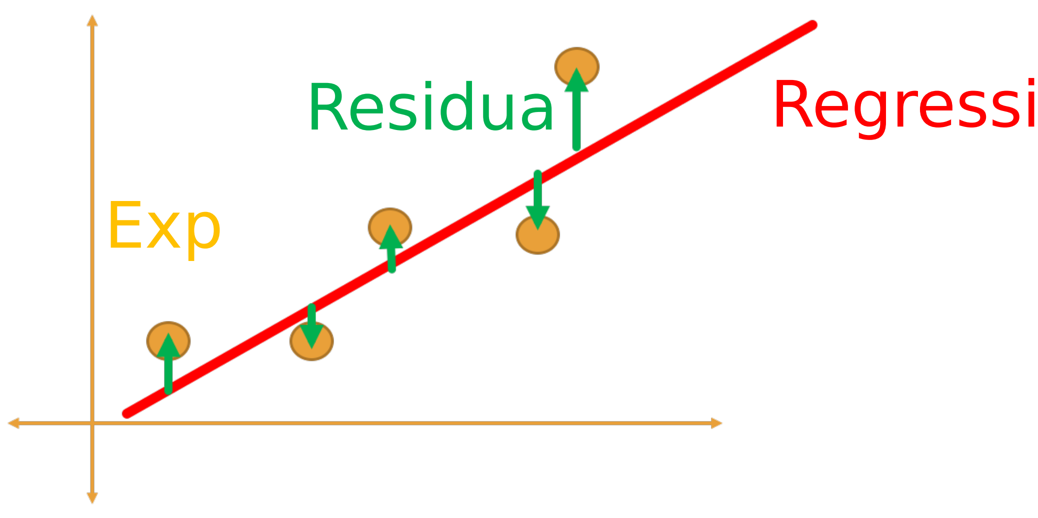
Curve Fitting (Regression Analysis)

Name: <u>Harsh Parekh</u> **Section: _______**
Date: <u>24 June, 2020</u>

There are many cases in experimental physics that you wish to determine or verify a quantity, such as the acceleration due to gravity, but it can not be directly calculated from the data that has been collected. For example, imagine the moment a ball was dropped, a stopwatch was started, and the time was recorded every 5 cm of displacement the ball descended. How could we determine the acceleration due to gravity given this information? This is exactly where regression analysis can come in handy.

What is Regression Analysis?

Imagine you have a set of data that you expect will obey a certain equation. For instance, let's say the equation is $y=mx+b$, which is a linear equation. We see *y* and *x* are variables, which correspond to the data that was collected, while *m* and *b* are unknown constants or parameters. Regression analysis is the process of finding the correct values of the parameters in the equation such that the plot of the equation and the plot of the data overlap as much as possible. We call this process 'fitting' the data. Since experiments are typically prone to error, the fitting is never exactly perfect. This process would be extremely difficult to due by hand, but many programs have tool kits that can perform regression analysis in a matter of seconds.



In the figure above, the data points are shown in orange, the regression line or line of best fit is shown in red, and the distance from the line of best fit to the data point otherwise known as a residual is shown in green. The goal of regression analysis is to reduce the average size of the residual to a minimum and then the equation describing the line of best fit will carry the unknown parameters.

Example

An example data set for the hypothetical experiment explained above is shown below.

Time (s)	Displacement (cm)
0.97	-5
1.35	-10
1.82	-15
2.05	-20
2.29	-25

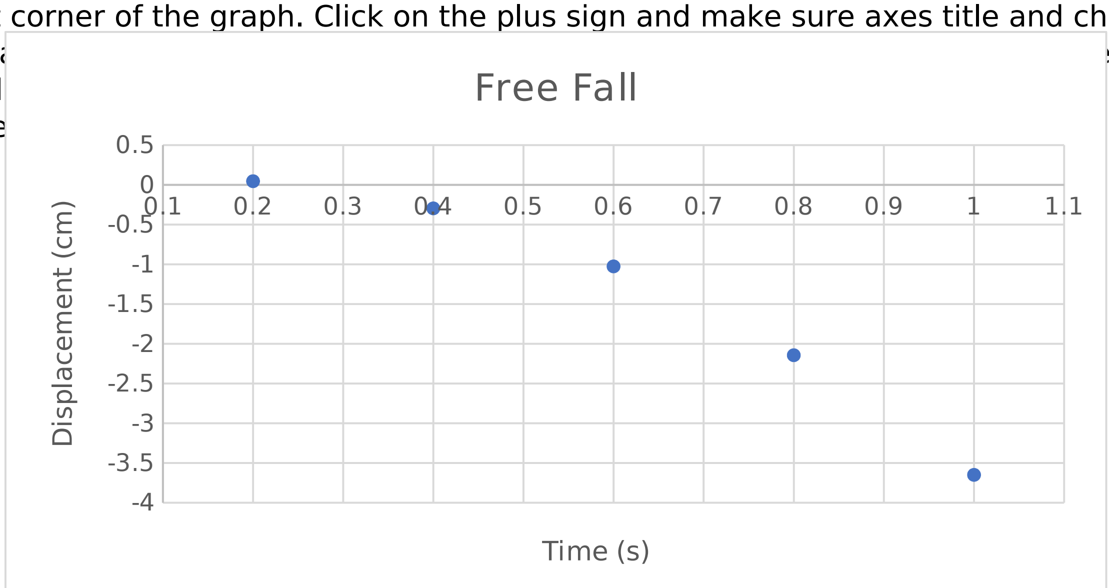
Notice that the displacement is negative because the object is falling. Also notice that we put time in the right column and displacement in the left column. This is because we want time on the x-axis and displacement on the y-axis. We know from lecture that an object falling due to gravity must obey a quadratic relationship

$$\Delta y = \frac{1}{2} g t^2 + v_o t$$

In this case, since the ball was released, the initial velocity is zero. Therefore, the equation becomes

$$\Delta y = \frac{1}{2} g t^2$$

If we plot the displacement vs. time of the data above, it should look quadratic. Let's check that this is true. To plot the data, copy paste the displacement and time into an excel spread sheet and highlight that data. Now at the top of the excel window, go to insert, and in the charts section, click on the *scatter chart*. You should see the graph pop up. If you click on the graph, a plus sign should appear at the top right corner of the graph. Click on the plus sign and make sure axes title and chart title a… e
label…
corre…



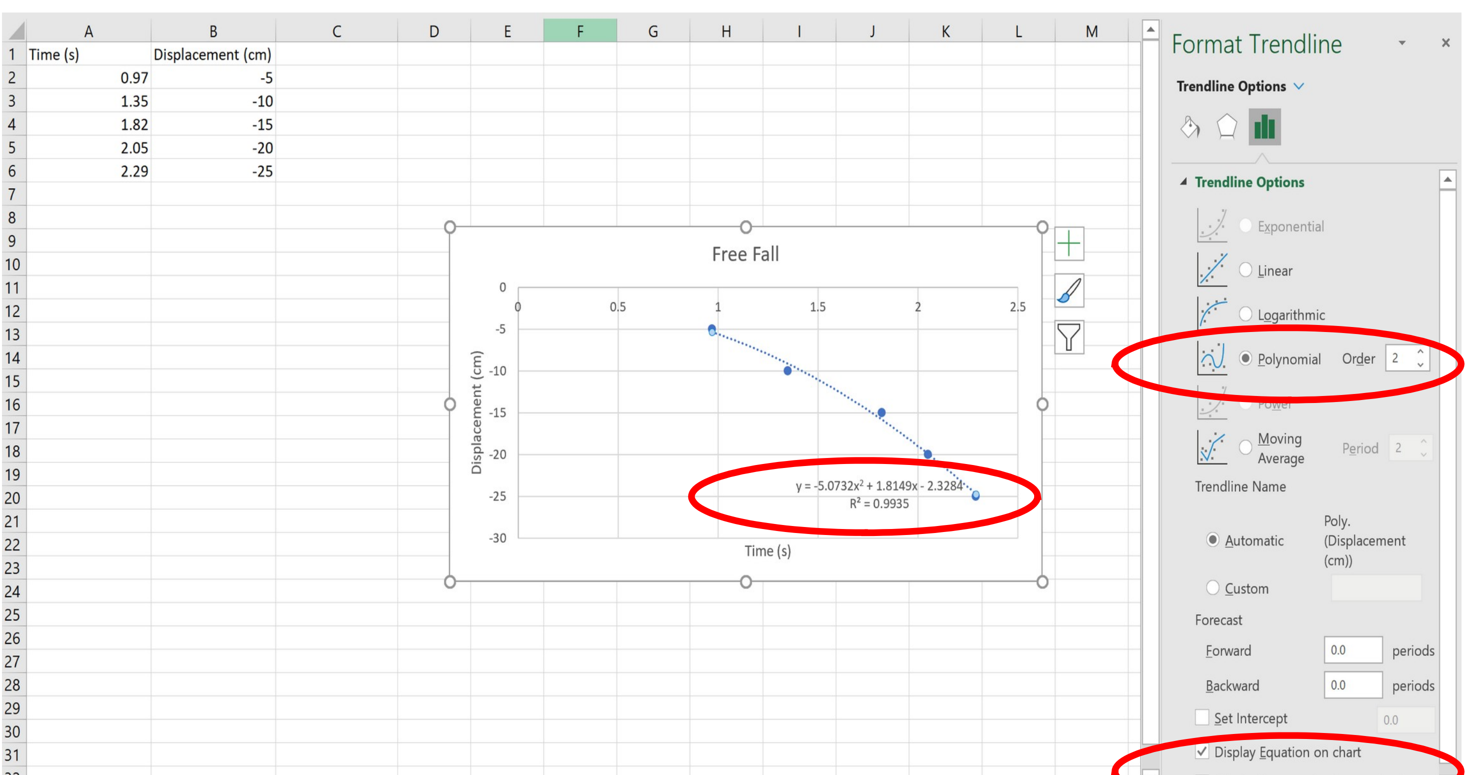
We see that this graph does seem to be quadratic, which is a good sign. However, we still do not have a way of determining the value of *g*. Luckily, Excel has a regression analysis tool kit build in. Since we expect this data to obey,

$$\Delta y=\frac{1}{2}gt^2+v_o t=\text{¿}\, y_f=\frac{1}{2}gt^2+v_o t+y_o$$

we will fit our data to a quadratic function $f(x)=Ax^2+Bx+C$. After completing the regression analysis, we will match our function obtained from the regression to our expected function. Where we would expect $A\approx\frac{1}{2}g=\frac{1}{2}9.81\frac{m}{s^2}, B\approx v_o\approx 0\left(\textit{because we realsed the ball}\right), \wedge C\approx y_o$. The parameter *C* is not important to us since we are only concerned with the displacement.

Using Excel for Regression Analysis

In Excel, click on the graph that we just plotted and then click on the plus sign on the top right corner. The list that pops up should say *Trendlines* at the bottom. Click on *Trendlines* and another menu should appear. At the bottom of this menu, you should see *More Options.* Click on *More Options* and you should see this window pop up.



Make sure *Polynomial* and *Degree 2* are selected (Equivalent to Quadratic). Secondly, make sure the '*Display Equation on Chart'* and *'Display R-squared value on Chart'* is selected. You should see both the equation and R^2 value appear on the chart along with the line of best fit. The R-squared value is a measure of how well the line of best fit actually fits the data. It ranges from 0 to 1 and the closer to 1 the value gets, the better the fit.

We see our equation that describes the line of best fit is given by **y = -5.0732x² + 1.8149x - 2.3284** where we see $A=-5.0732$. We can compare this to our expected value of $\frac{1}{2}g=-4.905$. To calculate the percent difference we use the equation

$$\frac{|expected-experimental|}{expected}\times 100 = Percent\ Difference$$

In this case we get about 3.4% difference, which shows our data fairly accurate.

Task

Below are three new data sets with the position versus time. For each data set complete the following:

1. Plot the Time vs Position graph
2. Fit the graph to a 2nd degree polynomial
3. Determine the equation of best fit
4. Determine the value of g from each data set
5. Calculate the percent difference
6. Was the object initially moving upward or downward

Data Set 1:

Time (s)	Position (cm)
0.2	0.046
0.4	-0.296
0.6	-1.026
0.8	-2.144
1	-3.65
1.2	-5.544

Data Set 2:

Time (s)	Position (cm)
0.1	-0.38955
0.2	-0.8782
0.3	-1.46595
0.4	-2.1528
0.5	-2.93875
0.6	-3.8238

Data Set 3:

Time (s)	Position (cm)
0.5	1.52875
1	0.615
1.5	-2.74125
2	-8.54

2.5	-16.78125
3	-27.465