Using Mathematical Modeling to Motivate the Study of Polynomial Functions

Rachelle Bouchat and Kristen Barnard
Berea College

Abstract: *The authors show how the online graphing tool Desmos can be used to motivate the study of polynomial end-behavior through the use of mathematical modeling of data sets.*

Keywords: *mathematical modeling, polynomials, technology*

Introduction

The question, *When are we ever going to use this?*, is all too familiar in mathematics classrooms. Unfortunately, too many concepts are presented in a way that obscures their applicability. This is particularly true of school algebra.

Additionally, students see complicated polynomial functions given in problem sets and wonder where they came from. Consider, for instance, the following problem from Beecher et al. (2016, p. 237):

> Since 2006, total admissions to state and federal prisons have been declining (Source: Bureau of Justice Statistics). The quartic function
>
> $$p(x) = 6.213x^4 - 432.347x^3 + 1922.987x^2 + 20503.912x + 638684.984,$$
>
> where x is the number of years after 2001, can be used to estimate the number of admissions to state and federal prisons from 2001 to 2012. Estimate the number of prison admissions in 2003, in 2006, and in 2011.

When posed questions such as this, students often ask about the origins of the modeling function. Students wishing to construct their own fit functions will only find raw data at the Bureau of Justice Statistics website—there are no quartic functions provided. In the discussion that follows, we present an activity that helps teachers address how such functions arise.

In our data-driven world, teachers and students at all levels should be looking for the applicability of the mathematics they study. Doing so enhances student interest and engagement and prepares students for careers in STEM-related fields. The United States Labor Department and the Bureau of Labor Statistics project high demand for statisticians, data scientists, epidemiologists, operations research analysts, actuaries, and other mathematical scientists in the foreseeable future (Kelly, 2021).

In the sections that follow, we illustrate how mathematical modeling can be introduced into the algebra curriculum in a fun and interesting way using the interactive graphing utility, Desmos (2015). Specifically, we use Desmos to promote active learning. We share an example activity that uses real-world data to make the study of the end-behavior of the graphs of polynomial functions more meaningful. Using real-world data helps showcase applications while helping students consider the importance of "goodness of fit" in their work. Teachers interested in using a data-based approach will find numerous data sources for classroom use online. For instance, we provide links to NCTM (2022) and Kaggle (2022) websites in our references at the end of this paper.

The use of Desmos requires students to understand the shape of a polynomial, including how the leading term effects the end-behavior of the graph of the polynomial, to determine a polynomial model for the data. This requires students to identify end-behavior of the model. Once a suitable polynomial is found, it can be used to make predictions about future trends.

As students construct and analyze polynomial models, they are active learners, engaging in "mathematical investigation, communication, and group problem-solving, while also receiving feedback on their work from both experts and peers" (CBMS, 2016). In mathematics, as well as other STEM fields, active learning has been shown to have a lasting (and positive) impact on students' learning and achievement, particularly with women and traditionally low-achieving students (Kogan and Laursen, 2014). Research supports the positive effects of active learning on examination performance in K-12 mathematics and STEM education (Freeman, et. al., 2014).

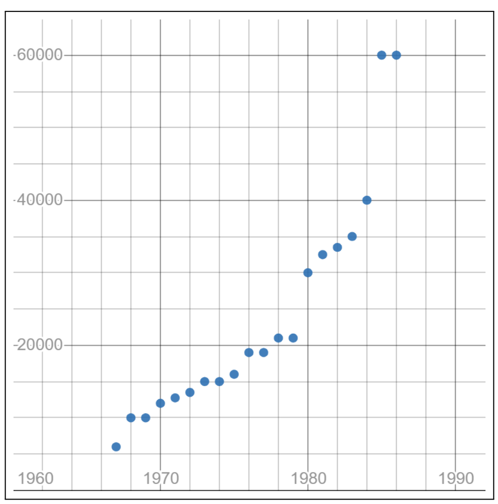
Using Desmos to Find Polynomial Models

After students find a data set, their first step is to enter data into a table. Let's consider the following data about minimum annual salaries in Major League Baseball (Baseball Almanac, 2022):

(1967, 6000)	(1968, 10000)	(1969, 10000)	(1970, 12000)	(1971, 12750)
(1972, 13500)	(1973, 15000)	(1974, 15000)	(1975, 16000)	(1976, 19000)
(1977, 19000)	(1978, 21000)	(1979, 21000)	(1980, 30000)	(1981, 32500)
(1982, 33500)	(1983, 35000)	(1984, 40000)	(1985, 60000)	(1986, 60000)

After inputting a data set into a table in Desmos, the data points are automatically plotted as shown in Figure 1.

Figure 1: *MLB player minimum wage salaries plotted using Desmos.*



We wish to find a polynomial model to fit this data and use it to make predictions about player salaries in subsequent years. With a scatterplot, students can analyze the shape of the data (including end behavior) to determine which type(s) of polynomial models are appropriate. Students can do this using regression features in Desmos.

Using default settings (x_1 and y_1), the syntax to generate linear, quadratic, and cubic fit curves is as follows:

$$y_1 \sim ax_1 + b, \tag{1}$$
$$y_1 \sim ax_1^2 + bx_1 + c, \tag{2}$$
$$y_1 \sim ax_1^3 + bx_1^2 + cx_1 + d. \tag{3}$$

Students use the tilde (~) rather than an equals sign when defining fit functions. The syntax for generating higher degree polynomial functions is analogous. However, using predefined constants such as e should be avoided as variable names since they are already assigned values in Desmos (e.g., $e \approx 2.718$).

When using Desmos to model the baseball salary data, we expect salaries to be strictly increasing. This means that the graph tends towards $-\infty$ on the left and $+\infty$ on the right. Using end-behavior to guide the construction of our polynomial model, we only consider polynomials of odd degree. The best fit linear function is

$$f(x) = 23860.65x + (-4.6932 \times 10^6),$$

with an associated $r^2 = 0.831$. Modeling the data with a cubic function yields

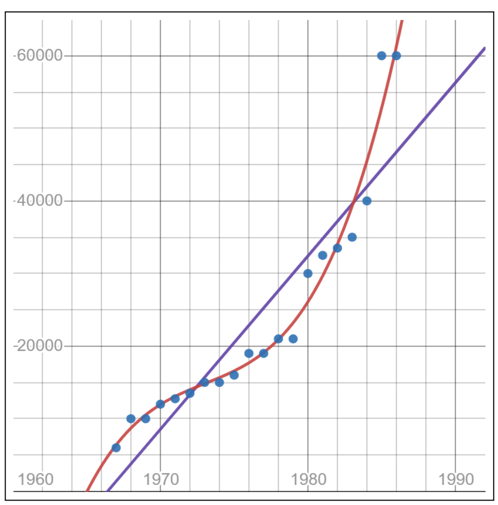
$$f(x) = 16.4124x^3 - 97146.1x^2 + (1.9167 \times 10^8)x - (1.2606 \times 10^{11}),$$

with an associated $r^2 = 0.9703$.

The r^2 statistic, also known as the coefficient of determination, provides a measure of how well the model should predict values outside the given data based on how well it fits the data set itself. The coefficient of determination falls between 0 and 1, where 0 is indicative of a poor fit and 1 is indicative of a good fit. When r^2 values are close to 1 are more likely to be appropriate for making predictions.

This is a naïve interpretation of the r^2 statistic, but it is appropriate for our students' level. Students use the r^2 value in Desmos to compare models when they can't tell which polynomial provides a better fit by visual inspection alone. The graphs of both curves against the data points are provided in Figure 2.

Figure 2: *The two regression curves plotted using Desmos.*



The r^2 value is closer to 1 with the cubic model than with the linear model (0.9703 versus 0.831, respectively). This provides evidence that the cubic model provides a better fit.

Now that we have a usable model, we can put it to the test and make predictions about future baseball salaries. In 1990, the minimum wage salary for Major League Baseball increased to \$100,000. Our cubic model predicts a value of $f(1990) = 16.4124(1990)^3 - 97146.1(1990)^2 + (1.9167 \times 10^8)(1990) - (1.2606 \times 10^{11}) = \$108,994.94$, which is much closer than the linear model's prediction of \$56,282.33.

Another Example

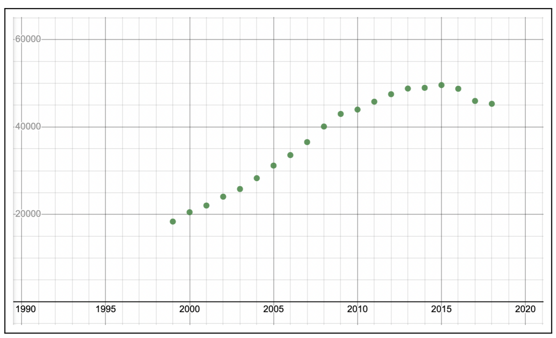
In this section, we provide an additional example, along with suggestions for implementing a data analysis approach to polynomials with Desmos.

Let's begin by considering the number of new cases of thyroid cancer from 1999 to 2018 as recorded by the Center for Disease Control and Prevention (CDC, Accessed May 2022):

(1999, 18359)	(2000, 20503)	(2001, 22036)	(2002, 24058)	(2003, 25812)	
(2004, 28295)	(2005, 31175)	(2006, 33575)	(2007, 36550)	(2008, 40119)	(4)
(2009, 43001)	(2010, 43995)	(2011, 45803)	(2012, 47518)	(2013, 48808)	
(2014, 48975)	(2015, 49605)	(2016, 48765)	(2017, 45971)	(2018, 45324)	

A graph of this data is provided in Figure 3.

Figure 3: *New cases of thyroid cancer from 1999 to 2018 plotted using Desmos.*

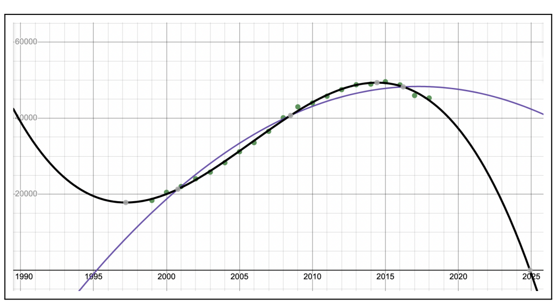


Given the shape of the graph, students conclude that a linear function is inappropriate for the data. Quadratic and cubic regression produce the following models:

$$y = -99.596x^2 + 401820x - 4.0524 \times 10^8, \quad r^2 = 0.9682,$$
$$y = -12.3671x^3 + 74418.2x^2 - (1.4927 \times 10^8)x + 9.9797 \times 10^{10}, \quad r^2 = 0.9979.$$

Graphs of the two fit curves are provided in Figure 4.

Figure 4: *Two cancer regression curves plotted using Desmos.*



With this information, students use the cubic model to make predictions about cases of thyroid cancer in 2019 and 2020 and compare their predictions to the known data. The model provides the following prediction for cases in 2025: $f(2025) = -12.3671(2025)^3 + 74418.2(2025)^2 - (1.4927 \times 10^8)(2025) + 9.9797 \times 10^{10} = -7118485.9375$. In such situations, we like to ask our students whether they believe the model's prediction is appropriate (or not) then have them justify their position with a mathematical argument.

Classroom Implementation

This activity will likely take 1-2 days of classroom time depending on the duration of your class period, but it can be scaled to best suit your needs.

Provide Students with a Data Set

Begin by providing students with a data set, such as the baseball salary data provided in this article. Showing your students a list of data values first will help them understand that mathematical models, such as the prison population function, are constructed using data. Engage your students with questions about what they see in the data values: *Do the values appear to be increasing or decreasing? Are the values changing slowly or rapidly? What would a graph of the data look like?* and so forth.

Generate a Plot of the Data with Desmos

To save time, provide your students with the data in a spreadsheet for easy copy-and-paste. With the data pasted into a table, students can check their earlier conjectures using the scatterplot automatically generated by Desmos. Note that when a table of values is plotted in Desmos, some data points may lie outside the graphing window. You may need to adjust the parameters of your x- and y-axes so that all data are visible.

Explore Regression in Desmos

Next, engage students in the process of fitting polynomial functions to data in Desmos by entering regression equations such as those provided in (1)-(3). After graphing several regression equations, ask your students which equations provide an appropriate fit for the data—for example, in the baseball salary data, students might recognize it would not be appropriate to use a quadratic or quartic model because this would suggest that at some point prior to 1967 the minimum wage salary was higher than it was in 1967. After students correctly identify appropriate fit functions, you can engage them in conversation about what makes one polynomial a better fit than others. It would be reasonable to discuss the r^2 statistic as above, but also encourage the students to make assessments based on visual inspections of the fit.

Small Group Exploration

After students have gone through the process of analyzing data, making predictions, and verifying or refuting them in Desmos, they are ready to work more independently. We assign small groups of students a new data set to explore in Desmos. For example, a group of 2-3 students might explore cases of thyroid cancer from 1999 to 2018 (see (4)).

As the groups discuss the data, make note of your students' predictions and observations. Next, have them input the data in Desmos and describe how the plots compare with their initial observations. As teachers check in on individual groups, some students might need to be reminded about resizing the graphing window. Others might struggle to articulate how plotted data either affirms or refutes their observations.

Perform Regression in Desmos

Once students have successfully plotted their data, it is time find a model for the data using regression tools in Desmos. As students examine end-behavior, they should be asked about the importance of models fitting data trends on both sides of the graph. It's important for students to know that appropriate end-behavior on the left may not be known. For instance, in the case of the thyroid cancer data presented in Figure 3, data collection may have begun in 1999—with no data occurring to the left of that point. Additionally, although data may appear to follow a trend, this may not be the case. For instance, referring once again to the thyroid data in (4), the number of new thyroid cancer diagnoses appears to be decreasing since 2015; however, this trend may not continue. Ask your students to use Desmos to explore different polynomial models until they find a suitable fit.

After students have had a chance to explore different polynomial regression lines, ask students which type of polynomial produces the best fit for their data and why they think so. For instance, Desmos generates the following quadratic and cubic polynomial models for the thyroid cancer data from (4).

$$y = -99.596x^2 + 401820x - 4.0524 \times 10^8, \quad r^2 = 0.9682,$$
$$y = -12.3671x^3 + 74418.2x^2 - (1.4927 \times 10^8)x + 9.9797 \times 10^{10}, \quad r^2 = 0.9979.$$

Students will see that the cubic polynomial is a better fit to the data visually, and that the r^2 statistic is very close to 1 which suggests that the model is a good fit for the data. Students can then use the cubic model to make predictions about 2019 and 2020 and compare it to the known data. Moreover, students can be asked whether they should believe the model's prediction for 2025.

Conclusions and Extensions

The activity we've discussed in this paper provides students with opportunities to use real-world data for mathematical modeling while helping them learn to identify the shape of the graph of a polynomial function. While we've focused much of our attention on matching a polynomial's end-behavior to data, students could also be asked to think about interior behavior of the polynomial or other defining characteristics. Numerous extensions and related activities could follow an introduction to regression in Desmos.

With the equations and the graphs on Desmos, students can discuss particular aspects of their models as well as interpretations based on contextual features of the data. For example, when students determine the x-intercept of the cubic model depicted in Figure 2, they are estimating the year in which the minimum wage for a Major League Baseball player was $0. Drawing on their knowledge of baseball and salaries, students should recognize a minimum wage of $0 as unrealistic.

Teachers are also encouraged to discuss the reasonableness of predictions made with models. Many data sets exhibit behavior that can be modeled accurately on a short interval, but that yields inaccurate predictions for values far beyond the scope of the observed data. For example, the y-intercept in our model of Major League Baseball player minimum wage salaries suggests that the minimum wage salary for a Major League Baseball player in the year 0 was below zero—perhaps players were paying to play during these years! In our thyroid cancer example, the model in Figure 4 predicts no new cases of thyroid cancer in 2025 (and a negative number of cases in 2026)!

Follow-up lessons could focus on similar investigations of data sets, but perhaps ones that are better modeled by a function that is not a polynomial. Data sets could be plotted and checked for exponential, logarithmic, and trigonometric functions to best model the data. This would be particularly useful for logarithmic data as it is often a topic where textbooks fail to adequately provide reasonable real-world applications.

This lesson could also be used to set up individual modeling projects in the classroom, having students collect or find data about an issue that is important to or interesting to them. Students then can construct a report or presentation about their data, the best function they can find to fit that data using Desmos, how good of a fit their model is for the data, and predictions they can make using their models.

References

Baseball Almanac. (n.d.). *Major League Baseball Minimum Wage: MLB Salary Research & MLB Minimum Wage Data*. https://tinyurl.com/mlbwagedata.

Beecher, J., Penna, J., & Bittinger, M. (2016). *College Algebra* (5th ed.). Pearson.

Centers for Disease Control and Prevention (CDC). (n.d.). United States Cancer Statistics: Data Visualizations. Available at: https://gis.cdc.gov/Cancer/USCS/#/Trends/.

Conference Board of Mathematical Sciences (CBMS). (2016), *Active learning in post-secondary mathematics education*. Available at: https://www.cbmsweb.org/2016/07/active-learning-in-post-secondary-mathematics-education/.

Desmos Graphing Calculator (Desmos). (2015). Available at: https://www.desmos.com/.

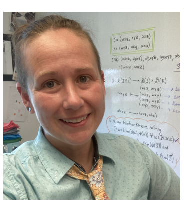
Freeman, S., Edy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.

Kaggle, Data Sets. (n.d.). Available at: https://www.kaggle.com/datasets.

Kelly, J. (September 16, 2021). *The 30 fastest-growing jobs and careers for the next 10 years*. Available at: https://tinyurl.com/forbescareers.

Kogan, M. & Laursen, S. (2014). Assessing long-term effects of inquiry-based learning: A case study from college mathematics. *Innovative Higher Education*, 39(3), 183–199.

National Council of Teachers of Mathematics (NCTM), Data Sets. (n.d.). Available at: https://www.nctm.org/classroom-resources/core-math-tools/data-sets/.



Rachelle Bouchat (bouchatr@berea.edu) is an Associate Professor at Berea College in Kentucky who is committed to making mathematics more accessible and inclusive through developing activities to show the power of mathematics and to emphasize the creativity needed to produce solutions to new problems.

Kristen Barnard (barnardk@berea.edu) is an Assistant Professor at Berea College. Her research interests lie primarily in the scholarship of teaching and learning in mathematics, alternative grading methods, and combinatorial game theory. One of her goals in every class she teaches is to help students understand that they are capable of learning math and that math is not only important, it's fun!