**AP Statistics**
Unit 04 – Probability
Day 05 Notes – Discrete & Continuous Random Variables

Name Key
Period

**Random Variable:** takes numerical values that describe the outcomes of some chance process.

**Probability Distribution:** gives possible values of a random variable and their corresponding probability.

**Example:**

A probability model describes the possible outcomes of a chance process and the likelihood that those outcomes will occur. For example, suppose we toss a fair coin 3 times. The sample space for this chance process is

HHH HHT HTH THH HTT THT TTH TTT

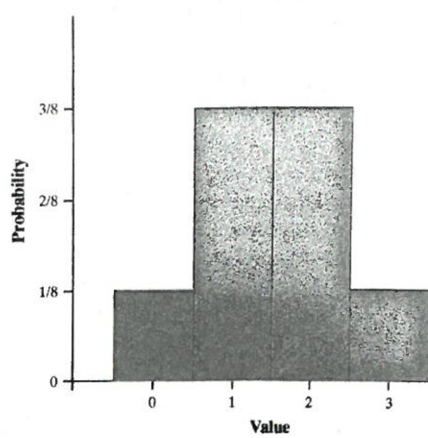
Because there are 8 equally likely outcomes, the probability is 1/8 for each possible outcome. Define the variable X = the number of heads obtained. The value of X will vary from one set of tosses to another but will always be one of the numbers 0, 1, 2, or 3. How likely is X to take each of those values? It will be easier to answer this question if we group the possible outcomes by the number of heads obtained:

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

We can summarize the probability distribution of X as follows:

| **Value:** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **Probability:** | 1/8 | 3/8 | 3/8 | 1/8 |

Here is the probability distribution of X in graphical form.
(3 coin tosses)




1. What's the probability that we get exactly one head?

   3/8

2. What's the probability that we get at least one head?

   $P(X \geq 1) = P(X=1) + P(X=2) + P(X)=3$
   $= 3/8 + 3/8 + 1/8 = 7/8$

3. What's the probability that we get, at most, 1 head?

   $P(X \leq 1) = P(X=0) + P(X=1)$
   $= 1/8 + 3/8 = 4/8 = 1/2$

*There are two main types of random variables:* ***discrete*** *and* ***continuous***.

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: $p_1 + p_2 + p_3 + \ldots = 1$.

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 five 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 newborn on each of the five criteria. A baby's Apgar score is the sum of the rating on each of the five scales, which gives a whole-number value from 0-10. Apgar scores are still used today to evaluate health of newborns.

What Apgar scores are typical? To find out, researchers recorded the Apgar scores of over 2 million newborn babies in a single year. Imagine selecting one of these newborns at random. (That's our chance process.) Define the random variable X = Apgar score of a randomly selected baby one minute after birth. 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.

$0.001 + 0.006 + 0.007 + 0.008 + 0.012 + 0.020 + 0.038 + 0.099 + 0.319 + 0.437 + 0.053$

$= 1.00$ and all probabilities are between 0 and 1.

b) Doctors decided that Apgar scores of 7 or higher indicate a healthy baby. What's the probability that a randomly selected baby is healthy?

$P(X \geq 7) = P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$= 0.099 + 0.319 + 0.437 + 0.053$

$= 0.908$

There is a 90.8% chance of randomly selecting a healthy baby.

**CHECK YOUR UNDERSTANDING:**

North Carolina State University posts the grade distributions for its courses online. Students in Statistics 101 in a recent semester received 26% A's, 42% B's, 20% C's, 10% D's, and 2% F's. Choose a Statistics 101 students at random. The student's grade on a four-point scale (with A = 4) is a discrete random variable X with this probability distribution:

| | F | D | C | B | A |
|---|---|---|---|---|---|
| **Value of X:** | 0 | 1 | 2 | 3 | 4 |
| **Probability:** | 0.02 | 0.10 | 0.20 | 0.42 | 0.26 |

1. Say in words what the meaning of $P(X \geq 3)$ is. What is this probability?

   $P(x \geq 3)$ is the probability that the student got either an A or a B. $P(X \geq 3) = 0.68$

2. Write the event "the student got a grade worse than C" in terms of values of the random variable X. What is the probability of this event?

   $P(X < 2) = 0.02 + 0.10 = 0.12$

**Mean of a Discrete Random Variable $\mu_X$:**

an average of the possible values of x but with an important change to take into account the fact that not all outcomes may be equally likely.

**Example: Winning (and Losing) at Roulette**

On an American roulette wheel, there are 38 slots numbered 1 through 36, plus 0 and 00. Half of the slots from 1 to 36 are red; the other half are black. Both the 0 and the 00 slots are green. Suppose that a player places a simple \$1 bet on red. If the balls land in a red slot, the player gets the original dollar back, plus an additional dollar for winning the bet. If the ball lands in a different-colored slot, the player loses the dollar bet to the casino.

Let's define the random variable X = net gain from a single \$1 bet on red. The possible values of X are -\$1 and \$1. (The player either gains a dollar or loses a dollar.) What are the corresponding probabilities? The chance that the ball lands in a red slot is 18/38. The chance that the ball lands in a different-colored slot is 20/38. Here is the probability distribution of X:

| | | |
|---|---|---|
| **Value:** | -\$1 | \$1 |
| **Probability:** | 20/38 | 18/38 |

What is the player's average gain? The ordinary average of the two possible outcomes -\$1 and \$1 is \$0 But \$0 isn't the average winnings because the player is less likely to win \$1 than to lose \$1. In the long run, the player gains a dollar 18 times in every 38 games played and loses a dollar on the remaining 20 of 38 bets. The player's long-run average gain for this sample bet is

$$\mu_X = (-\$1)(20/38) + (\$1)(18/38) = -\$0.05$$

You see that the player loses (and the casino gains) an average of five cents per \$1 bet in many, many plays of the game.

**Expected Value/Mean of a Discrete Random Variable:**

Suppose that X is a discrete random variable with probability distribution

| **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:

$$\mu_X = E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \ldots = \Sigma x_ip_i$$

**Example: Apgar Scores: What's Typical?**

In our earlier example, we defined the random variable X to be the Apgar score of a randomly selected baby. The table below gives the probability distribution for X once again.

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

Compute the mean of the random variable X. Interpret this value in context.

$E(X) = 0\cdot 0.001 + 1\cdot 0.006 + 2\cdot 0.007 + 3\cdot 0.008 + 4\cdot 0.012 + 5\cdot 0.020$
$+ 6\cdot 0.038 + 7\cdot 0.099 + 8\cdot 0.319 + 9\cdot 0.437 + 10\cdot 0.053 = 8.128$

The mean Apgar score of a randomly chosen newborn is 8.128.

**Variance $\sigma_X^2$ and Standard Deviation $\sigma_X$ of a Random Variable:**

Suppose that X is a discrete random variable with probability distribution

| **Value:** | $x_1$ | $x_2$ | $x_3$ | ... |
|---|---|---|---|---|
| **Probability:** | $p_1$ | $p_2$ | $p_3$ | ... |

And that $\mu_X$ is the mean of X. The **variance** of X is

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

The **standard deviation** of X, $\sigma_X$, is the square root of the variance.

$$\sigma_X = \sqrt{\Sigma(x_i - \mu_X)^2p_i}$$

**EXAMPLE:**

On Valentine's Day, the *Quiet Nook* restaurant offers a *Lucky Lovers Special* that could save couples money on their romantic dinners. When the waiter brings the check, he'll also bring the four aces from a deck of cards. He'll shuffle them and lay them out face down on the table. The couple will then turn one card over. If it's a black ace, they'll owe the full amount, but if it's the ace of hearts, the waiter will give them a $20 Lucky Lovers Discount. If they first turn over the ace of diamonds, they'll then get to turn over one of the remaining cards, earning a $10 discount for finding the ace of hearts this time. Based on the probability model for the size of Lucky Lovers discounts the restaurants will award, what's the expected discount for a couple?

1. List the expected value (X=the Lucky Lover's Discount)

2. List the probabilities of each event

   P(x=20) = P(of getting a heart)= $1/4$ (or $3/12$)

   P(x=10) = P(getting a diamond, then a heart)= $1/4 \cdot 1/3 = 1/12$

   P(x=0) = P(not getting $20 or $10)= $1 - 4/12 = 1 - 1/3 = 2/3$

3. Create a probability model (table or tree)

| **Outcome** | **Hearts** | **Diamond then heart** | **Black ace** |
|---|---|---|---|
| x | $20 | $10 | $0 |
| P(x) | 1/4 | 1/12 | 2/3 |

4. Calculate the expected value.

   $E(X) = 20(1/4) + 10(1/12) + 0(2/3)$

   $= 5 + 0.8\bar{3} + 0 \quad = \$5.8\bar{3}$

5. Explain in words what your calculation means.

   Couples dining at the Quiet Nook on Valentine's Day can expect an average discount of $5.83.

6. Calculate and interpret the standard deviation.

   $\sigma = \sqrt{(20-5.83)^2\frac{1}{4} + (10-5.83)^2\frac{1}{12} + (0-5.83)^2\frac{2}{3}}$

   $= \$8.62$

**Example: Apgar Scores: What's Typical?**

In the last example, we calculated the mean Apgar score of a randomly chosen newborn to be $\mu_X = 8.128$. The table below gives the probability distribution for X one more time:

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

Compute and interpret the standard deviation of the random variable X.

$\sigma_x = \sqrt{(0-8.128)^2 \cdot 0.001 + (1-8.128)^2 \cdot 0.006 + \ldots + (10-8.128)^2 \cdot 0.053}$

$= \sqrt{2.066} = 1.437$

A randomly selected baby's Apgar score will typically differ from the mean by about 1.437 units

**Example: How Many Languages?**

The mean number of languages spoken for a randomly selected US high school student to be $\mu_X = 1.457$. The table below gives the probability distribution for X:

| **Languages:** | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Probabilities:** | 0.630 | 0.295 | 0.065 | 0.008 | 0.002 |

Compute and interpret the standard deviation of the random variable X.

$\sigma_x^2 = (1-1.457)^2(0.630) + (2-1.457)^2(0.295) + (3-1.457)^2(0.065) + (4-1.457)^2(0.008) + (5-1.457)^2(0.002)$

$= 0.450$

$\sigma_x = \sqrt{0.450} = 0.671$

The number of language spoken by a randomly selected U.S. high school student typically varies by about 0.671 from the mean.

**CHECK YOUR UNDERSTANDING:**

A large auto dealership keeps track of sales made during each hour of the day. Let X = the number of cars sold during the first hour of business on a randomly selected Friday. Based on previous records, the probability distribution of X is as follows.

| **Cars sold:** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **Probabilities:** | 0.3 | 0.4 | 0.2 | 0.1 |

1. Compute and interpret the mean of X.

$E(x) = 0 \cdot 0.3 + 1 \cdot 0.4 + 2 \cdot 0.2 + 3 \cdot 0.1 = 1.1$

If many, many Fridays are randomly selected, the average number of cars sold will be about 1.1.

2. Compute and interpret the standard deviation of X.

$\sigma_x = \sqrt{(0-1.1)^2 \cdot 0.3 + (1-1.1)^2 \cdot 0.4 + (2-1.1)^2 \cdot 0.2 + (3-1.1)^2 \cdot 0.1}$

$= \sqrt{0.89} = 0.943$ The number of cars sold on a randomly selected Friday will typically vary from the mean by 0.943 cars.

**On a Calculator:**

1. Start by entering the values of the random variable in L1 and the corresponding probabilities in L2.
2. To graph a histogram of the probability distribution:
    a. Set up a stat plot with Xlist: L1 and Freq: L2.
    b. Press GRAPH.
    c. Adjust your window settings as necessary.
3. To calculate the mean and standard deviation of the random variable, use one-variable statistics with the values in L1 and the corresponding probabilities in L2.

**Continuous Random Variable:** (X) takes all values in an interval of numbers. The probability 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.

*All continuous probability models assign probability 0 to every individual outcome.*

**Example: Young Women's Heights**

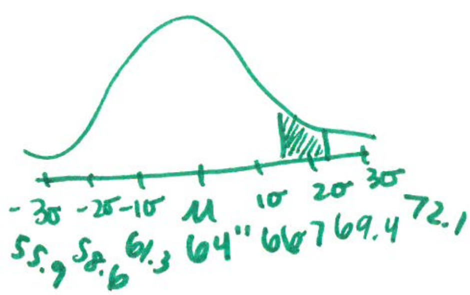
*Normal Probability Distributions*

The heights of young women closely follow the Normal distribution with mean $\mu = 64$ inches and standard deviation $\sigma = 2.7$ inches. This is a distribution for a large set of data. Now choose one young woman at random. Call her height Y. If we repeat the random choice very many times, the distribution of values of Y is the same Normal distribution that describes the heights of all young women. What's the probability that a randomly chosen young woman has height between 68 and 70 inches?

**STEP 1: State the distribution and the values of interest.**

The height Y of a randomly chosen young woman has the $N(64, 2.7)$ distribution. We want to find $P(68 \leq Y \leq 70)$.

**STEP 2: Perform calculations – show your work!**




$z = \dfrac{68-64}{2.7} = 1.48$

$z = \dfrac{70-64}{2.7} = 2.22$

$P(1.48 \leq z \leq 2.22) = 0.0562$

**STEP 3: Answer the question.**

There is about a 5.62% chance that a randomly chosen young woman has a height between 68" and 70."