Regression

Reading While Skimming the Lines


EVERY GROUP PROJECT
DOES 99% OF THE WORK
HAS NO IDEA WHAT'S GOING ON THE WHOLE TIME
SAYS HE'S GOING TO HELP BUT HE'S NOT
DISAPPEAR AT THE VERY BEGINNING AND DOESN'T SHOW UP AGAIN TIL THE VERY END
IN SCHOOL YOU HAVE EVER DONE

6.04
Nonlinear Regression

Nonlinear Regression Without Transformation of Data (Untransformed Data)

Who is this artist?

Saweetie

Doja Cat

Diamonté Quiava Valentin Harper

Icy Girl

Powered by Poll Everywhere

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Nonlinear Regression

Popular nonlinear regression models. Given n data pairs $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$

1. Exponential model: $y = ae^{bx}$
2. Power model: $y = ax^b$
3. Saturation growth model: $y = \frac{ax}{b+x}$
4. Polynomial model: $y = a_0 + a_1 x + a_2 x^2 + \cdots\cdots + a_m x^m$
5. Logistic model: $y = \frac{L}{1+e^{(b_0+b_1 x)}}$

Nonlinear Regression

Given (x_1, y_1), (x_2, y_2), . . . (x_n, y_n), best fit $y = f(x)$ to the data.

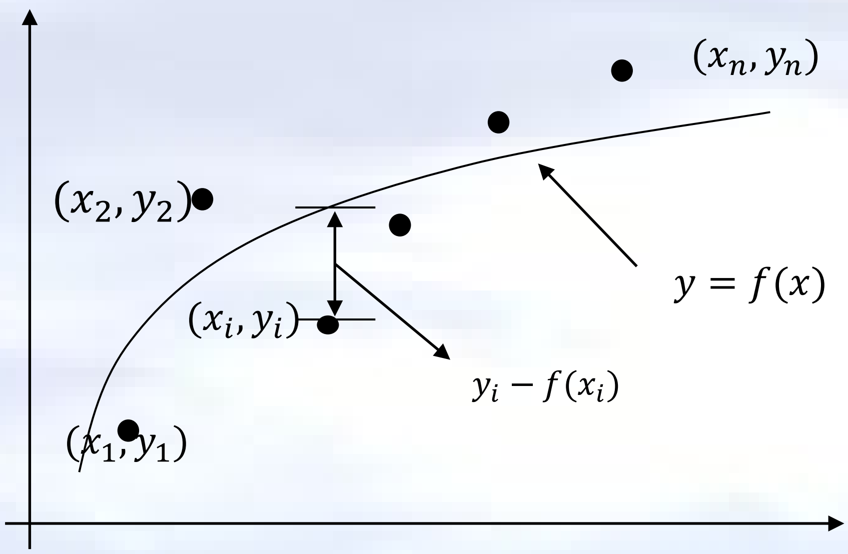


Figure. Nonlinear regression model for discrete y vs. x data

Exponential Model

Given $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$, best fit $y = ae^{bx}$ to the data. The variables a and b are the constants of the exponential model.

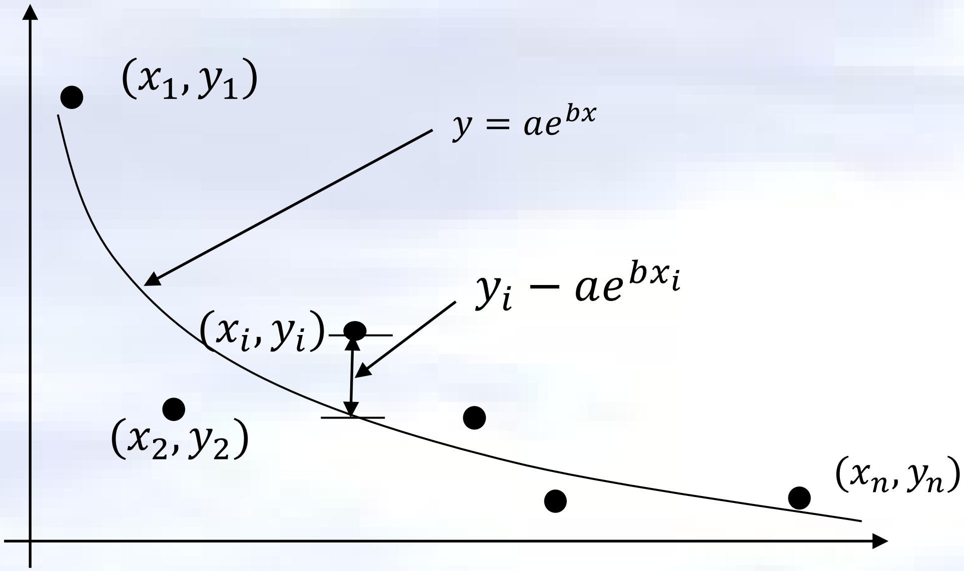


Figure. Exponential model of nonlinear regression for y vs. x data

Finding Constants of Exponential Model

$$S_r = \sum_{i=1}^{n} \left(y_i - a e^{b x_i} \right)^2$$

$$\frac{\partial S_r}{\partial a} = 0$$

$$\frac{\partial S_r}{\partial b} = 0$$

Follow this by
1) second derivative test to show solution corresponds to local minimum
2) showing above gives only one acceptable real solution
3) Recognizing that S_r is a differentiable function of a and b.

Finding Constants of Exponential Model

$$S_r = \sum_{i=1}^{n} \left(y_i - ae^{bx_i}\right)^2$$

Differentiate with respect to a

$$\frac{\partial S_r}{\partial a} = \sum_{i=1}^{n} 2\left(y_i - ae^{bx_i}\right)\left(-e^{bx_i}\right)$$

$$= \sum_{i=1}^{n} -2y_i e^{bx_i} + 2ae^{2bx_i}$$

$$= \sum_{i=1}^{n} -2y_i e^{bx_i} + \sum_{i=1}^{n} 2ae^{2bx_i}$$

$$\frac{\partial S_r}{\partial a} = 0$$

$$\sum_{i=1}^{n} -2y_i e^{bx_i} + \sum_{i=1}^{n} 2ae^{2bx_i} = 0$$ **Equation 1**

Finding Constants of Exponential Model

$$S_r = \sum_{i=1}^{n} \left(y_i - ae^{bx_i}\right)^2$$

Differentiate with respect to b

$$\frac{\partial S_r}{\partial b} = \sum_{i=1}^{n} 2\left(y_i - ae^{bx_i}\right)\left(-ax_i e^{bx_i}\right)$$

$$= \sum_{i=1}^{n} -2ay_i x_i e^{bx_i} + 2a^2 x_i e^{2bx_i}$$

$$= \sum_{i=1}^{n} -2ay_i x_i e^{bx_i} + \sum_{i=1}^{n} 2a^2 x_i e^{2bx_i}$$

$$\frac{\partial S_r}{\partial b} = 0$$

$$\sum_{i=1}^{n} -2ay_i x_i e^{bx_i} + \sum_{i=1}^{n} 2a^2 x_i e^{2bx_i} = 0$$ **Equation 2**

Finding Constants of Exponential Model

Equation 1 $$\sum_{i=1}^{n} -2y_i e^{bx_i} + \sum_{i=1}^{n} 2ae^{2bx_i} = 0$$

$$-2\sum_{i=1}^{n} y_i e^{bx_i} + 2a\sum_{i=1}^{n} e^{2bx_i} = 0$$

$$2a\sum_{i=1}^{n} e^{2bx_i} = 2\sum_{i=1}^{n} y_i e^{bx_i}$$

$$a = \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}}$$

Finding Constants of Exponential Model

$$a = \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}}$$

Equation 2

$$\sum_{i=1}^{n} -2ay_i x_i e^{bx_i} + \sum_{i=1}^{n} 2a^2 x_i e^{2bx_i} = 0$$

$$-2a\sum_{i=1}^{n} y_i x_i e^{bx_i} + 2a^2 \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

$$-\sum_{i=1}^{n} y_i x_i e^{bx_i} + a\sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

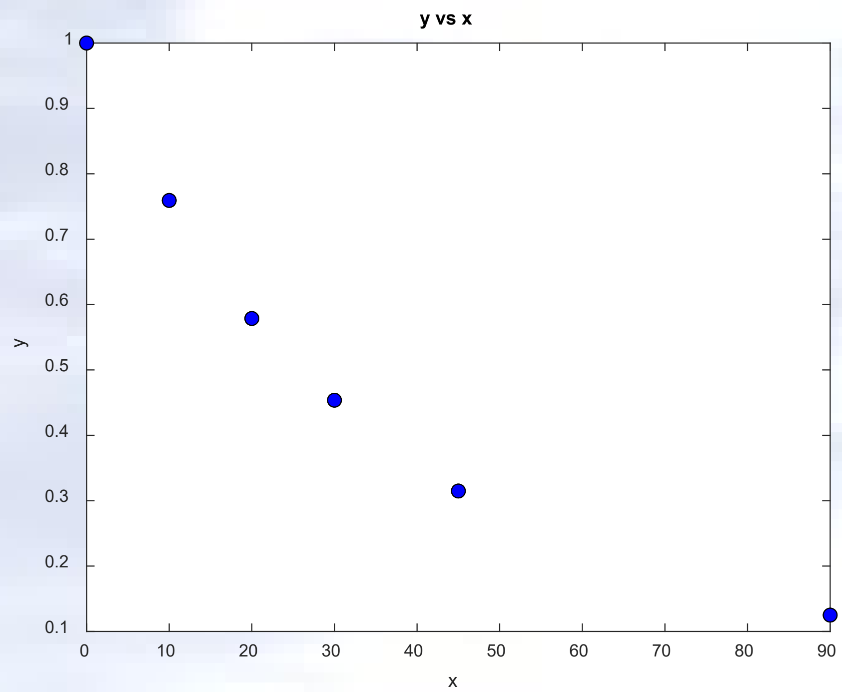
$$-\sum_{i=1}^{n} y_i\, x_i e^{bx_i} + \frac{\sum_{i=1}^{n} y_i\, e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

Example - Exponential Model

x	0	10	20	30	45	90
y	1.0	0.76	0.58	0.45	0.31	0.12

Use the regression model $y = ae^{bx}$. Estimate the regression constants a and b without transforming the data.

Plot of data



x	0	10	20	30	45	90
y	1.0	0.76	0.58	0.45	0.31	0.12

Constants of the Model

$$y = ae^{bx}$$

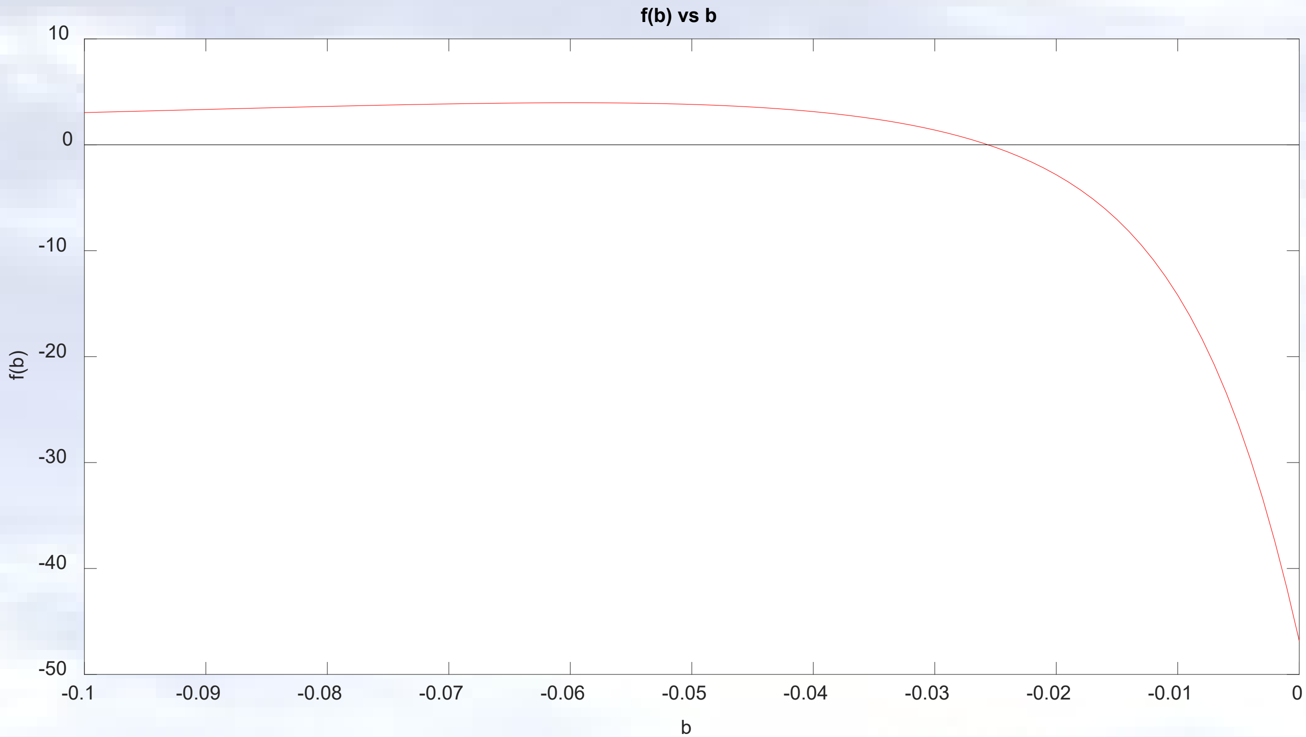
The value of b is found by solving the nonlinear equation

$$f(b) = -\sum_{i=1}^{n} y_i x_i e^{bx_i} + \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

$$a = \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}}$$

Setting up the Equation in MATLAB

$$f(b) = -\sum_{i=1}^{n} y_i x_i e^{bx_i} + \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$



x	0	10	20	30	45	90
y	1.0	0.76	0.58	0.45	0.31	0.12

Setting up the Equation in MATLAB

$$f(b) = -\sum_{i=1}^{n} y_i x_i e^{bx_i} + \frac{\sum_{i=1}^{n} y_i e^{bx_i}}{\sum_{i=1}^{n} e^{2bx_i}} \sum_{i=1}^{n} x_i e^{2bx_i} = 0$$

```
x=[0     10     20     30     45     90]
y=[1.0   0.76   0.58   0.45   0.31   0.12]
syms b real
sum1=sum(y.*x.*exp(b*x));
sum2=sum(y.*exp(b*x));
sum3=sum(exp(2*b*x));
sum4=sum(x.*exp(2*b*x));
f=-sum1+sum2/sum3*sum4;
b_soln=vpasolve(f,b);
```

$$b = -0.02561$$

Calculating the Other Constant

$$b = -0.02561$$

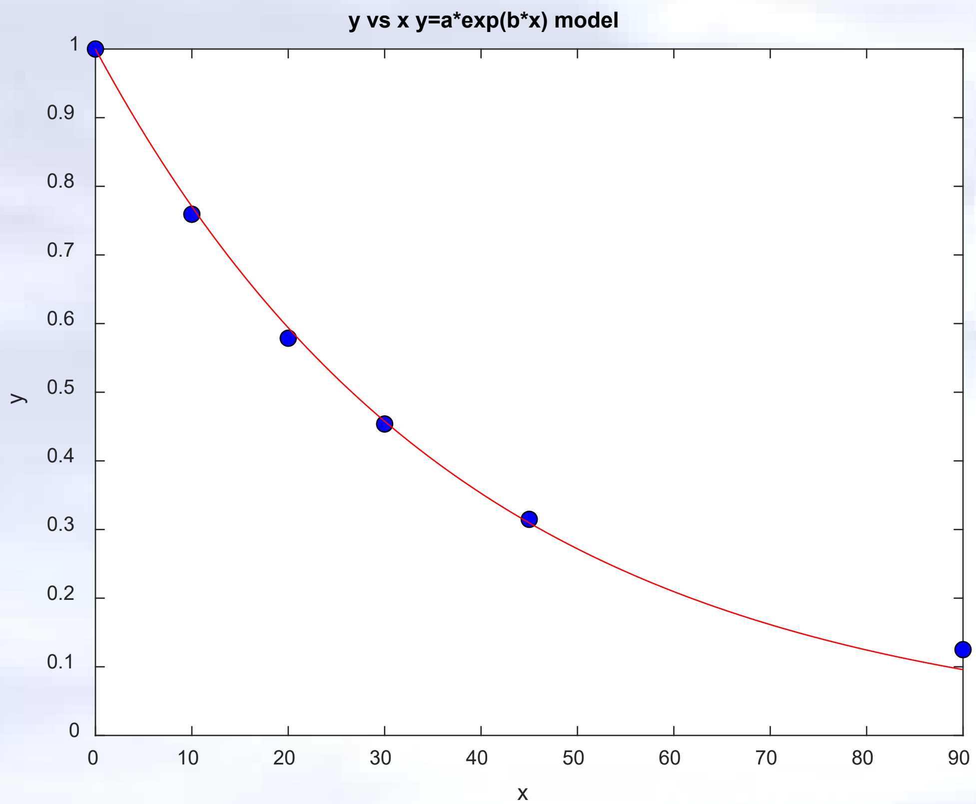
The value of a can now be calculated

$$a = \frac{\sum_{i=1}^{6} y_i e^{bx_i}}{\sum_{i=1}^{6} e^{2bx_i}} = 0.9885$$

The exponential regression model is

$$y = 0.9885\, e^{-0.02561x}$$

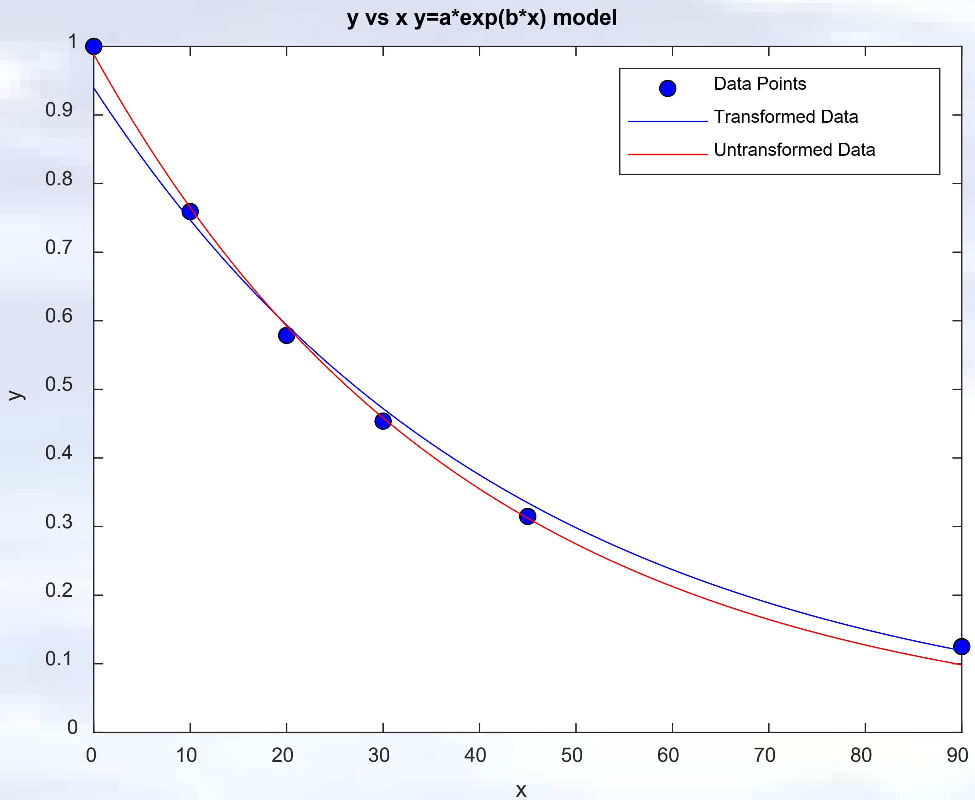
Plot of data and regression curve



$$y = 0.9885e^{-0.02561x}$$

Transformed vs Untransformed Data

x	0	10	20	30	45	90
y	1.0	0.76	0.58	0.45	0.31	0.12



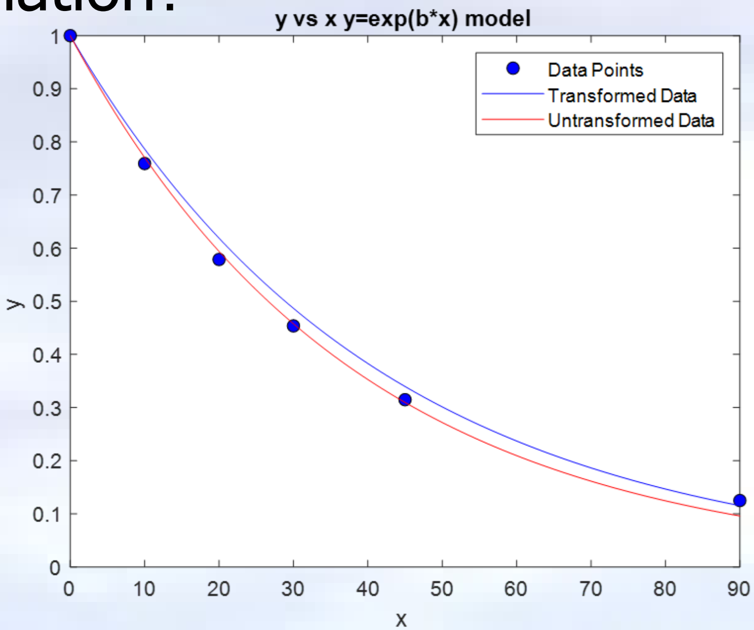
Untransformed data model $y = 0.9885\, e^{-0.02561x}$

Transformed data model $y = 0.9395e^{-0.02294x}$

Classwork/Homework: Transformed vs Untransformed Data

x	0	10	20	30	45	90
y	1.0	0.76	0.58	0.45	0.31	0.12

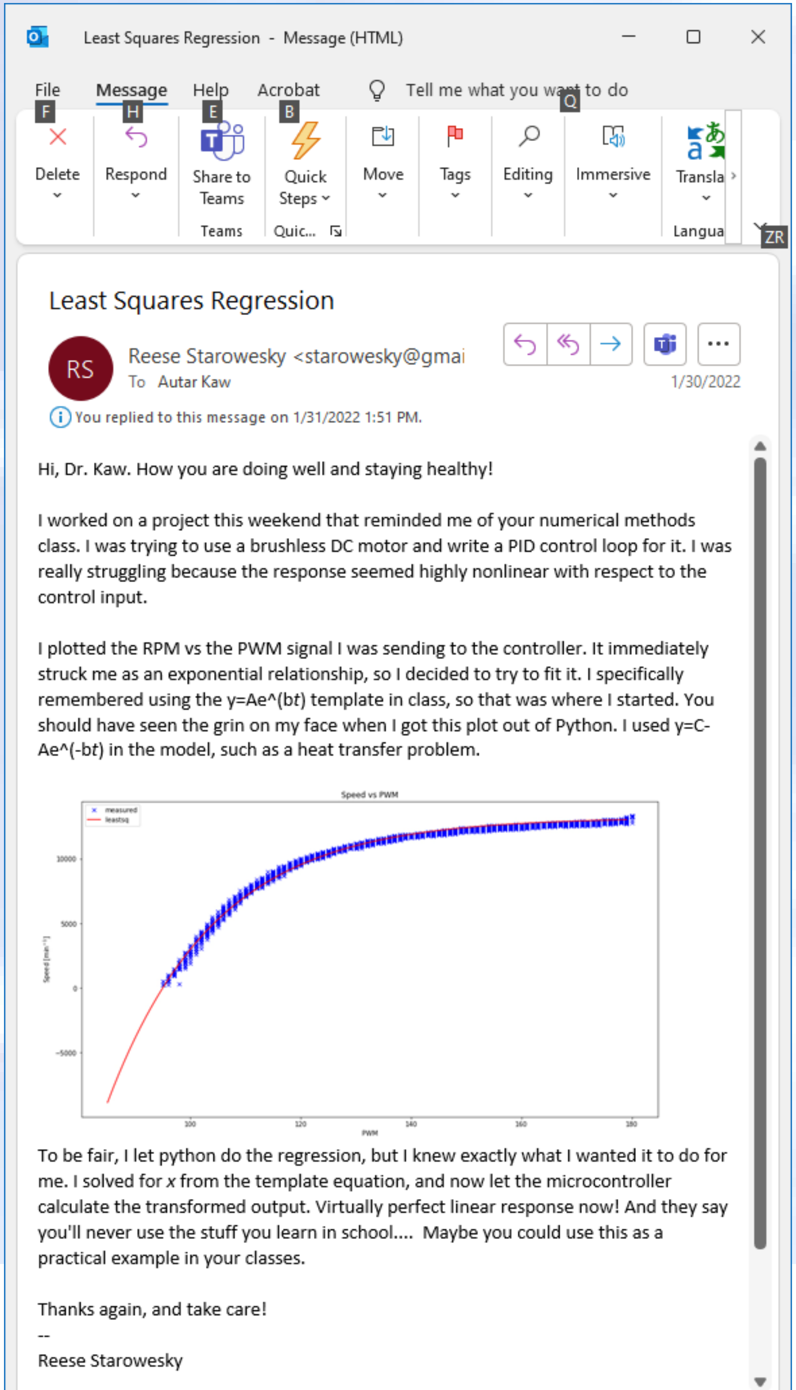
What if the model was $y = e^{bx}$? What is the solution with and without transformation?



Untransformed data model $y = e^{-0.02605x}$

Transformed data model $y = e^{-0.02400x}$

Start the presentation to see live content. Still no live content? Install the app or get help at **PollEv.com/app**



Least Squares Regression - Message (HTML)

File Message Help Acrobat Tell me what you want to do

Delete Respond Share to Teams Quick Steps Move Tags Editing Immersive Translate

Teams Quic... Langua

Least Squares Regression

Reese Starowesky <starowesky@gmai

To Autar Kaw

1/30/2022

You replied to this message on 1/31/2022 1:51 PM.

Hi, Dr. Kaw. How you are doing well and staying healthy!

I worked on a project this weekend that reminded me of your numerical methods class. I was trying to use a brushless DC motor and write a PID control loop for it. I was really struggling because the response seemed highly nonlinear with respect to the control input.

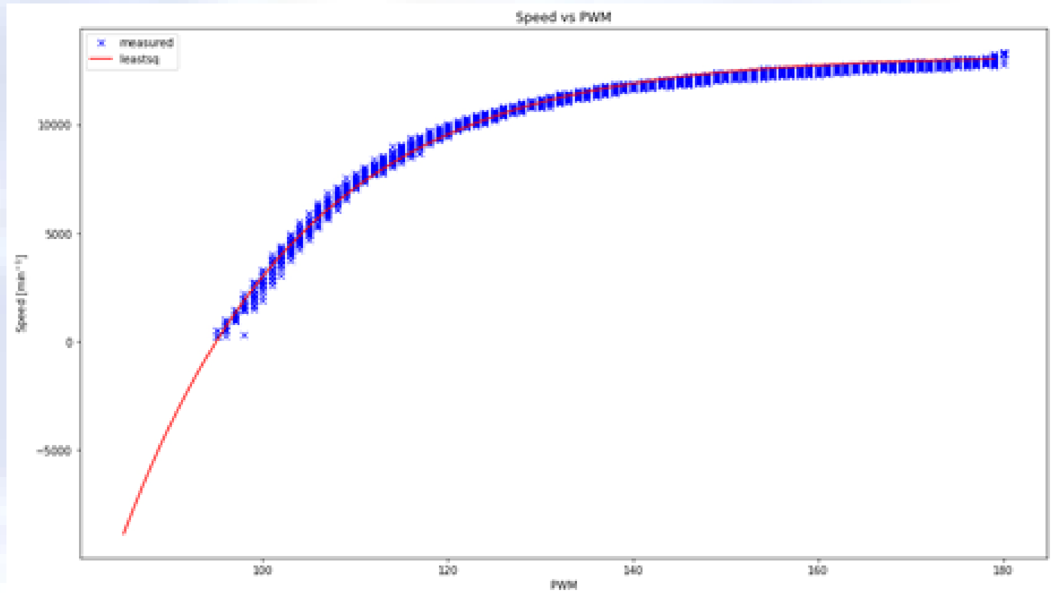
I plotted the RPM vs the PWM signal I was sending to the controller. It immediately struck me as an exponential relationship, so I decided to try to fit it. I specifically remembered using the y=Ae^(b*t*) template in class, so that was where I started. You should have seen the grin on my face when I got this plot out of Python. I used y=C-Ae^(-b*t*) in the model, such as a heat transfer problem.

To be fair, I let python do the regression, but I knew exactly what I wanted it to do for me. I solved for *x* from the template equation, and now let the microcontroller calculate the transformed output. Virtually perfect linear response now! And they say you'll never use the stuff you learn in school.... Maybe you could use this as a practical example in your classes.

Thanks again, and take care!

--

Reese Starowesky



Without transforming **the data to find the constants of the regression model** $y = ae^{bx}$ **to best fit**

$$(x_1, y_1), (x_2, y_2), \ldots\ldots\ldots\ldots\ldots\ldots(x_n, y_n),$$

the sum of the square of the residuals that is minimized is

$$\sum_{i=1}^{n}(y_i - ae^{bx_i})^2$$

$$\sum_{i=1}^{n}(\ln(y_i) - \ln a - bx_i)^2$$

$$\sum_{i=1}^{n}(y_i - \ln a - bx_i)^2$$

$$\sum_{i=1}^{n}(\ln(y_i) - \ln a - b\ln(x_i))^2$$

Powered by Poll Everywhere

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

When *transforming* the data to find the constants of the regression model $y = ae^{bx}$ to best fit $(x_1, y_1), (x_2, y_2), \ldots\ldots\ldots\ldots\ldots\ldots (x_n, y_n)$, the sum of the square of the residuals that is minimized is

$$\sum_{i=1}^{n}(y_i - ae^{bx_i})^2$$

$$\sum_{i=1}^{n}(\ln(y_i) - \ln a - bx_i)^2$$

$$\sum_{i=1}^{n}(y_i - \ln a - bx_i)^2$$

Powered by Poll Everywhere

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

When transforming the data for stress-strain curve $\sigma = k_1 \varepsilon e^{-k_2 \varepsilon}$ for concrete in compression, where σ is the stress and ε is the strain, the model is rewritten as

$$\ln \sigma = \ln k_1 + \ln \varepsilon - k_2 \varepsilon$$

$$\ln \frac{\sigma}{\varepsilon} = \ln k_1 - k_2 \varepsilon$$

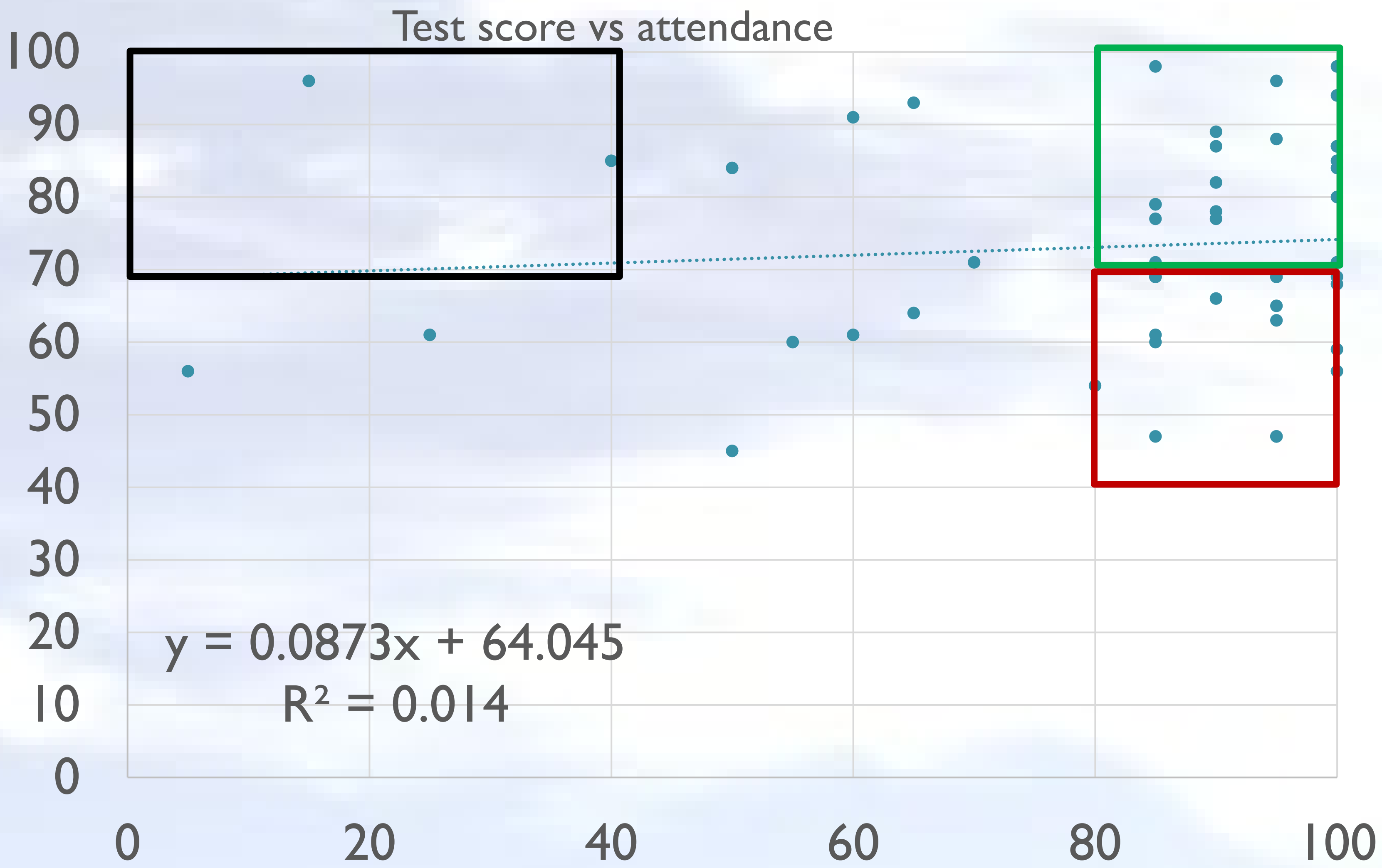
$$\ln \frac{\sigma}{\varepsilon} = \ln k_1 + k_2 \varepsilon$$

$$\ln \sigma = \ln(k_1 \varepsilon) - k_2 \varepsilon$$

Powered by Poll Everywhere

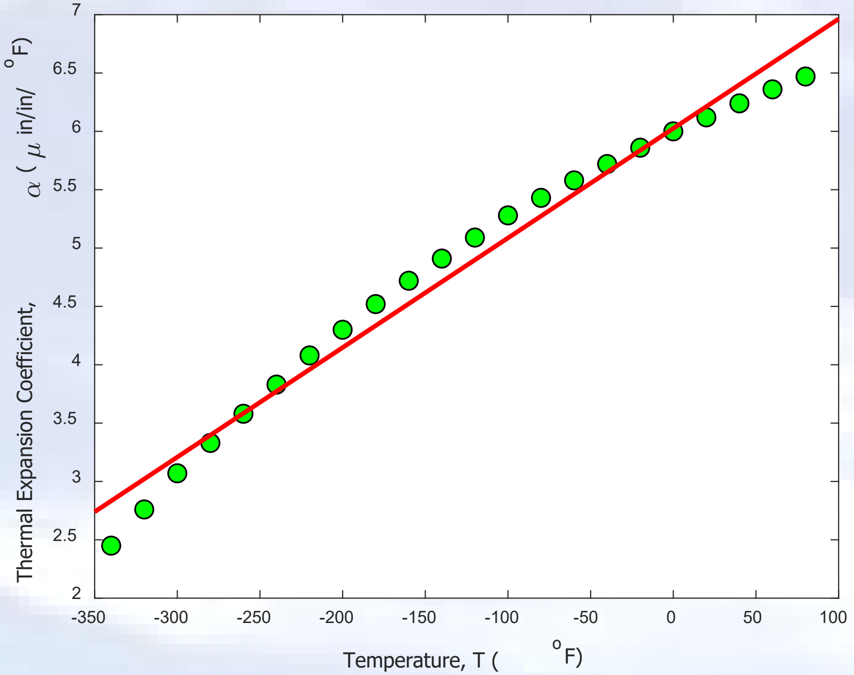
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Test 1 score vs attendance

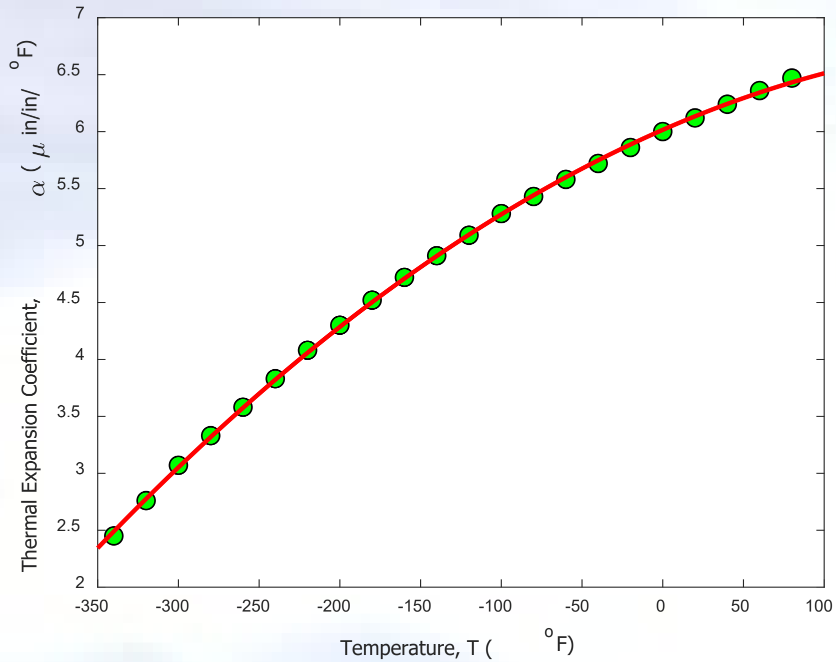


What polynomial model to choose if one needs to be chosen?

Which model to choose?

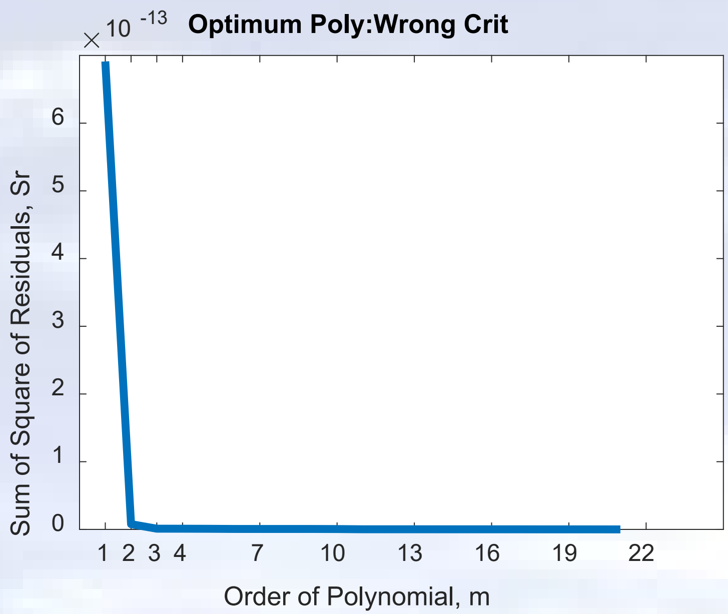


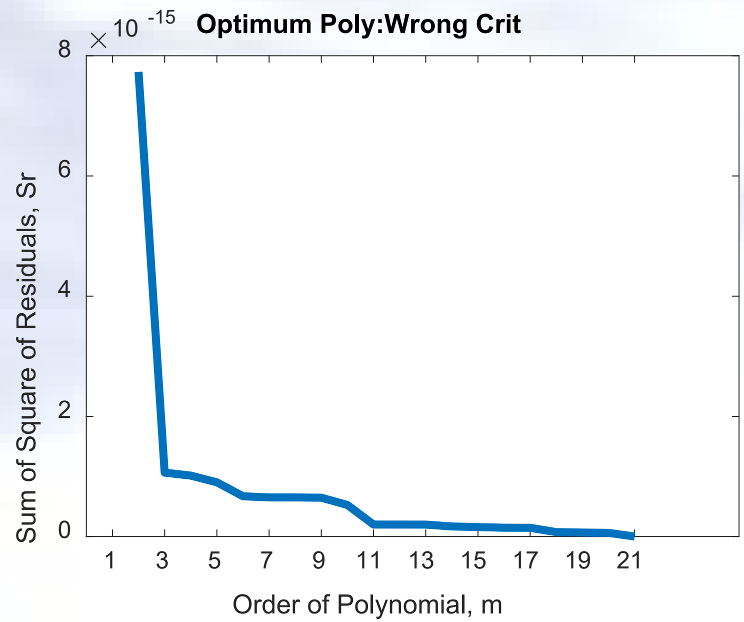
$$\alpha = 0.009387T + 6.025$$



$$\alpha = -0.00001228T^2 + 0.006196T + 6.015$$

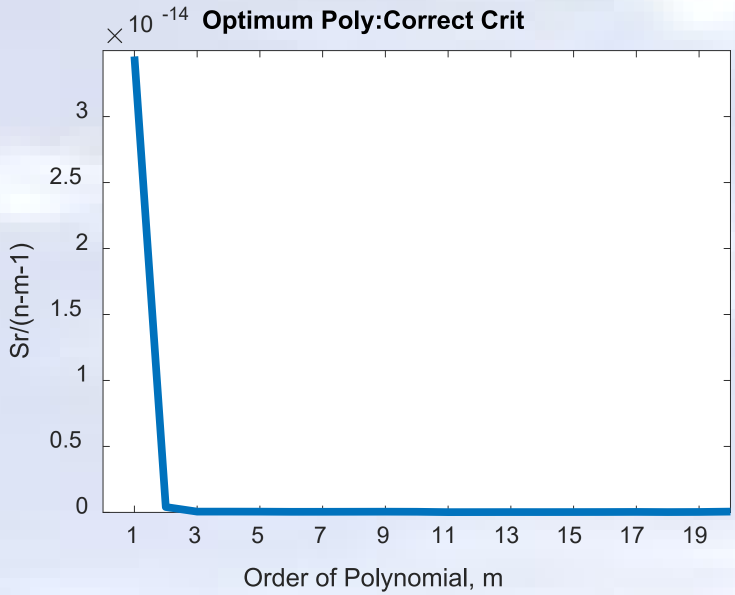
Optimum Polynomial: Wrong Criterion

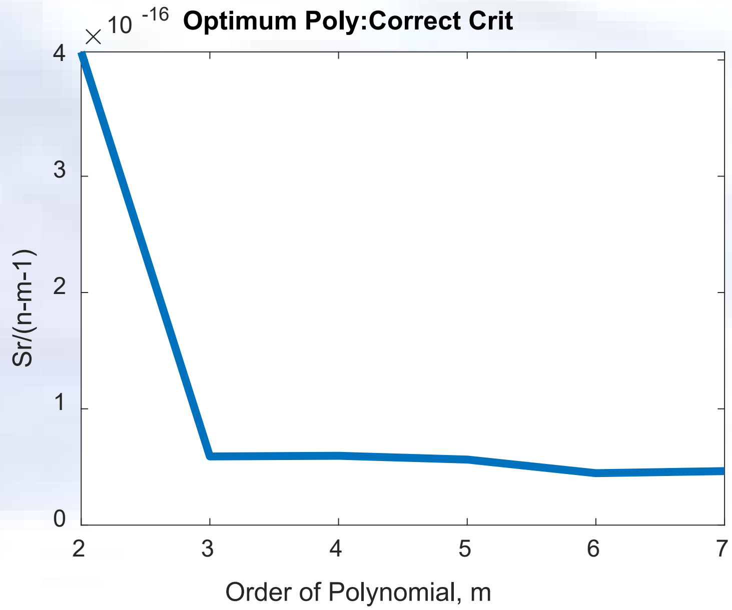




Both graphs are same
Left one starts at m=1
Right one starts at m=2

Optimum Polynomial: Correct Criterion

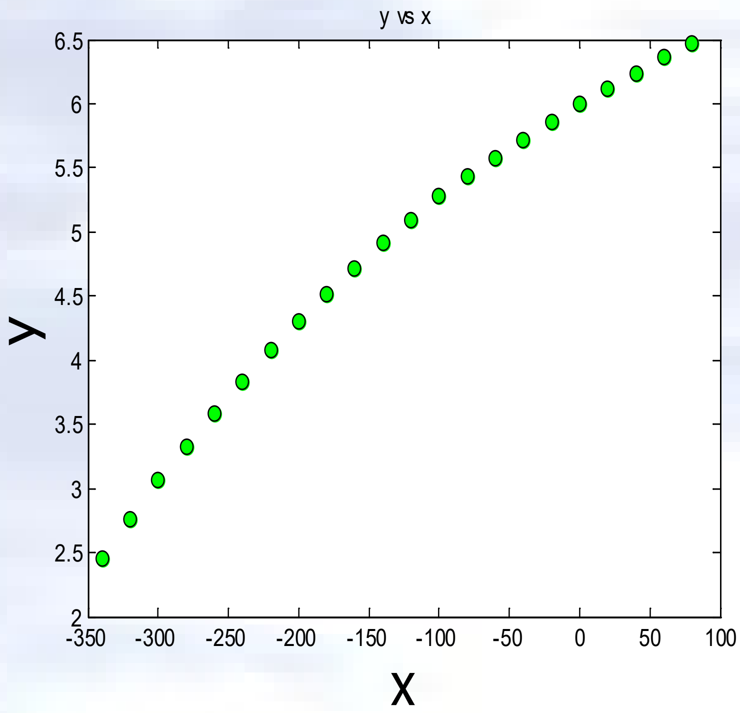




Both graphs are same
Left one starts at m=1
Right one starts at m=2

6.05
Adequacy of Linear Regression Models

Data



Therm exp coeff vs temperature

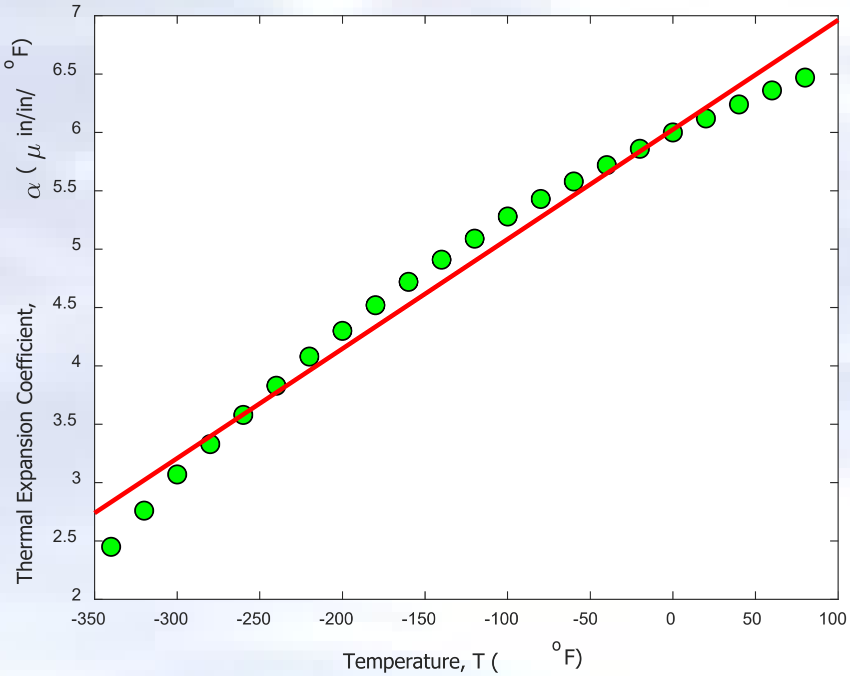
T	α
80	6.47
60	6.36
40	6.24
20	6.12
0	6.00
-20	5.86
-40	5.2
-60	5.58
-80	5.43
-100	5.28
-120	5.09

T	α
-140	4.91
-160	4.72
-180	4.52
-200	4.30
-220	4.08
-240	3.83
-260	3.58
-280	3.33
-300	3.07
-320	2.76
-340	2.45

T is in ^oF

α is in $\mu in/in/\,^oF$

Is this adequate?



Straight Line Model

Quality of Fitted Data

- Does the model describe the data adequately?
- How well does the model predict the response variable predictably?

Linear Regression Models

- **Limit our discussion to adequacy of straight-line regression models**

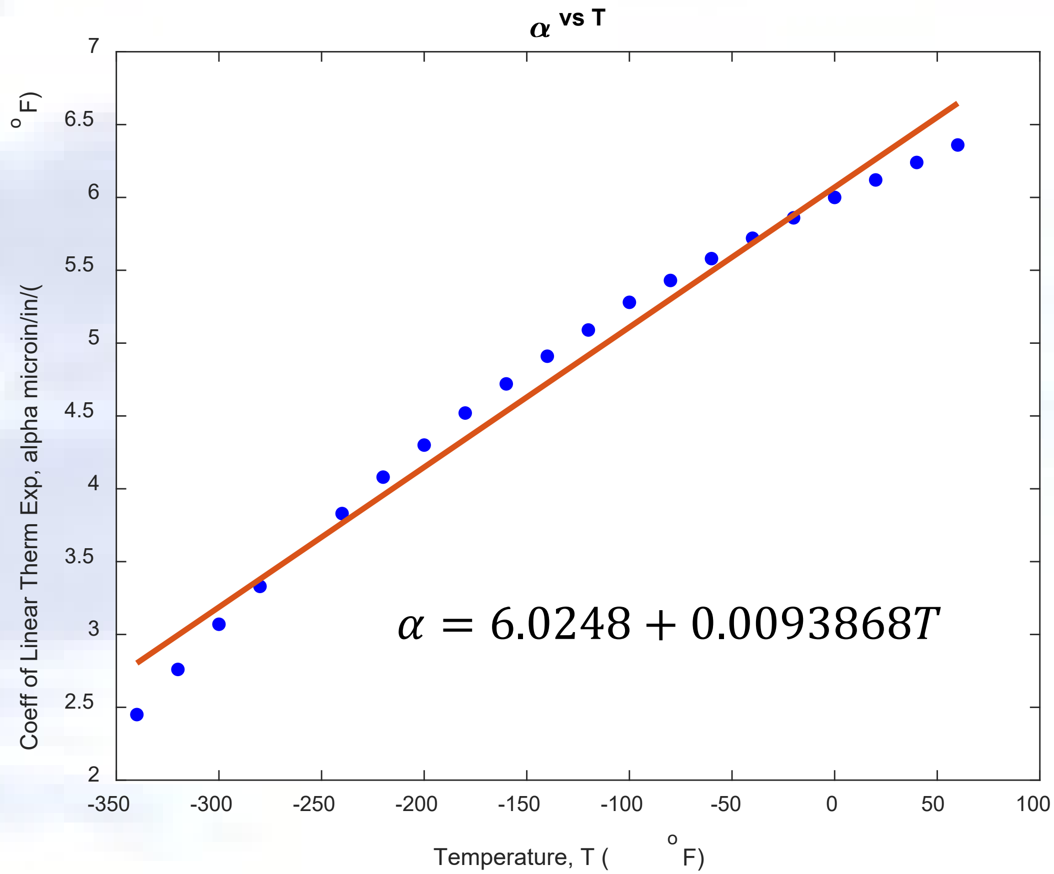
Four checks

1. Does the model look like it explains the data?
2. Do 95%of the residuals fall with ±2 standard error of estimate?
3. Is the coefficient of determination acceptable?
4. Does the model meet the assumption of random errors?

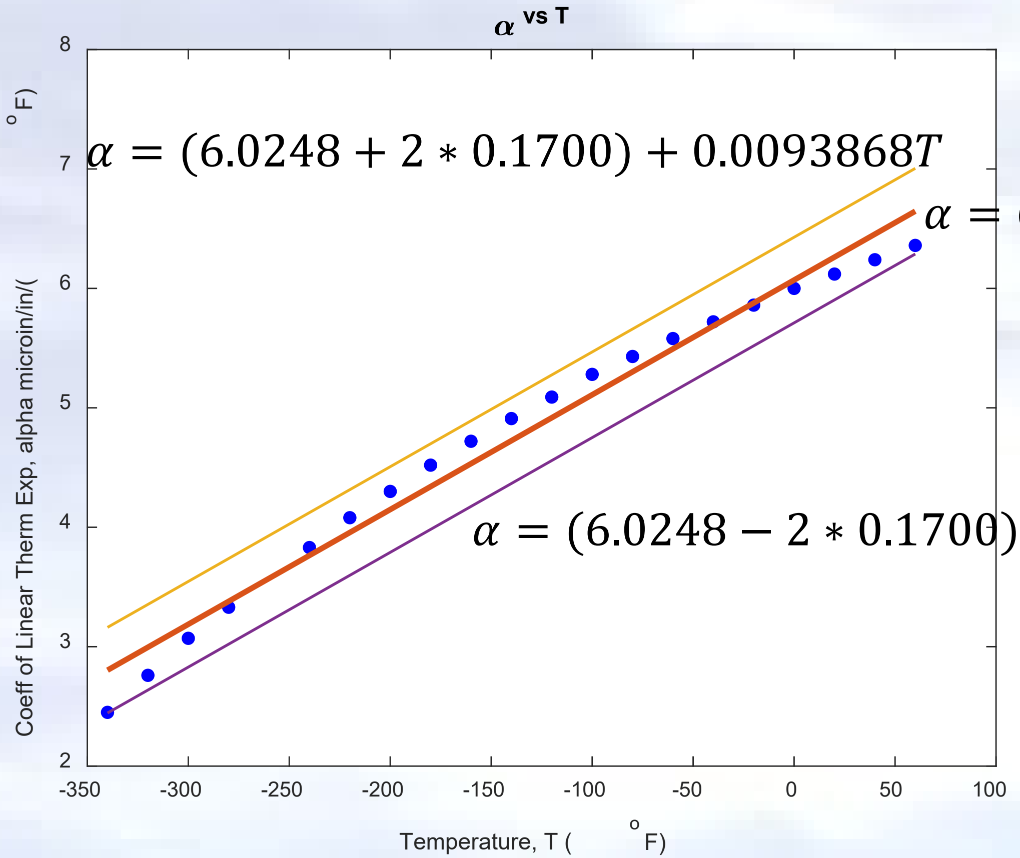
Check 1:Plot Model and Data

T	α
80	6.47
60	6.36
40	6.24
20	6.12
0	6.00
-20	5.86
-40	5.2
-60	5.58
-80	5.43
-100	5.28
-120	5.09

T	α
-140	4.91
-160	4.72
-180	4.52
-200	4.30
-220	4.08
-240	3.83
-260	3.58
-280	3.33
-300	3.07
-320	2.76
-340	2.45



Check 2: Using Standard Error of Estimate



$$\alpha = (6.0248 + 2 * 0.1700) + 0.0093868T$$

$$\alpha = 6.0248 + 0.0093868T$$

$$\alpha = (6.0248 - 2 * 0.1700) + 0.0093868T$$

$$s_{\alpha/T} = \sqrt{\frac{S_r}{n-2}}$$

$$s_{\alpha/T} = 0.1700$$

Problem Assigned

Given (2,4), (2,5), (3,5) and (3,6) as data points

1) Regress to a general straight line, $y = a_0 + a_1 x$. (Answer: y=1x+2.5)

2) Find the standard error of estimate (Ans: 0.7071).

3) Find the scaled residuals (Answer: -0.7071 0.7071 -0.7071 0.7071).

Check 3: Using Coefficient of Determination

$$r^2 = \frac{S_t - S_r}{S_t}$$

$$= \frac{27.614 - 0.5785}{27.614}$$

$$= 0.9791$$

Problem Assigned

Given (2,4), (2,5), (3,5) and (3,6) as data points (extension of previous problem)

1) Find the sum of the square of the differences with the mean (Ans: 2).

2) Find the sum of the square of the residuals. (Ans: 1)

3) Find the coefficient of determination (Ans: 0.5).

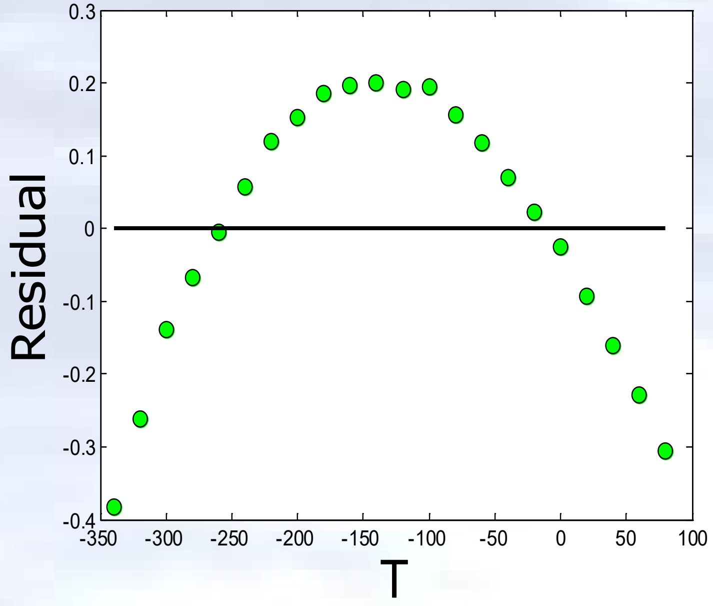
4) Find the correlation coefficient (Ans: 0.7071).

Check 4. Does the model meet assumption of random errors?

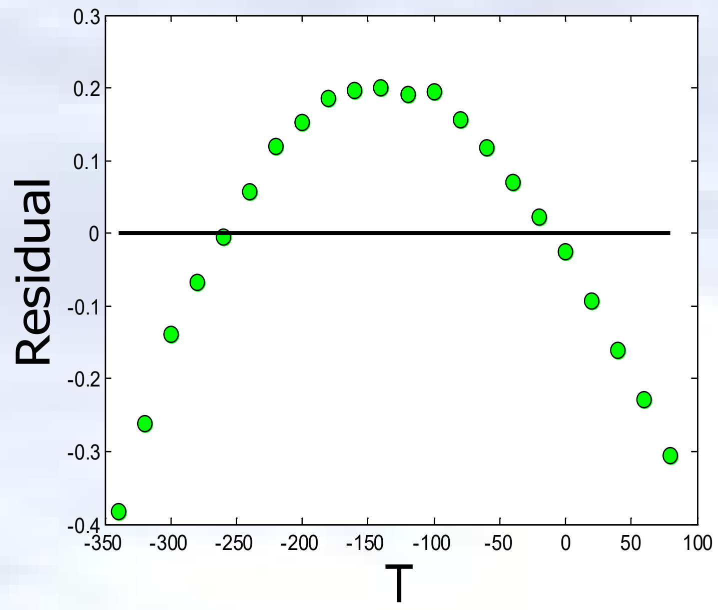
Model meets assumption of random errors

- Residuals are negative as well as positive
- Variation of residuals as a function of the independent variable is random
- Residuals follow a normal distribution
- ~~There is no autocorrelation between the data points.~~

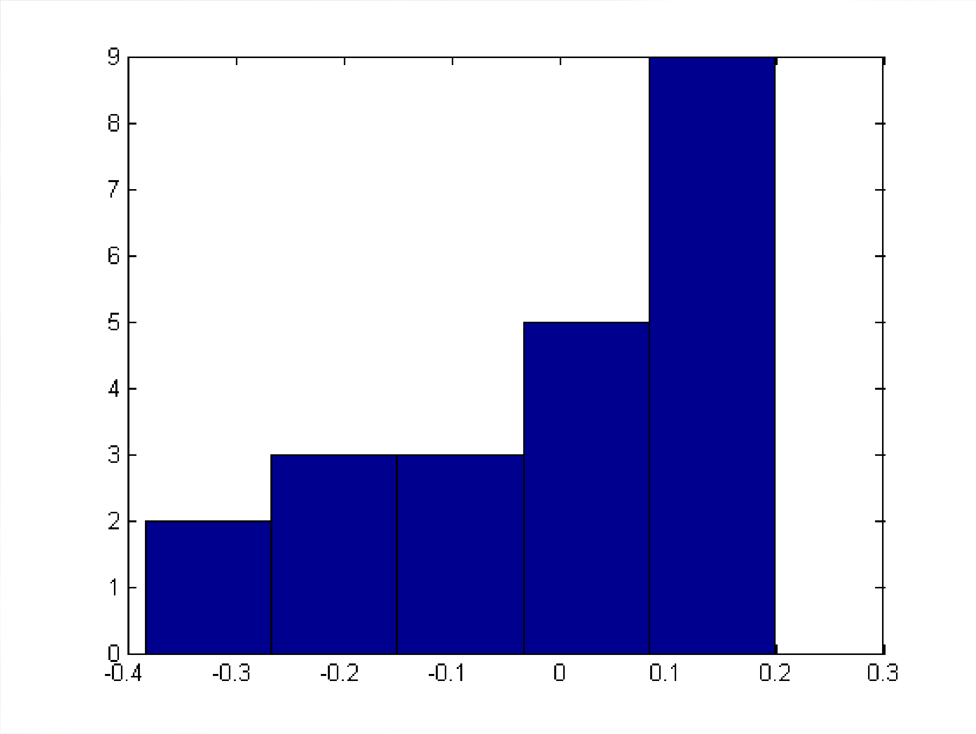
Are residuals negative and positive?



Is variation of residuals as a function of independent variable random?

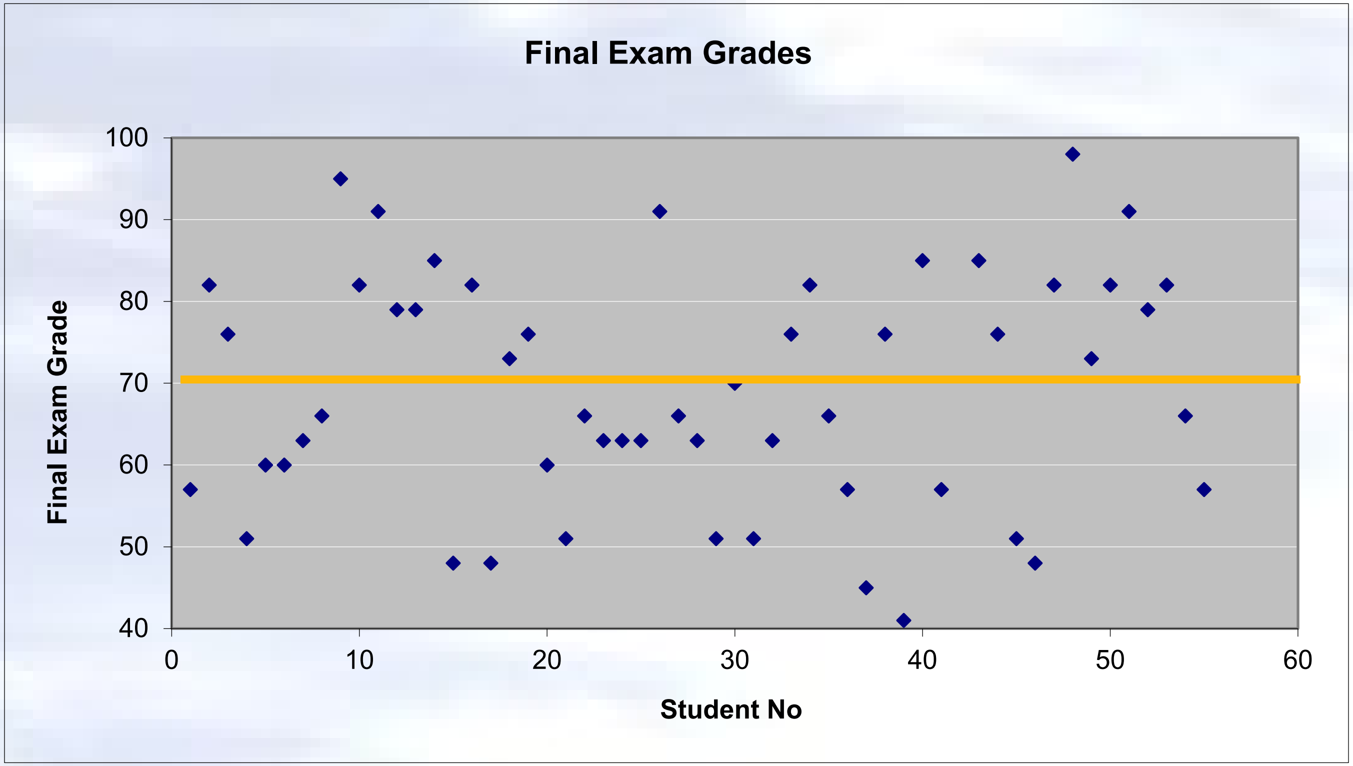


Do the residuals follow normal distribution?

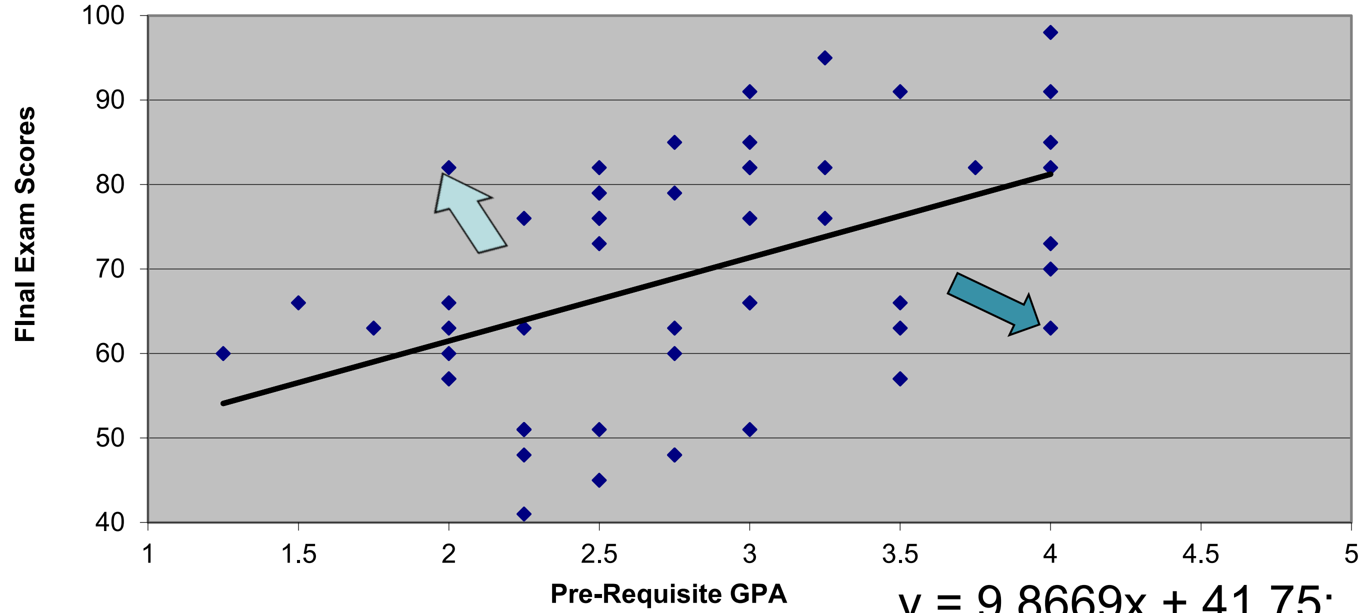


06.XX
Parting Thoughts

Final Exam Grade

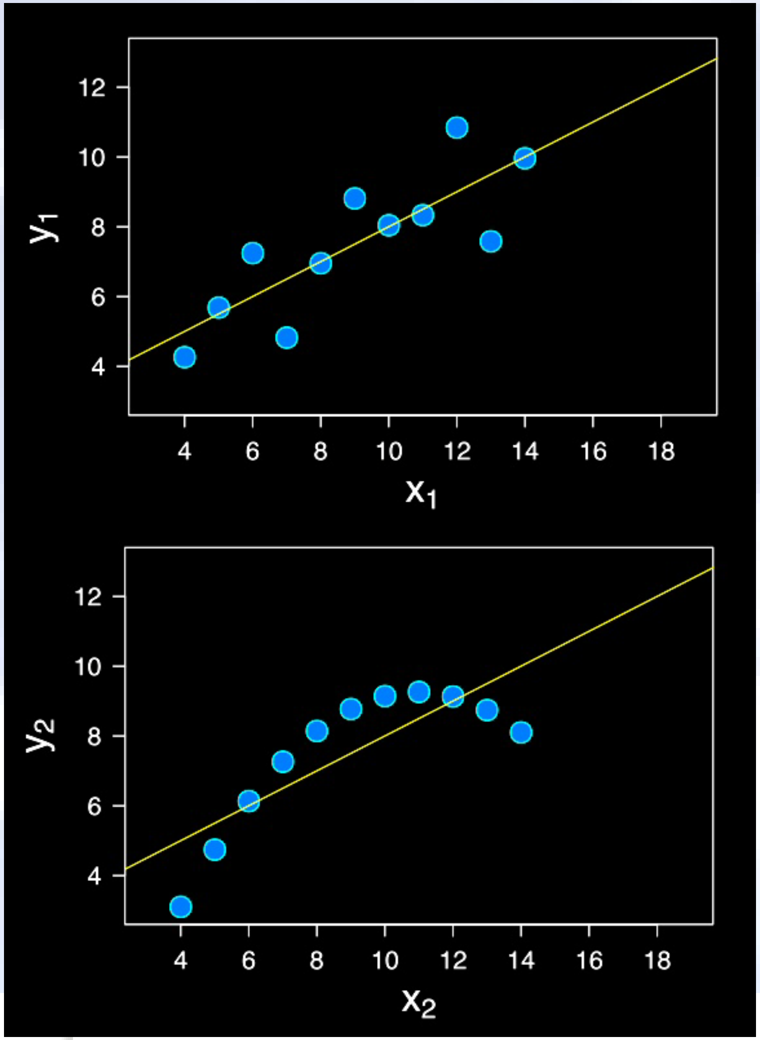


Final Exam Grade vs Pre-Req GPA



y = 9.8669x + 41.75;
R² = 0.2227
R=0.4719

Same but different

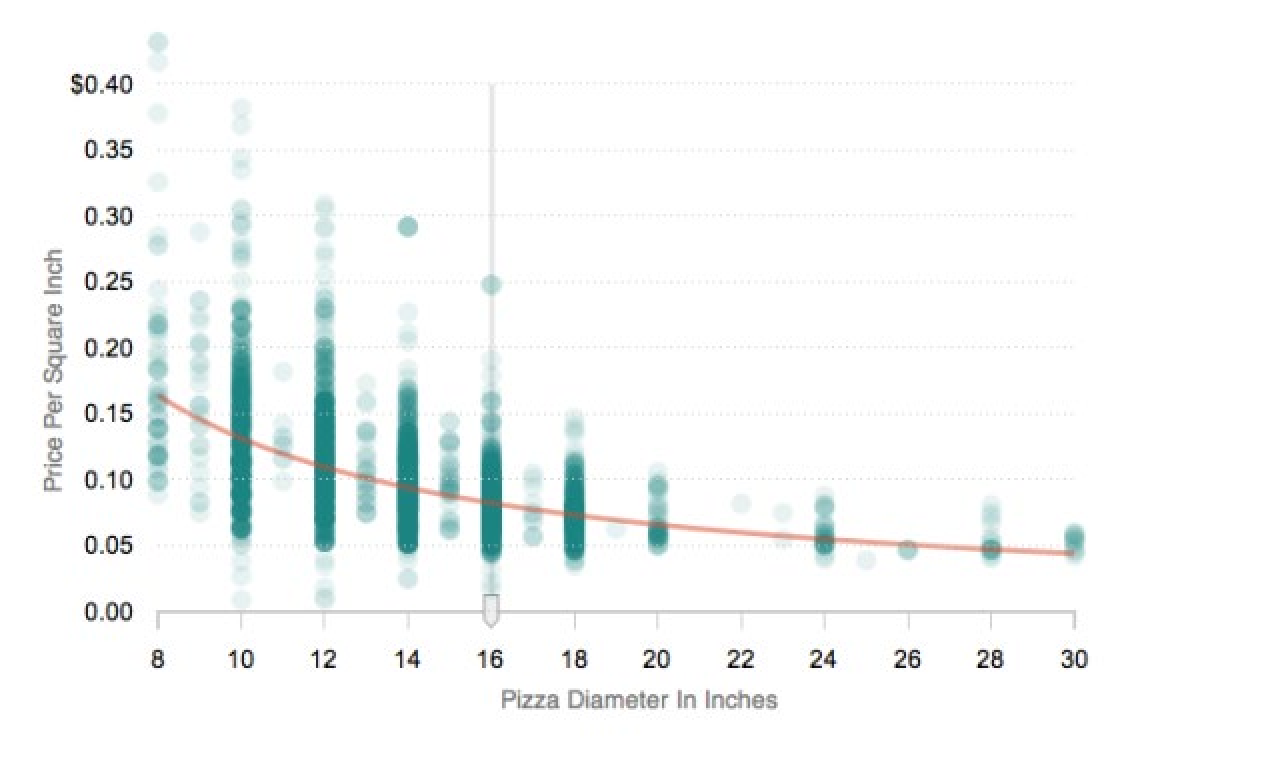


The following are same for both lines.

- Mean of x
- Sample variance of x
- Mean of y
- Sample variance of y
- Correlation between x and y
- Linear regression line
- Coefficient of determination of the linear regression

Source: Cliff Pickover: https://twitter.com/pickover/status/1576630790309707776/photo/1

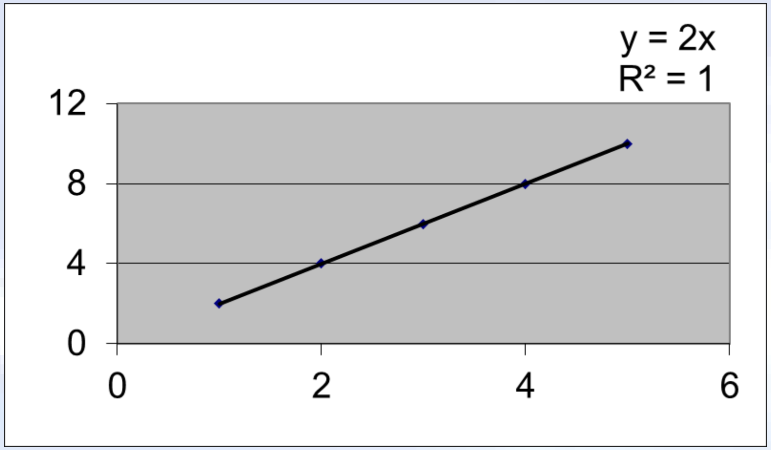
Pizza price vs Pizza Diameter



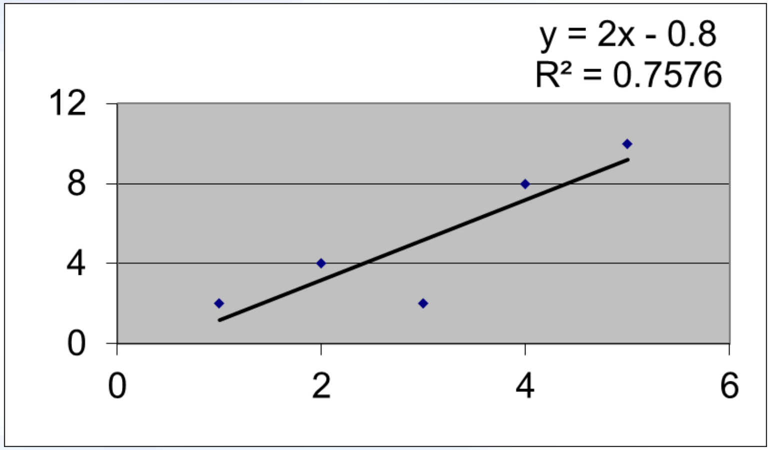
Sources: https://www.npr.org/sections/money/2014/02/26/282132576/74-476-reasons-you-should-always-get-the-bigger-pizza
https://www.themarysue.com/mpr-pizza-graph/

Effect of Outlier

1	2
2	4
3	6
4	8
5	10



1	2
2	4
3	2
4	8
5	10



END