Connecting Statistics with the Algebra Classroom

Doug Tyson

NCTM Annual Meeting 2019

San Diego, CA

Activity 1 – Modeling With Starbursts*

Activity 2 – When Are We Ever Going To Use Logarithms?

*Many thanks to statistics teacher Paul Myers (Atlanta, GA) for sharing his idea for a candy grab with Starbursts! In particular, Activity 2 is my adaptation of his wonderful idea.

copies of this handout can be found at MrTysonStats.com

(THIS PAGE INTENTIONALLY LEFT BLANK.)

GRABBING CANDY

Mr. Tyson is rewarding his students with a "candy grab." If a student earns a reward, Mr. Tyson will allow that student to reach in to a bowl and grab as many Starbursts as they can hold in their hand. Your job is to help Mr. Tyson predict how many Starbursts he can expect a student to grab. Students must grab "overhand" (like a claw) and may not use the side of the bowl, their other hand, or their body to assist in the grab.

1. Predict how many candies you will be able to grab: ________________

   Answers will vary by student.

2. What variable(s) might help you predict how many candies a student would grab?

   Answers will vary by class, but try to guide them to a hand measurement like handspan.

3. Collect data from the class on your predictor (explanatory) variable AND on the response variable. Record the class data here.

   Answers will vary by class. Data from one class are in the data file Starbursts.

WHAT'S THE BEST MODEL?

4. Graph your data with the *Least Squares Regression* applet at rossmanchance.com/applets. Click "Show Movable Line." Now move your line until it is a good fit (summary) of the data. Write the equation of your line here.

   Answers will vary by student.

5. Different people will choose different lines to fit the same set of data. What rule/criterion should we use to judge which line (model) does the BEST job for predicting the number of candies a person can grab?

   Answers will vary by student, but some common suggestions include ideas that the line is "closest" to the data (or vice versa). The catch is how do we measure "close." Let students try to come up with ways to measure distance from the line and then total or average all of those. In statistics, we use the sum of the squared residuals, but the sum of the absolute values of the residuals is another possible response.

THE LEAST-SQUARES CRITERION

6. One criterion for judging the "fit" of a model to data is to sum the squares of the residuals (prediction errors). Use the applet to calculate the sum of the squared residuals for your line and record it. Who has the best-fit model?

   Answers will vary by class. The best-fit model is the one with the smallest sum of squared residuals.

7. Can we do better? Let's have a contest. Try moving your line to minimize the sum of the squared residuals. What is the least sum of squares you can get?

   Answers will vary by class.

8. Use the applet to the least-squares regression line. Record this equation using variable names or abbreviations instead of x and y.

   Answers will vary by class. An example from one class: predicted candies = -27.9 + 2.74(handspan).

USING AND UNDERSTANDING THE MODEL

9. Use the least-squares regression equation to predict the number of candies that could be grabbed by student A, whose handspan is 21 cm.

   Using the equation above, predicted candies = 29.64 candies

10. Use the least-squares regression equation to predict the number of candies that could be grabbed by student B, whose handspan is 22 cm.

    Using the equation above, predicted candies = 32.38 candies

11. Calculate the difference (student B – student A) in the predicted number of candies for these two students. Does this number look familiar?

    Difference = 32.38 – 29.64 = 2.74 candies (this is the slope!)

12. Identify and interpret the slope.

    The slope is 2.74. For each additional 1 cm increase in handspan, we predict that 2.74 additional candies would be grabbed.

13. Predict the number of candies that could be grabbed by a student with a 0 cm handspan. Why is this prediction unreasonable?

    predicted candies = -27.9 + 2.74(0) = -27.9 candies

    This is unreasonable because the number of candies can't be negative and handspan can't be 0.

14. Identify and interpret the y-intercept.

    The predicted number of candies for someone with a handspan of 0 cm is –27.9 candies.

When Are We Ever Going To Use Logarithms?

Transforming One Variable

Open the data file *TwitterAccounts*. About half of these accounts are owned by real humans and about half are bot accounts.

1. Go to stapplet.com and use the *One Quantitative Variable, Multiple Groups* applet to make graphs comparing the average number of posts per day between humans and bots.
   a) How would you describe the shape of these distributions?

      Strongly skewed to the right.

   b) Why is it hard to see the shape?

      Because the values vary from the very small (less than 1) to the very large (thousands).

2. Transform the average number of posts per day by taking the logarithm (base 10) of each value.
   a) What does a logarithm of 3 mean in this situation? 0? –1?

   3 means 10^3 or 1000 posts per day, on average. 0 means 10^0 or 1 posts per day, on average. –1 means 10^{-1} or 0.1 posts per day, on average

   b) Make comparative graphs of the logarithms, comparing real humans to bots. What similarities/differences do you see?

   The shapes are similar (now they're approximately normal). The mean for bots is greater than the mean for humans. The variability is about the same in both distributions.

Transforming Two Variables: Light Intensity in a Lake

The following data were collected by college students who measured the intensity of light at various depths in a lake. The data are stored in a file called *LightIntensity*.

Depth (meters)	5	6	7	8	9	10	11
Light Intensity (lumens)	168.00	120.42	86.31	61.87	44.34	31.78	22.78

(Source: *The Practice of Statistics,* 4th edition. Starnes, Yates, Moore.)

3. Use technology to make a scatterplot of *light intensity* vs. *depth*. Describe the association.
   There is a curved (non-linear), strong, negative association between light intensity and depth.

When we have nonlinear data, but don't have a theoretical model for the relationship between two variables, we can try to use logarithms to straighten the data.

4. Use technology to find the common logarithm (logarithm base 10) of the y-variable. Graph log(*intensity*) versus *depth*. Do the transformed data look linear?

   Yes. The transformed data look very linear.

5. Perform a least-squares regression and record the equation.
   Predicted log(light intensity) = 2.9485 – 0.1446(depth)

6. Predict the *light intensity* at a depth of 14 meters. Show your work. Explain why the prediction in number 4 may not be very reliable.
   Predicted log(light intensity) = 2.9485 – 0.1446(14) = 0.9241 lumens.

   Predicted light intensity = 10^0.9241 = 8.40 lumens.

   This prediction isn't reliable because it is a big extraplolation.

Unknown to us, y (*light intensity*) is an exponential function of x (*depth*). Taking the logarithm (common or natural) straightens data that follow an exponential model in the form $y = ab^x$.

Transforming Two Variables: Weights and Lifespans

7. Use technology to make a scatterplot of *lifespan* versus *weight* and describe the association between these variables.
   Curved (non-linear), strong, positive.

8. Try plotting log(*lifespan*) versus *weight*. Did this transformation straighten the data?
   No, it's still curved.

9. Maybe this is a "backwards" exponential function. Try plotting *lifespan* versus log(*weight*). Did this transformation straighten the data?
   No. Now the curve is concave up.

10. Plot log(*lifespan*) versus log(*weight*). Does this transformation make these data look linear?
    Yes. They look much more linear.

Species	Weight (kg)	Lifespan (yr)
Baboon	32	20
Beaver	25	5
Cat, Domestic	2.5	12
Chimpanzee	45	20
Dog	8.5	12
Elephant	2800	35
Goat, Domestic	30	8
Gorilla	140	20
Grizzly Bear	250	25
Guinea Pig	1	4
Hippopotamus	1400	41
Horse	480	20
Lion	180	15
Mouse, House	0.024	3
Pig, Domestic	190	10
Red Fox	6	7
Sheep, Domestic	30	12

11. Perform a least-squares regression of the transformed data and record the equation.
    Predicted log(lifespan) = 0.7617 + 0.2182log(weight)

12. Predict the *lifespan* of a Capybara, whose adult mean weight is 50 kg (the Capybara is the largest rodent in the world).
    Predicted log(lifespan) = 0.7617 + 0.2182log(50)

    Predicted log(lifespan) = 1.1324

    Predicted lifespan = 10^1.1324 = 13.56 years

Unknown to us, y (*lifespan*) is a power function of x (*weight*). Taking the logarithm (common or natural) straightens data that follow a power model in the form $y = ax^p$.