



Regression

Reading While Skimming the
Lines

1

6.04 Nonlinear Regression

<http://nm.mathforcollege.com>

2

Nonlinear Regression Without Transformation of Data (Untransformed Data)

3

3

When *transforming* the data to find the constants of the regression model $y = ae^{bx}$ to best fit $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 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 poller.com/app

5

Nonlinear Regression

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

1. Exponential model: $y = ae^{bx}$
2. Power model: $y = ax^b$
3. Saturation growth model: $y = \frac{ax}{b+x}$
4. Polynomial model: $y = a_0 + a_1x + a_2x^2 + \dots + a_mx^m$

6

<http://nm.MathForCollege.com>

6

Nonlinear Regression

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

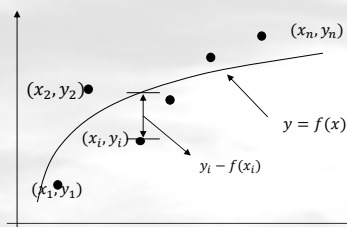


Figure. Nonlinear regression model for discrete y vs. x data

7

<http://nm.MathForCollege.com>

7

Exponential Model

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

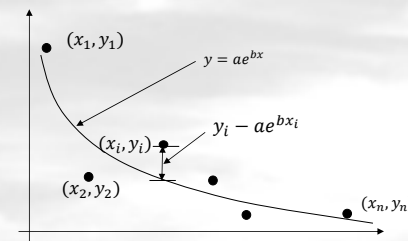


Figure. Exponential model of nonlinear regression for y vs. x data

8

<http://nm.MathForCollege.com>

8

Finding Constants of Exponential Model

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

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

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

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 .

9

<http://nm.MathForCollege.com>

9

Finding Constants of Exponential Model

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

Differentiate with respect to a

$$\begin{aligned} \frac{\partial S_r}{\partial a} &= \sum_{i=1}^n 2(y_i - ae^{bx_i})(-e^{bx_i}) \\ &= \sum_{i=1}^n -2y_i e^{bx_i} + 2ae^{2bx_i} \\ &= \sum_{i=1}^n -2y_i e^{bx_i} + \sum_{i=1}^n 2ae^{2bx_i} \end{aligned}$$

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

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

10

<http://nm.MathForCollege.com>

10

Finding Constants of Exponential Model

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

Differentiate with respect to b

$$\begin{aligned} \frac{\partial S_r}{\partial b} &= \sum_{i=1}^n 2(y_i - ae^{bx_i})(-ax_i e^{bx_i}) \\ &= \sum_{i=1}^n -2ay_i x_i e^{bx_i} + 2a^2 x_i e^{2bx_i} \\ &= \sum_{i=1}^n -2ay_i x_i e^{bx_i} + \sum_{i=1}^n 2a^2 x_i e^{2bx_i} \end{aligned}$$

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

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

11

<http://nm.MathForCollege.com>

11

Finding Constants of Exponential Model

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

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

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

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

12

<http://nm.MathForCollege.com>

12

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

$$\begin{aligned} \sum_{i=1}^n -2ay_i x_i e^{bx_i} + \sum_{i=1}^n 2a^2 x_i e^{2bx_i} &= 0 \\ -2a \sum_{i=1}^n y_i x_i e^{bx_i} + 2a^2 \sum_{i=1}^n x_i e^{2bx_i} &= 0 \\ -\sum_{i=1}^n y_i x_i e^{bx_i} + a \sum_{i=1}^n x_i e^{2bx_i} &= 0 \\ -\sum_{i=1}^n y_i x_i e^{bx_i} + \frac{\sum_{i=1}^n y_i e^{bx_i}}{\sum_{i=1}^n e^{2bx_i}} \sum_{i=1}^n x_i e^{2bx_i} &= 0 \end{aligned}$$

13

<http://nm.MathForCollege.com>

13

Example - Exponential Model

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

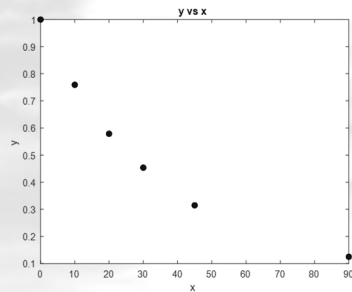
Use the regression model $y = ae^{bx}$. Estimate the regression constants a and b without transforming the data.

14

<http://nm.MathForCollege.com>

14

Plot of data



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

15

<http://nm.MathForCollege.com>

15

Constants of the Model

$$y = ae^{bx}$$

The value of b is found by solving the nonlinear equation

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

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

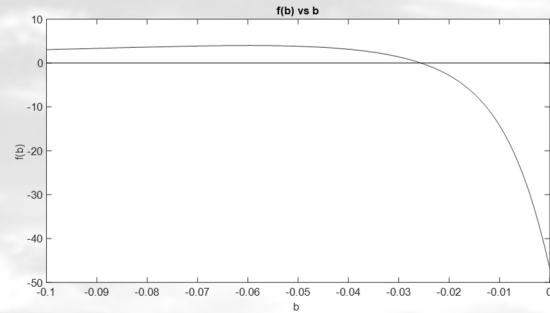
16

<http://nm.MathForCollege.com>

16

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

17

<http://nm.MathForCollege.com>

17

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);
```

$$b = -0.02561$$

18

<http://nm.MathForCollege.com>

18

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

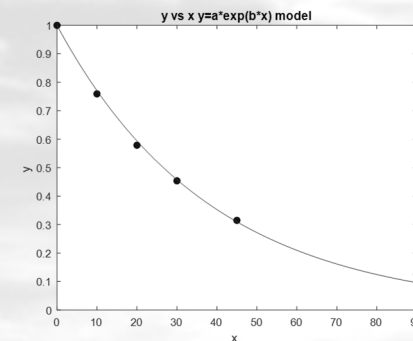
$$y = 0.9885 e^{-0.02561x}$$

19

<http://nm.MathForCollege.com>

19

Plot of data and regression curve



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

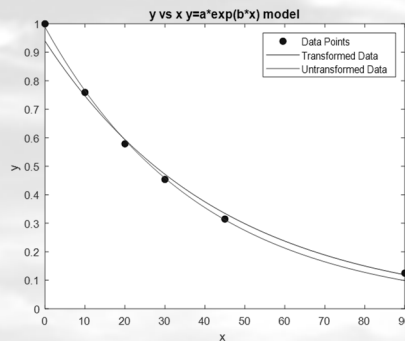
20

<http://nm.MathForCollege.com>

20

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

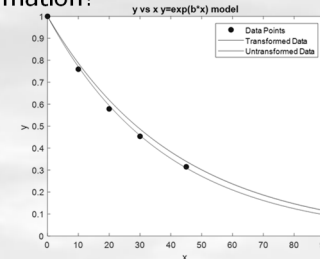
Transformed data model $y = 0.9395e^{-0.02294x}$

21

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?



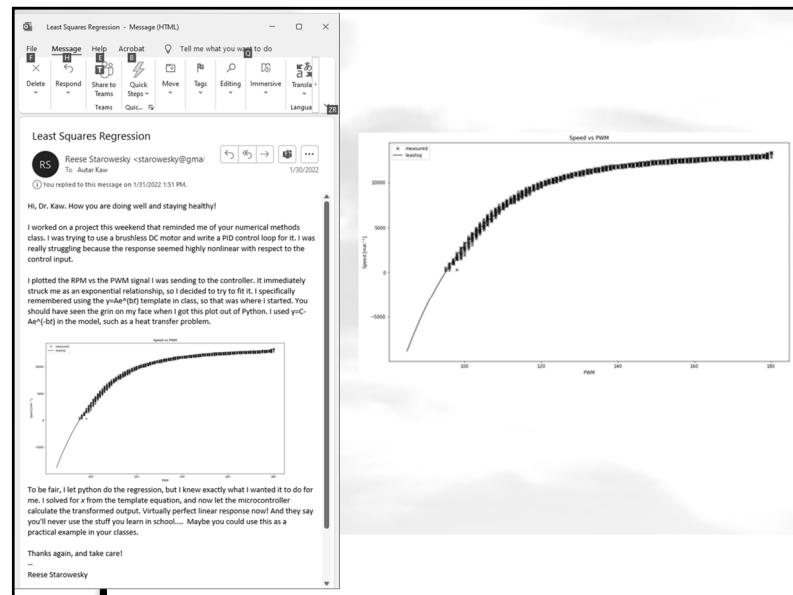
Untransformed data model $y = e^{-0.02605x}$

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

22

<http://nm.MathForCollege.com>

22



23

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

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

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

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

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

Powered by Poll Everywhere
Start the presentation to see live content. For screen share software, share the entire screen. Get help at poller.com/app

24

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

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

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

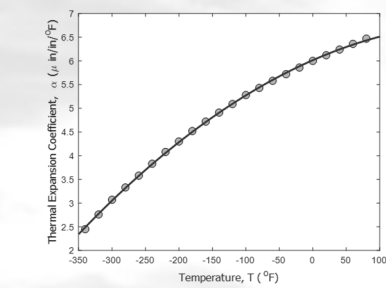
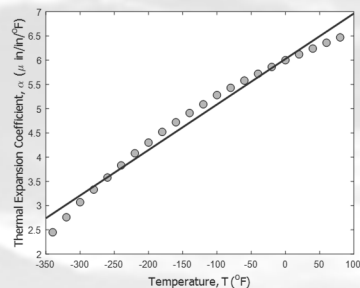
Start the presentation to see live content. For screen share software, share the entire screen. Get help at poll Everywhere.com/app

25

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

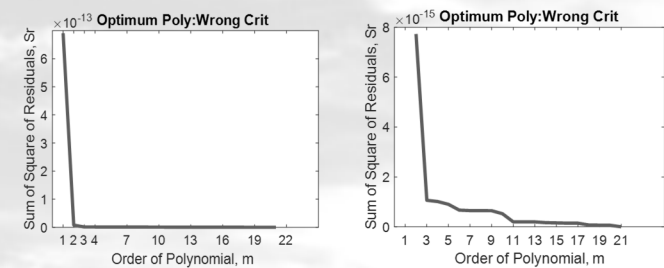
26

Which model to choose?



27

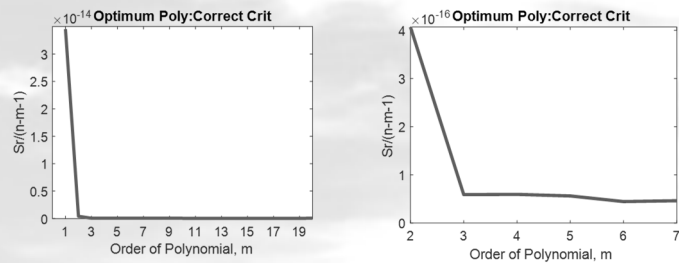
Optimum Polynomial: Wrong Criterion



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

28

Optimum Polynomial: Correct Criterion



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

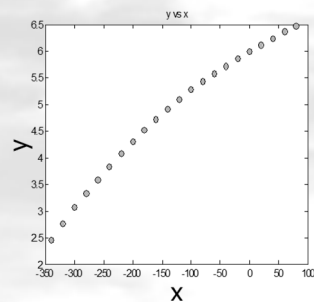
29

6.05 Adequacy of Linear Regression Models

<http://nm.mathforcollege.com>

30

Data



31

Therm exp coeff vs temperature

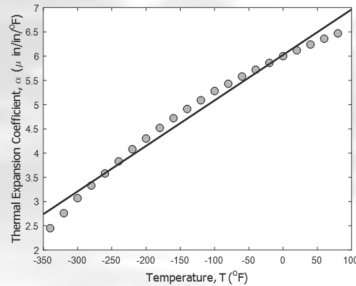
T	α
80	6.47
60	6.36
40	6.24
20	6.12
0	6.00
-20	5.86
-40	5.2
-60	5.58
-80	5.43
-100	5.28
-120	5.09

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 $^{\circ}F$
 α is in $\mu in/in/^{\circ}F$

32

Is this adequate?



Straight Line Model

33

Quality of Fitted Data

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

34

Linear Regression Models

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

35

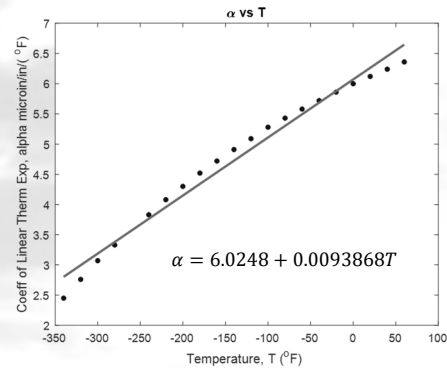
Four checks

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?

36

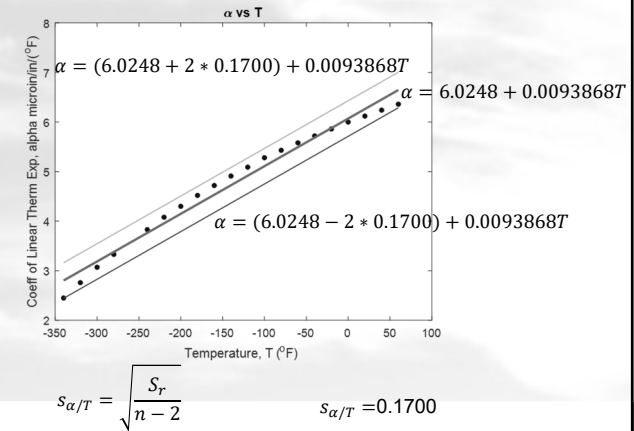
Check 1: Plot Model and Data

T	α	T	α
80	6.47	-140	4.91
60	6.36	-160	4.72
40	6.24	-180	4.52
20	6.12	-200	4.30
0	6.00	-220	4.08
-20	5.86	-240	3.83
-40	5.2	-260	3.58
-60	5.58	-280	3.33
-80	5.43	-300	3.07
-100	5.28	-320	2.76
-120	5.09	-340	2.45



37

Check 2: Using Standard Error of Estimate



38

Problem Assigned

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

1) Regress to a general straight line,

$$y = a_0 + a_1 x. \text{ (Answer: } y = 1x + 2.5 \text{)}$$

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

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

39

Check 3: Using Coefficient of Determination

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

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

$$= 0.9791$$

40

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).

41

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

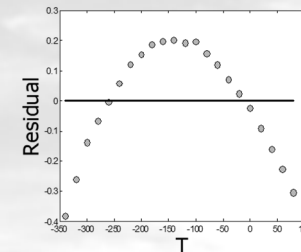
42

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.~~

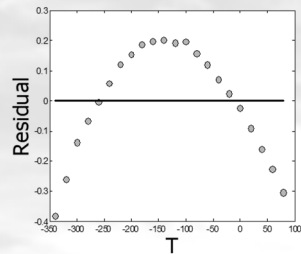
43

Are residuals negative and positive?



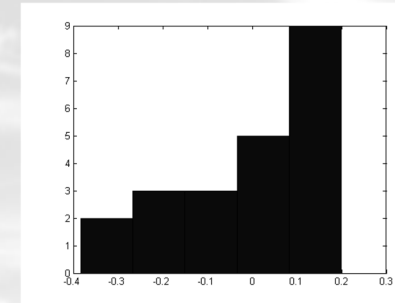
44

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



45

Do the residuals follow normal distribution?



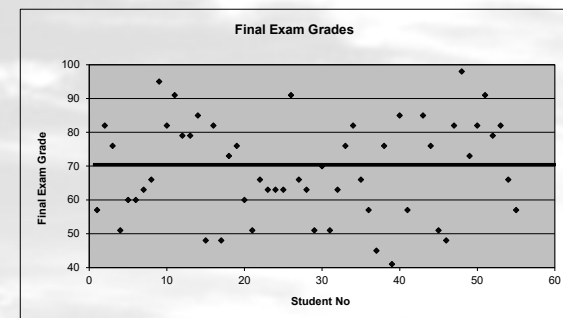
46

06.XX Parting Thoughts

<http://nm.mathforcollege.com>

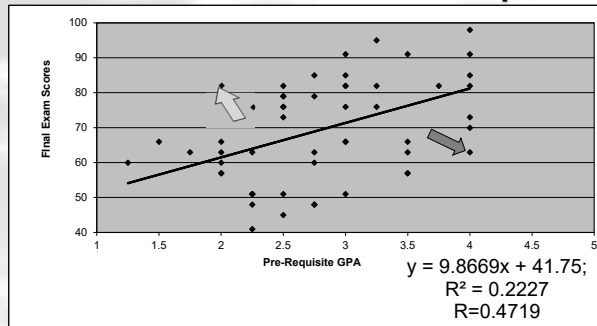
47

Final Exam Grade



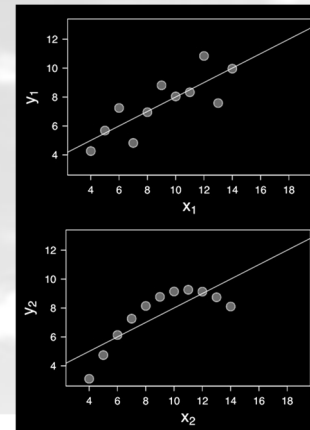
48

Final Exam Grade vs Pre-Req GPA



49

Same but different



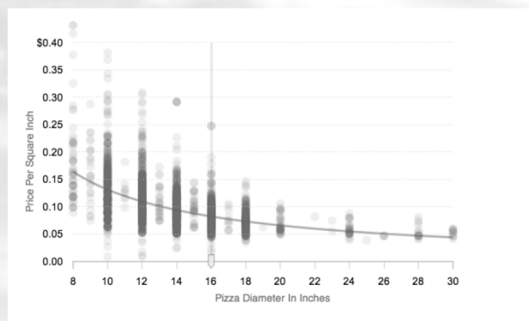
The following are same for both lines.

- 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>

50

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/>

51

END

52