**Mathematical Modeling**

Scientists often use mathematical equations (lines, polynomials, exponential functions, etc.) to describe a set of experimental data.  Due to ease of use and straightforward interpretations, scientist prefer linear relationships.  Scientists will take great care in designing experiments or data analysis methods when possible, to obtain a linear relationship.  So, let’s start by evaluating two data sets that we can fit to a linear equation.

| Data Set 1 | |
| --- | --- |
| **X-Values** | **Y-Values** |
| -9 | -4.5 |
| -4.5 | -2 |
| -2.5 | 1 |
| 1 | 5 |
| 7 | 8 |

| Data Set 2 | |
| --- | --- |
| **X-Values** | **Y-Values** |
| -9 | -5.5 |
| -4.5 | -3 |
| -2.5 | 3 |
| 1 | 7 |
| 7 | 7 |

Part I – Linear fit

1. Open the [Curve Fitting simulator,](https://phet.colorado.edu/sims/html/curve-fitting/latest/curve-fitting_en.html) **Plot Data Set 1**
   1. Click "values" in the box to the right of the graph so that you can see the coordinate of your points as you add them to the simulator
   2. Drag the orange points from the bucket on the bottom left to graph corresponding to **data set 1**. You will be adding five points.
2. Click "curve" in the box to the right of the graph to add a trendline to the data.
   1. Select "linear" from the middle right box and "best fit" from the bottom right box
   2. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
   3. Record the fit line equation in y=mx+b format from the top of the graph: **y = 0.83x + 2.8**
   4. Record the r2 value from the box to the left of the graph: **r2 = 0.96**
3. Reset the [Curve Fitting simulator](https://phet.colorado.edu/sims/html/curve-fitting/latest/curve-fitting_en.html) or open in second window, Plot **Data Set 2**
   1. Click "values" in the box to the right of the graph so that you can see the coordinate of your points as you add them to the simulator
   2. Drag the orange points from the bucket on the bottom left to graph corresponding to **data set 2.** You will be adding five points.
4. Click "curve" in the box to the right of the graph to add a trendline to the data.
   1. Select "linear" from the middle right box and "best fit" from the bottom right box
   2. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
   3. Record the fit line equation in y=mx+b format from the top of the graph: **y = 0.86x + 3.1**
   4. Record the r2 value from the box to the left of the graph: **r2 = 0.82**
5. An important aspect of using mathematical equations to describe data, is that it allows use to predict new measurements.  A line of best fit (calibration curve with y=mx + b format) is then used to describe the known data set.  Once the calibration curve is constructed, the instrument can be used to measure the absorbance of solutions with unknown concentrations.  The mathematical trend is the used to predict the unknown concentration (in other words, the measured y-value is used to predict or measure the x-value). Let’s practice this concept.  Assume that a y = 2.25 value was measured. Determine the x-value from the best fit line form data set 1 and data set 2.

**Data Set 1: x = -0.66**

**Data Set 2: x = -0.99**

**Part 2 – Linear vs. Nonlinear fits**

|  |  |
| --- | --- |
| Data Set 4 | |
| **x value** | **y value** |
| **-8** | -5.5 |
| **-6** | 3 |
| **-1.5** | 2.5 |
| **2** | -2.5 |
| **7.5** | 4 |
| **8** | 8 |

|  |  |
| --- | --- |
| Data Set 3 | |
| x value | **y value** |
| -5 | -6 |
| -3.5 | 1 |
| 1.5 | 9 |
| 6 | 1.5 |
| 8 | -6 |

1. Open the [Curve Fitting simulator,](https://phet.colorado.edu/sims/html/curve-fitting/latest/curve-fitting_en.html) Plot Data Set 3
   1. Click "values" in the box to the right of the graph so that you can see the coordinate of your points as you add them to the simulator
   2. Drag the orange points from the bucket on the bottom left to graph corresponding to **data set 3.** You will be adding five points.
2. Click "curve" in the box to the right of the graph to add a trendline to the data.
   1. Select "linear" from the middle right box and "best fit" from the bottom right box
      1. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
      2. Record the fit line equation in y=mx+b format from the top of the graph: **y = 0.63x + 1.4**
      3. Record the r2 value from the box to the left of the graph: **r2 = 0.39**
   2. Change the trendline to “quadratic” and “best fit”
      1. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
      2. Record the fit line equation in y=nx2+mx+b format: **y = -0.185x2 + 1.15x + 5.9**
      3. Record the r2 value from the box to the left of the graph: **r2 = 0.72**
3. Open the [Curve Fitting simulator,](https://phet.colorado.edu/sims/html/curve-fitting/latest/curve-fitting_en.html) Plot Data Set 4
   1. Click "values" in the box to the right of the graph so that you can see the coordinate of your points as you add them to the simulator
   2. Drag the orange points from the bucket on the bottom left to graph corresponding to **data set 3.** You will be adding five points.
4. Click "curve" in the box to the right of the graph to add a trendline to the data.
   1. Select "linear" from the middle right box and "best fit" from the bottom right box
      1. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
      2. Record the fit line equation in y=mx+b format from the top of the graph: **y = 0.48x + 1.4**
      3. Record the r2 value from the box to the left of the graph: **r2 = 0.45**
   2. Change the trendline to “quadratic” and “best fit”
      1. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
      2. Record the fit line equation in y=nx2+mx+b format: **y = 0.016x2 + 0.47x + 0.8**
      3. Record the r2 value from the box to the left of the graph: **r2 = 0.46**
5. Change the trendline to “cubic” and “best fit”
   * 1. Click “residuals” in the right top box to get a sense of how close the individual points are from the trendline (Important for later question.)
     2. Record the fit line equation in y=px3+nx2+mx+b format: **y = 0.040x3 + 0.011x2 – 1.76x + 0.4**
     3. Record the r2 value from the box to the left of the graph: **r2 = 0.99**

**Mathematical Modeling**

Data Sheet

Name: Sean Poston Date: 6/4/2021

**Part I.**

Linear fit of data set 1. Line equation: **y = 0.83x + 2.8**, r2 value: **0.96**

Linear fit of data set 2. Line equation: **y = 0.86x + 3.1**, r2 value: **0.82**

Which data set is better described by the linear trendline?  Why?

**Set one has a higher r2 which means the values have less variance.**

Using the trendline from data set 1, what is the predicted x-value (for the y=2.25 value): **-0.66**

Using the trendline from data set 2, what is the predicted x-value (for the y=2.25 value): **-0.99**

Which predicted x-value should you have more confidence in?  Why?

**I would stick with the data set 1 because of the reasons above.**

**Part II.**

Linear fit of data set 3. Line equation: **y = 0.63x + 1.4**, r2 value: **0.39**

Does the linear fit seem like a good fit of data set 3? (Y/N) Why?

**The r2 value is very low, so I would say it’s a bad fit.**

* + 1. Quadratic fit of data set 3. Line equation: **y = -0.185x2 + 1.15x + 5.9,** r2 value: **0.72**

Does the quadratic fit seem like a good fit of data set 3? (Y/N) Why?

**It’s definitely better than the linear fit, but it’s still not great.**

Linear fit of data set 4. Line equation: **y = 0.48x + 1.4**, r2 value: **0.46**

Does the linear fit seem like a good fit of data set 4? (Y/N) Why?

**Pretty bad fit. High variance / low accuracy.**

Quadratic fit of data set 4. Line equation: **y = 0.016x2 + 0.47x + 0.8**, r2 value: **0.46**

Does the quadratic fit seem like a good fit of data set 4? (Y/N) Why?

**It’s ever so slightly better, but still not a great fit.**

Cubic fit of data set 4. Line equation: **y = 0.040x3 + 0.011x2 – 1.76x + 0.4**, r2 value: **0.99**

Does the cubic fit seem like a good fit of data set 4? (Y/N) Why?

**It’s a near perfect fit. The values are all pretty much on the trendline.**