UNIT 1: Did LeBron James Choke in the Playoffs?
Exploring Categorical Data

In the 2008 playoffs, NBA basketball star LeBron James of the Cleveland Cavaliers was criticized in the media for not performing up to his usual standards. Was this a fair criticism or was the media being too harsh?

To begin to answer this question, we need to compare his performance in the playoffs with his performance in the regular season. There are many ways to assess a basketball player's performance, such as points, rebounds, assists, turnovers, and shooting percentage. Many of the articles during the 2008 playoffs focused on his low shooting percentage, so that is what we will focus on as well. Specifically, we will focus on his three-point shooting percentage and ask: "Did LeBron James' three-point shooting ability decrease in the playoffs?"

In each unit of this course, we will open with an interesting sports question and then attempt to answer it with an investigative process that involves four components:

1. Formulating Questions
2. Collecting Data
3. Analyzing Data
4. Interpreting Results

These four components are discussed in great detail in the document "Guidelines for Assessment and Instruction in Statistics Education" which is endorsed by the American Statistical Association. For more information see: www.amstat.org/education/gaise/

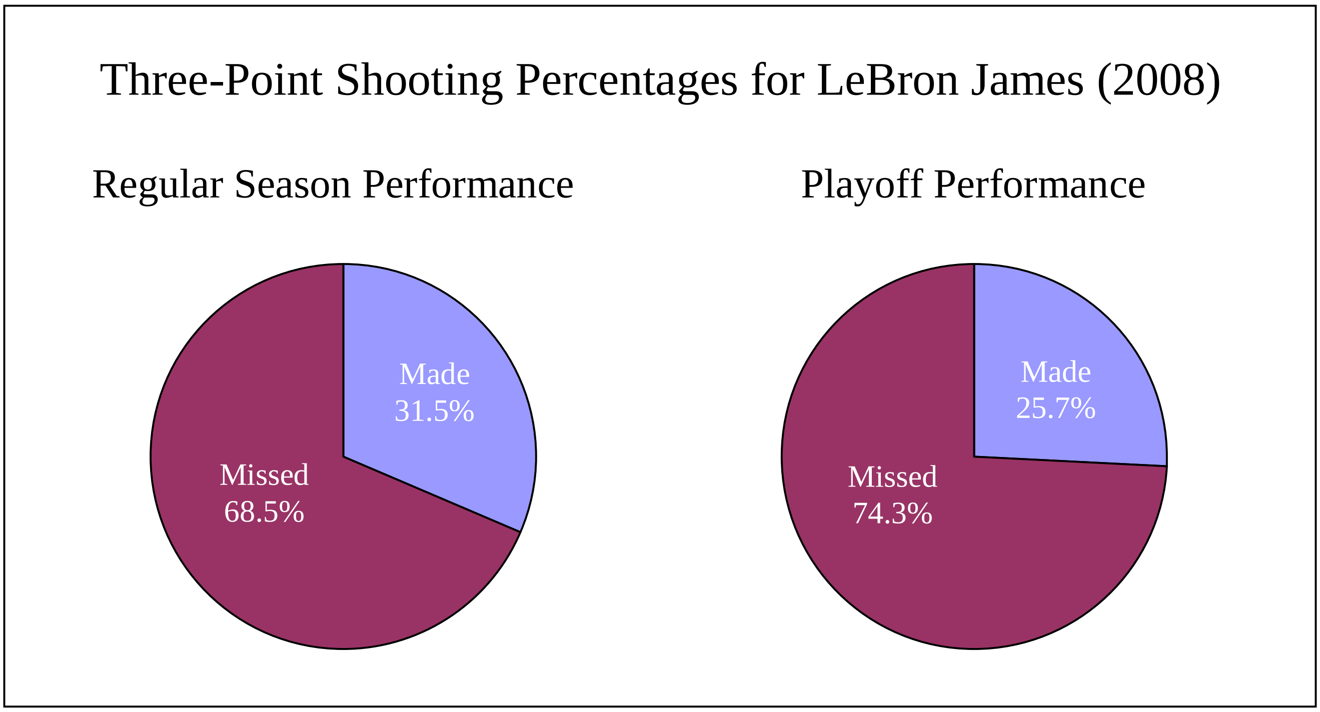
To answer the question "Did LeBron James' three-point shooting ability decrease in the playoffs?" we need to collect some relevant data. According to www.basketball-reference.com, LeBron James made 31.5% of his three-point attempts in the 2007-2008 regular season. However, in the 2008 playoffs, he only made 25.7% of his three-point attempts (18 made shots in 70 attempts).

Graphing Categorical Data

To begin the analysis of the data, we will learn a few different ways to visualize it using graphs. In fact, looking at graphs of the data should always be the first step of any data analysis. It is also one of the best ways to communicate the data with other people.

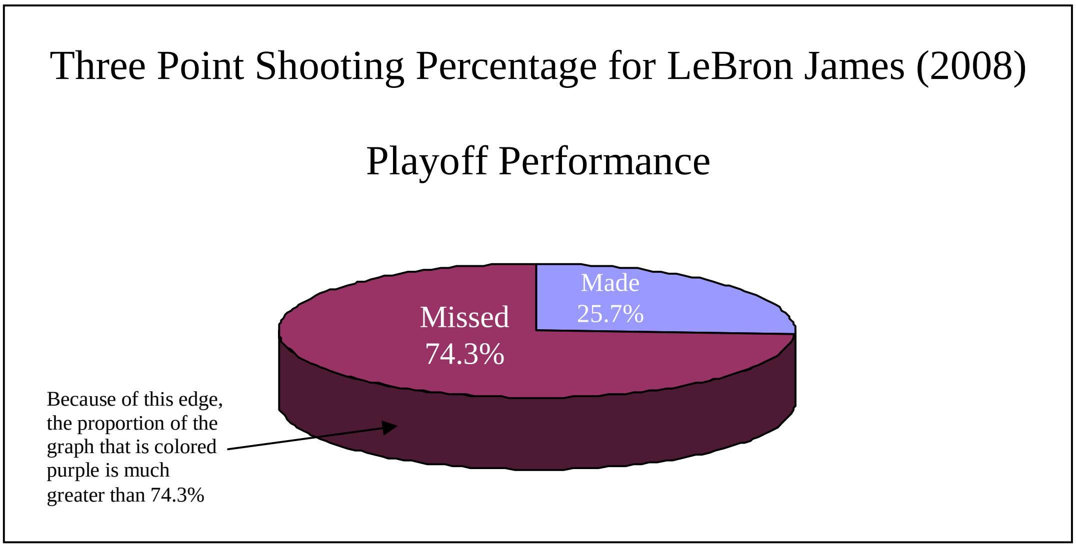
Data such as these are called **categorical data**, since the outcomes fall into categories. For example, in basketball the result of a shot is a categorical variable (made or missed). In baseball, the outcome of an at-bat is a categorical variable (hit, out, walk, etc.) Other types of data are called **numerical data**, since the outcomes are numerical in nature (the distance a golf ball travels, the number of rebounds a person has, etc.).

There are several ways to graph categorical data. A simple way is with a pie chart. Here are two pie charts, one for LeBron's three-point performance in the regular season and one for LeBron's three-point performance in the playoffs.

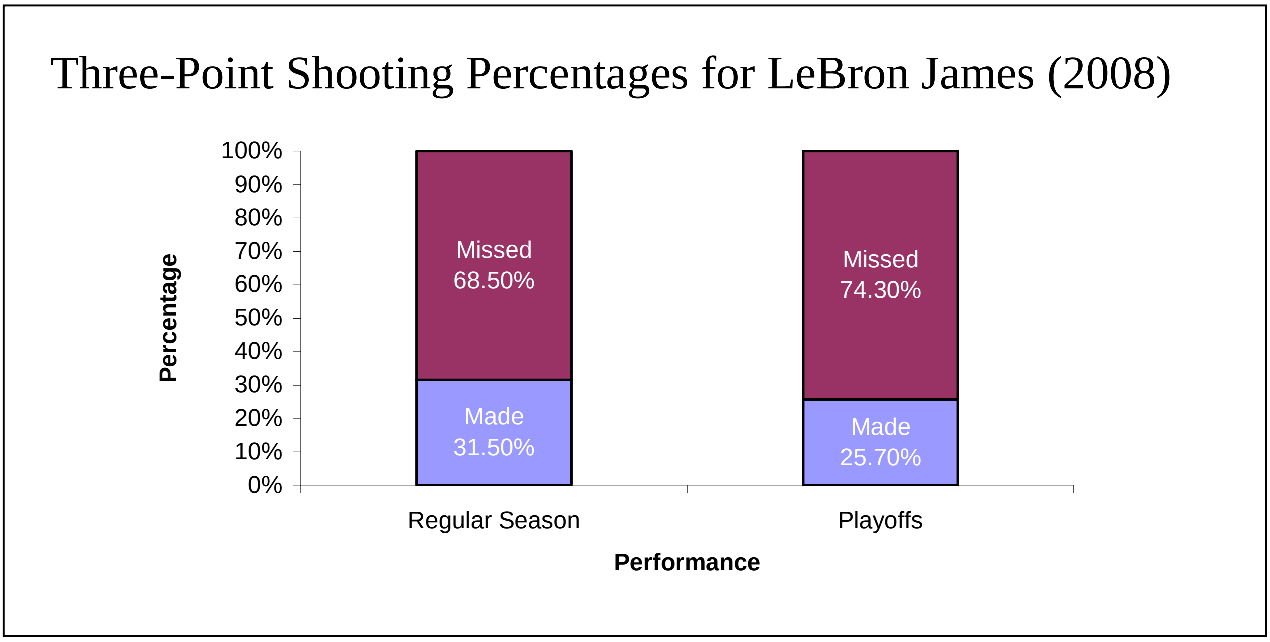


When making pie charts, especially when making comparisons, it is important to clearly label each chart, each category, and use consistent colors. For example, in both pie charts the same color should be used to represent a made shot. Also, the pie charts should be the same size for easy comparisons.

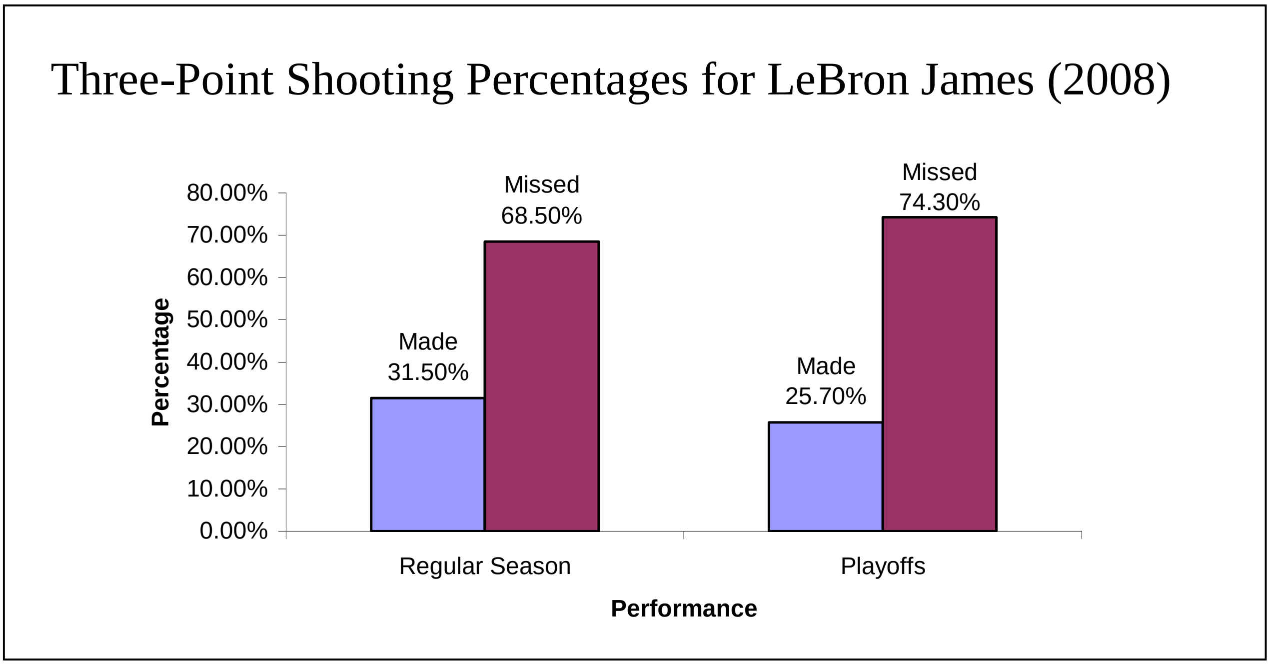
Finally, the charts should not violate the area principle. The **area principle** says that the area representing each category should be proportional to the number of observations in that category. A common type of pie chart that violates the Area Principle is a 3-D pie chart such as the one below. In this chart, the proportion of the chart that is colored purple is way more that 68.5% because of the 3-D view. So, avoid the temptation to make fancy looking graphs—they can be deceptive!



A rectangular version of a pie graph is called a segmented bar chart. Instead of dividing an area of a circle into regions proportional to the data, we divide a rectangle instead. This is much easier to do by hand and allows the graphs to be a little closer together and easier to compare. Remember not to violate the area principle by making the bars look 3-D!



The final way to graph categorical data is with a bar chart. In this type of chart, each category is given it's own bar and the height of the bar corresponds with the proportion of observations in that category. When comparing distributions, it is really important that the *proportion* in each category is displayed (rather than the number of observations in each category) so the graphs are on the same scale. Here is a comparative bar chart showing LeBron's distributions for the regular season and for the playoffs.



Note: All of these graphs were made using Microsoft Excel. Please see the Appendix for further details on how to construct these types of graphs.

Modeling Athletic Performance

Looking at the graphs, a simple and straight forward conclusion is easy to draw: LeBron James did shoot worse in the 2008 playoffs that he did in the regular season. However, this only focuses on LeBron's *Performance* in the playoffs. It doesn't tell us if his *Ability* also decreased.

Understanding the difference between *Ability* and *Performance* is central to any sophisticated analysis of sports data and will be the foundation of each unit in this course.

Consider a simple example: A basketball player is a 50% free-throw shooter. Suppose that during a game she gets fouled while attempting to shoot and is awarded two free throws. When she steps to the line, she misses both of them. Does this mean that her *Ability* as a free throw shooter has decreased from 50% to 0%? Of course not! This simply illustrates a fact that every athlete has experienced at one time or another: *Performance* isn't the same as *Ability*.

In this course, we will use the following model when addressing the issue of *Performance* versus *Ability*:

$$Performance = Ability + Random\ Chance$$

A model is simply a way to represent a particular phenomenon. In this case, we are trying to represent—or model—an athlete's *Performance*.

- *Performance* is what an athlete actually does in an athletic contest or series of contests. This is what is typically referred to as an athlete's statistics.
- *Ability* is a concept that imagines what an athlete would do if given an unlimited number of opportunities at a specific point in the athlete's career.
- *Random Chance* is an acknowledgement that there is variability in athletic *Performance*. In other words, even in identical conditions, a particular athlete's *Performance* will not always be the same.

To illustrate these concepts, let's return to the example of the 50% free throw shooter. The 50% in this context refers the shooter's *Ability*. To say that the shooter's *Ability* is 50% means that if this shooter were to be given the opportunity to take lots and lots of free throws, she should expect to make about 50% of them.

To simulate this shooter's *Performance* we can flip a coin letting heads represent a made free-throw and tails represent a missed free-throws. Assuming it is a fair coin, the coin has the *Ability* to land on heads 50% of the time just like the shooter has the *Ability* to make 50% of her free throws. Of course, the *Ability* to land on heads 50% of the time does not guarantee that the results of a series of coin flips will be exactly 50%.

For example, here are the results of 10 consecutive flips of a fair coin:

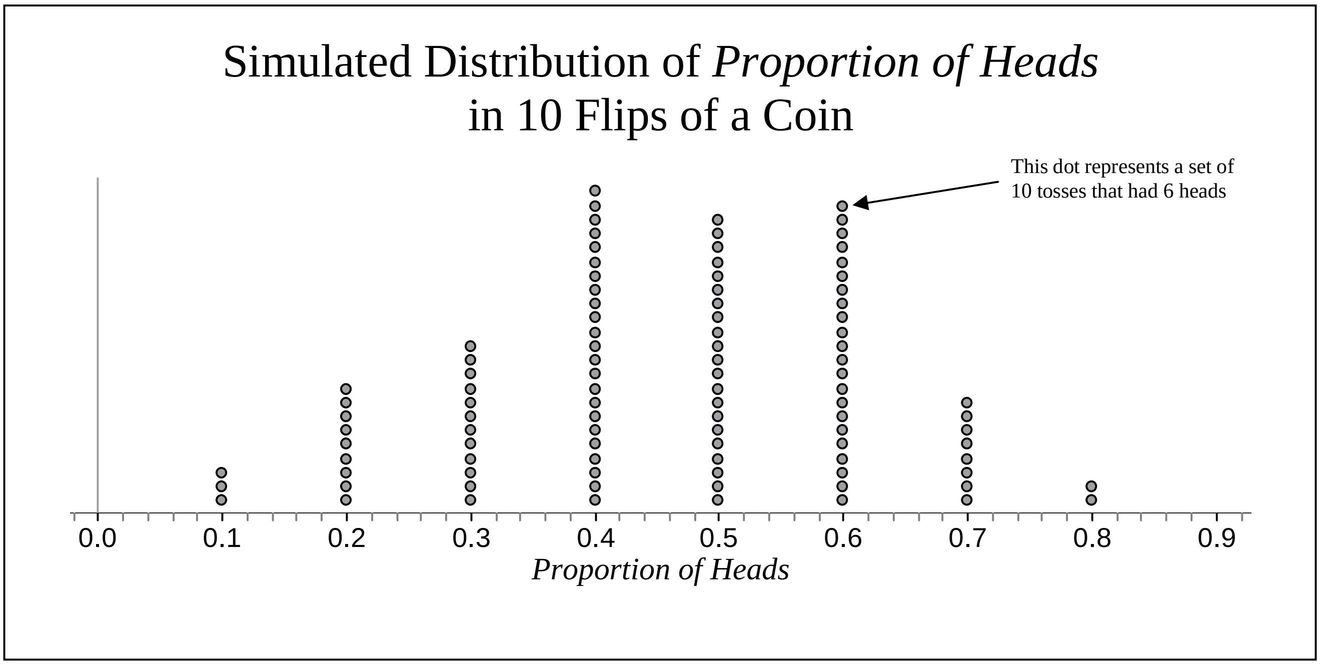
H H H T H T H H H H

As you can see, the *Performance* of the coin was 80% heads even when the *Ability* of the coin to land on heads was only 50%.

Here are the results of another 10 consecutive flips:

T T T H H T H T T T

In this case, the *Performance* of the coin was only 30% heads. This highlights one of the most important factors in any statistical analysis: the role of *Random Chance*. Below are the results of 100 sets of 10 consecutive coin flips. Each dot represents the proportion of heads in a particular set of 10 flips.



As you can see, the coin's *Performance* was exactly 50% in only 21 of the 100 sets. Furthermore, in some cases the *Performance* of the coin was quite different than 50% (in 3 of the sets, the coin's *Performance* was only 10% heads).

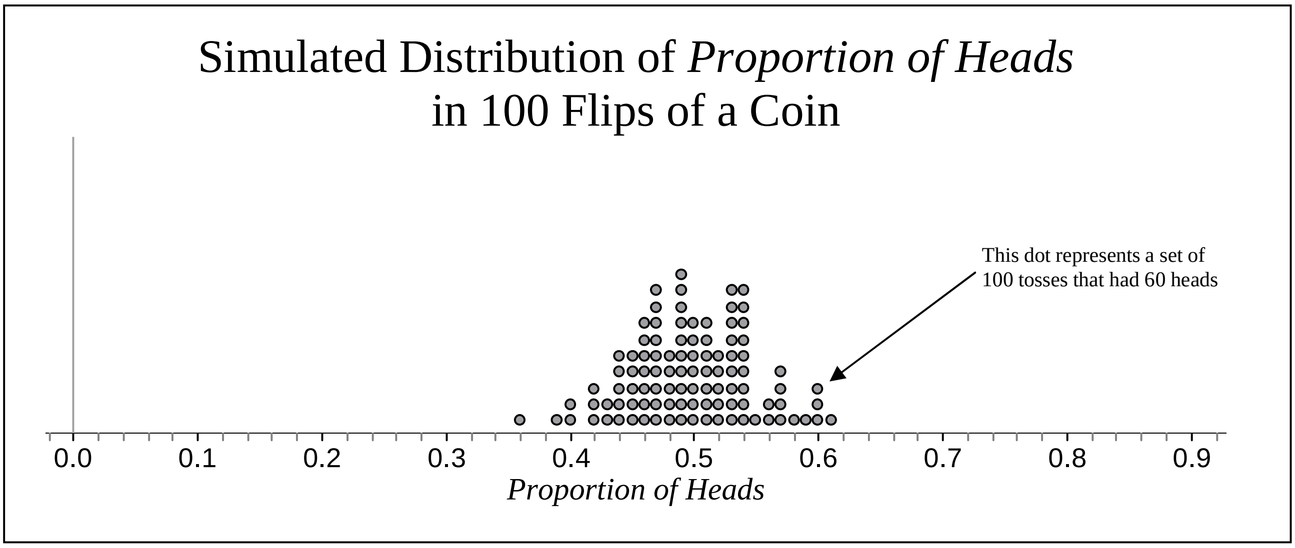
However, 10 flips (or 10 free throws) is not very many attempts. If we were to flip a coin many more times, its *Performance* will typically be much closer to its *Ability*.

Here are the results of 100 consecutive flips of a coin:

```
H H H H T H H T H T H T T T T H T T T T H H H T H
T H T T T H T T T H T H H T T T H H T T H H H H T
T T T T T H T H H T H T H H T H T T T H H T T T H
T H H T T H T H H T H H T H T T T T H H T T H H T
```

For this set of 100 flips, the coin's *Performance* was 46% heads (not very far from 50%).

Here are the results for 100 sets of 100 coin flips. Each dot represents the proportion of heads in a particular set of 100 coin flips.



As you can see, when flipping a coin 100 times its *Performance* is typically much closer to its *Ability* than when we flipped the coin only 10 times, even though the coin's *Performance* was rarely exactly 50%.

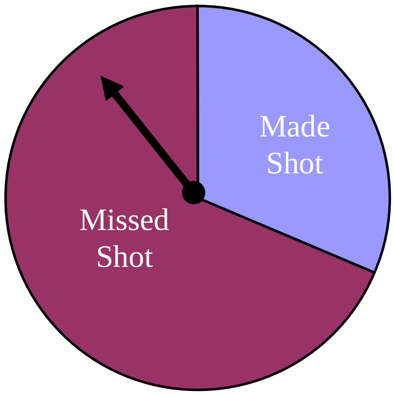
Relating this to sports, in the long-run, an athlete's *Performance* should be very similar to his or her *Ability*. However, in a small number of trials (shots, at-bats, passing attempts, shots on goal, etc.), an athlete's *Performance* can vary quite a bit from his or her *Ability*.

Back to LeBron

We are now ready to address the question that started the unit: "Did LeBron James' three-point shooting *Ability* decrease in the playoffs?"

Although it is impossible to know LeBron's true *Ability*, we will assume his true *Ability* to make three-point shots is 31.5%, which was his percentage in the regular season. What we want to know is how much variability we should expect to come from *Random Chance*. To find out, we need to simulate 70 shots (the number of three-point attempts LeBron had in the playoffs) assuming he has a 31.5% chance of making each shot (the same success rate as the regular season). Since there is no coin that will land on heads 31.5% of the time, we need another way to represent LeBron's shots.

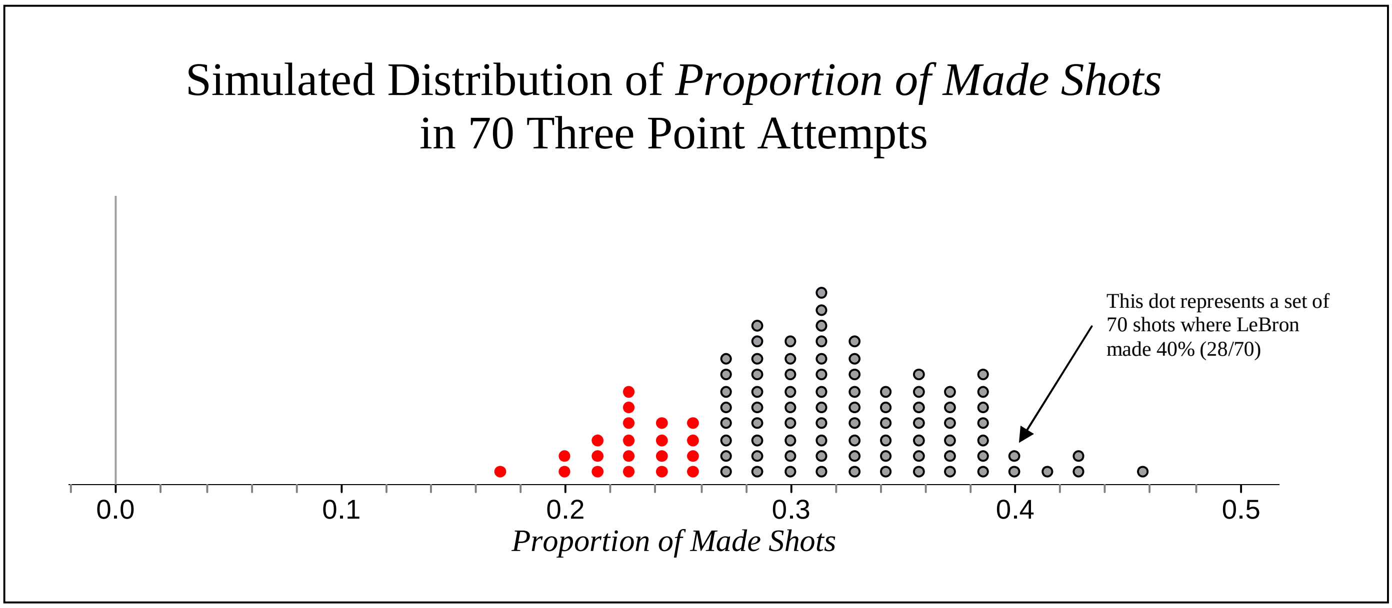
One way to do this is with a spinner. Below is a spinner with 2 regions: one with an area of 31.5% (representing a made shot) and the other with an area of 68.5% (representing a missed shot).



To simulate LeBron's playoff *Performance*, we will spin the spinner 70 times and keep track of the number of shots made. Here is one example (Y = made the shot, N = did not make the shot):

```
NNNYN YNNNN YNNNN NNNNY YNYNN YYNYN NYNNN
YNNNY NNNNN YYYNY YNYNN YYYYY NNYNN NNYNN
```

In this example, LeBron made 25 of 70 three-point shots which is 35.7% . His *Performance* was slightly better than his *Ability* and certainly better than his *Performance* in the actual 2008 playoffs (25.7%). Of course, this result is just one possibility for a shooter with a 31.5% *Ability* to make three-pointers. To see what other results are possible, we can do many more sets of 70 spins. Each dot in the plot below represents the outcome of a series of 70 shots for a 31.5% shooter. In other words, each dot represents what *could have* happened in the 2008 playoffs *if* LeBron's *Ability* stayed the same as in the regular season.



Even though we assumed LeBron James has a 31.5% *Ability* to make three-pointers for each set of 70 shots, his *Performance* varied quite a bit (from 17% to 46%). This illustrates what can happen in a short series of games such as the 2008 playoffs.

In the graph above, the red dots represent sets of 70 shots where LeBron's *Performance* was 25.7% or less, which was his actual *Performance* in the 2008 playoffs. Since 20 of the 100 the dots are red, we can conclude that shooting 25.7% is NOT an unusual outcome. In fact, we would expect him to shoot the same—or worse—in about 1 out of every 5 playoff appearances even if his *Ability* remained a constant 31.5%.

So, we are finally able to answer the question posed at the beginning of the unit, although it may not be a very satisfying answer. We asked "Did LeBron James' three-point shooting *Ability* decrease in the playoffs?" Our analysis tells us that while it is clear that his *Performance* was lower than his *Ability*, it is quite possible that his *Ability* didn't decrease and the sub-par *Performance* was due to *Random Chance*. In other words, we do not have convincing evidence that his *Ability* decreased.

A quick word of caution: saying that we don't have convincing evidence that his *Ability* decreased is NOT the same as saying that his *Ability* has stayed the same. For example, consider a man on trial for murder. If there is not convincing evidence that he is guilty, the jury must declare that he is not guilty. This does not mean that the defendant is innocent of the crime, it just means there is not enough evidence to convict him. Similarly, it is possible that LeBron's *Ability* did decrease in the 2008 playoffs, but not enough for us to have convincing evidence of the decrease.

This conclusion brings up a new question: What would be convincing evidence that LeBron's *Ability* has decreased? Looking at the graph above, it is possible that LeBron could shoot as poorly as 17% even when his *Ability* remained at 31.5%. Thus, any *Performance* lower than 17% would be fairly convincing evidence that the low percentage is due to something more than *Random Chance*—it would indicate that his *Ability* to make three-pointers really did decrease in the 2008 playoffs.

Even though we have already exonerated LeBron James from the criticisms he faced in the 2008 playoffs, there is another factor that we ignored in our analysis that works in his favor. During the playoffs, the opponents faced are, on average, better than the opponents faced in the regular season since only the better teams make the playoffs. So, it wouldn't be surprising at all to find that a player's playoff *Performance* was slightly worse that his regular season *Performance*. Adding this factor to the previous analysis gives LeBron another reason not to worry.

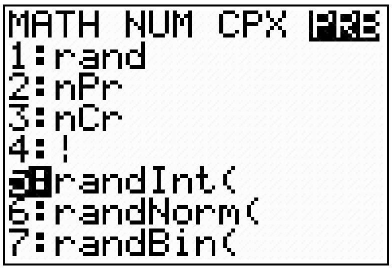
Conducting Simulations with Technology

Using a spinner to simulate 70 shots can be a little tedious, especially when you want to do many sets of 70 shots. Fortunately, we can use technology to help us make this process a little easier.

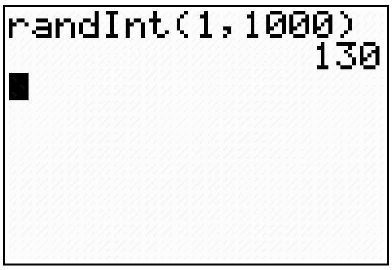
One way to do this is with a random number generator. A random number generator works like drawing numbers out of a hat. For example, suppose you had 10 ping pong balls in a hat and each had a number from 0 to 9. Then mix up the balls, draw one out at random, write the number down, and replace the ball. Then repeat this process over and over again. The results may look something like: 944783920394… If a random number generator is working correctly, then each number has an equal chance to show up on a given draw regardless of what numbers were drawn previously. In statistical language, this means that each number is independent of the others.

To simulate LeBron James' *Ability* to make three-pointers, we can use the numbers 1-1000. If a number from 1-315 comes up, this will represent making a three-pointer and if a number from 316 to 1000 comes up, this will represent missing a three-pointer. This way, there is still a 31.5% chance LeBron will make a three-pointer, which is what we assumed his *Ability* was based on his regular season statistics.

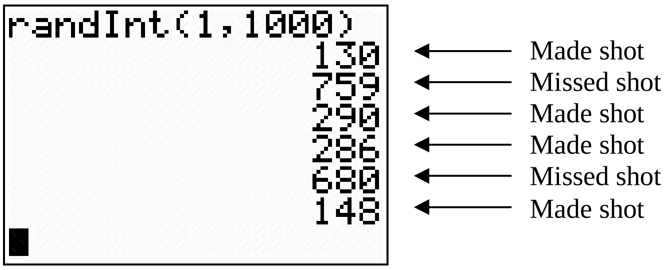
On the TI-84, we can generate random numbers using the Random Integer command. Press Math: Prb: 5 RandInt:



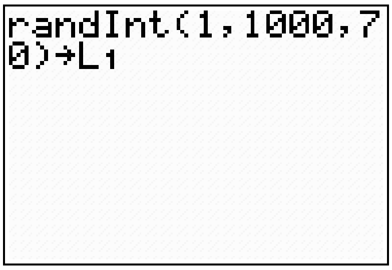
For example, entering RandInt(1,1000) will generate a single integer from 1 to 1000.

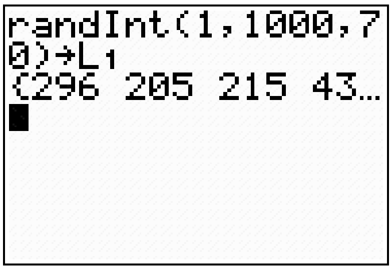


Doing this repeatedly can produce results such as these. Of course, your calculator will probably give you different random numbers!

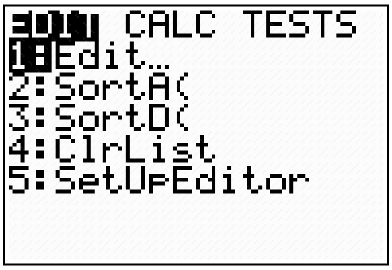


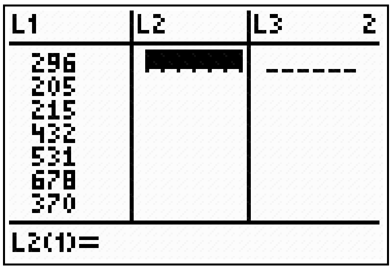
If we want to generate 70 integers at a time we would enter RandInt(1,1000,70). Rather than having all 70 numbers scrolling across the home screen, it may be easier to store these values to a list by entering the command RandInt(1,1000,70)→L1. The arrow can be found just above the On button (it is called Sto for store). L1 is the name of a specific list and can be found by pressing the 2nd button followed by the number 1.





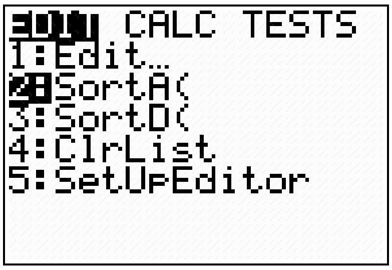
To view the list, press the Stat button and press Enter to choose Edit. This will display the first 3 lists: L1, L2, and L3 (scrolling to the right will reveal 3 more).

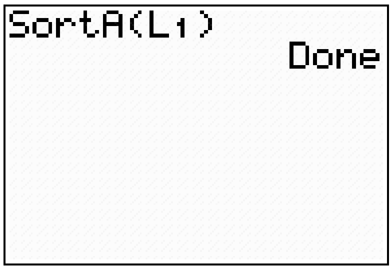


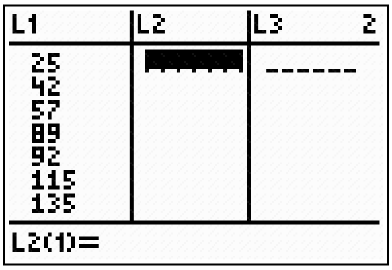


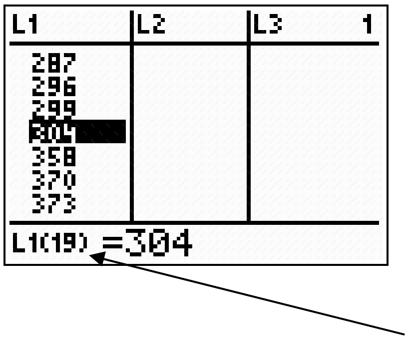
Scrolling down this list will reveal all 70 random numbers. To find out how many he made, simply count how often the numbers 1-315 showed up in the list.

To make the counting easier, it is possible to sort a list so the numbers are in order. To do this, press Stat and scroll down to SortA. This will sort the chosen list in ascending order (smallest to largest). For example, entering SortA(L1) will sort the values in list 1 from smallest to largest, making it much easier to count the number of made shots.









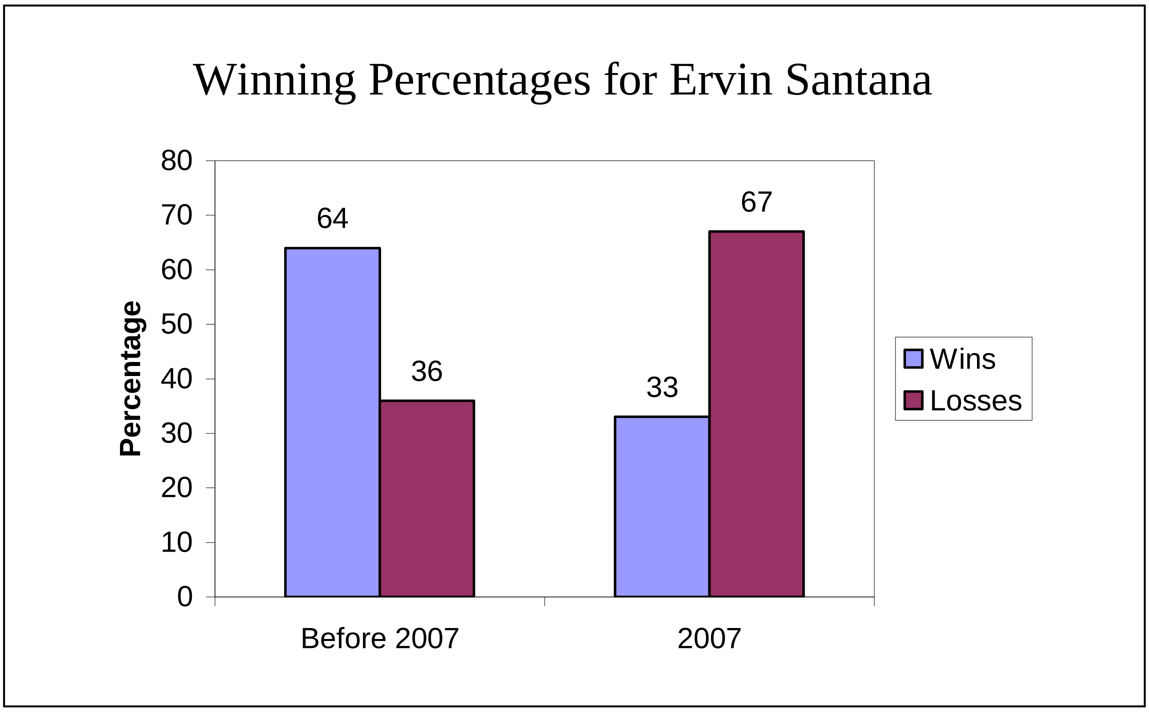
Note: Since the largest random number less than or equal to 315 was in the 19th position in the list, we can tell that LeBron made 19 of the 70 shots in this trial.

Another way to generate random integers is on the internet: http://www.random.org/integers/

Example: Ervin Santana, 2007 Los Angeles Angels

At the beginning of the 2007 season, fans of the Los Angeles Angels were excited about the prospects of one of their young pitchers, Ervin Santana. In 2005 Santana won 12 games while losing 8 (12-8) and in 2006 he improved to 16 wins and 8 losses (16-8). However, in 2007 he took a big step backwards winning only 7 games while losing 14 (7-14).

As always, we begin our analysis with a graph:



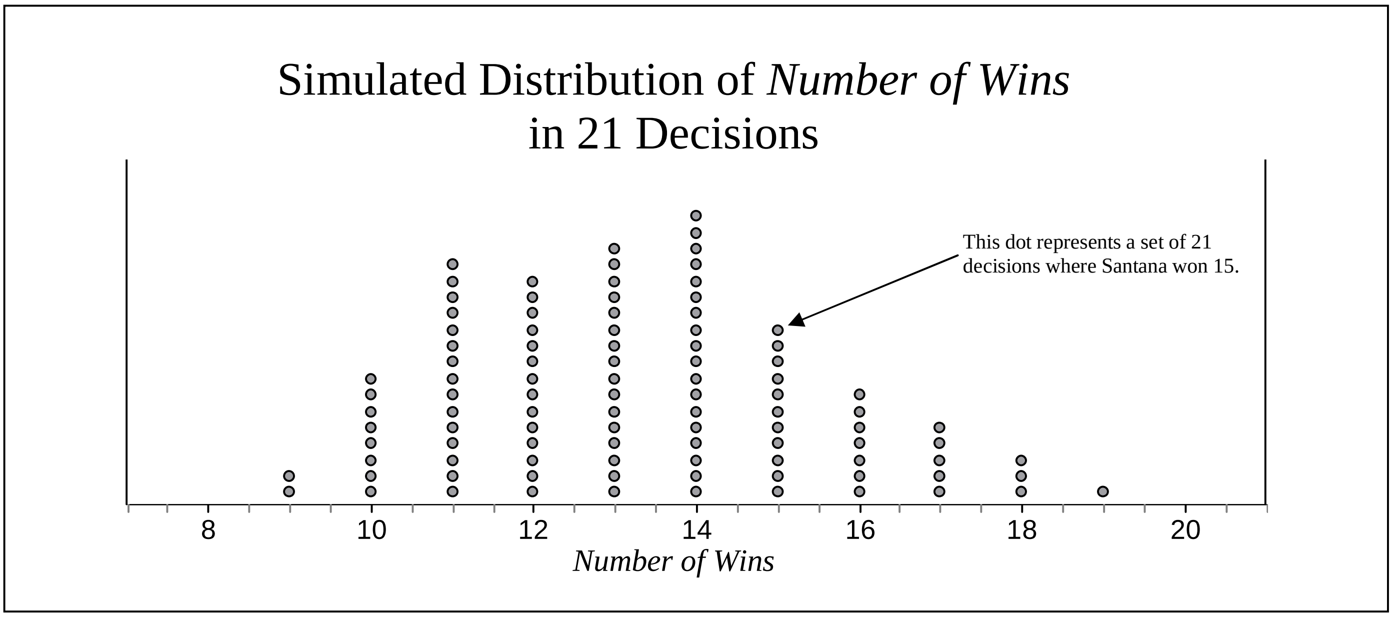
It is obvious that his *Performance* in 2007 was worse than expected, but did Santana's *Ability* to win games decrease in 2007? To continue our analysis, we need to make an assumption about Santana's *Ability* to win games. Based on his 28-16 W-L record in two seasons in the major leagues, we will assume he has the *Ability* to win 64% (28/44) of his games. Of course, this is really just an *estimate* of his true *Ability*, but it is the best estimate we have.

Next, we need to simulate Santana's 21 games in the 2007 season. Using random numbers from 1 to 100, we will let 1-64 represent a win (since Santana had a 64% chance of winning) and 65-100 represent a loss. Now, we will generate 21 random numbers from 1 to 100 to represent one possible season:

24 W	35 W	94 L	44 W	73 L	15 W	11 W
33 W	40 W	26 W	5 W	65 L	4 W	36 W
11 W	92 L	100 L	70 L	30 W	31 W	91 L

In this simulated season, Santana won 14 games and lost only 7, which was much better than his actual *Performance* in 2007 (7 wins and 14 losses).

To finish the simulation, we need to repeat this process many times to see how unusual it would be to win only 7 games in 21 decisions. Here are the results of 100 simulated seasons, assuming a 64% *Ability* to win games:



Note: In this example, *Number of Wins* is recorded instead of the *Percentage of Wins.* The graphs would look identical, except for the scale on the x-axis.

In 100 simulated seasons, the worst *Performance* was 9 wins and 12 losses. In 2007, Ervin Santana was still 2 wins short of even the lowest value we got by *Random Chance*! This analysis indicates that it wasn't just *Random Chance* that caused Santana to *Perform* worse than expected. Rather, we must conclude that his *Ability* to win games had decreased in 2007.

Still, we need to be careful about rushing to blame Santana. There are many reasons why his *Ability* may have decreased. It may have been poor conditioning or a poor mental approach during the 2007 season. However, it could have been an undisclosed injury (many young players try to "tough it out" so as to not look wimpy or lose their place in the lineup) or it could have been poor support from his teammates. After all, to win games the offense needs to score runs and the defense needs to make outs. It is possible that the Angels as a team were less supportive in 2007 than in the previous two seasons.

Connections: Looking Forward…Looking Back

In every unit of this course there will be a section dedicated to making connections with other parts of the course, both in the past and in the future. Since this is unit 1, we can't really look back! However, we can look forward to future units.

A recurring theme in future units will be our model for athletic *Performance*:

Performance = Ability + Random Chance

We will continue to use this model as a basis for our simulations and a way to explain the variability we see in athletic *Performances*.

Also in the future, we won't limit ourselves to comparing a single categorical variable to an assumed value. We will also learn how to compare two categorical variables with each other (such as *Winning Percentage at Home* vs. *Winning Percentage on the Road*), how to compare two quantitative variables (such as *Runs Scored with a Designated Hitter* vs. *Runs Scored without a Designated Hitter*), and how to analyze relationships between variables (such as *Average Driving Distance* and *Scoring Average* in the LPGA).

Stats 101: The Traditional Approach

In every unit of the course there will also be a section dedicated to making the connection between what we learned in the unit and how it might be presented in a more traditional statistics course. This way, things won't look so different if you have ever taken another statistics course or if you may take another statistics course in the future. Remember, the reasoning is the same, it's just that some of the vocabulary and methods may be different.

For example, in a traditional statistics course, an important distinction is made between a parameter and a statistic. A **parameter** is the true value of some quantity, such as the mean weight of *all* US college students or the true proportion of *all* US high school students with a tattoo. Unfortunately, without doing a census of the entire population, parameters are generally unknowable. Instead, we take a sample of the population and calculate a **statistic**—an estimate of the true mean weight or true proportion with a tattoo. This estimate will probably not be exactly correct, but it should be reasonably close if we collect data in a proper way.

Hopefully this sounds somewhat familiar. In the world of sports, the true value (parameter) we are interested in is an athlete's *Ability*. Like other parameters, we can never know an athlete's true *Ability* since we can never fully eliminate the effects of *Random Chance*. However, to estimate an athlete's *Ability* we can use his or her *Performances* just like we use data from a sample to estimate a parameter. This also means that our estimate of an athlete's *Ability* is also probably not exactly correct, but reasonably close to the truth.

Another concept we covered in this unit is the **Law of Large Numbers** which says that our estimates will typically be closer to the truth when we have a larger sample of data. Quoting from earlier in the text: "Relating this to sports, in the long-run, an athlete's *Performance* should be very similar to his or her *Ability*. However, in a small number of trials (shots, at-bats, passing attempts, shots on goal, etc.), an athlete's *Performance* can vary quite a bit from his or her *Ability.*"

Finally, simulations are not the only way to estimate how likely it is to see certain events such as LeBron James making 25.7% or fewer of his three-pointers. The Binomial Distribution is one way to calculate the probability of this kind of outcome (see box below). Another way is to approximate the Binomial Distribution with a Normal Distribution and do a "One sample *z* test for a proportion" also called a "one proportion *z* test."

MATH NOTE: The Binomial Distribution

Instead of using simulations to estimate the probability that a basketball player makes a certain number of shots, we can use the Binomial Distribution. To use this distribution, 4 conditions must be satisfied:

1. There are a fixed number of trials (in the case of LeBron James, a fixed number of shots: n = 70)
2. There are only 2 outcomes for each trial (in this case, he either makes the shot or he doesn't)
3. Each trial is independent (in this case, this means that the probability he makes a shot isn't affected by the results of previous shots. Whether or not this is a safe assumption is investigated more thoroughly in unit 3)
4. The probability of success remains the same (in this case, we assume that his *Ability* to make three-pointers is not changing and always equals 31.5%)

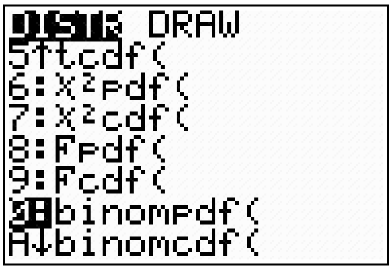
If these conditions are all satisfied, then the probability a player makes x shots is given by the following formula where p = the probability of success and n = the number of attempts.

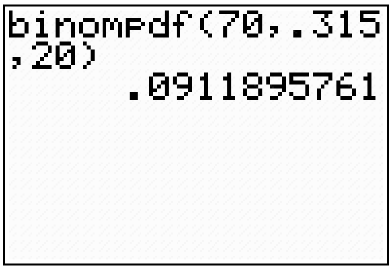
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ where } \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

For example, to find the probability that LeBron makes exactly 20 three-pointers in 70 attempts, we would calculate:

$$P(X = 20) = \binom{70}{20} 0.315^{20} (1-0.315)^{70-20} = 0.091$$

To do this calculation on the TI-84, press DISTR (2^{nd} VARS): 0 binompdf(70, .315, 20)





The TI-84 and many computer software packages can also randomly generate values from a binomial distribution. For example, to simulate the number of shots LeBron makes in 70 attempts on the TI-84, go the Math:Prb menu and enter RandBin(70,0.315). The output of this command will be the number of made shots in 70 attempts. This is a much quicker, but more abstract, method than spinning a spinner 70 times or looking at 70 random numbers.

For Practice

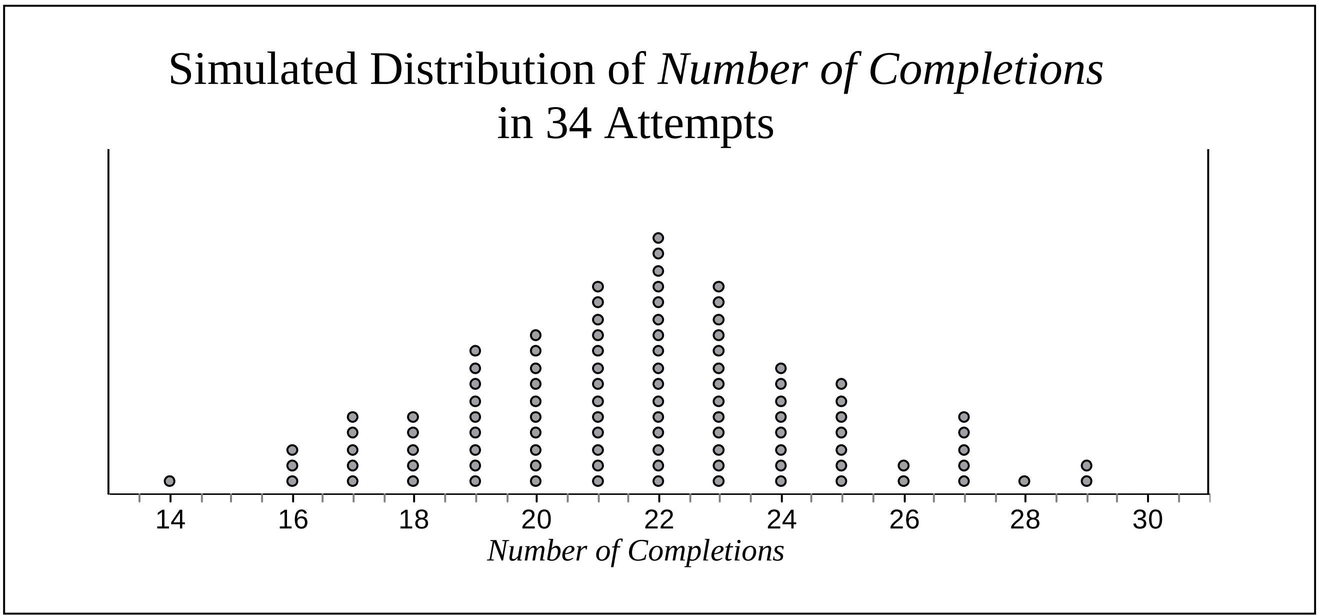
1. In the 2007 season baseball player Alex Rodriguez hit 54 homeruns in 583 at-bats. In the 2007 playoffs, he hit 1 homerun in 15 at-bats. Use an appropriate graphical display to compare these *Performances*. Make the display by hand or use computer software.

2. Over his career in the regular season, hockey goalkeeper Patrick Roy saved 25,807 shots in 28,353 attempts. In his career in the playoffs, he saved 6565 shots in 7149 attempts. Use an appropriate graphical display to compare these *Performances*. Make the display by hand or use computer software.

3. Over his football career, quarterback Peyton Manning has completed 64% of his pass attempts. Explain how you would simulate a football game where he makes 34 pass attempts. Give an example of one simulated game.

4. In 2008, golfer Annika Sorenstam had a driving accuracy of 71%. That is, on par 4 and par 5 holes, her tee shot landed in the fairway 71% of the time. Explain how you would simulate a round of golf where she attempts 15 drives. Give an example of one simulated round.

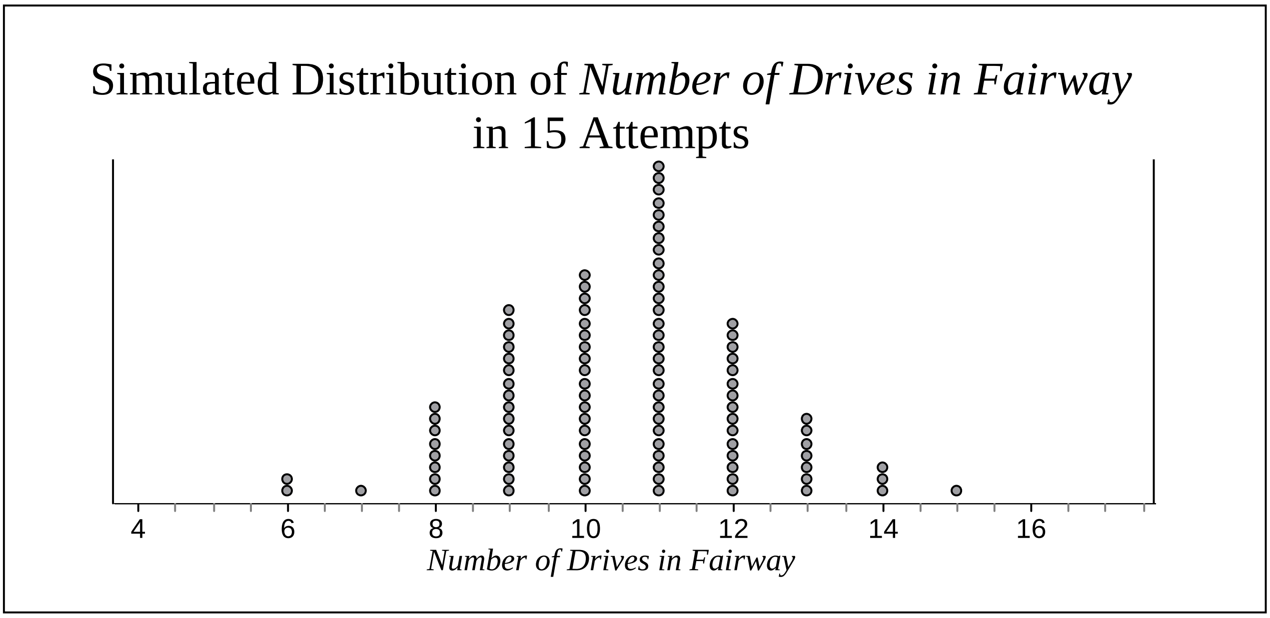
5. Here are the results of 100 simulated football games where Peyton Manning attempts 34 passes, assuming he has the *Ability* to complete 64% of his passes. Note: The Number of Completions is recorded here instead of the percentage of completions so the values on the horizontal axis will be integers.



a) In the 100 simulated football games, what was Manning's best *Performance*? Worst *Performance*?
b) In this context, explain the difference between *Ability* and *Performance*.

c) Suppose that in his first actual game next season Manning attempts 34 passes and completes 25 of them. Would this be a surprising *Performance*? Explain.
d) Suppose that he completed only 10 of his 34 attempts in his first actual game next season. Would this be a surprising *Performance*? Explain.
e) What actual outcomes would lead you to believe that Manning's *Ability* had changed?
f) Based on the results of this suppose that you decide that Manning's *Ability* has changed. What are some possible causes of this change?

6. Here are the results of 100 simulated rounds of golf where Annika Sorenstam attempts 15 drives, assuming she has the *Ability* to hit 71% of her drives in the fairway. Note: The Number of Completions is recorded here instead of the percentage of completions so the values on the horizontal axis will be integers.



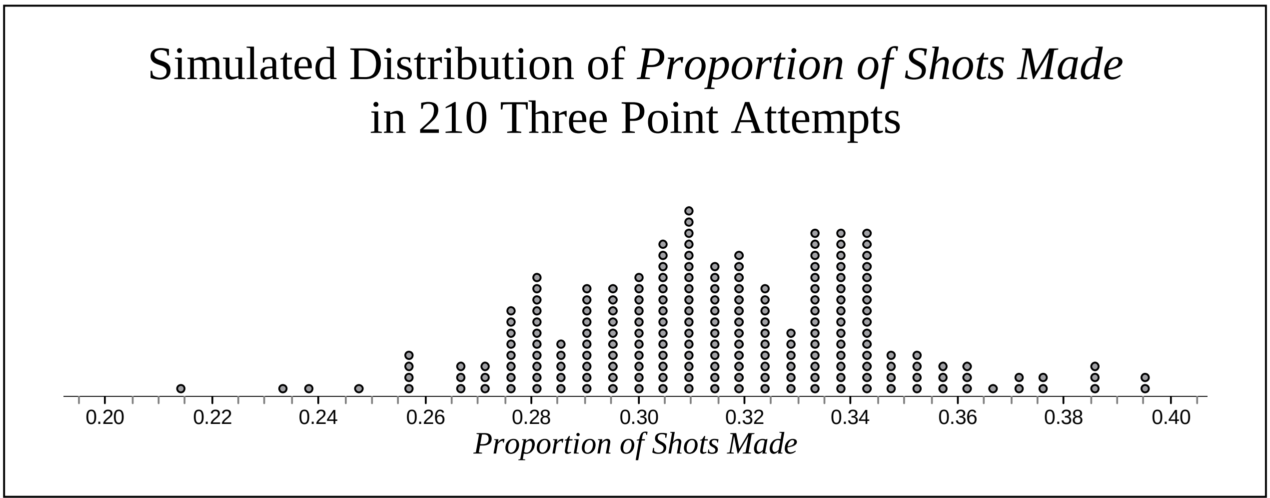
a) In the 100 simulated rounds, what was Sorenstam's best *Performance*? Worst *Performance*?
b) In this context, explain the difference between *Ability* and *Performance*.
c) Suppose that in her next actual round Sorenstam attempts 15 drives and lands 12 of them in the fairway. Would this be a surprising *Performance*? Explain.
d) Suppose that she lands only 5 of her 15 drives in the fairway in her next actual round. Would this be a surprising *Performance*? Explain.
e) What actual outcomes would lead you to believe that Sorenstam's *Ability* had changed?
f) Based on the results of this suppose that you decide that Sorenstam's *Ability* has changed. What are some possible causes of this change?

7. Suppose that a high school volleyball player has made 85% of her serves so far this season. In a particular match, she only makes 5 of her 10 serves and is criticized by her coach. Is the criticism deserved? Use the methods of this unit to formulate a response.

8. Suppose that a high school basketball player has made 60% of his free throws this season. In a particular game, he only makes 1 out of 5 and is criticized by his coach. Is the criticism deserved? Use the methods of this unit to formulate a response.

9. Returning to the LeBron James example from the beginning of the unit, suppose that he attempted 210 three-point shots in the playoffs (instead of 70) and made 54 (instead of 18). This results in the same shooting percentage as the actual 2008 playoffs (25.7%). Assuming LeBron's *Ability* is still 31.5%, how likely is it for his *Performance* to be 25.7% or lower? To help answer this question, 100 trials of a simulation were conducted. For each trial, 210 shots were simulated assuming that there is a 31.5% chance of making each shot. Then, the proportion of made shots was recorded.

a) Based on the results of the simulation below, is it possible that LeBron's *Ability* stayed the same and his poor *Performance* was just due to random chance? Explain.



b) How did the increase in the number of attempts (from 70 to 210) change the chances that LeBron makes 25.7% or fewer shots?

c) Based on your answer to (b), what would you conclude about LeBron's ability if he made only 257 shots in 1000 attempts?

For Investigation:

Using a website such as www.baseball-reference.com and the methods of this unit, investigate if a certain player performs better (or worse) in particular circumstances (such as the playoffs or against a certain team). Simply enter a players name in the search box. Once you have found that players home page, you will see all sorts of statistics, including regular season and playoffs. Clicking on "splits" will break down the player's statistics into many subcategories.

Other similar websites include:

- www.pro-football-reference.com
- www.basketball-reference.com
- www.hockey-reference.com
- www.sports-reference.com/olympics/

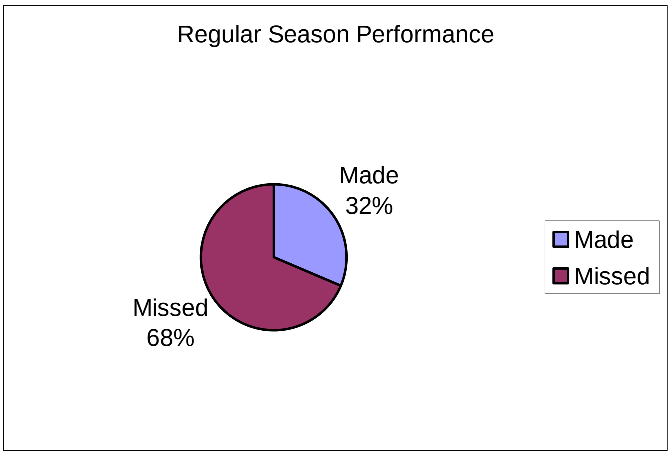
Making Pie Charts on Microsoft Excel:
Using LeBron's Regular Season *Performance* (31.5% Made and 68.5% Missed):

1. Enter the data so it looks like the following:

	A	B
1	31.5	Made
2	68.5	Missed

2. Highlight the numbers in column A and press the Chart Wizard Icon or open the Insert menu and choose Chart. This will open up the Chart Wizard.
3. Choose Pie from the list of Standard Types and select the first chart sub-type. Press Next.
4. In step 2, select the Series Tab at the top and in the box for Name, enter "Regular Season Performance". Also, in the box for Category Labels, use your cursor to highlight the cells in column B (Made, Missed). Press Next.
5. In step 3, select the Data Labels Tab and check the Category Labels and Percentage boxes. Press Next.
6. In step 4, press Finish.

Here is what your graph should look like:



Making Segmented Bar Charts on Microsoft Excel:
Comparing LeBron's Regular Season Performance (31.5% made, 68.5% missed) with his Playoff Performance (25.7% made, 74.3% missed):

1. Enter the data so it looks like the following:

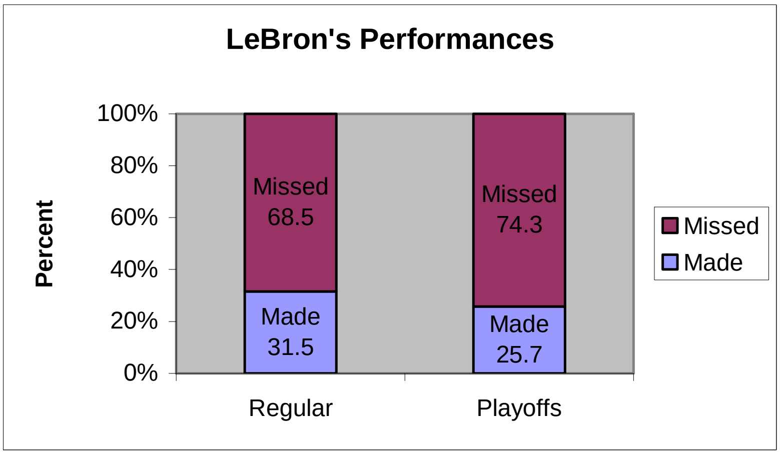
	A	B
1	31.5	25.7
2	68.5	74.3
3	Regular	Playoffs

2. Highlight the numbers in rows 1 and 2 and press the Chart Wizard Icon or open the Insert menu and choose Chart. This will open up the Chart Wizard.
3. In step 1, choose the first Standard Chart type (Column) and the third Chart sub-type (100% stacked column). Press Next.
4. In step 2, select the Series tab at the top. While Series 1 is highlighted, enter "Made" in the Name box. Now, highlight series 2 and enter "Missed" in the Name box. At the bottom, in the box for

Category (X) axis labels, use your cursor to highlight the names in row 3 of the spreadsheet. Press Next.

5. In step 3, enter "LeBron's Performances" in the Title Box and "Percent" in the Value (Y) axis box. Choose the gridlines tab and turn the gridlines off. In the Data Labels tab, check the boxes for Series Name and Value. For the Separator, choose New Line. Press Next.
6. In step 4, press Finish.

Here is what your graph should look like:



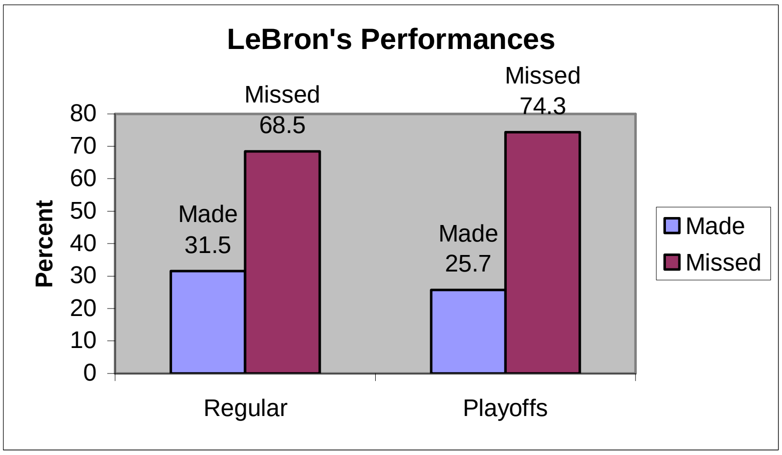
Making Comparative Bar Charts on Microsoft Excel:
Comparing LeBron's Regular Season Performance (31.5% made, 68.5% missed) with his Playoff Performance (25.7% made, 74.3% missed):

1. Enter the data so it looks like the following:

	A	B
1	31.5	25.7
2	68.5	74.3
3	Regular	Playoffs

2. Highlight the numbers in rows 1 and 2 and press the Chart Wizard Icon or open the Insert menu and choose Chart. This will open up the Chart Wizard.
3. In step 1, choose the first Standard Chart type (Column) and the first Chart sub-type (clustered column). Press Next.
4. In step 2, select the Series tab at the top. While Series 1 is highlighted, enter "Made" in the Name box. Now, highlight series 2 and enter "Missed" in the Name box. At the bottom, in the box for Category (X) axis labels, use your cursor to highlight the names in row 3 of the spreadsheet. Press Next.
5. In step 3, enter "LeBron's Performances" in the Title Box and "Percent" in the Value (Y) axis box. Choose the gridlines tab and turn the gridlines off. In the Data Labels tab, check the boxes for Series Name and Value. For the Separator, choose New Line. Press Next.
6. In step 4, press Finish.

Here is what your graph should look like:



Note: If you double click on the graph, you can change other aspects of the display, including colors, etc.