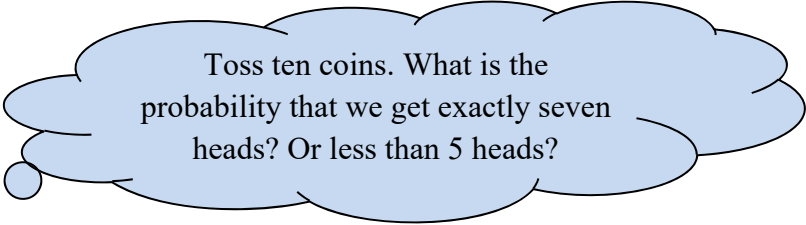General Education Statistics
Class Notes
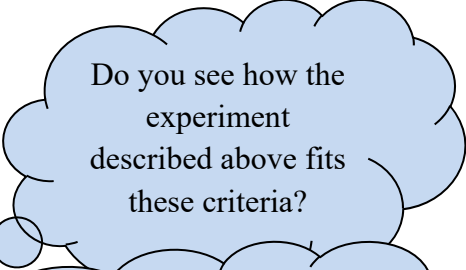The Binomial Probability Distribution (Section 6.2)

Toss ten coins. What is the probability that we get exactly seven heads? Or less than 5 heads?
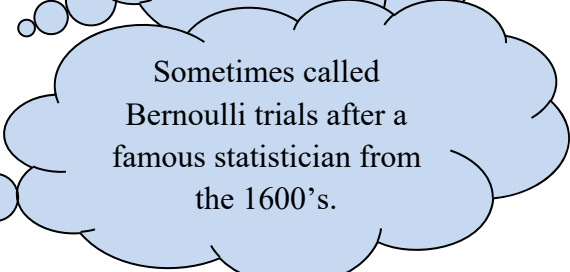
We studied discrete distributions in the last section. Here, we see a special discrete distribution called the Binomial Probability Distribution. The "bi" part of its name refers to the two mutually exclusive outcomes for each trial. We usually call them **success** and **failure**.

Definition: An experiment is said to be a **binomial experiment** if it fits *all* four of these criteria.

1. The experiment is performed a fixed number of times. Each repetition of the experiment is called a **trial**.

2. The trials are independent. This means the outcome of one trial will *not* affect the outcome of the other trials.

3. For each trial, there are two mutually exclusive (or disjoint) outcomes, success or failure.

4. The probability of success is fixed for each trial of the experiment.

Do you see how the experiment described above fits these criteria?

Sometimes called Bernoulli trials after a famous statistician from the 1600's.

Notation: We will say there are n independent trials.

The probability of success is p. Since failure is the only other option, we know that $1 - p$ will be the probability of failure.

The random variable X will count how many successes we get out of n trials. So, we know x ranges from 0 to n.
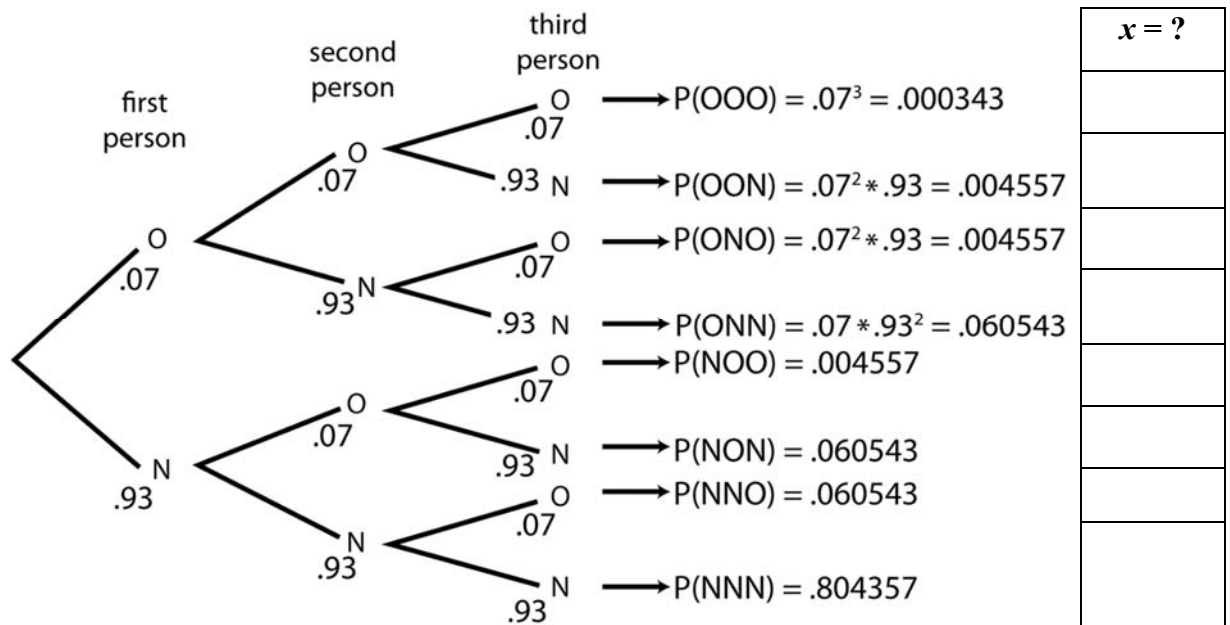
expl 1: Which of these are binomial experiments? If so, give the values of n, p, and possible values that X could take on.

a.) A football kicker attempts ten kicks (through the uprights). The probability that he makes any one kick is 75%. The number of kicks, out of ten, is recorded.

b.) A bag has ten marbles including three red, three blue, and four yellow. A marble is chosen, with replacement, and the color is recorded. We will stop when a yellow marble is chosen.

c.) A survey revealed that 28% of Americans choose chocolate as their favorite ice cream. A sample of size 100 is obtained and the number of respondents who choose chocolate is recorded.

1

expl 2: According to the American Red Cross, 7% of people in the U.S. have the blood type O-negative. A simple random sample of size three is obtained and the number of people with O-negative blood X is recorded. We will inspect a probability distribution for the random variable X. Follow the steps below.

a.) Consult the tree diagram which shows the possible outcomes for a sample of size three. I used O and N as the symbols for "O-negative" and "Not O-negative". We can assume independence so I computed the probabilities of each outcome by multiplying the probabilities along each branch.

third person

second person

first person

$P(OOO) = .07^3 = .000343$

$P(OON) = .07^2 * .93 = .004557$

$P(ONO) = .07^2 * .93 = .004557$

$P(ONN) = .07 * .93^2 = .060543$

$P(NOO) = .004557$

$P(NON) = .060543$

$P(NNO) = .060543$

$P(NNN) = .804357$

$x = ?$

b.) Now, X counts the number of O-negative people in each sample of three. Use the tree diagram to determine the value of X for each outcome and record it next to the diagram in the space provided. Notice that the number of times that a certain x value appears is $C(n, x)$.

c.) Inspect the probabilities (found by adding the probabilities above) given here.

We add the probabilities for all outcomes with x successes.

Number of respondents	
x	$P(x)$
0	0.804357
1	0.181629
2	0.013671
3	0.000343

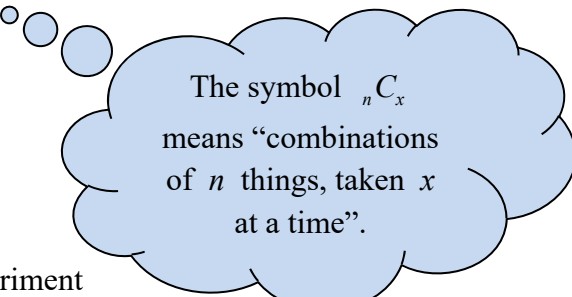So, 80% of samples of size 3 have *no* O-neg. people.

2

Next, we will see a formula that will find these probabilities without drawing a tree diagram. It involves the concept of combinations, a counting technique from chapter 5.

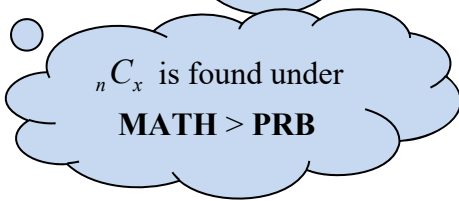**Definition: Binomial Probability Distribution Function:**
The probability of obtaining x successes in n independent trials of a binomial experiment is given by $P(x) = {}_nC_x \cdot p^x \cdot (1-p)^{n-x}$ where x can be any whole number from 0 to n. Here, recall that p is the probability of success.

Remember, before using this formula, you have to be sure that the criteria on page 1 are actually met.

The symbol ${}_nC_x$ means "combinations of n things, taken x at a time".

expl 3: Use this formula to find $P(2)$ for the previous experiment about blood type. It should match what we have in part c.

${}_nC_x$ is found under
MATH > PRB

Calculator Instructions:

To find $P(X = x)$ on the calculator, press **2ⁿᵈ VARS** to access the **DISTR** menu. Arrow down until you get to **binompdf(** and select it to put it on the home screen.

You'll need to enter the values for n, p, and x, following each with a comma (directly above the 7 button). Use the decimal form of p. End the expression with a parenthesis and press **ENTER**. Newer calculators may have a different interface.

To find $P(X \leq x)$ on the calculator, press **2ⁿᵈ VARS** to access the **DISTR** menu. Arrow down until you get to **binomcdf(** and select it to put it on the home screen. [This time, we use the **binomcdf(** for the *cumulative* distribution function. This finds $P(X \leq x)$ as opposed to $P(X = x)$.]

You'll need to enter the values for n, p, and x, following each with a comma (directly above the 7 button). Use the decimal form of p. End the expression with a parenthesis and press **ENTER**. Newer calculators may have a different interface.

Use the calculator to find the following probabilities.

3

We will see many problems that involve phrases like "at least" or "no more than". Be sure you know their meaning. Here are several phrases and their corresponding inequality symbols.

Phrase	Math Symbol
"at least" or "no less than" or "greater than or equal to"	$\geq$
"more than" or "greater than"	$>$
"fewer than" or "less than"	$<$
"no more than" or "at most" or "less than or equal to"	$\leq$
"exactly" or "equals" or "is"	$=$

expl 4: According to the Air Travel Consumer Report, the 11 largest air carriers had an on-time percentage of 79.0% in May 2008. Suppose that 10 flights are randomly selected from May 2008 and the number of on-time flights X is recorded.

a.) What values can the variable X take on?

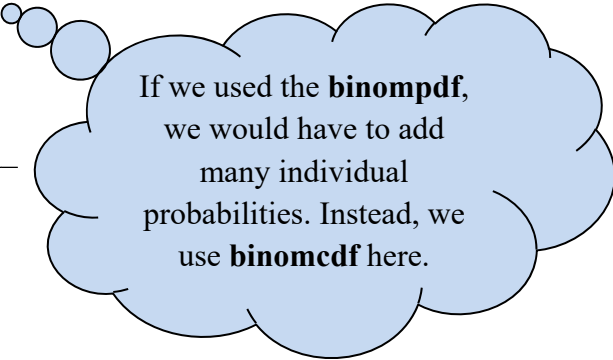b.) What is the probability that there were seven on-time flights out of the ten surveyed?

First, organize your values of $n,$ $p,$ and $x.$

$n =$ _____ $p =$ _____ $x =$ _____

c.) What is the probability that there were no more than five on-time flights out of the ten surveyed?

First, organize your values of $n,$ $p,$ and $x.$
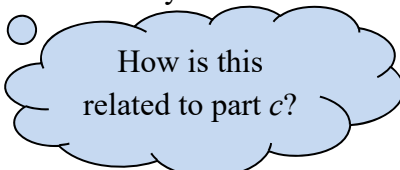
$n =$ _____ $p =$ _____ $x \leq$ _____

If we used the **binompdf**, we would have to add many individual probabilities. Instead, we use **binomcdf** here.

4

Probabilities Other Than "Less than or Equal to":

So the calculator will give us $P(X \leq x)$ [using the **binomcdf(** function] and $P(X = x)$ [using the **binompdf(** function]. However, we may be asked other, related probabilities. Often the complement rule or some other logical reasoning must be used.

expl 4 continued: According to the Air Travel Consumer Report, the 11 largest air carriers had an on-time percentage of 79.0% in May 2008. Suppose that 10 flights are randomly selected from May 2008 and the number of on-time flights X is recorded.
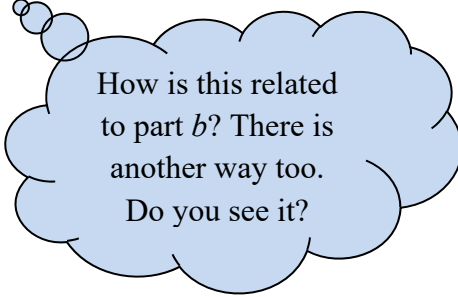
d.) What is the probability that there were at least six on-time flights out of the ten surveyed?

How is this related to part c?

e.) What is the probability that there were three or four on-time flights out of the ten surveyed?

f.) What is the probability that there were less than seven on-time flights out of the ten surveyed?

How is this related to part b? There is another way too. Do you see it?

Instructions for StatCrunch:

We will also use StatCrunch for these calculations.

Select **Stat** > **Calculators** > **Binomial**. It gives you a place to enter the values of n and p. It also has a spot to form the probability (with x) you want to find. Play around with the tabs **Standard** and **Between** to see what they do. You can change the inequality symbols as needed.

expl 5: According to the American Red Cross, 7% of people in the U.S. have the blood type O-negative. Use StatCrunch to answer these questions. If we select twenty people at random, what is the probability that …

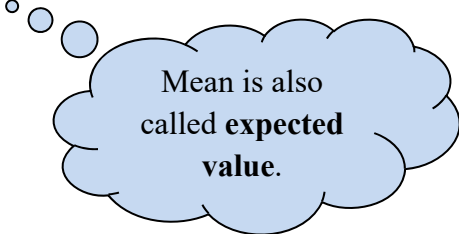a.) exactly 5 of them are O-negative?   c.) all 20 are O-negative?

b.) between 10 and 15, inclusive, are O-negative?   d.) less than 15 are O-negative?

Mean and Standard Deviation of Binomial Distributions:

We could use the formulas given in the last section, but there are easier formulas just for binomial distributions. Be aware and only use the following for binomial distributions!
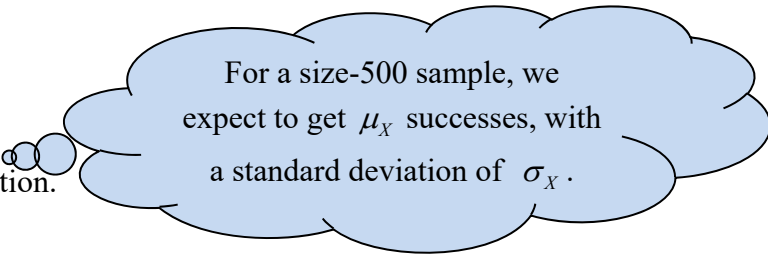
Definition: A binomial experiment with n independent trials and probability of success p has a **mean** and **standard deviation** given by the formulas $\mu_X = n \cdot p$ and $\sigma_X = \sqrt{np(1-p)}$.

> Mean is also called **expected value**.

expl 6: According to the American Red Cross, 7% of people in the U.S. have the blood type O-negative. A simple random sample of size 500 is obtained and the number of people with O-negative blood X is recorded.

a.) Find the mean (or expected value) for this distribution.

> For a size-500 sample, we expect to get μ_X successes, with a standard deviation of σ_X.

b.) Find the standard deviation for this distribution.

Worksheet: Mean and Standard Deviation of a Binomial Distribution:
We will review what makes a distribution binomial. We will find probabilities in the form $P(X=x)$ as well as the mean and standard deviation (and their meanings) of a sample binomial distribution.

Referring to the previous example, we can find that the probability that 45 people out of this sample of 500 are O-negative is about 0.015. So would it be **unusual** for us to get 45 out of 500 people that are O-negative? We have a slightly different metric to use for this question when it concerns binomial distributions.

Symmetry and the Binomial Distribution:
It turns out that, for a fixed value of p, as the number of trials n increases, the distribution becomes bell-shaped, or symmetric. As a rule of thumb, if $np(1-p) \geq 10$, the probability distribution will be approximately bell-shaped. What that means is that we can use the Empirical Rule we saw back in chapter 3. Recall, this states that for bell-shaped distributions, 95% of all outcomes lie between $\mu - 2\sigma$ and $\mu + 2\sigma$.

Empirical Rule for Unusual Events in Binomial Distributions:
First, be sure that $np(1-p) \geq 10$. If this is met, then you can say that 95% of all outcomes lie between $\mu - 2\sigma$ and $\mu + 2\sigma$. Any observation that lies outside this interval can be labeled as **unusual** since it occurs 5% of the time or less.

expl 7: Again, let's look at the sample of size 500 to determine blood type. Recall, 7% of people in the U.S. are O-negative. A simple random sample of size 500 is obtained and the number of people with O-negative blood X is recorded.

a.) Calculate $np(1-p)$ to verify it is greater than or equal to 10. Otherwise, this metric does *not* apply.

b.) Use your values from the previous example to calculate $\mu - 2\sigma$.

c.) Use your values from the previous example to calculate $\mu + 2\sigma$.

d.) What is the interval of "usual" outcomes? Would we consider a sample of 500 that yielded 45 O-negative people to be unusual?

e.) If you sampled 500 people and recorded that 120 were O-negative, what might you conclude?

7