Name:_____________________________

Class Period:______________________

Date: ______________________________

Statistics Activity 2, Part 1

Objective:
Students will be able to:
Collect data and display it in a table and on a scatterplot
Estimate and determine a line of best fit for the data and interpret the equation in the context of the data
Design an experiment that will produce reliable data
Determine the experimental and theoretical probabilities of events

Materials:
Number cubes (one per group)
Stopwatch (one per group)
Markers (two colors per group)
Optional: Regression line calculators or graphing calculators; Excel and a projector for the class

Directions:
Step 1:
In your group of four people, assign each person a role:

____________________ will be the time keeper (Person A)

____________________ will be the data recorder for ONEs (Person B)

____________________ will be the number cube roller (Person C)

____________________ will be the data recorder for ALL rolls (Person D)

Step 2:
For the next two minutes:
Person A will start a timer and let the group know every time 10 seconds pass.

Person B will make tally marks (in column B_0 and the appropriate row) for every ONE that is rolled in each 10-second interval.

Person C will roll the number cube and tell person B every time a ONE is rolled.

Person D will make tally marks (in column D_0 and the appropriate row) for EVERY roll in each 10-second interval.

Step 3:
Ready? GO!

Comment [WS1]: The length of time for this activity can vary from one long class period to a multi-class-period project, depending on your needs.
It is recommended that students complete the Part 1 handout (and associated questions) before receiving the Part 2 handout.

Also, if the handouts are printed on double-sided paper, ensure that the tables and scatterplots are on different sheets of paper (as they are now) rather than back-to-back.

Comment [WS2]: Because there are so many things happening at the same time during the two-minute activity, we recommend it be done in groups of four. However, the number of people per group can be reduced to three or even to two, depending on your classroom needs. Possible modifications are below:
The teacher can take the timekeeper role for the whole class and ring a bell or call out something like "next" every time 10 seconds passes.
The data recorder roles can be combined into one role: instead of trying make two tally marks every time a one is rolled (one in column B and one in column D), we recommend that the data recorder make a tally mark for every non-one roll and make a circle or other symbol every time a one is rolled. The tally of ones and total rolls for each ten-second interval can then be sorted out after the two-minute interval.
Additionally, instead of spending two uninterrupted minutes on this part of the activity, the timer could be stopped at the end of each 10-second interval, and 12 10-second trials could be completed instead.

Comment [WS3]: The letters assigned to the roles roughly correspond to the columns each person is responsible for in the table. Also, the columns with the bold text are the columns that should be filled out during the two-minute activity; the rest can be filled out immediately after. It may be useful to point this out to students.
Also, the columns that use the subscript 0 are for making tally marks, which we found was the easiest way to keep track of 1s and total rolls while the timer was going. For ease of reading data after the two-minute activity is complete, there are separate columns to record the numeric value that corresponds to the number of tally marks made. The columns that contain values relevant to Part 1 contain subscript 1; in Part 2, the subscript of 2 is used (and 0 is used again for recording tally marks).

Comment [WS4]: A helpful suggestion to make to students is that person A say something like "next", rather than the row number, each time a 10-second interval passes. (This will prevent confusion between the number of the row and the number being rolled).

Comment [WS5]: Before the two-minute activity begins, a verbal reminder to move to the next row of the table every ten seconds may be helpful.

Name:____________________________

Class Period:_______________________

Date: _____________________________

Statistics Activity 2, Part 1

Objective:
Students will be able to:
Collect data and display it in a table and on a scatterplot
Estimate and determine a line of best fit for the data and interpret the equation in the context of the data
Design an experiment that will produce reliable data
Determine the experimental and theoretical probabilities of events

Materials:
Number cubes (one per group)
Stopwatch (one per group)
Markers (two colors per group)
Optional: Regression line calculators or graphing calculators; Excel and a projector for the class

Directions:
Step 1:
In your group of four people, assign each person a role:

____________________ will be the time keeper (Person A)

____________________ will be the data recorder for ONEs (Person B)

____________________ will be the number cube roller (Person C)

____________________ will be the data recorder for ALL rolls (Person D)

Step 2:
For the next two minutes:
Person A will start a timer and let the group know every time 10 seconds pass.

Person B will make tally marks (in column B_0 and the appropriate row) for every ONE that is rolled in each 10-second interval.

Person C will roll the number cube and tell person B every time a ONE is rolled.

Person D will make tally marks (in column D_0 and the appropriate row) for EVERY roll in each 10-second interval.

Step 3:
Ready? GO!

Part 1 Table:

	Column A_0	Column A_1	**Column B_0**	Column B_1	Column C_1	**Column D_0**	Column D_1	Column E_1
	Time interval (t seconds)	Number of seconds since start	**Number of 1s in <u>this</u> time interval (tally marks)**	Number of 1s in <u>this</u> time interval (numeral) (Color #1)	<u>Total</u> number of 1s since start (numeral) (Color #2)	**Number of rolls in <u>this</u> time interval (tally marks)**	Number of rolls in <u>this</u> time interval (numeral)	<u>Total</u> number of rolls since start (numeral)
Row 1	$0 \le t \le 10$	10						
Row 2	$10 < t \le 20$	20						
Row 3	$20 < t \le 30$	30						
Row 4	$30 < t \le 40$	40						
Row 5	$40 < t \le 50$	50						
Row 6	$50 < t \le 60$	60						
Row 7	$60 < t \le 70$	70						
Row 8	$70 < t \le 80$	80						
Row 9	$80 < t \le 90$	90						
Row 10	$90 < t \le 100$	100						
Row 11	$100 < t \le 110$	110						
Row 12	$110 < t \le 120$	120						

Now, work together to fill in the total number of 1s rolled since the start of the activity (column C_1), the total number of rolls since the start of the activity (column E_1), and the numerals (in B_1 and D_1) that correspond to the number of tally marks in columns B_0 and D_0.

Step 4:

Using marker color #1, work with your group to plot the ordered pairs formed by the values in columns A_1 and B_1 on the scatterplot on the next page.

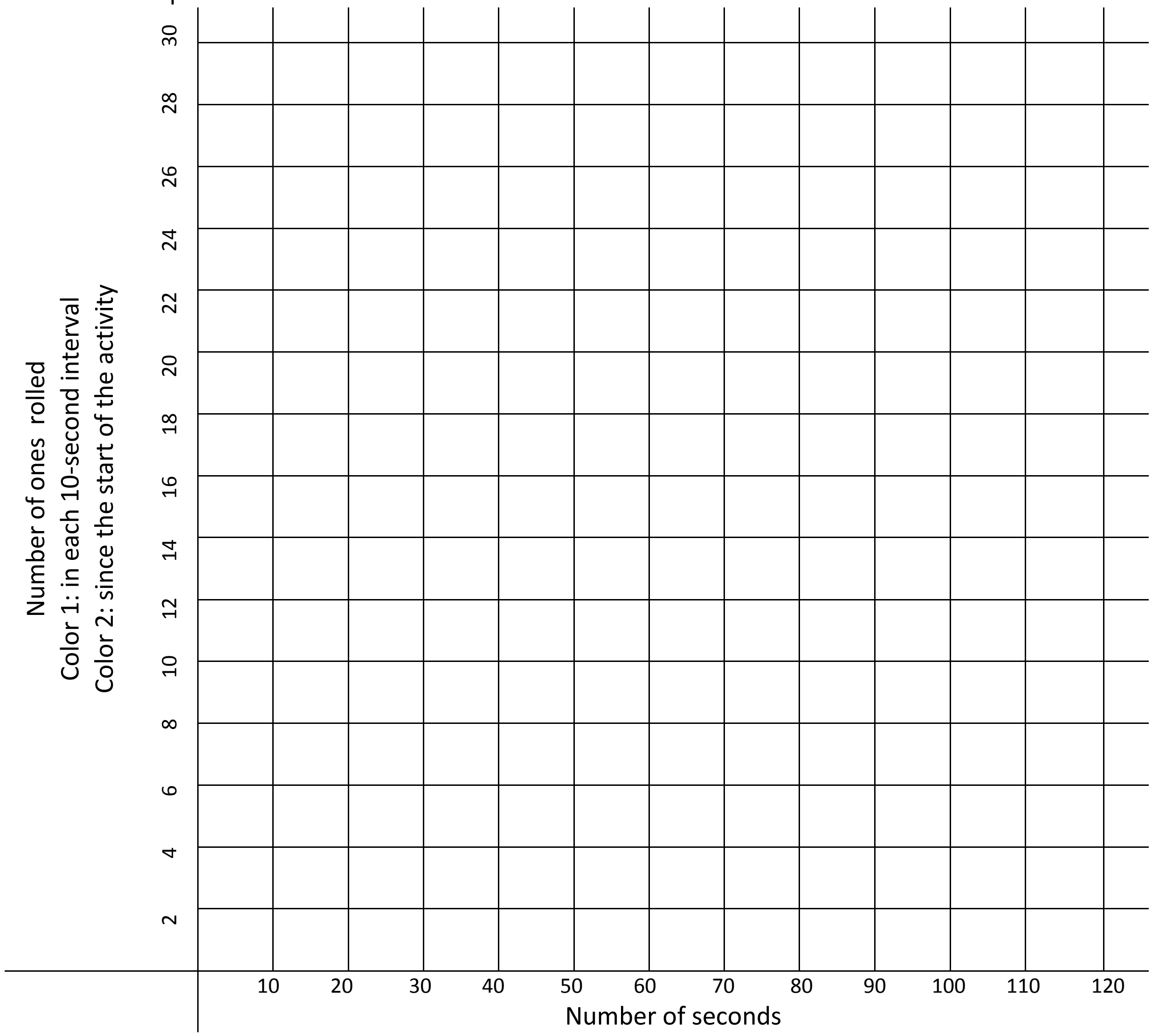
Find or estimate the equation for the line of best fit for this data. Using color #1, draw the line of best fit on the scatterplot and label it as line b_1. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Step 5:

Using marker color #2, work with your group to plot the ordered pairs formed by the values in columns A_1 and C_1 on the same scatterplot (below).

Find or estimate the equation for the line of best fit for this data. Using color #2, draw the line of best fit on the scatterplot and label it as line c_1. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Part 1 Scatterplot:



Equation for line b_1 (Part 1, columns B_1 vs A_1):	Equation for line c_1 (Part 1, columns C_1 vs A_1):
What this line represents:	What this line represents:

Name:___________________________

Class Period:______________________

Date: ____________________________

Statistics Activity 2, Part 2

Step 6:
In your groups, discuss the following question:
How might the results of the experiment from Part 1 have been different if the independent variable was the number of rolls rather than the number of seconds?
Record your answer:

Next, we are going to complete a similar activity to find out exactly how the results would be different.

Directions:
Step 7:
In your group of four people, assign each person a role (which may be different from that person's role in Part 1):

___________________ will be the number cube roller (Person A)

___________________ will be the data recorder for ONEs in each interval(Person B)

___________________ will total the tally marks and fill in the remaining columns (Person C)

___________________ will keep track of the number of 10-roll intervals (Person D)

Step 8:
For this part of the experiment:
Person A will roll the number cube and let person B know every time a ONE is rolled.

Person B will make tally marks (in column B_0 and the appropriate row) for every ONE that is rolled.

Person C will, at the end of each 10-roll interval, record the total number of ONEs (in column C_2 and the appropriate row) that have been rolled from the time the experiment started. This person will also complete columns D_2 and E_2, which can be done before or after the experiment begins.

Person D will keep track of the TOTAL number of number cube rolls and let the group know every time 10 rolls have been completed.

Part 2 Table:

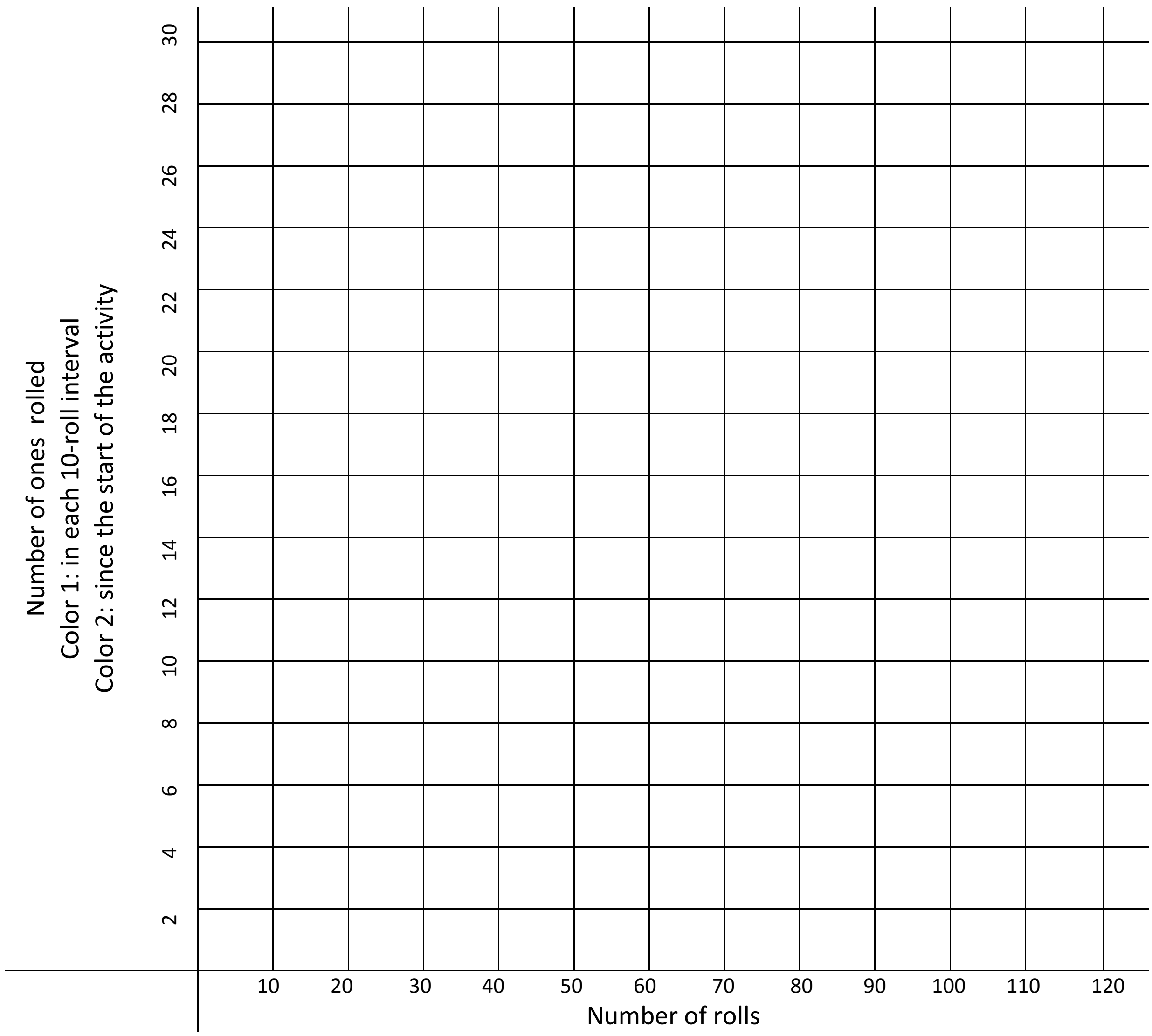
	Column A_0	Column A_2	**Column B_0**	Column B_2	Column C_2	Column D_2	Column E_2
	Roll interval (*r* rolls)	Number of rolls since start	**Number of 1s in <u>this</u> roll interval (tally marks)**	Number of 1s in <u>this</u> time interval (numeral) (Color #1)	<u>Total</u> number of 1s since start (numeral) (Color #2)	Number of rolls in <u>this</u> time interval (numeral)	<u>Total</u> number of rolls since start (numeral)
Row 1	$0 \leq r \leq 10$	10					
Row 2	$10 < r \leq 20$	20					
Row 3	$20 < r \leq 30$	30					
Row 4	$30 < r \leq 40$	40					
Row 5	$40 < r \leq 50$	50					
Row 6	$50 < r \leq 60$	60					
Row 7	$60 < r \leq 70$	70					
Row 8	$70 < r \leq 80$	80					
Row 9	$80 < r \leq 90$	90					
Row 10	$90 < r \leq 100$	100					
Row 11	$100 < r \leq 110$	110					
Row 12	$110 < r \leq 120$	120					

Now, fill in any remaining columns.

Step 9:
Using marker color #1, work with your group to plot ordered pairs formed by the values in columns A_2 and B_2 on the scatterplot on the next page.
Find or estimate the equation for the line of best fit for this data. Using color #1, draw the line of best fit on the scatterplot and label it as line b_2. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Step 10:
Using marker color #1, work with your group to plot ordered pairs formed by the values in columns A_2 and C_2 on the scatterplot on the next page.
Find or estimate the equation for the line of best fit for this data. Using color #1, draw the line of best fit on the scatterplot and label it as line c_2. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Part 2 Scatterplot:



Equation for line b_2 (Part 2, columns B_2 vs A_2):	Equation for line c_2 (Part 2, columns C_2 vs A_2):
What this line represents:	What this line represents:

Possible discussion questions for Part 1:

1. How are the scatterplots and lines of best fit different when the values in column B_1 (the number of 1s rolled in each <u>time interval</u>) are plotted or if the values in column C_1 (the cumulative number of 1s rolled <u>since the start</u> of the experiment) are plotted? Explain.

2. Does the line of best fit based on the data from column B_1 mean that there is no correlation? Does it mean there is no association? Explain.

3. In your group only, why is there a difference in the number of 1s rolled in each time interval? Why might that be?

4. How do you think the results in your group will compare to the results of the whole class? Explain.

5. Is it better to look just at your group's data, or just at another group's data, or at data from all of the groups in the class, or at some other subset of data? Explain.

6. Compare your results with the results of two groups next to you. Is there a difference in the number of 1s rolled in your group compared to other groups? In each interval? Overall? Why or why not? Use details about the design of the experiment to explain.

7. How do your lines of best fit for your data compare to the lines of best fit in the groups next to you? Why might this be the case?

8. How do your lines of best fit compare to the lines of best fit for the data from the whole class? What trends or patterns do you notice? Explain.

9. What is your experimental probability of rolling a 1 during the entire two-minute activity? What is the range in your experimental probabilities in each interval? Based on theoretical probability, how many 1s would you expect to roll in each interval? Is this the same as what you would expect from other groups? Why?

10. Is there a better way the experiment could have been designed to reduce variation between groups? Explain. (Hint: look at the values in columns D_1 and E_1 and compare your group's numbers to other groups.)

11. What might you expect to happen to the correlation coefficient if the experimental design had been different?

Possible discussion questions for Part 2:
12. How does your data from Part 2 compare to the data you collected in Part 1? Explain.

13. How does the correlation coefficient of the data from Part 1 compare to the correlation coefficient of the data from Part 2? Based on the experimental design, is this what you would expect? Why or why not?

14. How do you think your data from part 2 will compare to the data that other groups collected in Part 2? When comparing across different groups, would you expect the Part 1 data or the Part 2 data to be more similar? Why?

15. Compare your results with two other groups. Is there a difference in the number of 1s rolled in your group compared to other groups? In each interval? Overall? Why might this be the case?

16. How does the variation in the number of 1s rolled in each interval and each group in Part 2 compare to the variation in Part 1? Explain.

17. How do the lines of best fit for the data in your group compare to the lines of best fit for the data from other groups?

18. Are your lines of best fit more similar or less similar to those of other groups for the data from Part 2 or for the data in Part 1? Is this what you thought would happen? Explain.

19. What do you expect the lines of best fit for the whole class to look like? How might this compare to Part 1?

20. What is the slope of your lines of best fit? What is the meaning of the slopes of the line of best fit? (Hint: refer to the axis labels.) How does this compare to the slope of the line of best fit for the whole class?

21. What is the value of the y-intercept of the lines of best fit? What does this mean? Can a line of best fit predict that a certain number of ones will be rolled even if the number cube has been rolled zero times?

21. Compare your actual values to the values predicted by your line of best fit. How accurate are they? Explain.

22. Do you think your group's data, scatterplot, and/or line of best fit from Part 2 would be noticeably different if your group had counted twos or sixes instead of ones? What about the scatterplot for the data from the whole class? Explain.

23. Assuming the number cubes are fair, what is the theoretical probability of rolling a one? How many ones would you expect to roll in each interval? How does this compare to your response in part 1?

24. What was your experimental probability of rolling a one, in part 2, after 10 rolls? After100 rolls? What was the experimental probability of rolling a one based on all of the rolls in the class?

25. What is the slope of these lines of best fit when all of the data from the class is used? How does this relate to the theoretical probability of rolling a one? What about experimental probability? How does this relate to correlation and association? Why is the slope of the line of best fit for the data in column C not the same as the experimental probability of rolling a 1?

Part 1 Table:

	Column A_0	Column A_1	**Column B_0**	Column B_1	Column C_1	**Column D_0**	Column D_1	Column E_1
	Time interval (t seconds)	Number of seconds since start	**Number of 1s in this time interval (tally marks)**	Number of 1s in this time interval (numeral) (Color #1)	Total number of 1s since start (numeral) (Color #2)	**Number of rolls in this time interval (tally marks)**	Number of rolls in this time interval (numeral)	Total number of rolls since start (numeral)
Row 1	$0 \leq t \leq 10$	10						
Row 2	$10 < t \leq 20$	20						
Row 3	$20 < t \leq 30$	30						
Row 4	$30 < t \leq 40$	40						
Row 5	$40 < t \leq 50$	50						
Row 6	$50 < t \leq 60$	60						
Row 7	$60 < t \leq 70$	70						
Row 8	$70 < t \leq 80$	80						
Row 9	$80 < t \leq 90$	90						
Row 10	$90 < t \leq 100$	100						
Row 11	$100 < t \leq 110$	110						
Row 12	$110 < t \leq 120$	120						

Now, work together to fill in the total number of 1s rolled since the start of the activity (column C_1), the total number of rolls since the start of the activity (column E_1), and the numerals (in B_1 and D_1) that correspond to the number of tally marks in columns B_0 and D_0.

Step 4:
Using marker color #1, work with your group to plot the ordered pairs formed by the values in columns A_1 and B_1 on the scatterplot on the next page.
Find or estimate the equation for the line of best fit for this data. Using color #1, draw the line of best fit on the scatterplot and label it as line b_1. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

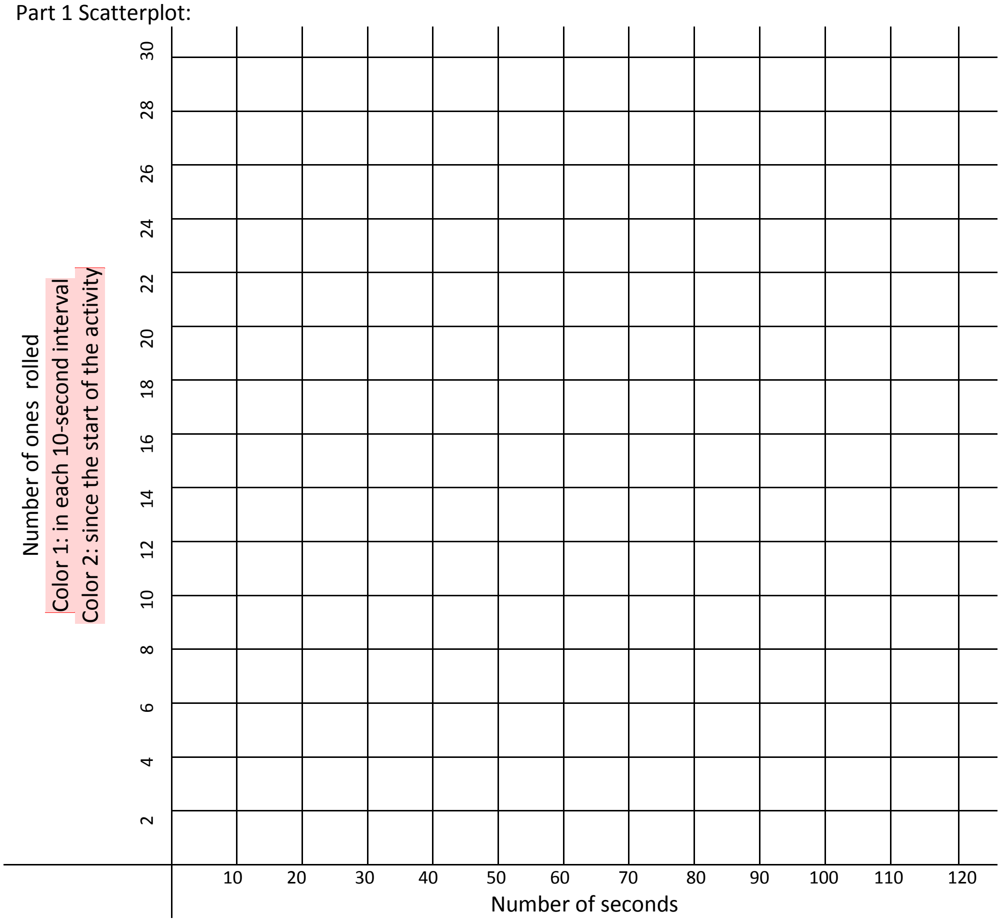
Comment [WS6]: The variables are not defined on the scatterplot. The variable *t* is used in the table above, so t should be used for the horizontal axis. Discussing the variables and the merits of using function notation might be a useful extension at this point in the activity. (If one variable is used for line b1 and another variable is used for line c1, the vertical axis would not have a clear variable label; function notation allows for plotting y = f(t) and y = g(t) on the same coordinate plane. (Or, a different variable could be used instead of y; "o" should probably be avoided, even though it is the first letter of what we are counting (ones), since it is easily confused with a zero and often denotes the origin, not an axis; "r" is used in part 2 to denote number of rolls, so it should be avoided as well.)

Step 5:

Using marker color #2, work with your group to plot the ordered pairs formed by the values in columns A_1 and C_1 on the same scatterplot (below).

Find or estimate the equation for the line of best fit for this data. Using color #2, draw the line of best fit on the scatterplot and label it as line c_1. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Part 1 Scatterplot:



Equation for line b_1 (Part 1, columns B_1 vs A_1):	Equation for line c_1 (Part 1, columns C_1 vs A_1):
What this line represents:	What this line represents:

Comment [WS7]: Depending on time and your students' previous exposure to lines of best fit, it may be useful to delve into a deeper explanation of scatterplots and lines of best fit at this point.

There are a number of excellent tools for comparing estimated to actual equations for lines of best fit (http://illuminations.nctm.org/Activity.aspx?id=4186), and a discussion of the purpose and meaning of a line of best fit may fit well here.

Additionally, it would be very helpful (either at this point in the lesson or during the follow-up questions) to discuss how close the data points are to the line of best fit and to give a general overview of correlation coefficients. Comparing data from other groups and for the whole class can be useful for this discussion.

Additional notes and suggested questions are at the end of this document.

Comment [WS8]: To keep better track of different sets of data, suggest that students highlight or underline each of these labels with the colors used for the lines of best fit. Additionally, they can highlight or put boxes around the corresponding columns in the table on the previous page and on the table below.

Comment [WS9]: Possible response:
The line estimates the of 1s rolled in every 10-second interval based on the number of seconds since the activity began.
A recommended discussion related to the meaning of this line is included in the list of questions at the end.

Comment [WS10]: Possible response: The line estimates the total number of 1s rolled t seconds after the activity began.
A recommended discussion related to the meaning of this line is included in the list of questions at the end.

Comment [WS11]: After the groups have found the equations for the lines of best fit and described how the lines relate the two variables, recommend putting data from all of the groups in the class into the pre-prepared Excel sheet. A list of possible questions for discussion, as well as the types of connections and knowledge they are intended to elicit, is at the end of this document.
The way the questions are presented to students is up to you: students can answer them on their own, in groups, as a class discussion, or whatever works best. The questions cover a wide range of topics but are not exhaustive; you can choose to delve more deeply into certain topics and not discuss or ask about others, or have different students or different groups explore different topics and possibly share their information with the rest of the class.

The included list of questions is meant to be a suggestion, but they should be adjusted so that they are most useful to you and your students' learning needs.

Name:____________________________

Class Period:____________________

Date: _____________________________

Statistics Activity 2, Part 2

Step 6:
In your groups, discuss the following question:
How might the results of the experiment from Part 1 have been different if the independent variable was the number of rolls rather than the number of seconds?
Record your answer:

Next, we are going to complete a similar activity to find out exactly how the results would be different.

Directions:
Step 7:
In your group of four people, assign each person a role (which may be different from that person's role in Part 1):

____________________ will be the number cube roller (Person A)

____________________ will be the data recorder for ONEs in each interval(Person B)

____________________ will total the tally marks and fill in the remaining columns (Person C)

____________________ will keep track of the number of 10-roll intervals (Person D)

Step 8:
For this part of the experiment:
Person A will roll the number cube and let person B know every time a ONE is rolled.

Person B will make tally marks (in column B_0 and the appropriate row) for every ONE that is rolled.

Person C will, at the end of each 10-roll interval, record the total number of ONEs (in column C_2 and the appropriate row) that have been rolled from the time the experiment started. This person will also complete columns D_2 and E_2, which can be done before or after the experiment begins.

Person D will keep track of the TOTAL number of number cube rolls and let the group know every time 10 rolls have been completed.

Comment [WS12]: Some notes about how to use data collected in columns D1 and E1 to shorten or extend the activity are below:

In part 1, columns D and E were used primarily to highlight the variability in the number of rolls in each interval, both within one group and from group to group. They may also be used to find the experimental probability of rolling a 1 during each interval or over the two-minute activity.
A comparison of column D1 to D2 and E1 to E2 (in parts 1 and 2) should help to highlight how much more consistent (reliable) the data collected in Part 2 should be. Additionally, there is a benefit of comparing data sets for groups of 10 seconds to groups of 10 rolls. The two different parts of the activity were designed with these benefits in mind, but completing both parts as they exist now requires a lot of time.

The following modification can be made if time is limited:
The activity can be adjusted to omit the number cube rolling activity in part 2 and instead use the data collected in Part 1: on the Part 2 scatterplot, the plot the values in columns B1 vs E1 (rather than B2 vs A2) and C1 vs E1 (rather than C2 vs A2). By doing this, the independent variable changes from seconds to rolls, which allows for more reliable data, but the number of rolls completed by each group during the two-minute activity may still vary widely, and the comparison of intervals of 10 cannot be made. Additionally, if some groups had very few rolls during the two-minute trial, the information gleaned from their group's data (using this modification) will not be very robust. ...

Comment [WS13]: Sharing this question before Part 1 is complete may give away some of the answers to questions posed in part 1. Part 2 works best in a second class period.

Comment [WS14]: Part 2 can also be modified to be done in groups of 2 or 3; since time is not being tracked, it is much less fast-paced than the part 1 activity. Also, more information can be filled in on the table before the activity begins, or as the activity takes place.

Comment [WS15]: Again, the names of the roles correspond to the names of the columns that each person is expected to fill in, the columns with a subscript of 0 are for tally marks, and the bold column is the only one that has to be filled in as the activity is being completed.

Comment [WS16]: If multiple number cubes are available in each group, the following modification can be made to save time:
Instead of rolling one number cube 120 times, a group of 5 number cubes can be rolled; after two rolls of the group of 5 cubes, the number of ones in the two groups can be recorded. This reduces number of rolls to 24 rather than 120.

Comment [WS17]: Again, recommend suggesting that this person say "next" rather than the number of rolls or the row number.

Part 2 Table:

	Column A_0	Column A_2	**Column B_0**	Column B_2	Column C_2	Column D_2	Column E_2
	Roll interval (*r* rolls)	Number of rolls since start	**Number of 1s in this roll interval (tally marks)**	Number of 1s in this time interval (numeral) (Color #1)	Total number of 1s since start (numeral) (Color #2)	Number of rolls in this time interval (numeral)	Total number of rolls since start (numeral)
Row 1	$0 \leq r \leq 10$	10					
Row 2	$10 < r \leq 20$	20					
Row 3	$20 < r \leq 30$	30					
Row 4	$30 < r \leq 40$	40					
Row 5	$40 < r \leq 50$	50					
Row 6	$50 < r \leq 60$	60					
Row 7	$60 < r \leq 70$	70					
Row 8	$70 < r \leq 80$	80					
Row 9	$80 < r \leq 90$	90					
Row 10	$90 < r \leq 100$	100					
Row 11	$100 < r \leq 110$	110					
Row 12	$110 < r \leq 120$	120					

Now, fill in any remaining columns.

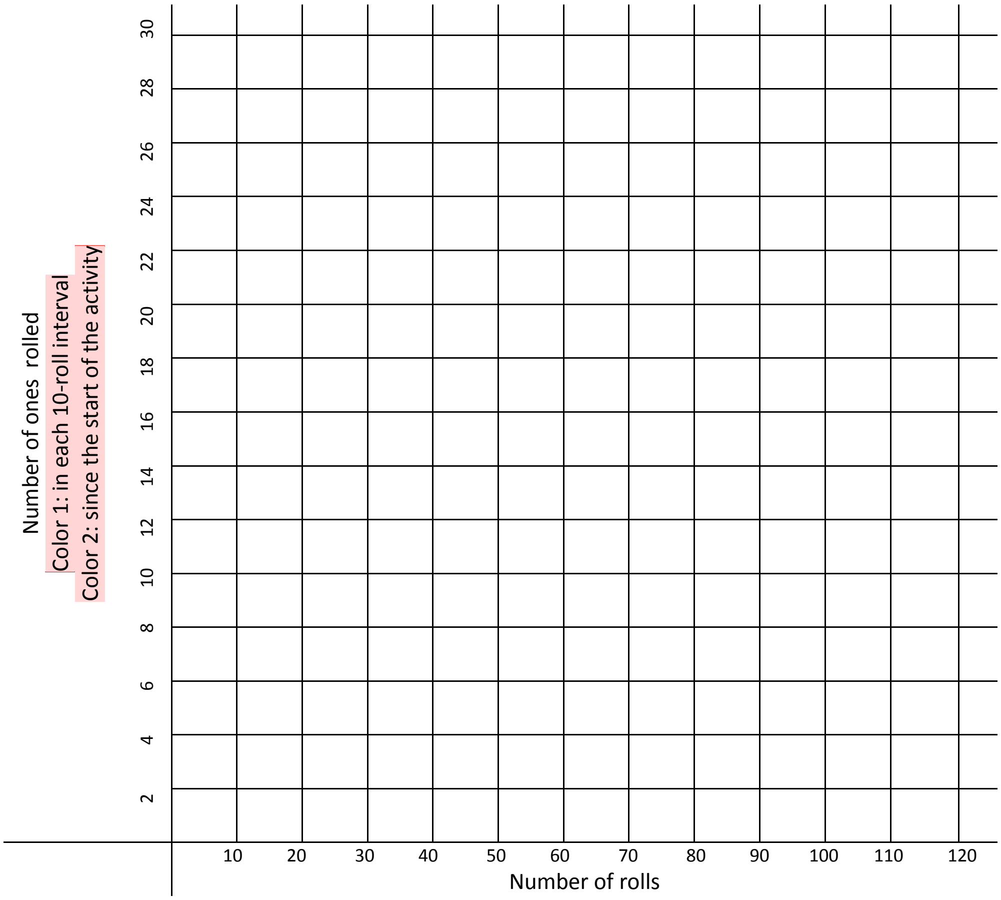
Step 9:
Using marker color #1, work with your group to plot ordered pairs formed by the values in columns A_2 and B_2 on the scatterplot on the next page.
Find or estimate the equation for the line of best fit for this data. Using color #1, draw the line of best fit on the scatterplot and label it as line b_2. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Step 10:
Using marker color #1, work with your group to plot ordered pairs formed by the values in columns A_2 and C_2 on the scatterplot on the next page.
Find or estimate the equation for the line of best fit for this data. Using color #1, draw the line of best fit on the scatterplot and label it as line c_2. Record the equation of the line in the box below the scatterplot. Then, in the box below that, describe how the line relates the two variables.

Comment [WS18]: Again, this is a good place to take a deep dive into scatterplots and lines of best fit.

Comment [WS19]: Again, the variables are not defined, but r is given in the table as the number of rolls.

Part 2 Scatterplot:



Equation for line b_2 (Part 2, columns B_2 vs A_2):	Equation for line c_2 (Part 2, columns C_2 vs A_2):
What this line represents:	What this line represents:

Comment [WS20]: To keep better track of different sets of data, suggest that students highlight or underline each of these labels with the colors used for the lines of best fit. Additionally, they can highlight or put boxes around the corresponding columns in the table on the previous page and on the table below.

Comment [WS21]: Possible response: The line estimates the number of 1s rolled in each group of 10 rolls, based on the number of rolls since the activity began.

A recommended discussion related to the meaning of this line is included in the list of questions at the end.

Comment [WS22]: Possible response: The line estimates the total number of 1s rolled based on the number of rolls that have been completed since the activity began.

A recommended discussion related to the meaning of this line is included in the list of questions at the end.

Possible discussion questions for Part 1:

1. How are the scatterplots and lines of best fit different when the values in column B_1 (the number of 1s rolled in each time interval) are plotted or if the values in column C_1 (the cumulative number of 1s rolled since the start of the experiment) are plotted? Explain.

2. Does the line of best fit based on the data from column B_1 mean that there is no correlation? Does it mean there is no association? Explain.

3. In your group only, why is there a difference in the number of 1s rolled in each time interval? Why might that be?

4. How do you think the results in your group will compare to the results of the whole class? Explain.

5. Is it better to look just at your group's data, or just at another group's data, or at data from all of the groups in the class, or at some other subset of data? Explain.

6. Compare your results with the results of two groups next to you. Is there a difference in the number of 1s rolled in your group compared to other groups? In each interval? Overall? Why or why not? Use details about the design of the experiment to explain.

7. How do your lines of best fit for your data compare to the lines of best fit in the groups next to you? Why might this be the case?

8. How do your lines of best fit compare to the lines of best fit for the data from the whole class? What trends or patterns do you notice? Explain.

9. What is your experimental probability of rolling a 1 during the entire two-minute activity? What is the range in your experimental probabilities in each interval? Based on theoretical probability, how many 1s would you expect to roll in each interval? Is this the same as what you would expect from other groups? Why?

10. Is there a better way the experiment could have been designed to reduce variation between groups? Explain. (Hint: look at the values in columns D_1 and E_1 and compare your group's numbers to other groups.)

11. What might you expect to happen to the correlation coefficient if the experimental design had been different?

Comment [WS23]: This is a good place to talk about how variables are defined. Because the values in column B are the number of 1s *per interval*, this value is NOT dependent on the number of seconds that have passed: it doesn't matter which interval it is, and the slope of the line of best fit is approximately horizontal. The values in column C ARE somewhat dependent on the number of seconds since the start, because we are counting total 1s rolled since the start. The scatterplot and ...

Comment [WS24]: A horizontal line of best fit generally means that there is no correlation, which makes sense because the number of 1s rolled in each interval is not dependent on which interval we are examining. However, the number of 1s rolled in each interval is likely to be fairly consistent from ...

Comment [WS25]: There could be many reasons: maybe the number cube rolled off of the table, maybe the counting was off, maybe sometimes a roll that was started in one interval was counted in that interval and sometimes it was counted in the next, etc.; the goal is that students ...

Comment [WS26]: Again, students should realize that using time (in seconds) leads to more variability: maybe their group rolled more quickly or less quickly than other groups, for example. Again, referring students to columns D and E may be helpful.

Comment [WS27]: This question is getting at the fact that generally, a larger sample is better. Some students may not trust the data collected by other groups or in their group, based on the way they (or other groups) carried out the activity. In some cases, these concerns may be valid.

Comment [WS28]: This (and the next two questions) is an opportunity to check to see if the hypotheses about how data will compare between groups are correct. It is also a time to point out interesting similarities and differences between data from different groups, using Excel.

Comment [WS29]: This is an opportunity to compare lines of best fit for each groups to the line of best fit for the data from the whole class, and see what happens when all of the data is used/how it is affected by groups that have abnormal data.

Comment [WS30]: Depending on time and focus, you can dive more deeply into experimental probability here, and experimental vs. theoretical. At this point, it might be difficult for students to determine the expected number of 1s in each interval; this is another chance to emphasize the ...

Comment [WS31]: Students should conclude that the independent variable should be rolls rather than time in seconds.

Comment [WS32]: Because we are reducing the variability when we use the number of rolls rather than the number of seconds, the correlation coefficients for the lines of best fit should be stronger in the redesigned experiment.

Possible discussion questions for Part 2:

12. How does your data from Part 2 compare to the data you collected in Part 1? Explain.

13. How does the correlation coefficient of the data from Part 1 compare to the correlation coefficient of the data from Part 2? Based on the experimental design, is this what you would expect? Why or why not?

14. How do you think your data from part 2 will compare to the data that other groups collected in Part 2? When comparing across different groups, would you expect the Part 1 data or the Part 2 data to be more similar? Why?

15. Compare your results with two other groups. Is there a difference in the number of 1s rolled in your group compared to other groups? In each interval? Overall? Why might this be the case?

16. How does the variation in the number of 1s rolled in each interval and each group in Part 2 compare to the variation in Part 1? Explain.

17. How do the lines of best fit for the data in your group compare to the lines of best fit for the data from other groups?

18. Are your lines of best fit more similar or less similar to those of other groups for the data from Part 2 or for the data in Part 1? Is this what you thought would happen? Explain.

19. What do you expect the lines of best fit for the whole class to look like? How might this compare to Part 1?

20. What is the slope of your lines of best fit? What is the meaning of the slopes of the line of best fit? (Hint: refer to the axis labels.) How does this compare to the slope of the line of best fit for the whole class?

21. What is the value of the y-intercept of the lines of best fit? What does this mean? Can a line of best fit predict that a certain number of ones will be rolled even if the number cube has been rolled zero times?

21. Compare your actual values to the values predicted by your line of best fit. How accurate are they? Explain.

22. Do you think your group's data, scatterplot, and/or line of best fit from Part 2 would be noticeably different if your group had counted twos or sixes instead of ones? What about the scatterplot for the data from the whole class? Explain.

23. Assuming the number cubes are fair, what is the theoretical probability of rolling a one? How many ones would you expect to roll in each interval? How does this compare to your response in part 1?

24. What was your experimental probability of rolling a one, in part 2, after 10 rolls? After100 rolls? What was the experimental probability of rolling a one based on all of the rolls in the class?

Comment [WS33]: The questions below can be answered for both the data from column B and from column C, or they can be focused on the data from column C, which is the more appropriate data set to use to represent the data.

Comment [WS34]: This is a good time to have students point out interesting trends and differences between the two sets of data. There will likely be more 1s rolled in Part 2 because there are likely more total rolls; the correlation coefficient will be stronger, etc.

Comment [WS35]: Again, this is a chance to check hypotheses and emphasize the importance of repeatability of an experiment in order to get more reliable data.

Comment [WS36]: Students should realize that their data should be more similar to other groups in Part 2 since the design of the activity reduced variation.

Comment [WS37]: In this case, the difference in the number of 1s rolled in different groups is due to chance (assuming a fair number cube and fair rolling) rather than other factors.

Comment [WS38]: Students should understand that some variability is natural, but that sometimes some variability (such as in part 1) is due to things like bad experimental design.

Comment [WS39]: This is a good chance to get students to notice interesting trends, similarities, and differences. For example, what happens when there are a few intervals with a large or small number of 1s in a row? What if those intervals occur at the beginning of the activity? How do the equations for the lines of best fit differ?

Comment [WS40]: The lines of best fit should be much more similar across groups in Part 2 than in Part 1. This goes back to the design of the activity.

Comment [WS41]: Students should note that because of the design of the experiment, the whole-class data from Part 2 has more trials and should ...

Comment [WS42]: The slope of the line of best fit represents the number of 1s rolled (the rise) divided by the total number of rolls (the run). Thi ...

Comment [WS43]: It is likely that the line of best fit will have a y-intercept that is a value other than 0. If there are a large number of 1s rolled at ...

Comment [WS44]: Some data points may be closer to the line of best fit than others. This may be a good time to discuss why this is the case.

Comment [WS45]: While there may be some differences that might be noticeable in small sets of data, the probability of rolling a 2 or a 6 is the sar ...

Comment [WS46]: The theoretical probability is 1/6. For ten rolls, one would expect 1/6 *10 ones ...

Comment [WS47]: This is a good place to compare theoretical and experimental values and to highlight the fact that the experimental probabili ...

25. What is the slope of these lines of best fit when all of the data from the class is used? How does this relate to the theoretical probability of rolling a one? What about experimental probability? How does this relate to correlation and association? Why is the slope of the line of best fit for the data in column C not the same as the experimental probability of rolling a 1?

Comment [WS48]: The slope of the line of best fit for the data in column C should be close to the experimental and theoretical probabilities of rolling a 1. The slope of the line of best fit for the data in column B should be close to zero, since there is no correlation between the number of ones rolled in an interval and the number of rolls that have occurred since the start of the experiment. As discussed in part 1, even though there is no correlation for this data, there is a correlation for the line using data from column C, and there is a linear association for both sets of data. As mentioned above, the slope of the line of best fit is not the same as the experimental or theoretical probability because the line of best fit likely has a nonzero y-intercept.

An interesting extension activity might be determining the equation for the line of best fit if the data performed exactly as expected: what would be the equation if a 1 was rolled only on the first of every six rolls? What if the 1 were the third roll or the 6th roll of every 6 rolls? What if it appeared randomly in each group of 6 rolls?