

# 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](http://MrTysonStats.com)

(THIS PAGE INTENTIONALLY LEFT BLANK.)

---

# MODELING WITH STARBURSTS

## 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: \_\_\_\_\_
2. What variable(s) might help you predict how many candies a student would grab?
3. Collect data from the class on your predictor (explanatory) variable AND on the response variable. Record the class data here.

## WHAT'S THE BEST MODEL?

4. Graph your data with the *Least Squares Regression* applet at [rossmanchance.com/applets](http://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.
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?

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

## 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.
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.
11. Calculate the difference (student B – student A) in the predicted number of candies for these two students. Does this number look familiar?
12. Identify and interpret the slope.
13. Predict the number of candies that could be grabbed by a student with a 0 cm handspan. Why is this prediction unreasonable?
14. Identify and interpret the  $y$ -intercept.

---

# 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](http://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?
  - b) Why is it hard to see the shape?
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?
  - b) Make comparative graphs of the logarithms, comparing real humans to bots. What similarities/differences do you see?

## 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*, 4<sup>th</sup> edition. Starnes, Yates, Moore.)

3. Use technology to make a scatterplot of *light intensity* vs. *depth*. Describe the association.

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  $\log(\text{intensity})$  versus *depth*. Do the transformed data look linear?

- Perform a least-squares regression and record the equation.
- 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.

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

- Use technology to make a scatterplot of *lifespan* versus *weight* and describe the association between these variables.
- Try plotting  $\log(\textit{lifespan})$  versus *weight*. Did this transformation straighten the data?
- Maybe this is a “backwards” exponential function. Try plotting *lifespan* versus  $\log(\textit{weight})$ . Did this transformation straighten the data?
- Plot  $\log(\textit{lifespan})$  versus  $\log(\textit{weight})$ . Does this transformation make these data look 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

- Perform a least-squares regression of the transformed data and record the equation.
- Predict the *lifespan* of a Capybara, whose adult mean weight is 50 kg (the Capybara is the largest rodent in the world).

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  $y = ax^p$ .