



Games, Chance, and Predictions

Explorations with Statistics and Probability



Presentation

by

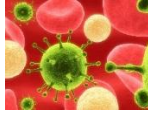
Linda Bridges, M.Ed., NBCT
Alabama Math, Science, Technology Initiative
Secondary Mathematics Specialist
University of Alabama in Huntsville
linda.bridges@uah.edu

NCTM 2014 Annual Meeting

New Orleans, Louisiana

April 2014

SPREAD of DISEASE EXPONENTIAL DECAY EXPERIMENT



OBJECTIVE: Collect and analyze data for which an exponential equation would be the model. (Provides a hands-on activity for exponential equations).

CONTENT STANDARDS:

- *Distinguish between situations that can be modeled with linear functions and with exponential functions.* [HSF-LE.A.1]
Recognize situations in which a quantity grows or decays by a constant percent rate per unit interval relative to another. [HSF-LE.A.1c]
- *Interpret the parameters in a linear or exponential function in terms of a context.* [HSF-LE.B.5]
- *Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.* [HSS-ID.B.6]
Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models. [HSS-ID.B.6a]
Informally assess the fit of a function by plotting and analyzing residuals. [HSS-ID.B.6b]

CONTEXT: Suppose there is a disease which is affecting about $\frac{1}{6}$ of the population each day. Those affected are removed from the community and taken to a hospital for specialized treatment in a neighboring town. If there is a small community with population 100, how long will it take for all the 100 people in the community to be affected?



SIMULATION: Pass out 100 number cubes (dice) to the participants. These 100 number cubes represent the population of the small community. It does not matter if everyone has the same number of number cubes or not, as long as the total number of number cubes used in the experiment equals 100. Since approximately $\frac{1}{6}$ of the population is affected by the disease each day, have the class to select ONE number (1, 2, 3, 4, 5 or 6) on the number cubes to represent the “diseased number”. This number will be the one which, when rolled, will mean that the person has contracted the disease. The class must select only one number to be used by all participants. Suppose the class chooses the number “3”.

Everyone rolls his number cubes and removes all the “3’s”. Count the number of number cubes remaining. Then enter the number of the roll and the number of number cubes remaining in an x-y table with the number of the roll (days since the disease started in the community) being the “x” value and the number of number cubes remaining (people remaining) after that roll being the “y” value. For instance, suppose after the first roll, 85 number cubes remain (after all the 3’s are taken out). Then on the second roll only these 85 number cubes are rolled, the 3’s are removed, and the number of number cubes remaining are counted. Suppose after the second roll,

there are 72 number cubes remaining. Then roll these 72 number cubes, remove the 3's and count how many are left. Suppose after the third roll, there are 62 number cubes remaining. Continue in this manner until all the number cubes are gone.

Then your x-y table would have ordered pairs such as (1, 85) (2, 72), (3, 62), etc.....

Remember: x represents the number of days since the onset of the disease and y represents the number of healthy people remaining at the end of that day.

(It usually takes somewhere between 20 and 30 rolls (days) to end the activity)

Put the x-values in "L1" in a graphing calculator and put the y-values in "L2". Observe the scatterplot of *y versus x*. The graph should make a curve which closely resembles exponential decay. Then have the students find the curve of best fit. (Hint: Try exponential regression features on the calculator). Note: Do not enter a point in the calculator where the y value is 0. This will result in an error when you attempt to find the exponential regression equation, since for an exponential equation, y can never be zero for a non-zero value of *a*. Talk to students about what "a" and "b" represent in the exponential equation $y = ab^x$ found in the calculator. Lead students to discover that by the laws of probability, approximately one-sixth of the number cubes should be removed each time and thus five-sixths should be remaining after each roll. The numbers are decreasing by one-sixth of the starting number each time---thus an example of an exponential equation. The exponential equation of best fit should closely resemble the theoretical equation of $y = 100 * 0.833^x$ since y represents the number of people remaining after the number of days, x. Thus, 100 was the initial number of people and approximately 5/6 or 83.33% of the population remain after each day. Since the numbers remaining are decreasing, it is an example of exponential decay rather than exponential growth.

QUESTIONS:

1. By observing the scatterplot of the data, what type of regression model should be used to closely fit this data? Why would a linear model not work in this situation? (Students must recognize that the expectations are not for the y-values to decrease by the same constant number each day, rather to decrease by about the same percentage each day--- thus not linear, but exponential model).
2. Theoretically, what is the equation which should be used to model this situation? (See discussion in previous paragraph).
3. How closely related are the exponential regression equation derived from the simulation and the theoretical equation?
4. In an exponential equation of the form $y = ab^x$, why would y never be zero?
5. Using your regression equation, predict when the disease would affect 50, 60, or all members of the town with population 100. (Hint: If 60 people are affected, the number remaining would be 40; substitute $y = 40$ and find x, etc)
6. Observe the correlation coefficient for the regression model and determine its meaning. (See handout for explanation of how to use technology to find and interpret the correlation coefficient).
7. Calculate the residuals for *y versus x*. Observe a residual plot. (See handout for calculator instructions). How does the residual plot help us to informally assess the fit of the equation?

EXTENSION: EXPONENTIAL DECAY EXPERIMENT AND LINEARIZING DATA

In order to linearize the data, one should take the logarithm (either natural or common) of the y values and graph $\log y$ versus x . Using linear regression features on the calculator, find the line of best fit. Given a value of x , make predictions for y based upon the exponential equation of best fit of y versus x and then make the same prediction based upon the line of best fit of $\log y$ versus x . The answers should be the same.

Sample Data—100 Dice—Spread of Disease

Number of Roll (Days) (x)	Number of Dice (Healthy People) Remaining (y)
0	100
1	87
2	74
3	63
4	50
5	47
6	34
7	28
8	24
9	22
10	18
11	16
12	12
13	12
14	11
15	10
16	9
17	9
18	8
19	8
20	6
21	6
22	4
23	3
24	3
25	2
26	2
27	2
28	2
29	1

Other applications of exponential functions: Functions such as the one found in the “Spread of Disease” activity, **half-life applications**, and **drug concentrations** in the body after certain amounts of time are examples of exponential decay. Exponential functions can also model a pattern of **exponential growth**. **Compound interest investment problems where an initial deposit of money grows by a given percentage each year is an example of exponential growth. One of the most common examples of exponential growth deals with bacteria.** Bacteria can multiply at an alarming rate. For example, if we start with only one bacteria which can double every hour, by the end of one day we will have over 16 million bacteria.

Conditional Probability and the Rules of Probability

CONTENT STANDARDS:

Understand independence and conditional probability and use them to interpret data.

Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent. [HSS-CP.A.2]

Understand the conditional probability of A given B as $P(A \text{ and } B)/P(B)$, and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B. [HSS-CP.A.3]

For two events (whether they are independent or not), $P(A \text{ and } B) = P(A) \cdot P(B/A)$, where $P(B/A)$ means probability of B given that A has already occurred. The formula can also be written:

$P(A \text{ and } B) = P(B) \cdot P(A/B)$, where $P(A/B)$ means the probability of A given that B has already occurred. Solving for $P(A/B)$ gives the formula, $P(A/B) = \frac{P(A \text{ and } B)}{P(B)}$

Likewise, $P(B/A) = \frac{P(A \text{ and } B)}{P(A)}$ If two events A and B are independent events, then

$P(A/B) = P(A)$ and likewise, $P(B/A) = P(B)$.

For instance, suppose you are drawing two cards from a standard 52-card deck. Suppose that you are replacing the first card before you draw the second card from the deck. In other words, you are drawing from a “full deck” each time. Then if you calculate the probability of drawing an ace on the first draw and then a king on the second draw, the second draw is independent of the first draw, since you are replacing the card after each draw. It really does not matter what you drew the first time. It does not affect the probability on the second draw. Therefore, the probability of drawing an ace and then a king = $P(A \text{ and } B) =$

$P(A) \cdot P(B) = \frac{4}{52} * \frac{4}{52}$. However, if you used the general formula

$P(A \text{ and } B) = P(A) * P(B/A)$, then you would have $\frac{4}{52} * \text{probability of drawing a king given that you drew an ace on the first draw}$. However, it does not matter what you drew first; thus, you would still have $\frac{4}{52} * \frac{4}{52}$.

Now let’s consider that you are going to draw two cards and not put the first one back before drawing the second card. If we once again consider the probability of drawing an ace on the first draw and then a king on the second draw, these are not independent events. Thus, we must use the formula $P(A \text{ and } B) = P(A) * P(B/A) = \frac{4}{52} * \frac{4}{51}$, since the probability of drawing a king on the second draw GIVEN that you drew an ace on the first draw means that there are still 4 kings in the deck but only 51 cards.

Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities. [HSS-CP.A.4]

Example: Collect data from a random sample of students in your school on their favorite subject among mathematics, science, and English. Estimate the probability that a randomly selected student from your school will favor science, given that the student is in tenth grade. Do the same for other subjects and compare the results.

Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. [HSS-CP.A.5]

Observe the following example of a two-way frequency table.

		FAVORITE SUBJECT			
GRADE	Math	Science	English	TOTAL	
10th	40	9	11	60	
11th	15	16	19	50	
12th	6	10	25	41	
TOTAL	61	35	55	151	

- What is the probability of selecting at random a student who is a 10th grader and says his favorite subject is math?
ANSWER: $\frac{40}{151}$
- What is the probability of selecting at random a student who says math is his favorite subject, given that the student is a 10th grader? (The condition that the student is a 10th grader narrows our choices down to 60 students.) Thus the probability is $\frac{40}{60} = \frac{2}{3}$.
- Assuming the student chose math as his favorite subject, what is the probability that he is a 10th grader? The assumption that the student favors math narrows our choices down to 61 students. Thus the probability is $\frac{40}{61}$.
- What is the probability that the student will favor science, given that the student is in the 10th grade. The answer is $\frac{9}{60} = \frac{3}{20}$.
- ARE TWO EVENTS INDEPENDENT? In the example above are grade level of student and choice of favorite subject independent events? In other words, is there a relationship between grade level in high school and favorite subjects? If there is a relationship, they are not independent events. If there is no significant relationship, the two events are independent.

Understand the conditional probability of A given B as $\frac{P(A \text{ and } B)}{P(B)}$

and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A , and the conditional probability of B given A is the same as the probability of B . [HSS-CP.A.3]

Consider $P(\text{Math favorite subject given that the student is a } 10^{\text{th}} \text{ grader}) = \frac{2}{3}$

Now, consider $P(\text{Math favorite subject}) = \frac{61}{151}$ (does not equal) $\frac{2}{3}$. Thus the events are not independent.

Use the rules of probability to compute probabilities of compound events in a uniform probability model.

Find the conditional probability of A given B as the fraction of B's outcomes that also belong to A, and interpret the answer in terms of the model. [HSS-CP.B.6]

- Example: Suppose you roll a number cube with sides labeled 1, 2, 3, 4, 5, or 6. What is the probability that you roll an even number, given that the number rolled is greater than 3?

Solution: Since you know that the number rolled is greater than three, it is given that it is either a 4, 5, or 6. Thus the probability of even is $\frac{2}{3}$. The formula is:

$P(\text{even/greater than 3})$ “read probability of rolling an even number, given that the number rolled is greater than 3” = $P(\text{even and greater than three}) / P(\text{greater than 3}) =$

$$\frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

Apply the Addition Rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, and interpret the answer in terms of the model. [HSS-CP.B.7]

The general rule of addition is $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, regardless of whether the two events are mutually exclusive or not. If two events are mutually exclusive (both events cannot occur at the same time), then the formula could be $P(A \text{ or } B) = P(A) + P(B)$. For instance, suppose you are going to draw a card from a standard 52-card deck. Consider the probability that the card is either a jack or a 5. These are mutually exclusive, since a card cannot be both a jack and a 5. The formula, $P(A \text{ or } B) = P(A) + P(B)$, could be used. Thus $P(\text{jack or } 5) =$

$P(\text{jack}) + P(\text{a five}) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52}$. If we use the general formula, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, we would get the same answer, since $P(A \text{ and } B) = 0$, since the events cannot both occur at the same time. For two mutually exclusive events, $P(A \text{ and } B)$ will always be 0. Thus $P(\text{jack or five}) = P(\text{jack}) + P(5) - P(\text{jack and } 5) = \frac{4}{52} + \frac{4}{52} - 0 = \frac{8}{52}$. (Same answer as before).

Now, let's consider the probability of drawing one card from a standard 52 card deck and getting either a jack or a red card. These are not mutually exclusive events, since a card can be BOTH a jack and a red card—jack of hearts or jack of diamonds. Thus, we must use the formula, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

$$P(\text{jack or red}) = P(\text{jack}) + P(\text{red}) - P(\text{both jack and red}) = \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

If you think about this, there are only 28 cards in the deck that are either a jack or a red card: Jack of spades, Jack of clubs, and 26 red cards (which will include the two red jacks).

Thus, the general addition rule is $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

(+) *Apply the general Multiplication Rule in a uniform probability model, $P(A \text{ and } B) = P(A)P(B/A) = P(B)P(A/B)$, and interpret the answer in terms of the model. [HSS-CP.B.8]*

Refer to notes under [HSS-CP.A.3]

(+) *Use permutations and combinations to compute probabilities of compound events and solve problems. [HSS-CP.B.9]*

- A permutation of a set of objects is an arrangement of some or all of these objects in a specific order. The order of the objects is important. For example, FER is a different permutation than RFE, etc. A combination of a set of objects is a selection of some (or all) of these objects with no regard to order. Order is NOT important. The combinations FER and RFE are the same combination.
- Formulas: ${}_n P_r = \frac{n!}{(n-r)!}$ and ${}_n C_r = \frac{n!}{r!(n-r)!}$
- Example: Find the probability that a committee of 10 students chosen from an organization consisting of 40 male and 35 female students will include 3 males and 7 females. First of all, how many committees of size 10 from a group of 75 can be chosen? ORDER IS NOT IMPORTANT, so use Combinations. Thus ${}_{75}C_{10} = 8.289311064E11$ or 828,931,106,400 This gives us the total number of possible committees.

You may use the TI-83/84 calculators to find permutations and combinations. In the case of trying to find ${}_{75}C_{10}$, you first type in 75 and then press the MATH key, scroll over to PRB and then down to nCr, press ENTER and then type in 10 and press ENTER for the answer. You will find permutations the same way, except you will choose nPr rather than nCr.

We wish to compute the probability for having 3 males and 7 females on the committee. Again order is not important, so we use combinations.

$${}_{40}C_3 * {}_{35}C_7 = 9880 * 6,724,520 = 66,430,257,600$$

We now find the probability by dividing the number of ways to have the desired committees divided by the total number of possible committees = $\frac{66,430,257,600}{828,931,106,400} = 0.08$

ACTIVITY: Get two different flavors of STARBURST candy and place 6 of flavor A and 4 of flavor B in a brown paper bag. Let participants reach into the bag and draw out 2 pieces of candy? What is the probability that both pieces of candy are the same flavor? (Replace the candies after each draw and see how many participants got a matched set as compared with total number of participants. How does this experimental probability compare with the theoretical answer of roughly 47%?

Probability that both are flavor A or both are flavor B is: $\frac{C(6,2) + C(4,2)}{C(10,2)} = 0.47 \text{ or } 47\%$

- Example: In how many ways can a code sequence be formed if it must consist of 3 letters followed by 3 digits and none of the digits nor letters may repeat? Since we cannot repeat and order does make a difference, we are choosing 3 different letters from 26 and 3 different digits from 10, AND order is important. Thus, we use ${}_{26}P_3$ for the letters and then ${}_{10}P_3$ for the digits. The answer is ${}_{26}P_3 * {}_{10}P_3 = 11,232,000$.
- PLEASE NOTE: If the letters and digits could have been repeated, we would not necessarily be choosing 3 distinct letters from 26, etc. Thus, we could not use permutations, since permutations and combinations require that the subset chosen has no repetitions. We would then have used the counting principle and the answer would have been $26 * 26 * 26 * 10 * 10 * 10 = 17,576,000$

Using Probability to Make Decisions

CONTENT STANDARDS:

Calculate expected values and use them to solve problems.

(+) *Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions. [HSS-MD.A.1]*

(+) *Calculate the expected value of a random variable; interpret it as the mean of the probability distribution. [HSS-MD.A.2]*

(+) *Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value. For example, find the theoretical probability distribution for the number of correct answers obtained by guessing on all five questions of a multiple-choice test where each question has four choices, and find the expected grade under various grading schemes. [HSS-MD.A.3]*

(+) *Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value. For example, find a current data distribution on the number of TV sets per household in the United States, and calculate the expected number of sets per household. How many TV sets would you expect to find in 100 randomly selected households? [HSS-MD.A.4]*

Expected Value = Sum of all the products of the outcomes multiplied by their respective probabilities.

Mean of a Discrete Random Variable : $E(X) = \mu_x = \sum [x_i * P(x_i)]$

1. Calculate Expected Gross Winnings of a Lottery Ticket

Example Calculate the expected gross winnings for the \$5 BIG GAME ticket with probabilities given below.

Prize Probability

\$1 prize with probability $\frac{1}{10}$

\$10 prize with probability $\frac{1}{20}$

\$50 prize with probability $\frac{1}{50}$

\$100 prize with probability $\frac{1}{100}$

\$1000 prize with probability $\frac{1}{1000}$

\$5,000 prize with probability $\frac{1}{500,000}$

The expected gross winnings are obtained by adding all of the products of the prize multiplied by the probability. This is calculated as

$$(\$1 \times \frac{1}{10}) + (\$10 \times \frac{1}{20}) + (\$50 \times \frac{1}{50}) + (\$100 \times \frac{1}{100}) + (\$1000 \times \frac{1}{1000}) + (\$5000 \times \frac{1}{500,000}) = \$3.61$$

Thus, for each \$5 ticket you purchase, you can expect to win \$3.61 back on the average. However, since you paid \$5 for the ticket, you can expect to lose on the average \$1.39 on each ticket.

2. Card Draw

Draw a card randomly from a deck of 52 playing cards, and you will receive the number of dollars indicated by the rank (Ace counts "1," face cards count "10" and all other cards count their face value).

What is the expected payoff? In other words, on the average, how much money do you expect to win or lose per game?

Simulation: Using a deck of cards, have each student play the game above and record his/her earnings. Then find the average earnings for the class.

My winnings (loss) _____ Class average _____

Expected Value: (calculated theoretically): _____

How do the experimental and theoretical values compare?

3. Toss a coin

Toss a coin 5 times and let T be the number of times heads appears. What is the expected value of T?

4. Which Game???

You can choose to play one of two games. Each game costs one dollar to play.

Game 1: A wheel with three numbers on it--zero, one, and two--is spun so that there is a 50% chance that the wheel lands on zero, a 10% chance the wheel lands on one, and a 40% chance the wheel lands on two. You get back the amount in dollars of the number that the wheel lands on.

Game 2: A different wheel with three numbers on it--zero, one, and two--is spun so that there is a 5% chance that the wheel lands on zero, a 70% chance the wheel lands on one, and a 25% chance the wheel lands on two. You get back the amount in dollars of the number that the wheel lands on.

Problems:

- If you had \$10 to spend playing this game, which one would you choose to play and why? What would be your expected net earnings (or loss) after playing 10 games?
- If your math team was trying to raise money at a carnival, which game board should they choose to set up—game 1 or game 2? Why? If 300 students paid one dollar each and played the game you chose to set up at the carnival, how much profit (or loss) would you expect your club to earn?

Use probability to evaluate outcomes of decisions.

(+) Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values. [HSS-MD.B.5]

Find the expected payoff for a game of chance. For example, find the expected winnings from a state lottery ticket or a game at a fast-food restaurant. [HSS-MD.B.5a]

Evaluate and compare strategies on the basis of expected values. For example, compare a high-deductible versus a low-deductible automobile insurance policy using various, but reasonable, chances of having a minor or a major accident. [HSS-MD.B.5b]

(+) Use probabilities to make fair decisions (e.g., drawing by lots, using a random number generator). [HSS-MD.B.6]

(+) Analyze decisions and strategies using probability concepts (e.g., product testing, medical testing, pulling a hockey goalie at the end of a game). [HSS-MD.B.7]

SCIENCE/MEDICAL EXAMPLES:

1. *The probability of having a certain disease is 0.05. The probability of testing positive if you have the disease is 0.98. The probability of testing positive if you do not have the disease is 0.10. What is the probability that you have the disease if you test positive for it?*

Draw a tree diagram to illustrate this problem.

What is the probability that you have the disease, if you test positive for the disease?

2. *Suppose that a drug test for an illegal drug is such that it is 98% accurate in the case of a user of that drug (e.g. it produces a positive result with probability .98 in the case that the tested individual uses the drug) and 90% accurate in the case of a non-user of the drug (e.g. it is negative with probability .9 in the case the person does not use the drug). Suppose it is known that 10% of the entire population uses this drug.*

- a. *You test someone and the test is positive. What is the probability that the tested individual uses this illegal drug?*
- b. *What is the probability of a false positive with this test (e.g. the probability of obtaining a positive drug test, given the person tested is a non-user)?*
- c. *What is the probability of obtaining a false negative for this test (e.g. the probability that the test is negative, given that the individual tested is a user)?*
3. **Observe the following table and indicate if the given disease and smoking are independent events?**

	% who are Smokers	% who are Non-Smokers	Total Percent
% with the disease	5		8
% Do not have disease			
Total Percent	25		100

4. **Observe the following table and indicate if the ability to memorize and the noise level (loud music or silence) within the room are independent events.**

	Able to memorize selection	Not able to memorize selection	Total Students
Loud Music	9	36	45
Silence	3	12	15
Total Students	12	48	60

ANSWERS to problems on previous page

1. A tree diagram is helpful, but it can be solved with a formula
 $P(\text{have the disease/ test positive})$ which is read “the probability that you have the disease given that you test positive” = $P(\text{have the disease and test positive})$ DIVIDED BY $P(\text{test positive})$ =

$$\frac{(0.05)(.98)}{(0.05)(.98) + (0.95)(.10)} = 0.34$$

2. A. $0.52 \quad \frac{.098}{.098 + .09} = 0.52$

B. $0.1 \quad \frac{.09}{0.9} = 0.1$

C. $0.02 \quad \frac{.002}{.1} = 0.02$

3. Not independent, since for instance:

$P(\text{having disease}) = 8\%$ and $P(\text{having disease, given a smoker}) = 20\%$

4. Independent, since for instance: $P(\text{able to memorize}) = 12/60 = 1/5$ and

$P(\text{Probability(able to memorize, given silence)}) = P(\text{able to memorize, given loud music}) = 1/5$

Graphing Regression Equations on the TI-83/84 Calculators

1. **First of all, clear all previous graphs.** Hit the “y=” button and clear all equations entered in the y_1, y_2, \dots graph lines. At the same screen, use the “up” arrow and scroll up to the Plot 1, Plot 2, Plot 3 line at the top. Make sure that none of these is highlighted. If they are highlighted, hit the “enter” key as you scroll over to each one and this will toggle them off.
2. To get prepared to enter data into lists, hit the “STAT” key and “enter”. If L1 and L2 contain data, **clear these lists out.** There are several ways to clear lists, but one of the easiest ways is to use the “up” arrow and highlight the list name, L1, and then hit “clear” and “enter”. Scroll over to L2, go up to the list name (L2), highlight it, and hit “clear” and “enter”. You can clear any list by using this method.
3. Now you are ready to **enter data into L1 and L2.** Go to L1 and type the first number, press “enter”, type the next number in the list, press enter, etc. After you have typed all the “x” values into L1, go to L2 and type all the “y” values into this list by the same method.
4. **To see a graph of your data,** hit “2nd” and then “y=” to access the STATPLOT menu. Push “enter” to get to the first statplot menu. Turn it on. To do this, push “enter” until only “on” is highlighted. Scroll down to the menu of graphs and choose the one with the scatterplot (the first one). Scroll down and for the Xlist: type in L1 (“2nd” and then “1” key). Scroll down to Ylist: and type in L2 (“2nd” and then “2” key). Choose your mark and then hit “ZOOM” on the top row keys and choose ZOOMSTAT and push “enter”. Your scatterplot graph should appear.
5. **Turn your diagnostics on to see correlation coefficient.** This will allow you to see “r” and “ r^2 ” values after you find the regression equation. Reminder: The closer the absolute value of “r” (the correlation coefficient) is to one, the better the fit. To turn the diagnostics on, hit “2nd” and “Zero” key. Notice the zero key has “catalog” written above it. This will access the catalog feature. You must scroll down to “diagnostics on”--the best way is to hit the “D” key and this will get you to the d’s and you can find “diagnostics on” quicker. Once you find it, hit “enter” twice until it says “done”.
6. Now go to the STAT key and then scroll over to CALC and then you can scroll down to whatever **regression equation** you think would most likely fit the curve or line. For instance, if you think that a linear equation is the best fit for your data, choose LinReg. Hit “enter” once and then the cursor is blinking---it is waiting for you tell it which lists to perform the linear regression on. So you type in L_1, L_2, y_1 . **Important:** To type in L1, hit the “2nd” key and the “1” key, since you see L1 above the numeral 1 key. Likewise, to type in L2, hit the “2nd” key and the “2” key, since you see L2 above the numeral 2 key. To type in Y1, hit the VARS key, scroll over to Y-vars, hit “enter”, and then hit “enter” again. Now you should see, LinReg L_1, L_2, y_1 on the screen. (Remember to insert commas). Hit “enter” and you will see your equation appear.

CAUTION: When observing the regression equation, be sure to notice if any of the coefficients are written in scientific notation format. Sometimes they will be of the form $a = 9.56021345 \text{ E } -5$, meaning 0.0000956021345.

7. Next, push the **GRAPH** key on the top row and the graph will appear. The graph will contain the scatterplot as well as the graph of the regression equation.
8. **To evaluate the function at a particular value of x**, you can use function notation just as you would for the following example---if $f(x) = 2x$, then $f(5) = 10$. Thus, suppose you want to evaluate the value of your function at $x = 216$. You need to find $y_1(216)$, since your function is located in y_1 . Remember to get to y_1 , you hit VARS, scroll over to y-vars, then "function", then y_1 . Then after y_1 appears on your screen, type $y_1(216)$ and push enter and the function located in y_1 will be evaluated at $x = 216$.
9. **To see a residual plot.** Immediately after you have completed a regression equation for a set of ordered pairs, you may graph the residual plot. A residual for a given x-value is the difference between the observed y value and the predicted y value (the y-value which the regression equation predicts for a given x value). To graph a residual plot, go to 2nd y= to access the stat plot menu. Proceed as before to graph a scatterplot. This time the x-list will be the list where the x values are located. The y-list should be the list of residuals. To access the residual list, go to 2nd STAT and scroll down until you see RESID and press enter. This will place RESID into the ylist under the stat plot menu. Now press ZOOM STAT and the resulting graph should be the residual plot for this regression. In order to assess the fit of a regression equation for a set of data by looking at a residual plot, one must look to see that the residuals are relatively small, scattered with no apparent pattern in the plot, and have approximately an equal number of positive and negative residuals.

Regression, Residuals, and Correlation Coefficients

With TI-NSPIRE CX Calculators

Consider the information found in the table below. The table indicates the number of annual home burglaries since 2003 in a given city. Note: "03" represents the year 2003.

Year	03	04	05	06	07	08	09	10	11
Home Burglaries	350	162	43	11	55	86	361	640	992

1. Open a new document and add "Lists and Spreadsheets". In the cell with the "A" in column A, name the first column "year". In column B, name the column "thefts". Skip down to cell A1 and begin typing the years in that column and then go over to cell B1 and start typing the number of burglaries in that column.) Press ENTER after each entry.
(Note: Since the numbers in column A are consecutive, you could enter 3 in the first cell, then enter "=A1+1" in cell 2, then hit ENTER. Then highlight cell A2 and go to MENU, DATA, FILL, click. Scroll down and highlight the cells below and then push ENTER to have a quick entry of all the data in column A.)
2. Hit the DOC key, choose INSERT and then choose DATA AND STATISTICS. Points will appear in the next window. Scroll down to the bottom where it says "Click to add variable" and choose "year" from the menu, since this is the X- variable. Now scroll over to the right of the graph to where it says "Click to add variable" and choose "thefts" from the menu, since this is the y-variable. A scatter plot will appear.
3. Even though the scatter plot appears to be quadratic in nature, let's examine some regression models to illustrate some very important points. Hit the MENU button and then choose ANALYZE and then choose REGRESSION, then SHOW LINEAR, and click. (You should see a line with the equation $y = 78.55x - 249.85$). Ask students if the line appears to fit the data?? (No)

To assess the residual plot, hit MENU, choose ANALYZE, RESIDUALS, and SHOW RESIDUAL PLOT, then click. Notice the residuals have a distinct pattern, which is another indication that the linear model is not a good fit.

To check the correlation coefficient for the linear regression (r), go back to the table found on page 1.1 by scrolling up to the 1.1 tab at the top of the calculator screen and click on it. Hit the MENU button again, choose STATISTICS, then STAT CALCULATIONS, then LINEAR REGRESSION, then click. A screen will appear. Input "year" in the x list cell and "thefts" in

the y-list cell. Go down and hit OK. You will notice that the table now shows the regression equation and the r value (0.650411) and the value of r^2 (0.423035). The closer the r value is to 1 or -1, the better the fit. So again, accessing the actual graph of the line with the data, the residual plot (which contains a distinct pattern), and the r value---all indicate that the linear model may not be the best fitting model for the data.

4. To see another potential model of best fit, go back to the page with the graph on it. Hit the MENU button, then choose ANALYZE, then REGRESSION, then Show QUADRATIC. You will see the parabola passing through the data points, as well as the quadratic equation $y = 40.2695x^2 - 485.223x + 1454.89$ on the screen. Notice the residual plot also changes. Assess the fit....How well does this curve fit the model? (Very well) Assess the residual plot—Are the residuals scattered and relatively small? Yes Do you see a distinct pattern in the residuals (no). Thus, this appears to be a good fitting model for the data.

Let's look at the R^2 value for the quadratic regression. Go back to page 1.1 and open it again. Then hit the MENU button again, choose STATISTICS, then STAT CALCULATIONS, then QUADRATIC REGRESSION, then click. A screen will appear. Input "year" in the x list cell and "thefts" in the y-list cell. Go down and hit OK. You will notice that the table now shows the regression equation and the R^2 value of 0.993771. This R^2 value is called the coefficient of determination and shows the percentage of variation in the number of burglaries which is attributed to the variation in the year. Of course, there are probably underlying reasons for why certain years had such fewer numbers of home burglaries—perhaps change in law enforcement practices, neighborhood watches, etc. A high R^2 value does not indicate causation.