Unit 3: Does "The Zone" Exist?
Investigating Independence

If you listen to announcers, fans, or even athletes themselves, you will often hear the phrase "in the zone." What exactly does this mean? To an athlete it might mean that everything is working well and the chances of success are higher than normal. An announcer might use this phrase when an athlete has several successful plays in a row.

A statistician might use this phrase if the probability of success increases based on previous successes. That is, a player is in the zone if his or her *Ability* to be successful is higher following a successful *Performance* than following an unsuccessful *Performance*.

In symbolic terms, we can say a player is "in the zone" if:

$$P(\text{success} \mid \text{success}) > P(\text{success} \mid \text{failure})$$

Note: The vertical bar in the parentheses is read "given" so that P(A | B) is read the probability of event A occurring given than event B has occurred.

Using more statistical language, if a player is "in the zone" then the chances of being successful <u>depend</u> on the outcome of previous *Performances*. Therefore, if attempts are <u>independent</u>, then the *Ability* to be successful on a given outcome does *not* change based on previous *Performances*.

In symbolic terms, two attempts are independent if:

$$P(\text{success} \mid \text{success}) = P(\text{success} \mid \text{failure})$$

That is, the probability of success is the same following a success and following a failure.

One simple way to illustrate this is with free throw shooting, since free throws usually come in pairs. If the two shots are independent, then the *Ability* to make the second shot is the same, no matter if the first shot is made or the first shot is missed. In symbolic terms, the two shots are independent if:

$$P(\text{make the } 2^{nd} \mid \text{made the } 1^{st}) = P(\text{make the } 2^{nd} \mid \text{missed the } 1^{st})$$

Example: Shaquille O'Neal

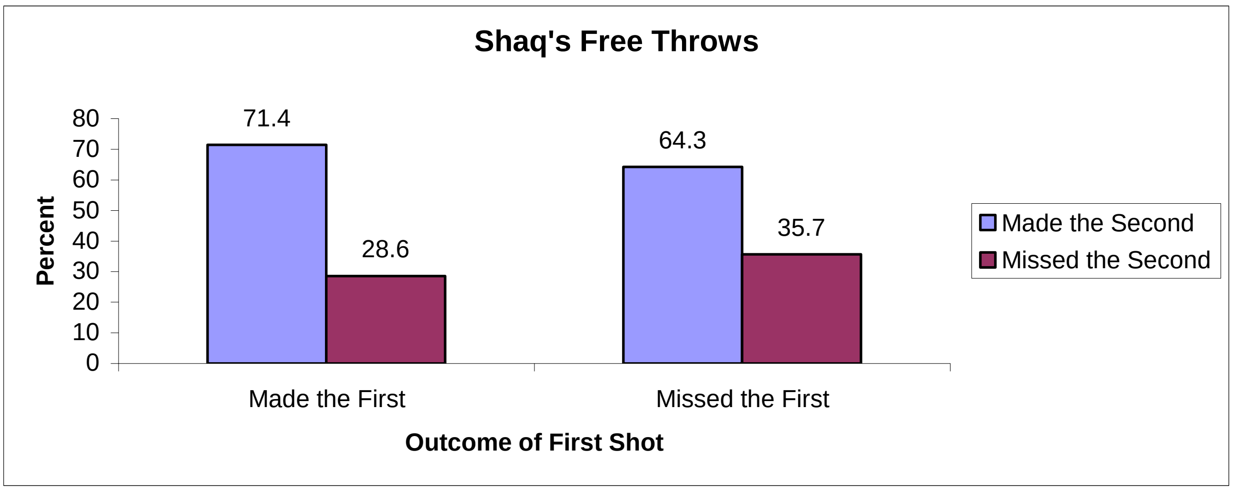
Over his career, Shaquille O'Neal has been known for his problems at the free throw line. Even though he has a lower percentage than most, this information doesn't tell us anything about whether or not his free throw attempts are independent. To investigate, we will look at Shaq's *pairs* of free throws in a 10 game stretch (February 26 through March 15, 2009). This means we will only be looking at the times when he attempted two free-throws, so the possible outcomes are YY if he makes both, YN if he makes the first and misses the second, etc.

Here are the results:

YY	YY	NY	YY	NY	YY	NY	YN	NN	NY	YY	YY
NY	NN	YY	NN	YN	YY	NN	YN	YN	NY	YY	YY
NY	NY	NN	NY								

The following table and graph summarizes these results:

	Made the first	Missed the first	
Made the second	10	9	19
Missed the second	4	5	9
	14	14	28



Checking our rule for independence we see that:

P(made the 2^{nd} | made the 1^{st}) = 10/14 = 71.4%
P(made the 2^{nd} | missed the 1^{st}) = 9/14 = 64.3%
Difference = 7.1%

Since his *Performance* is not the same in these two situations, it might seem that Shaq's *Ability* to make the second free throw is higher if he makes the first free throw. However, it is possible that his *Ability* does not depend on the outcome of the first shot and the difference in *Performance* we see (7.1%) could be due to *Random Chance*.

To see if Shaq gets "in the zone" when shooting free throws, we want to test:

H_0 : Shaq's *Ability* to make a free throw is the same following a made free throw and following a missed free throw

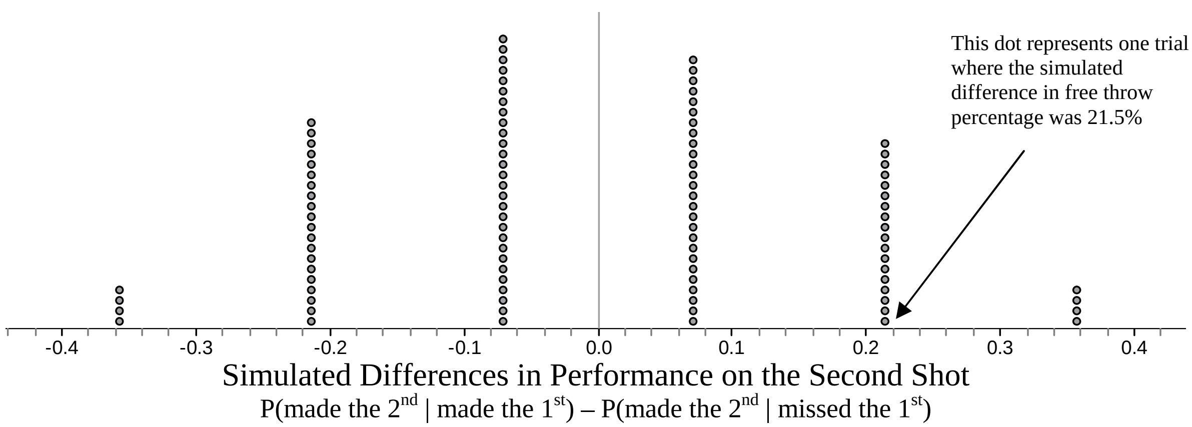
H_a : Shaq's *Ability* to make a free throw is greater following a made free throw than following a missed free throw.

To see how likely it is to get a difference of 7.1% by chance alone, we will use the same simulation technique we learned in Unit 2. To do this by hand, take 28 equally sized cards and write Y on 19 of them (to represent the 19 times Shaq made the second free throw) and write N on 9 of them (to represent the 9 times Shaq missed the second free throw). Shuffle these cards up and make two piles of 14: the first pile represents the 14 times Shaq made the first free throw and the second pile represents the 14 times Shaq missed the first free throw.

Here an example of one trial of the simulation:

After making the first shot:	YNYYY YYYYY YNNY	11/14 = 78.6%
After missing the first shot:	YYNNY NNYYY YNNY	8/14 = 57.1%
		Difference = 21.5%

Using technology to complete 100 trials of the simulation, we get the following distribution of possible differences:



Recalling that Shaq's actual difference in *Performance* was 7.1%, we can see that we do not have strong evidence that Shaq's *Ability* to make free throws increases when he makes the first free throw. When we start with the assumption that his *Ability* is the same in both cases, a difference of 7.1% or higher was not unusual at all (p-value $\approx$ .48). Thus, the apparent increase could have easily been due to *Random Chance* alone and we should not conclude that Shaq gets "in the zone" based on this data.

Note: To estimate the p-value we count how many of the 100 trials gave differences at least as extreme as the 7.1% Shaq had in real life. In this case, 48 of the 100 trials produced differences of 7.1% or greater just by *Random Chance*.

Finally, it is important to notice that in the alternative hypothesis we specified that we were looking for an increase in *Ability*. If we were simply testing if the shots are not independent, we would be looking for evidence of an increase *or* decrease in *Ability*.

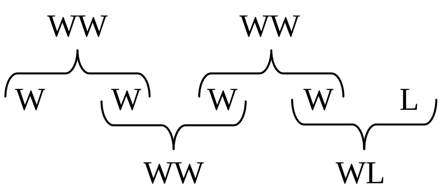
Another example: Can a Team be "In the Zone?"

One way to assess whether or not a team is "in the zone" is to consider their ability to win a game following a victory compared to their ability to win a game after a loss.

Here are the game results for the 2008 Michigan State softball team (in order)

WWWWL	LLLLL	LLLWW	WWWWL	LWWLW	LWLLW
WWWWW	LLWLL	WLWLL	WWLLW	LLLLW	L

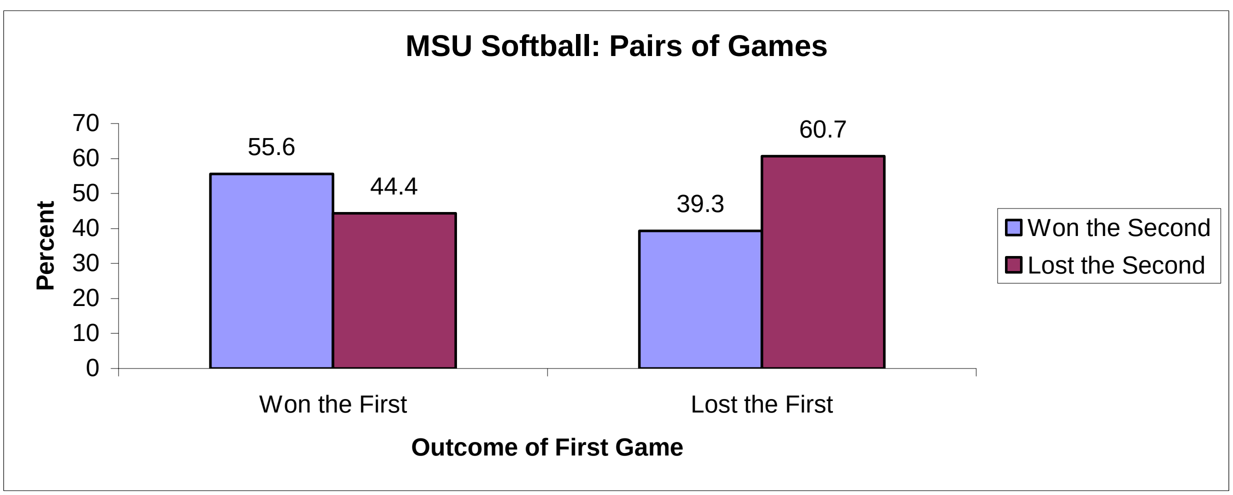
To investigate independence, we want to look at *pairs* of games. For example, if we consider the first five games, we actually have 4 *pairs* of games:



etc.

The following table and graph summarizes the results for each *pair* of consecutive games:

	Won the first	Lost the first	
Won the second	15	11	26
Lost the second	12	17	29
	27	28	55



Checking our rule for independence we see that:

P(won the 2^{nd} | won the 1^{st}) = 15/27 = 55.6%
P(won the 2^{nd} | lost the 1^{st}) = 11/28 = 39.3%
Difference = 16.3%

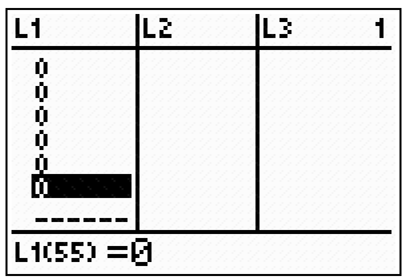
Since the MSU softball team's *Performance* after a victory was much better after a victory than after a defeat, it appears that the team does get "in the zone." However, like always, it is possible that their *Ability* to win games does not depend on the outcome of the previous game and the difference in *Performance* was due to *Random Chance*.

So, we will test the following hypotheses:

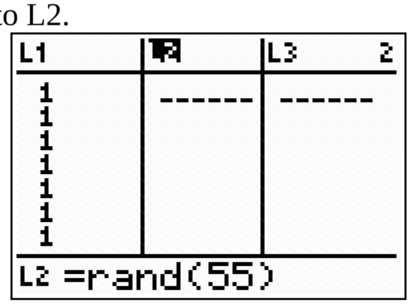
H_0 : The *Ability* of the MSU softball team to win a game is the same following a victory and following a loss.
H_a : The *Ability* of the MSU softball team to win a game is greater following a victory than following a loss.

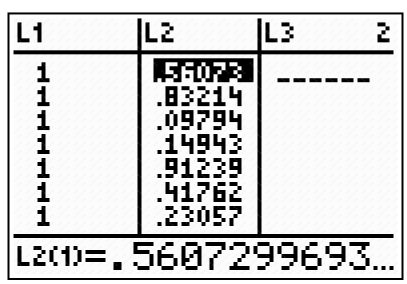
To see how likely it is to get a 16.3% difference in *Performance* we will conduct a simulation, this time with a TI-84:

1. Open the list editor (Stat: Edit) and enter 26 1's and 29 0's into L1. In this case, the 1's represent the 26 second-game wins and the 0's represent the 29 second-game losses.

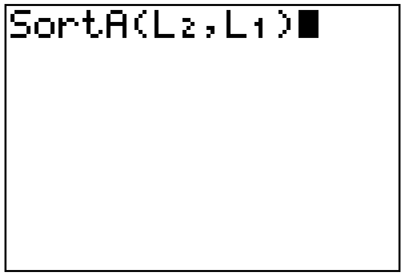


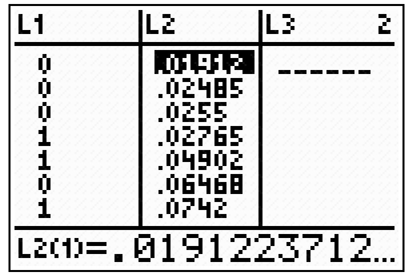
2. Next, we need to randomly shuffle these 1's and 0's into two groups: the first group should be size 27 to represent the 27 times the first game was a win and the second group should be size 28 to represent the 28 times the first game was a loss. To do this, go to the heading of L2 and enter the following formula: Rand(55) and press enter (the Rand command is in the Math:Prb menu). This puts 55 random numbers into L2.





3. Then, from the home screen, enter the command SortA(L2,L1) and press enter (the SortA command is in the Stat:Edit menu). This will put the random numbers in order and shuffle the 1's and 0's at the same time.





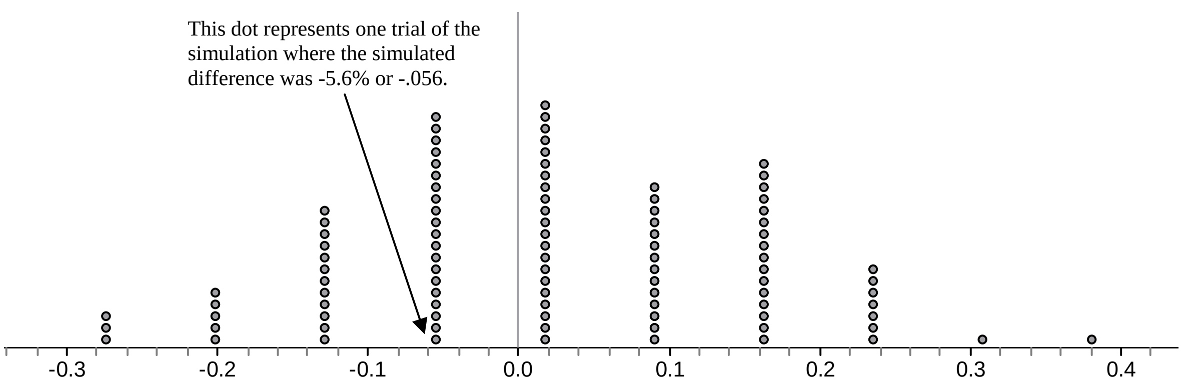
4. Finally, count how many 1's there are in the first 27 positions (the first-game wins) and how many 1's there are in the last 28 positions (the first-game losses) in L1 and compute the difference in the winning percentage: P(won the 2^{nd} | won the 1^{st}) – P(won the 2^{nd} | lost the 1^{st}). Note: It is possible that this difference will be negative which means that the team won a higher percentage of games after losses than after wins.

Note: In the technology appendix you will find the code for a TI-84 program which automates this simulation process. You may also use an online applet for two-way tables.

An example of one trial is shown below:

Following a victory: 12 W and 15 L → 12/27 = 44.4%
Following a loss: 14 W and 14 L → 14/28 = 50.0%
Difference = -5.6%

The results of 100 trials of the simulation are shown below:



Simulated Differences in Performance in the Second Game
P(won the 2^{nd} | won the 1^{st}) – P(won the 2^{nd} | lost the 1^{st})

Although it may be possible for a team to get "in the zone," the 2008 Michigan State softball team does not provide us convincing evidence. Although their *Performance* was definitely better following a win than following a loss (16.3%), according to the simulation it is fairly likely to get a difference in *Performance* this large or larger by *Random Chance* (*p*-value = .25) when their *Ability* to win games is the same following a victory or following a loss.

Another Way to Assess Independence: Streakiness

In the game on September 14, 2008, Tony Romo of the Dallas Cowboys completed 24 of his 32 passes (75%). Here are the outcomes of each passing attempt (in order) where C = pass was complete and I = pass was incomplete (or intercepted). The attempts are written in groups of 5 to make them easier to read.

CCCCC CCCCI CICCI CICIC ICCCI CCCIC CC

You probably noticed that Romo made nine consecutive passes to begin the game. Does this streak of nine consecutive completions give evidence that Romo was "in the zone?" You can just imagine a broadcaster raving about how "hot" Romo is by the 5th or 6th straight completion.

We want to test the following hypotheses for Romo's game on September 14, 2008:

H_0 : Tony Romo's *Ability* to complete a pass is the same following a completed pass and following an incomplete pass

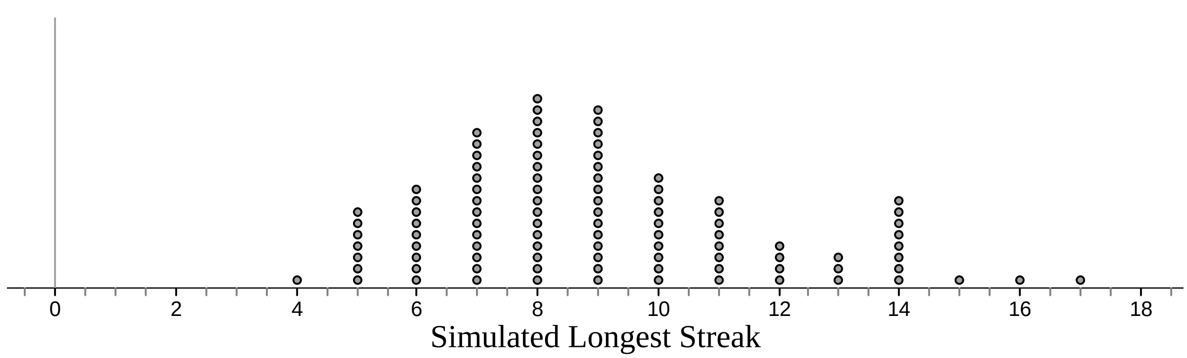
H_a : Tony Romo's *Ability* to complete a pass is higher following a completed pass than following an incomplete pass

These hypotheses are virtually the same as the hypotheses for the Shaq example. However, instead of comparing the probability of making a completion after a completion with the probability of making a completion after an incomplete pass, we are going to focus on the streak of nine consecutive completions. How unusual would it be to get a streak of *Performances* like this if Romo's *Ability* to complete passes is independent of the outcomes of previous attempts?

To answer this question, we need to know what streak lengths are possible when pass attempts are independent. To do this, we will use 32 cards, 24 labeled C (completion) and 8 labeled I (incompletion). Then, we will shuffle the cards and lay them out in a random order. Then, we record the longest streak in each simulated game. Here is the result of one trial:

CCCCC ICCCC CCCII CICIC CICCC CCICC CI

In this trial, the longest streak of completions was 7 (underlined above). Here are the results of 100 trials:



As you can see, having a streak of 9 completions is not unusual at all! According to this simulation, the *p*-value is approximately 0.52, which means that in slightly over half of the simulated games Romo had streaks of 9 or more consecutive completions. Thus, a streak of 9 consecutive passes does *not* give us convincing evidence that Romo's *Ability* to complete a pass increases after a completed pass. It simply could have been *Random Chance*.

Note: A long streak of consecutive incomplete passes is also evidence that pass attempts are not independent. After all, if a player is more likely to complete a pass after a completion, then he is also more likely to throw an incomplete pass if the previous pass is incomplete. So, a streak of either variety is evidence that a player's *Ability* to be successful is higher following a success than following a failure.

Caution: Type I Errors

Suppose that we looked at 100 different basketball players and decided that if the *p*-value was less than 5%, we would conclude that "the zone" exists for that player. However, even if "the zone" doesn't exist, we would still expect 5 of the 100 players to have *p*-values this small, just by *Random Chance*. Statisticians call this a Type I Error: concluding that "the zone" exists for a particular player when in reality it does not exist. In general, a Type I error is committed whenever we conclude that the alternative hypothesis is true when in fact the null hypothesis is true.

In this example, the probability of making a Type I error for any particular player is 5%. So, avoiding this error is fairly likely when assessing a single player (probability of avoiding a Type I error in one test = 0.95). However, when looking at many players the probability of making at least one Type I error is bigger.

For example, when analyzing 10 players, the probability we avoid a Type I error for all the players is $(.95)(.95)\ldots(.95) = (.95)^{10} = 0.599$. Thus, the probability of making at least one Type I error is $1 - 0.599 = 0.401$.

Conclusion: If you look at enough players, you will find some with *Performances* that are unlikely to happen by chance. People in the media often are guilty of "cherry-picking" these performances to illustrate something that may be entirely due to chance.

Caution: Type II Errors

Although the examples in this Unit were selected without knowing what the conclusions would be, none of them gave convincing evidence that "The Zone" exists. This doesn't mean that it doesn't exist, however. It is also possible that we made what statisticians call a Type II Error: failing to conclude that "the zone" exists when it really does. In general, a Type II error is committed when we do not conclude that the results could be due to *Random Chance* alone, when in reality the alternative hypothesis is true.

In many cases the alternative hypothesis is really true, but we simply don't have enough evidence to be convinced the null hypothesis is false. To avoid this error, we want to gather as much data (evidence) as possible. If the alternative hypothesis is true, we have a much better chance to make that conclusion when we look at more *Performances.*

There is definitely a psychological aspect to this investigation as well. Humans tend to underestimate the amount of variability in independent trials and also tend to see evidence that fits with their preconceived ideas. For example, for someone who believes in "The Zone," a streak of any length may seem like strong evidence that a player is in the zone.

Finally, many researchers have investigated "The Zone" (also called "The Hot Hand") and come up with varied conclusions. If you are interested in reading more about this, the following website has a great collection of research on this topic: http://thehothand.blogspot.com/.

Connections: Looking Forward…Looking Back

The idea of independence has been and will continue to be very important to us. Without saying so, in the first two units we assumed that athletic *Performances* were independent when we conducted our simulations.

For example, when we used a spinner to simulate LeBron's *Performance* in the playoffs in unit 1, the outcomes of each spin were independent since the spinner has no way increasing the likelihood of landing in the "made" region based on where it landed on the previous spin.

In future units we will continue to make the assumption of independence for a couple reasons: 1. it seems to fit well with the data we have and 2. it will make our analysis much easier!

Stats 101: The Traditional Approach

Many of the concepts and terms we used in this unit are identical to the concepts and terms you would encounter in a traditional statistics course. For example, the definitions of independence and Type I and Type II errors are the same.

However, there are many different ways to test independence. We chose to use simulation, which is also a frequently used tool of statisticians. However, just like in unit 2, other methods exist for such tests.

Finally, another concept related to Type I and Type II errors is the concept of power. In statistics, the power of a test is the probability that we decide the alternative hypothesis is true when in reality it is true. For example, if a team really has a greater *Ability* to win following a win than following a loss, the power of the test is the probability that the results of the test also indicate that they have a higher *Ability*. In other words, it is the probability of correctly acknowledging that a team's outcomes are not independent. This probability is the complement of the probability of a Type II error, since in both cases the alternative hypothesis is really true, but we come to opposite conclusions. The best way to increase power is to look at more data.

For Practice

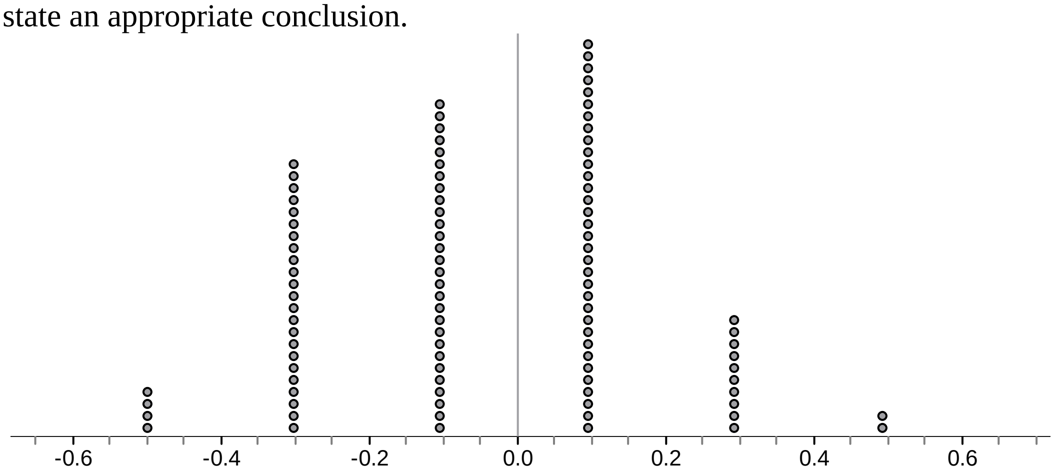
1. Suppose you and your friends are at the bowling alley and one of your friends claims he is "in the zone." What might this mean in the context of bowling?

2. Suppose you and your friends are playing darts. In this context, what does it mean for tosses to be independent?

3. Are consecutive at-bats independent in baseball? Suppose that we looked at the first two at-bats in a game for a particular player over a 25-game stretch. Looking at only the first two at-bats means that it is very likely the hitter is facing the same pitcher in the two at-bats. Here are the results:

	Hit in the first at-bat	Out in the first at-bat	Total
Hit in the second at-bat	3	6	9
Out in the second at-bat	4	12	16
Total	7	18	25

a) How much better did the hitter *Perform* after getting a hit in the first at-bat than after getting an out in the first at-bat?
b) State the hypotheses we are interested in testing.
c) Describe how you could use index cards to simulate this situation.
d) A simulation was conducted assuming this player has the same *Ability* to get a hit in the second at bat following a hit in the first at-bat or following an out in the first at-bat. Based on the results of the 100-trial simulation shown below, estimate the *p*-value and state an appropriate conclusion.

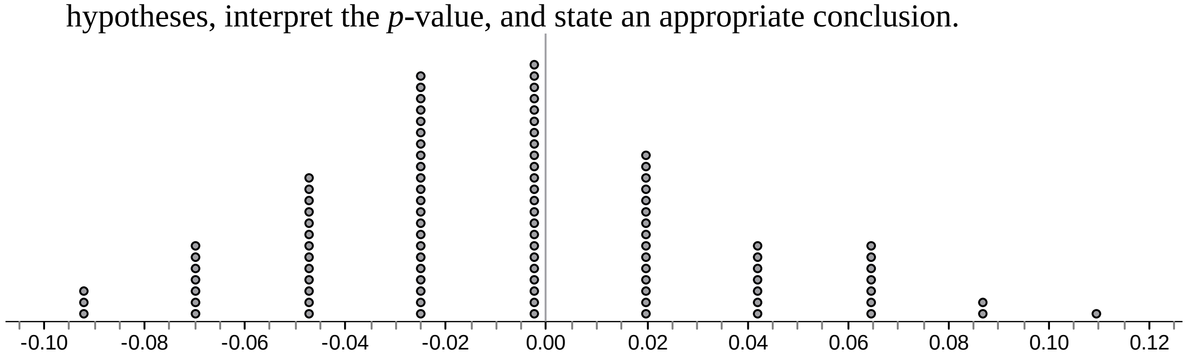


P(hit in the 2nd at bat | hit in the 1st at bat) – P(hit in the 2nd at bat | out in the 1st at bat)

4. Here is a two-way table summarizing the outcomes for pairs of free throws for Larry Bird in the 1981-1982 season (Simpson's Paradox and the Hot Hand in Basketball Author(s): Robert L. Wardrop Source: The American Statistician, Vol. 49, No. 1, (Feb., 1995), pp. 24-28)

	Made the first	Missed the first	
Made the second	251	48	299
Missed the second	34	5	39
	285	53	338

a) Surprisingly, Bird's performance on the second free throw was actually better following a missed shot than following a made shot. How much better did he *Perform* after a missed shot?
b) Can you think of a reason why he might do better after a missed shot than a made shot?
c) Assuming that Bird's *Ability* is the same following a made free throw and following a missed free throw, 100 trials of a simulation were conducted and the simulated difference in his *Performance* after a made shot and after a missed shot was calculated. Does the data give convincing evidence that his *Ability* is higher after a miss? State your hypotheses, interpret the *p*-value, and state an appropriate conclusion.



Simulated Difference in Performances on the Second Shot
P(made the 2nd | made the 1st) – P(made the 2nd | missed the 1st)

5. In the 2008-2009 NCAA basketball season, the Duke Blue Devils men's basketball team finished with 11 wins and 5 losses in ACC conference play. Here are the outcomes of each game in order: W W W W W L W L W L L W W W W L.

a) What was Duke's longest streak during the season?
b) Does the length of their longest streak give convincing evidence that Duke's *Ability* to win is higher after a win than after a loss? State the hypotheses we are interested in testing.
c) Describe how you could use the methods of this unit to simulate this situation.
d) Conduct a simulation to with at least 10 trials and use a dotplot to display the length of the longest streak in each trial.
e) Based on your simulation in part (d), interpret the *p*-value and make an appropriate conclusion.

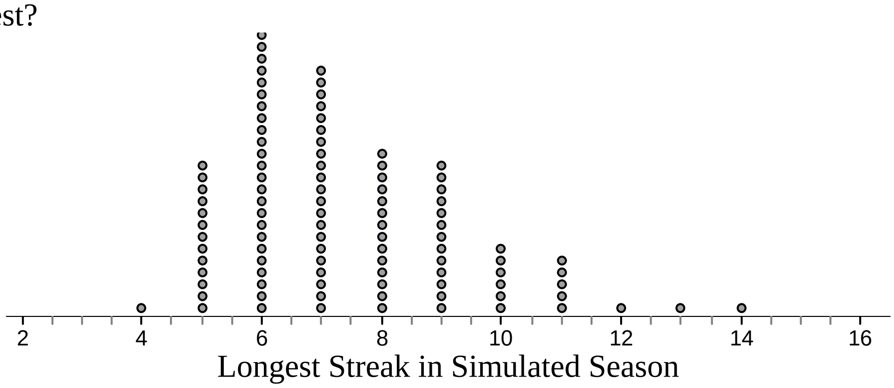
6. In the 2008 NFL season, the Tennessee Titans finished with 13 wins and 3 losses, but started the season with 10 wins in a row. Does this streak give convincing evidence that the Titans were a streaky team (in other words, that the Titans' *Ability* to win is higher after a win than after a loss)?

a) State the hypotheses we are interested in testing.
b) Describe how you could use the methods of this unit to simulate this situation.

c) Conduct a simulation to with at least 10 trials and use a dotplot to display the length of the longest streak in each trial.
d) Based on your simulation in part (c), interpret the *p*-value and make an appropriate conclusion.

7. In 2007, the Colorado Rockies surprised the world of baseball with a late season charge to capture a spot in the playoffs and an overall record of 90-73. During this tremendous month they had a streak of 11 consecutive wins. Does this streak give convincing evidence that the Rockies were a streaky team (in other words, that the Rockies' *Ability* to win is higher after a win than after a loss)?

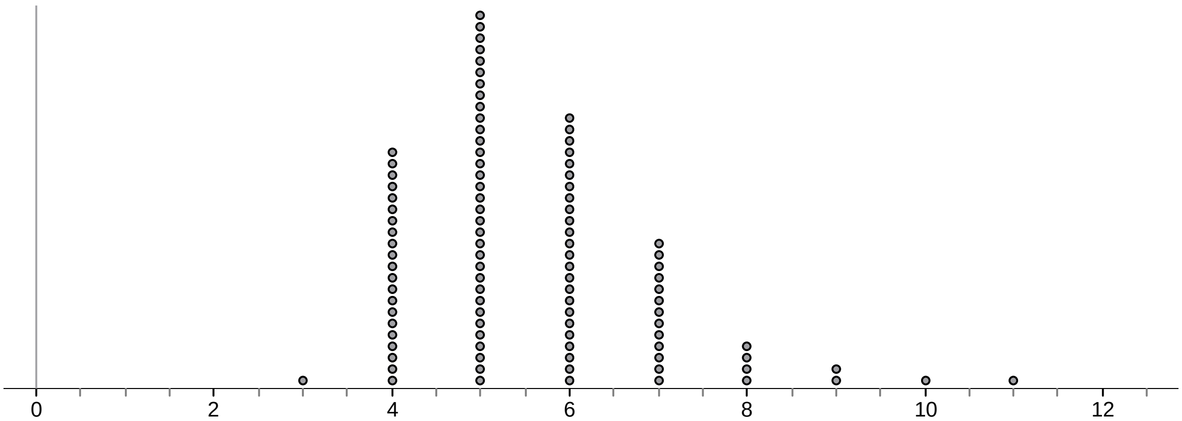
a) State the hypotheses we are interested in testing.
b) Describe how you could use the methods of this unit to simulate this situation.
c) A 100-trial simulation was conducted assuming that the Rockies have the same *Ability* to win after a win and after a loss. In each trial, the longest streak was recorded. The results are shown in the dotplot below. What was the Rockies shortest streak in a season? Longest?



d) Interpret the *p*-value and make an appropriate conclusion.

8.) Earlier in this unit we looked at the 2008 Michigan State softball team. In the example we investigated independence by considering pairs of games. However, we could also test for independence by investigating their streakiness. Overall, the team had 27 wins and 29 losses and their longest streak was a string of 9 losses in a row. Does this streak give convincing evidence that their *Ability* to win a game is higher following a win (and consequently that their *Ability* to lose is higher following a loss)?

a) State the hypotheses we are interested in testing.
b) Describe how you could use the methods of this unit to simulate this situation.
c) A 100-trial simulation was conducted assuming that the MSU softball team has the same *Ability* to win after a win as win after a loss. In each trial, the longest streak was recorded. The results are shown in the dotplot below. What was MSU's shortest streak in a season? Longest?



d) Estimate the *p*-value and make an appropriate conclusion.

9. Using the hypotheses from problem 7, describe a Type I error and a Type II error in context.

10. Using the hypotheses from problem 8, describe a Type I error and a Type II error in context.

For Investigation:

1. Can you get "in the zone?" Go to a basketball court with your class and take turns shooting pairs of free throws (it is important to have a short break in between each set of 2 free throws). Fill out a two-way table for your results like the one for Shaq in the first example and use it to investigate if your free throws are independent.

2. How long does it take to get "in the zone?" In the MSU softball example in this Unit, we only considered the outcome of the previous game. What about the previous two games? For example, is the ability to win a game following 2 straight victories higher than the ability to win following some other combination? What about the previous 3 games? Compare the results at least 2 different methods to investigate independence.

3. Another way to investigate streakiness is to consider the number of streaks in a season. For example, count how many streaks of 5 or more wins or losses during a season and see if the number of streaks is unusual. Research a team on the internet or gather data from a team at your school and use this method to investigate independence.