected baby

  The table below gives the probability distribution for X:

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) 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 $P(X \geq 7)$?

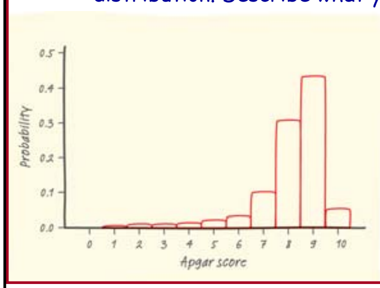
Example: Apgar Scores: Babies' Health at Birth

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

- All probabilities are between 0 and 1 and
- They add up to 1.
- This is a legitimate probability distribution.

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

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



(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.

(c) Apgar scores of 7 or higher indicate a healthy baby.

$P(X \geq 7) = .908$

We'd have a 91 % chance of randomly choosing a healthy baby.

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:

$$\mu_x = E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \ldots$$
$$= \sum x_ip_i$$

Discrete and Continuous Random Variables

Example: Apgar Scores – What's Typical?

Consider the random variable *X* = Apgar Score

1)Compute the mean of the random variable *X* and 2) 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

$$\mu_x = E(X) = \sum x_i p_i \quad = (0)(0.001) + (1)(0.006) + (2)(0.007) + \ldots + (10)(0.053)$$
$$= 8.128$$

Context: The mean Apgar score of a randomly selected newborn is 8.128. This is the long-term average Apgar score of many, many randomly chosen babies.

Note:

- The expected value (8.128) does not need to be a possible value of *X* or an integer!
- It is a long-term average over many repetitions.

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.

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

- **The heights of young women closely follows the normal distribution with μ=64 inches and the standard deviation σ=2.7 inches.**
- **This is a distribution for a large set of data.**
- **Now choose one young women at random. If we repeat the random choice very many times, the distribution of values will be the same.**

Let's define the random variable:

Y= the height of a randomly chosen young woman.

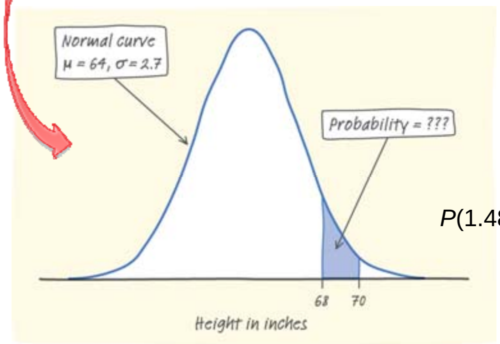
Y is a continuous random variable whose probability distribution is N(64, 2.7).

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

Example: Young Women's Heights

- Continuous Random variable defined: Y = 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?



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

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

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

$= 0.9868 - 0.9306$

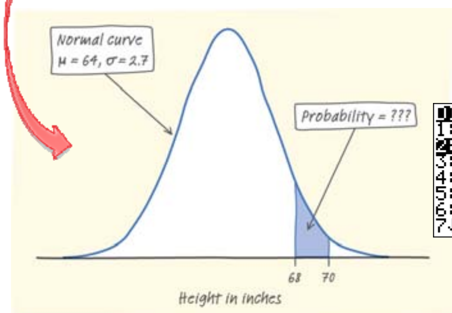
$= 0.0562$

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

Using the TI84Plus to find probabilities for a normal distribution

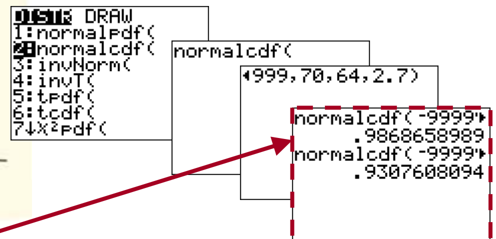
- The syntax is normalcdf (lowerbound, upperbound, μ, σ)
- But the disadvantage is it does not provide a picture of area you are trying to find. Always draw a picture!

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



$N(64, 2.7)$

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



$\boxed{= 0.9868 - 0.9306 = 0.0562}$

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.

+

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…

In the next Section…

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**