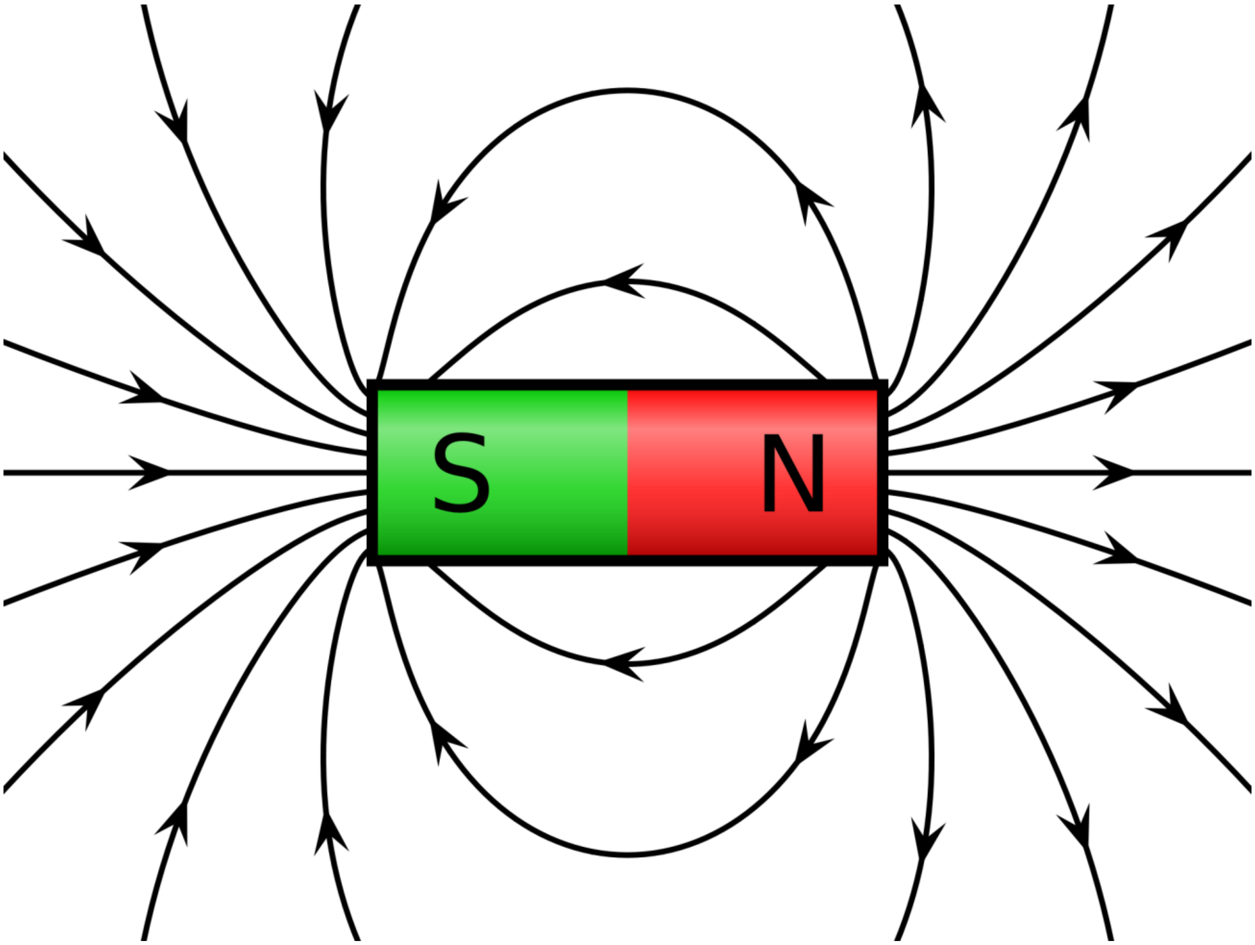
# Regression Project

**Part I. Magnetism**




Magnetism is "a physical phenomenon produced by the motion of electric charge, resulting in attractive and repulsive forces between objects." (Google)

Magnetism applies force to objects through a vector field: each point in space is associated with a vector that measures the force (direction and magnitude) at that point. We often describe the field with field lines as shown above, showing the direction of the field, but losing information about magnitude. We will not be working directly with the lines of force (this is a topic better suited to multivariable calculus), but instead we will be using magnetic properties and the strength of the field for various magnets to collect data for a regression application.

**Materials Needed for the Project:**

- Magnets, of varying strengths (the magnets we are using are bar magnets with a pole on each end as shown in the top illustration)

- Uniform metal balls (like BBs) or small nails – must be large enough to be countable, but small enough that the smallest magnet you have can pick one up. They should **not** be pre-magnetized.
- Technology (like a calculator or MatLab) for analysis
- Colored pencils for graphing

We want to use our magnets to see how many BBs we can pick up in a strand (as shown on the right) with each magnet. The strength of the force on each magnet is labeled. You will record the strength of the magnet you are using and the number of BBs you are able to pick up. The force is propagated through the metal of the BBs, but there will be an upper limit on the weight of BBs can be picked up in this manner.

A word of caution about the really strong magnets:

- If they stick together, they are very hard to pull apart.
- You will have to take more care to get the BBs to hang in a strand.
- Keep the strong magnets away from mass of BBs as they will also be hard to pull off. Start slowly and add one BB at a time.

A. What is the independent variable in this experiment? ___________________________________

B. What is the dependent variable in this experiment? _____________________________________

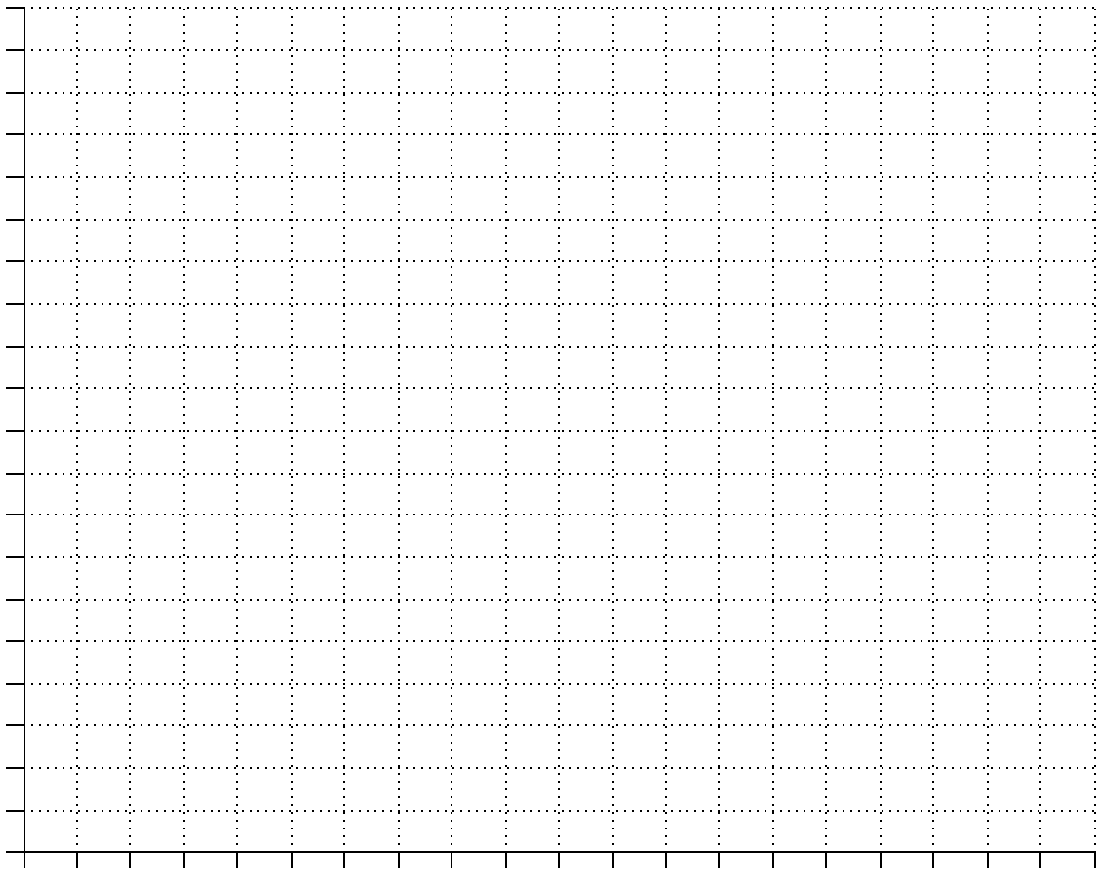
C. What units are you using in each case? ______________________ _________________________

You will need to collect at least 10 data points for different strength magnets. Record your data in the table below.

| **Independent Variable ($x$)** | **Dependent Variable ($y$)** |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

D. Look at the data you've collected. What is the general trend in the data? Does it appear to be increasing or decreasing? Does it appear approximately linear, or does some other pattern appear to jump out?

E. Now that you've collected the data, we need to analyze it. We will do this by graphing the data in a scatterplot. Label each axis appropriately to allow all the magnetic forces to fit on the $x$-axis, and the number of BBs on the horizontal axis. (A blank graph is on the top of the next page.)



F. Does the graph of the data support or contradict your observations from part D? What pattern, if any, does there appear to be?

G. We will start with a linear regression model of the form $\beta_0 + \beta_1 x = y$. To calculate the values of $\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, we will solve the equation $(A^T A)^{-1} A^T \vec{y} = \vec{\beta}$. To construct $A$, we substitute our $x$-values (independent variable) into the equation, and then enter the coefficient values into one row of a matrix. In the same order, we construct $\vec{y}$ using the $y$-values (dependent variable) in the same order we used the $x$-values. If we have taken ten measurements, our matrix $A$ should be $10 \times 2$, and our vector $\vec{y}$ should be $10 \times 1$. Calculate our estimate for $\vec{\beta}$. Record the result below.

H. What is the equation of the line obtained?

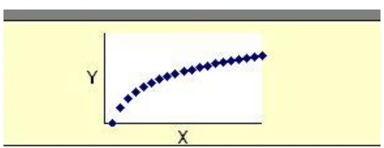
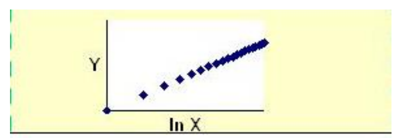
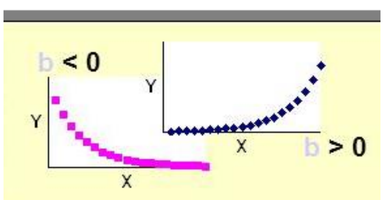
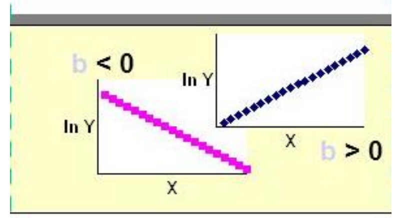
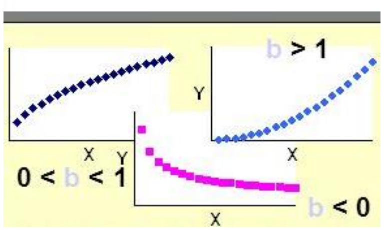
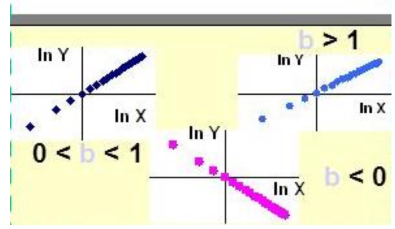
I. Add the line to the graph of the data. Describe how well the line fits to the data.

J. Consider the end-run behavior of your equation compared to the data. If the trend we see in the data was to continue, and the line was to continue as is, does it appear like the line would stay close to future data, or does it appear to be doing less well for large (or small) values? Does the line predict behavior that is logically impossible?

K. Do you think the point $(0,0)$ makes sense in this data? If our "magnet" had no magnetic force, is it reasonable to think it would be able to pick up zero BBs? If so, would be it reasonable to add this free point to the graph and our data set? If not, why not?

L. Add (0,0) to the dataset, and recalculate $\vec{\beta}$. Add the value to the line in a different color and compare your results. Did it improve the fit? What is the equation you came up with in this case?

M. In order to calculate non-polynomial graphs which are not linear, we will have to do a transformation of the variables. Some graphs become linear after transformation. These are illustrated in the chart below.

<table>
<tr><th>Before Transformation</th><th>Model</th><th>After Transformation</th></tr>
<tr><td>
</td><td>Logarithm model<br>$y = a + b \ln x$<br><br>Transform $x$</td><td>
</td></tr>
<tr><td>
</td><td>Exponential Model<br>$y = ae^{bx}$ $or$ $y = ab^x$<br>becomes<br>$\ln y = \ln a + bx$ or<br>$\ln y = \ln a + (\ln b)x$<br><br>Transform $y$</td><td>
</td></tr>
<tr><td>
</td><td>Power Model<br>$y = ax^b$<br>becomes<br>$\ln y = \ln a + b \ln x$<br><br>Transform both $x$ and $y$</td><td>
</td></tr>
<tr><td colspan="3">We can also consider reciprocal models $y = a + \frac{b}{x}$ as intrinsically linear, after transforming $x$.</td></tr>
</table>

N. We are going to try a power model to fit the data. To do this, find the natural log of all the values in our model, and record the data in the table below. [Note: Since we are using logs, we can't use the (0,0) point used in the second linear model.]

| Independent Variable ($ln\ x$) | Dependent Variable ($ln\ y$) |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

O. Construct your matrices for $A$ and $\vec{y}$ just as you did before. These three models are called intrinsically linear because they become linear after the transformation. Find $(A^T A)^{-1} A^T \vec{y} = \vec{\beta}$. Using the coefficients in the graphics, $\vec{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}$. Record your solution here.

P. What is the equation you came up with? Convert it back to a power model on the original variables. Add the equation (in a different color) to the original graph of your data.

Q. Analyze your equation. How well does it fit the data? Are there any obvious concerns with the equation? For instance, what does the long-run behavior suggest will happen as the force increases, and does that seem reasonable? Are any of the coefficients close to zero or one? Does this suggest one of the coefficients is an artifact or the model is really linear?

R. Choose one more equation to model the data. It can be one of the intrinsically linear graphs, or a polynomial model if that seems appropriate. Calculate $A, \vec{y}, \vec{\beta}$ for the equation. Record which model you are using, the values of $\vec{\beta}$, and the final form of the equation.

S. Add the model to the graph (in a different color). Analyze the graph compared to the data as before. What does the long-run behavior of the model suggest will happen as the force increases? Does this make sense? Are any of the coefficients in the model near zero (or one in the log/exponential models)?

T. Based on the models you've created, which model appears to fit the data the best? Why? Write the model you've selected here.

U. Use the model you've selected to predict how many BBs you can lift with a magnetic force value you did not use to create your model. State the force you chose, and the prediction from the model.

V. Test your prediction. How close did you get?

W. If we have a magnet whose force we don't know, how can we solve the equation we found to measure the force of the magnet using BBs?

X. Using the equation you found in W, measure the number of BBs the unlabeled magnet can pick up, and estimate the force in pounds.

**Part II. Your Data**

Follow the model of what we did in Part I to model the data you found online. You need to include the following:

- Name the independent and dependent variables
- Graph the data
- Select at least three types of graphs to use to model the data and your reasons for selecting each
- Solutions to the coefficients for each graph type and the final equations.
- A graph of the data compared to the regression models (this can be one graph, or three separate graphs)
- Determine which of the models is best, and explain your reasoning in terms of the meaningfulness of the coefficients, end-run behavior, etc.
- Use your model to make a prediction. Describe the kind of measurement that would need to be taken to verify your results. (You don't need to make the observation, just discuss how it could be done.)

Regression can be used for a variety of nonlinear models as long as the coefficients of the model is linear. One type of model we have not use in this project so far is a sinusoidal model. Models of the form $y = a\sin(\omega x) + b\cos(\omega x) + c$ is linear in $a, b, c$, but nonlinear in $\omega$. If you are using a sine model, you will need to first find the frequency of wave function (by measuring the distance between peaks), and use that figure when constructing $A$. We can build up extremely complex models this way.

We can also model the relationship between several input variables and a single output variable. Models of this type can be linear, generally polynomial or intrinsically linear. If you are modeling multiple variables, it may help to model each input variable separately, and then put the graph together, or begin with a complete linear model such as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$, and then add on quadratic factors, which can include interaction terms like $\beta_{ij} x_i x_j$ or $\beta_k x_i^2$. Determining if the interaction terms make sense can be observed from graphs. Remember to exclude terms whose coefficients are near zero.

You can construct these models on your calculator, or use MatLab. If you use your calculator, you will have to draw your graphs by hand or use a TI-emulator and take screenshots, as well as include any by-hand work. If you are using MatLab, submit your code along with your models and graphs.