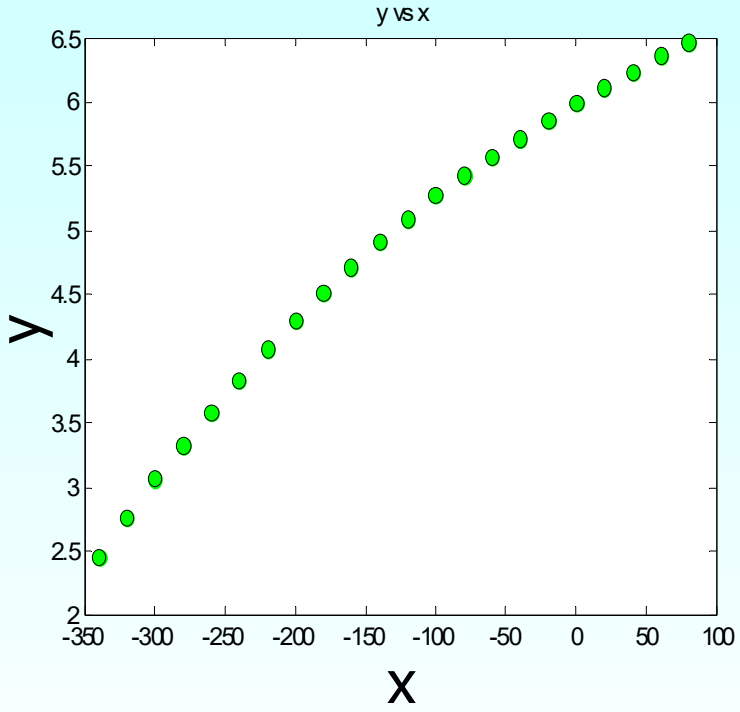
Adequacy of Linear Regression Models

http://numericalmethods.eng.usf.edu

Transforming Numerical Methods Education for STEM Undergraduates

Data



Therm exp coeff vs temperature

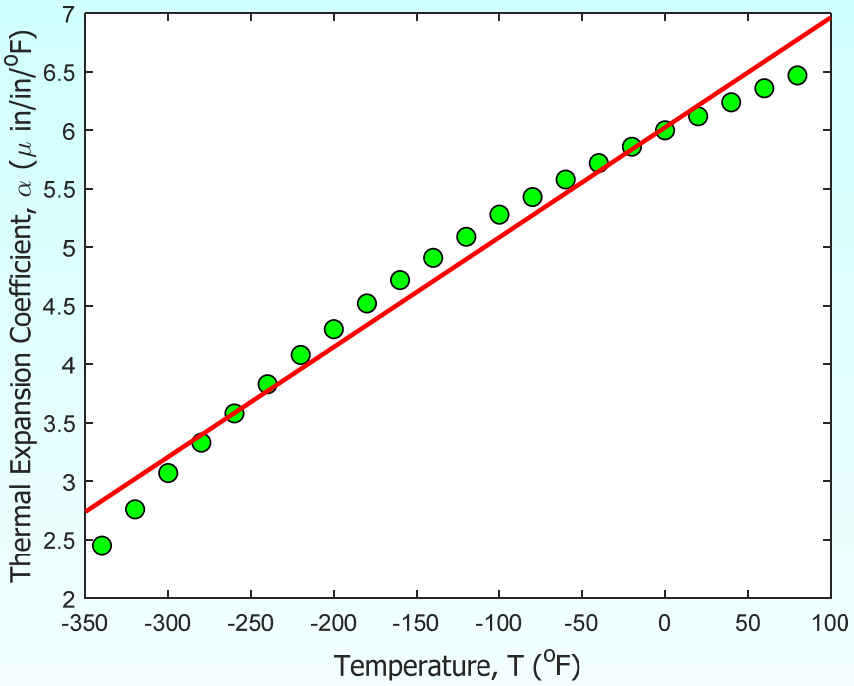
T	α
80	6.47
60	6.36
40	6.24
20	6.12
0	6.00
-20	5.86
-40	5.2
-60	5.58
-80	5.43
-100	5.28
-120	5.09

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 ^{o}F

α is in μin/in/ ^{o}F

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**

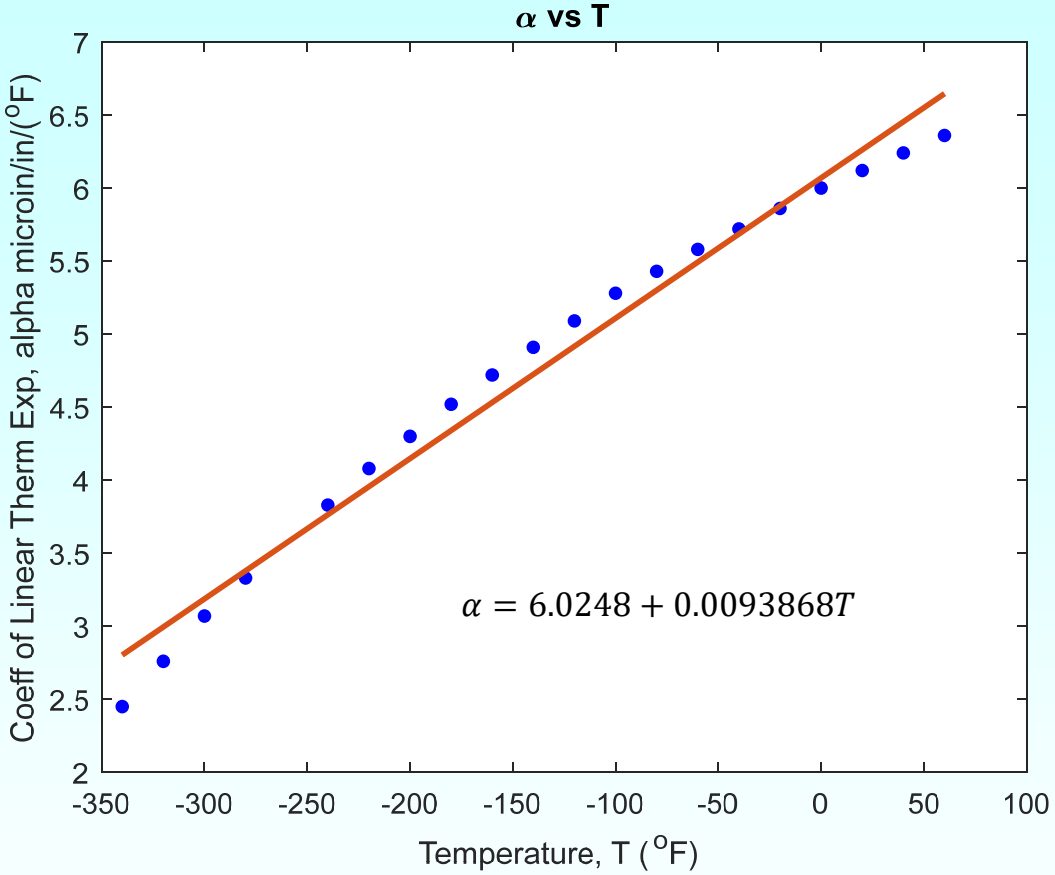
Four checks

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?

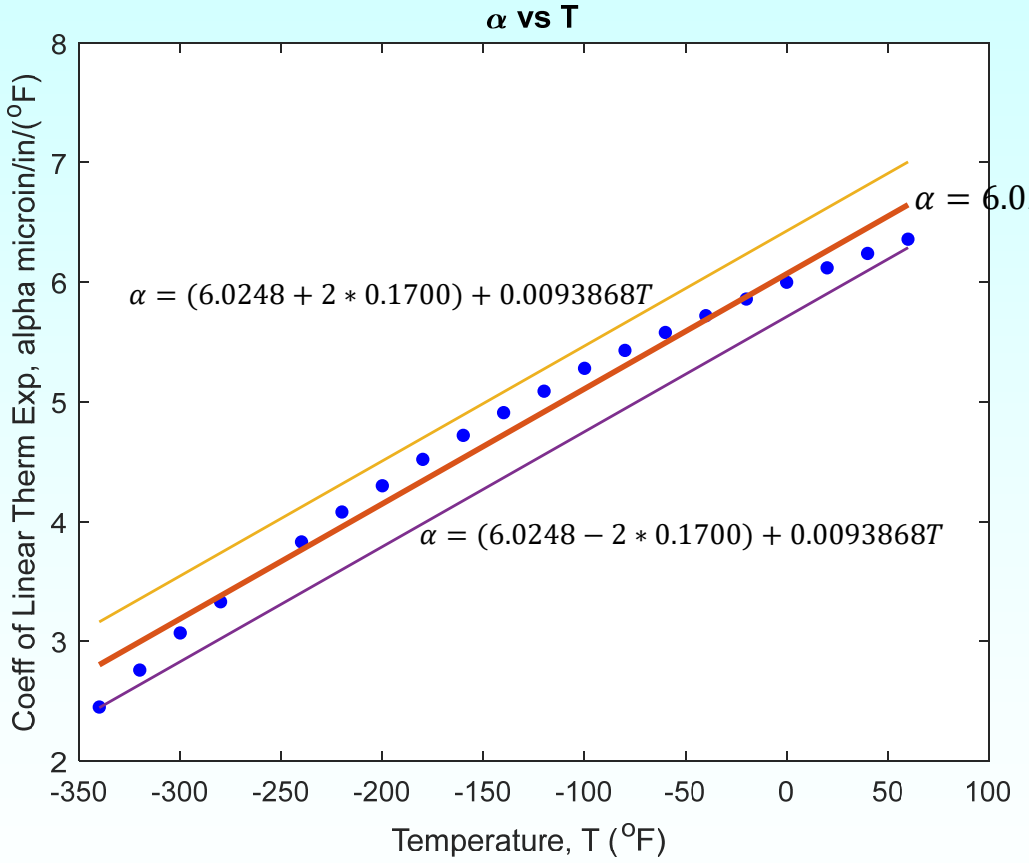
Check 1:Plot Model and Data

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



$s_{\alpha/T}$ =0.1700

Check 3: Using Coefficient of Determination

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

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

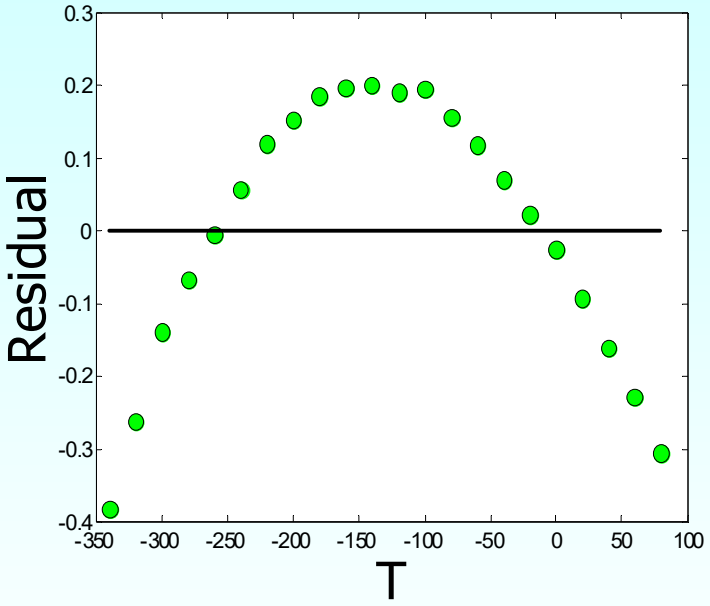
$$= 0.9791$$

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

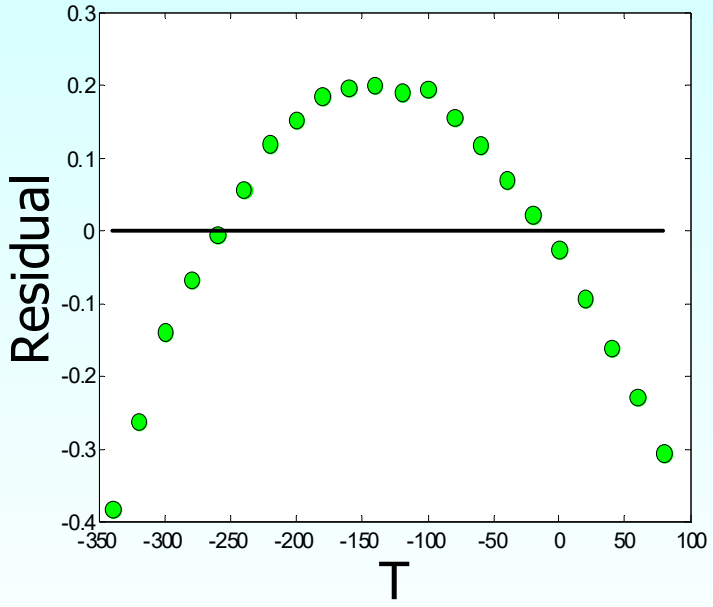
Model meets assumption of random errors

- 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.~~

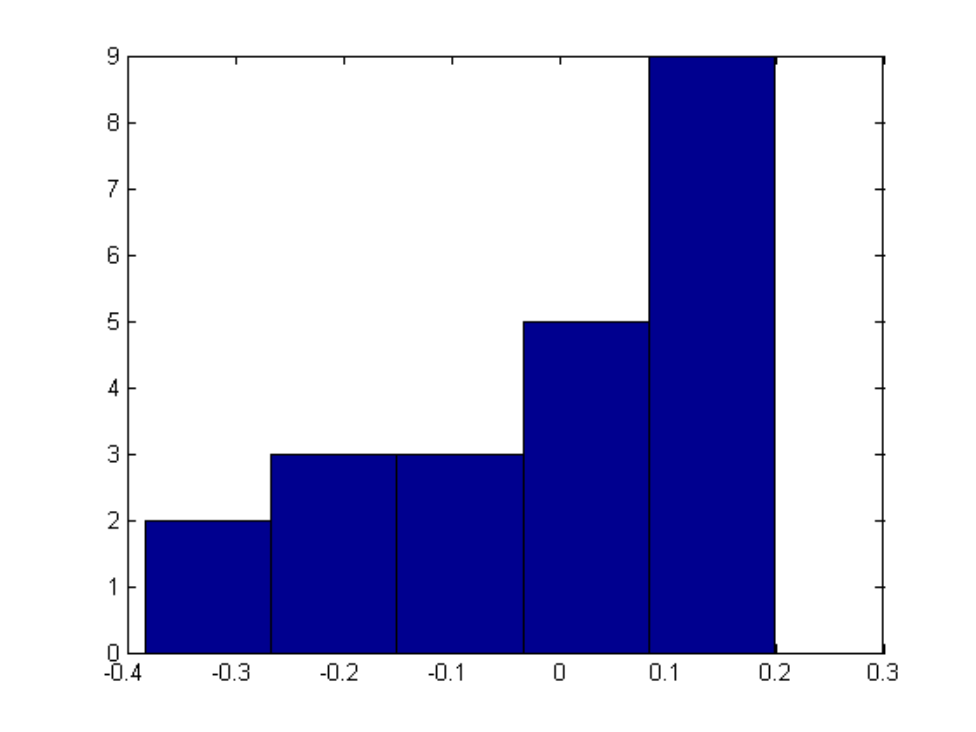
Are residuals negative and positive?



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



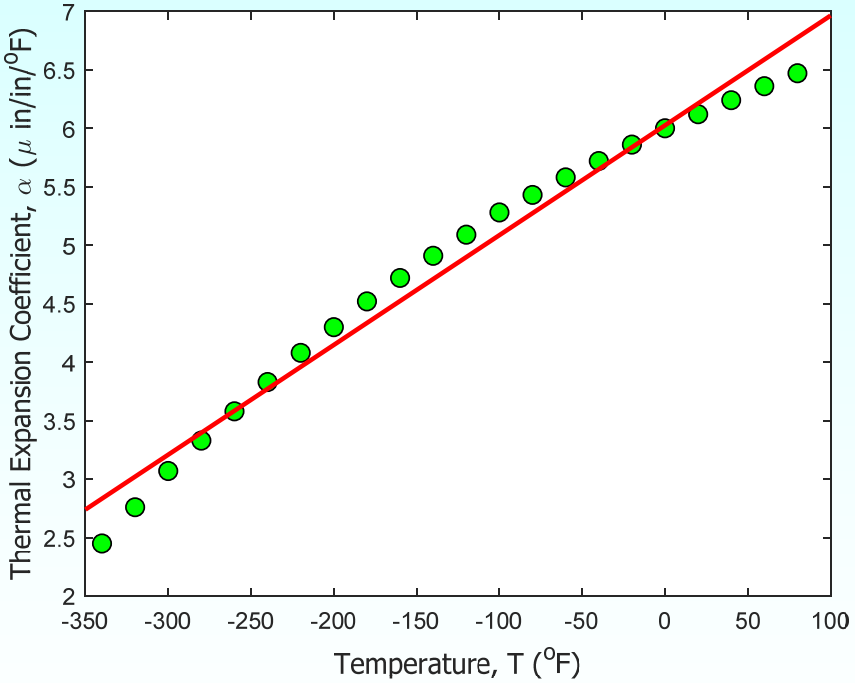
Do the residuals follow normal distribution?



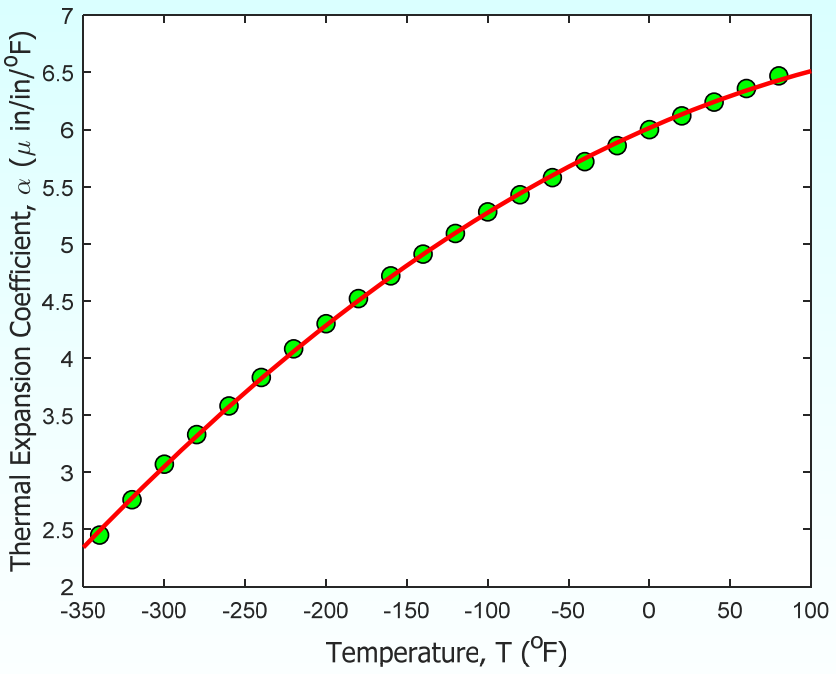
END

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

Which model to choose?

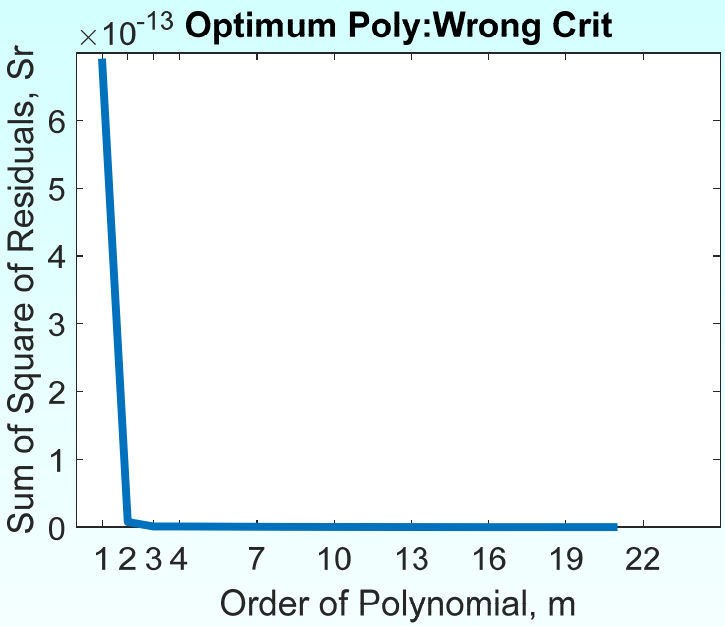


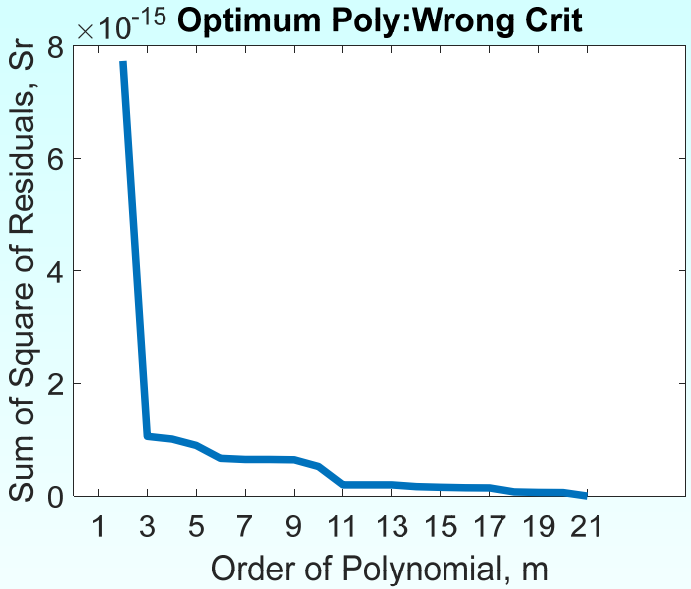
$$\alpha = 0.009387T + 6.025$$



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

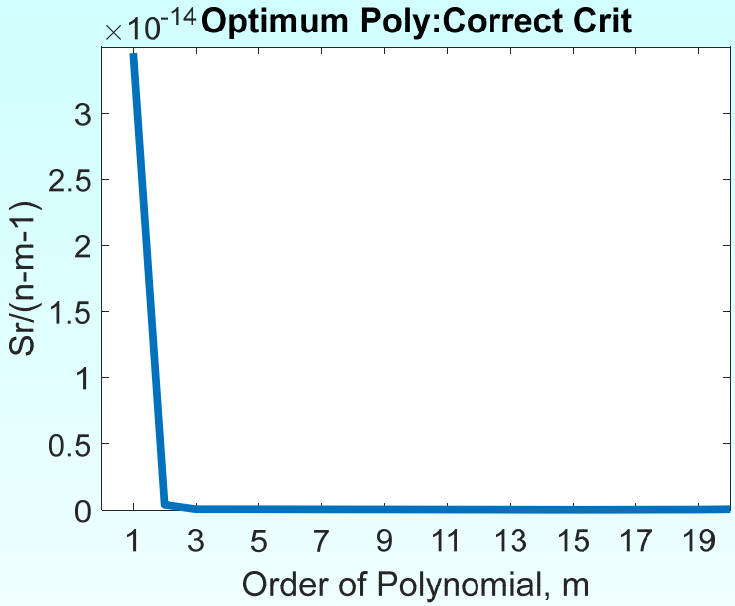
Optimum Polynomial: Wrong Criterion

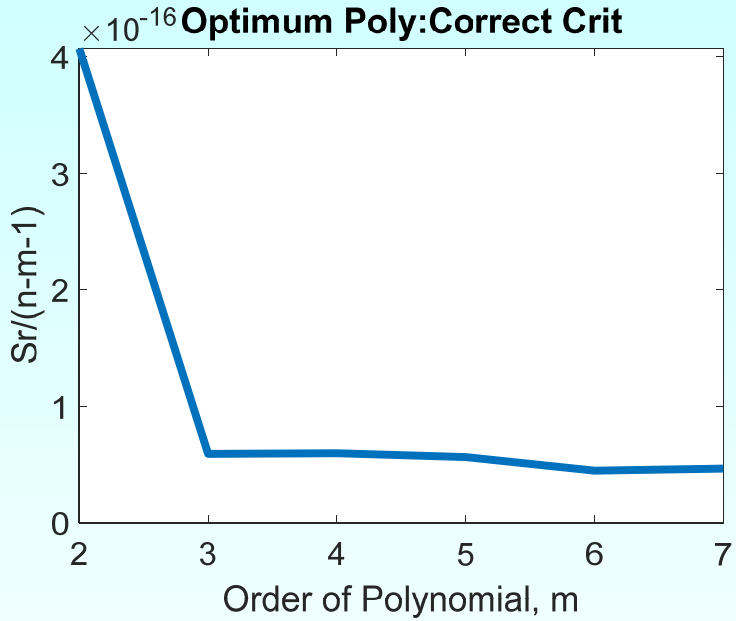




Both graphs are same
Left one starts at m=1
Right one starts at m=2

Optimum Polynomial: Correct Criterion



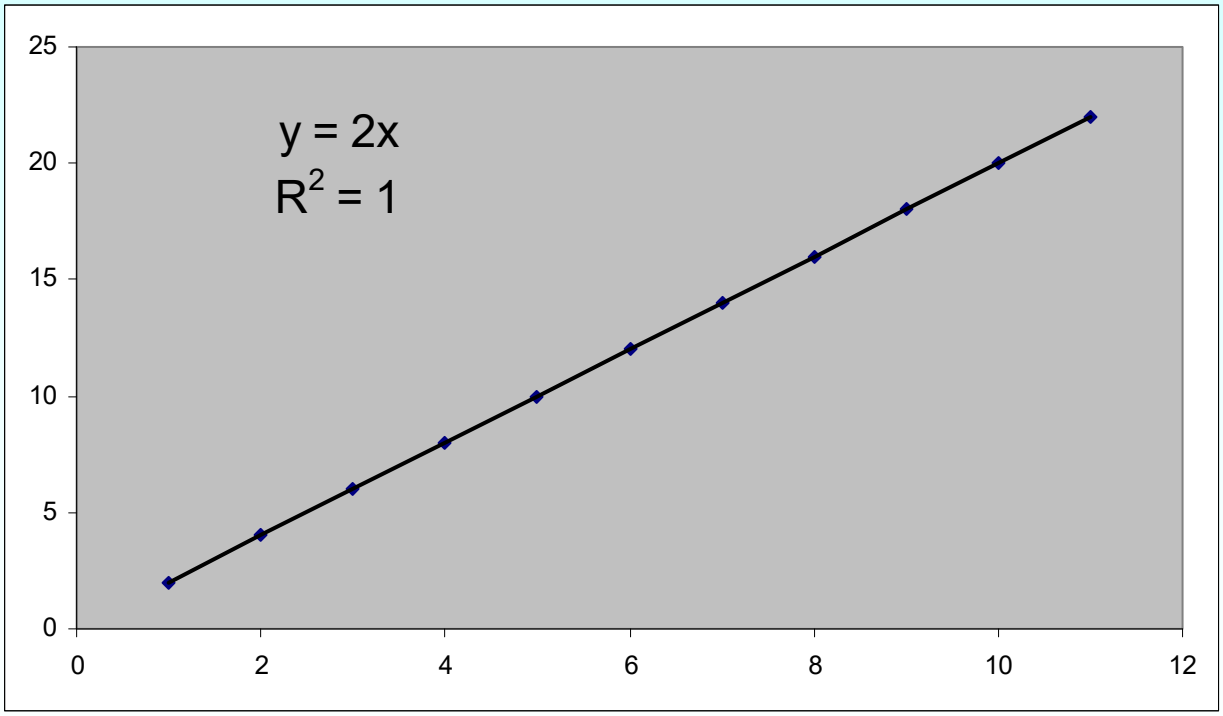


Both graphs are same
Left one starts at m=1
Right one starts at m=2

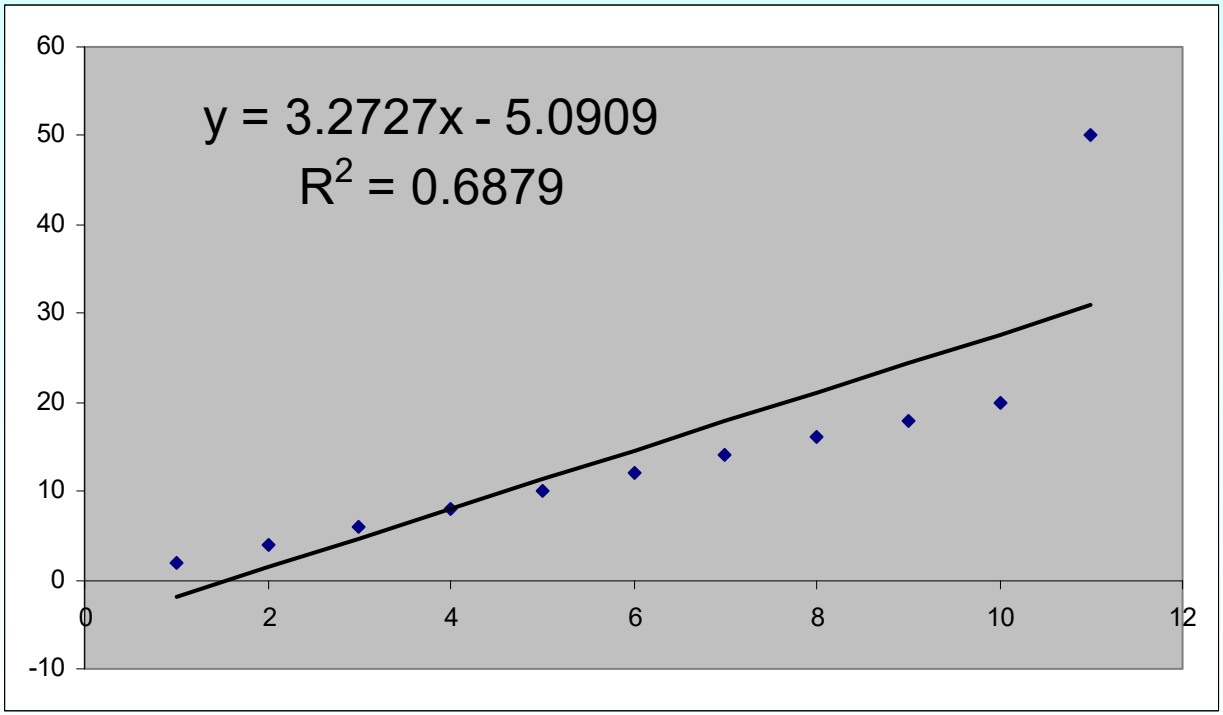
END

Effect of an Outlier

Effect of Outlier

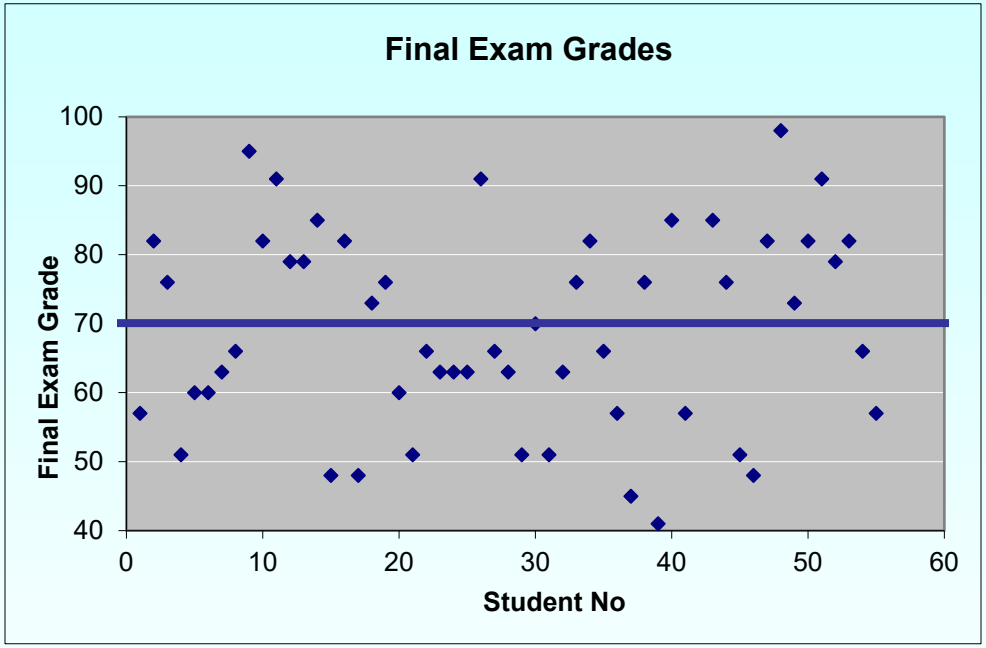


Effect of Outlier



END

Final Exam Grade



Final Exam Grade vs Pre-Req GPA