Empirical Modeling:
Residuals & Polynomial Models

Objectives: by the end of this class you should be able to:

- Calculate, plot and interpret residuals
- perform polynomial regressions and
- evaluate polynomial regressions using residual plots
- Begin Preparing for the fina

Palm:

Sections 5.6 & 5.7

Please download: polyregr2.mat from the file transfer web page

Review: Function Discovery for 2 parameter models
1. Identifying simple functional forms

	Equation	Linerized	Plot
linear	$y = mx + b$	$y = mx + b$	linear (plot)
power	$y = bx^m$	$\log(y) = \log(b) + m\log(x)$	log-log (loglog)
exponential	$y = be^{mx}$	$\ln(y) = \ln(b) + mx$	semilog (semilogy)
	$y = b10^{mx}$	$\log(y) = \log(b) + mx$	

Review: Function Discovery for 2 parameter models
2. Fitting Parameters (m & b)

	Model	Setup for Fit	Untransform Parameters
linear	$y = mx + b$	x vs. y	m = p(1) b = p(2)
power	$y = bx^m$	log(x) vs. log(y)	m = p(1), b = 10^p(2)
exponential	$y = be^{mx}$	x vs. ln(y)	m = p(1), b = e^p(2)
	$y = b10^{mx}$	x vs. log(y)	m = p(1), b = 10^p(2)

Pairs Exercise:

Please plot this data and determine:
- the likely model
- the parameters (m&b)

(data is available in polyregr2.mat)

plot resulting data and model. Prepare a print out of your MATLAB commands and plot

(Keep your data and graph in MATLAB for later in class)

x	y
1	5
2	8
3	10
4	20
5	21
6	29
7	34
8	36
9	45

Create x & y vectors (notice transpose operator):

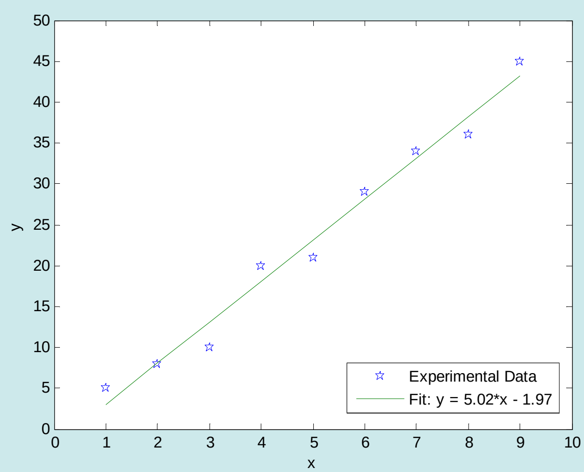
```
>> x = (1:9)'; y =[5, 8, 10, 20, 21, 29, 34, 36, 45]';
```

Model: Check linear plot

```
>> plot(x, y, 's')
```

Graph is linear, fit a simple linear model and calculate predicted values

```
>> X = [x./x, x];  Y = y;    % Set up:  design and response matrix
>> B = X \ Y                     % Fit: parameters
B =
  -1.9722
   5.0167
>> yhat = X*B;               % Predict: calculate fitted values (the fits)
```



A Reminder of Some Nomenclature:

y ➔ response (dependant variable) vector
y_i ➔ an individual response

x ➔ predictor (independent variable) vector
x_i ➔ an individual predictor value

$\hat{y}$ ➔ the predicted value (the fits)
$\hat{y}_i$ ➔ an individual predicted value (fit)

Residuals:

- difference between model and the actual data:
  residuals = yhat - y
- Represents what is not fit by the model
- Ideal model should capture all systematic information
- Residuals should contain only random error
- Plot residuals and look for patterns

Preparing a residuals plot

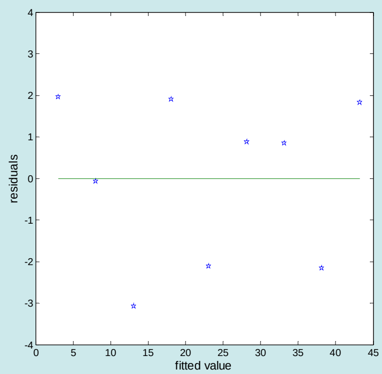
Calculate residuals

```
>> res = yhat – y
```

Plot residuals vs. fits

```
>> plot(yhat, res, 'p', yhat, 0*res)
>> ylim([-4 4])
>> xlabel 'fitted value'
>> ylabel 'residuals'
```

Generally you make the y axis equal in both positive and negative direction (hence the ylim command



What to look for in a residual plot:

1. Does the residual plot look correct?
   data should vary about zero
   sum of residuals must equal zero
2. Are there any patterns in the residuals?, e.g.,
   curvature: high center, low ends or
   low center, high ends
   changes is variability: the spread of the data in the y direction should be constant
3. How big are the residuals?
   (what is the magnitude of the y axis)

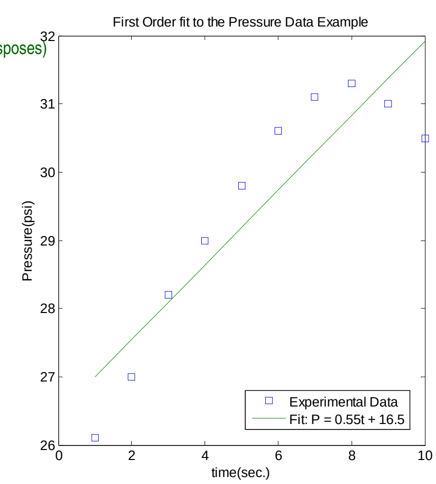
Plot these two vectors

(load polyregr2.mat and they will be in your workspace)

t1 = time (sec.)	P= Pressure (psi)
1	26.1
2	27.0
3	28.2
4	29.0
5	29.8
6	30.6
7	31.1
8	31.3
9	31.0
10	30.5

Model: y = b + mx

```
>> % Setup X & Y arrays (notice transposes)
>>  X = [ones(length(t1), 1), t1'];
>> Y = P' ;
>> % Fit:  parameters
>> B = X\Y
B =
  26.4533
   0.5467
>> % Predict:
>> Phat = X*B
>> % Plot: and label graph
>> plot(t, P, 's' , t1', Phat)
>> xlabel 'time(sec.)'
>> ylabel 'Pressure(psi)'
>> title 'First Order fit to the
          Pressure Data Example'
>> legend( 'Experimental Data',
     'Fit: P = 0.55t + 16.5' )
```



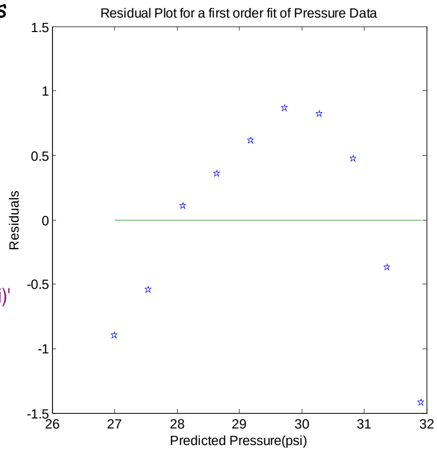
Fitted Model is: P = 0.55t + 26.5

Plotting the residuals

```
>> % Calc residuals & plot
>> res = P - Phat;
>> plot(Phat, res, 'p', Phat, res*0)

>> % Balance y axis
>> ylim([-1.5 1.5])

>> % add appropriate labels
>> xlabel 'Predicted Pressure (psi)'
>> ylabel 'residuals'
>> title 'Residual Plot for a first
   order fit of Pressure Data'
```



Notes on first order fit to pressure data

- Scatter Plot: Clearly this is a poor fit, a linear equation has no chance of fitting a curve with a maximum. Power and exponential models will not fit this data because of the maximum
- Residual Plot: However, it is instructive to notice the residual plot. This plot clearly amplifies the curvature that was not captured by the model and is anything but random.
- So we will continue our quest and try the a second order polynomial.

Second Order Polynomial Design Matrix (X)

Model: $\hat{y} = b(1) + m_1 x + m_2 x^2$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ & \cdots & \end{bmatrix}$$

Setup: `>> X = [ones(length(x),1), x, x.^2]`

Fit Polynomial Equation

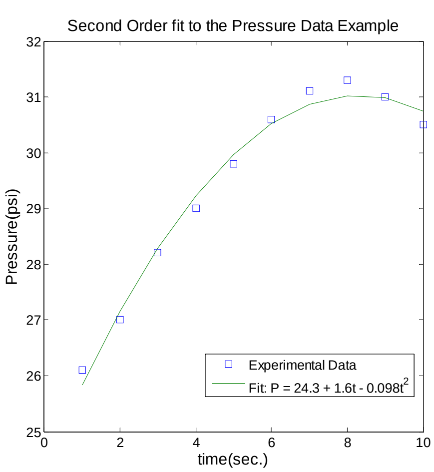
- **Model:** $\hat{y} = b(1) + m_1 x + m_2 x^2$
-
- **Setup** (this case): `>> X = [ones(length(t1),1), t1', (t1').^2]`
  `>> Y = P'`
- **Fit:** `>> B = X \ Y`
- **Predict:** `>> Yhat = X*B`
  + Residuals `>> res  = Y - X*B`
- **Plot:**
  Scatter Plot: `>> plot( X(:,2),Y, 'p', X(:,2), Yhat)`
  + labels
  Residual Plot: `>> plot(Yhat, res, 'p'. Yhat, res*0)`
  + labels

Scatter Plot – Second order model for pressure data:

```
>> % Plot and label graph
>> plot(t, P, 's', t1', Yhat)
>> xlabel 'time(sec.)'
>> ylabel 'Pressure(psi)'
>> legend('Experimental Data',...
   'Fit: P = 24.3 + 1.6t - 0.098t^2')
>> title 'Second Order fit to the
   Pressure Data Example'
```

Model is:

$P = 24.3 + 1.6t - 0.098t^2$

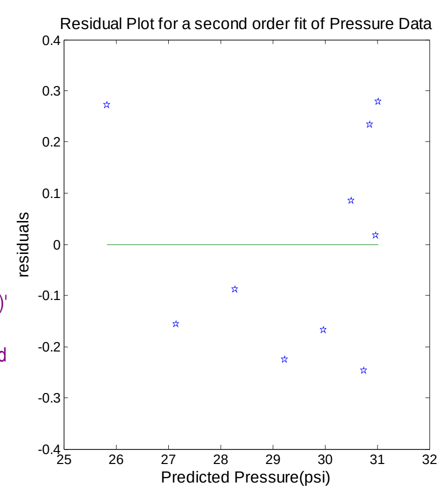


Residual Plot – Second order model for pressure data

```
>> % plot residuals
>> plot(Yhat, res, 'p', Yhat, res*0)

>> % Balance y axis
>> ylim([-0.4 0.4])

>> % add appropriate labels
>> xlabel 'Predicted Pressure(psi)'
>> ylabel 'residuals'
>> title 'Residual Plot for a second
   order fit of Pressure Data'
```



Notes on 2nd order fit to pressure data

- Scatter Plot: This plot looks quite a bit better. The second order model does capture the maximum.
- Residual Plot: The residual plot is much better and almost acceptable but still does show some signs of curvature.
- Notice that the only positive residuals are at low and high values and none in the middle.
- On the positive side notice that the range of the y-axis has been reduced from 3 to 0.8.

Third Order Polynomial Design Matrix (X)

Model: $\hat{y} = b(1) + m_1 x + m_2 x^2 + m_3 x^3$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ & & \dots & \end{bmatrix}$$

Setup: >> X = [ones(length(x),1), x, x.^2, x.^3]

or >> X = [X, x.^3]

Fitting 3rd Order Polynomial Equation – Exactly the same as second order case

Fit: >> B = X \ Y

Predict: >> Yhat=X*B

+ Residuals >> res = Y - X*B

Plot:

Scatter Plot: + labels >> plot(Y, X(:,2), 'p', Yhat, X(:,2))

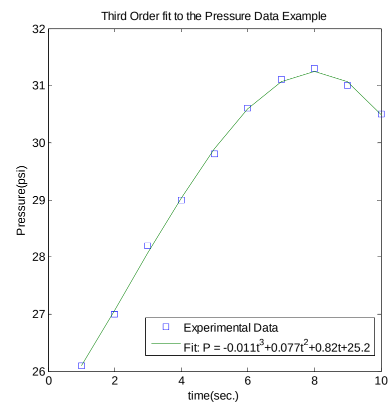
Residual Plot: + labels >> plot(Yhat, res, 'p'. Yhat, res*0)

Scatter Plot – Third order model for pressure data:

```
>> % Plot and label graph
>> plot(t, P, 's', t, Phat)
>> xlabel 'time(sec.)'
>> ylabel 'Pressure(psi)'
>> legend('Experimental Data', 'Fit: P =
   -0.011t^3+0.077t^2+0.82t+25.21 ')
>> title 'Third Order fit to the
   Pressure Data Example'
```

Model is:

P = 24.3 + 1.6t - 0.098t²

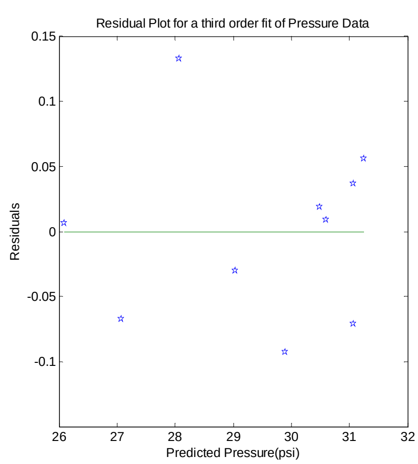


Residual Plot – Third order model for pressure data:

```
>> % Calc residuals & plot
>> res = Y - Yhat;
>> plot(Yhat, res, 'p', Yhat, res*0)

>> % Balance y axis
>> ylim([-0.4 0.4])

>> % add appropriate labels
>> xlabel 'Predicted Pressure(psi)'
>> ylabel 'residuals'
>> title 'Residual Plot for a second
   order fit of Pressure Data'
```



Notes on third order fit to pressure data

- Scatter Plot: The scatter plot is now looking pretty good.
- Residual Plot: The residual plot is also looking pretty good. There are no clear signs of curvature and the range of the y-axis has been reduced from 0.8 to 0.3
- We could look at fourth order just to be sure. However, fourth order rarely contributes much to a polynomial regression (x^4 is too similar to x^2).

Fitting a Polynomial

- Same steps as for linear model (Model, Setup, Fit, Predict, Plot)
- Polynomial model is $\hat{y} = b(1) + m_1 x + m_2 x^2 + m_3 x^3 + \cdots$
- The Pressure (Y) data must a column vector
- The X matrix therefore must contain:
  - A column of ones (constant term)
  - A column of the x data
  - A column of the x data squared
  - A column of the x data cubed

  …
- The Resulting parameter vector is: $B = \begin{bmatrix} b \\ m_1 \\ m_2 \\ m_3 \\ \cdots \end{bmatrix}$

Thermocouple Calibration Data ➔ is it linear?

mV (mV)	T(°C)
0	0
0.3910	10.0000
0.7900	20.0000
1.1960	30.0000
1.6120	40.0000
2.0360	50.0000
2.4680	60.0000
2.9090	70.0000
3.3580	80.0000
3.8140	90.0000
4.2790	100.0000

- Plot this data
  Does it look linear?
- Fit a linear model
- Determine the residuals
  Prepare a residuals plot
- Is it linear?

Model: Start with Linear Model - Plot data to see if it is linear

```
>> plot(mV, T, 'p')
>> ylabel 'Temperature(C)', xlabel 'Voltage(mV)'
```

Plot looks linear so try model y = b + m x

Setup: `>> X = [ones(length(mV),1), mV]`

Fit: `>> B = X \ T`

```
B =
        1.4875
       23.3509
```

Untransform: not necessary for a linear model

Predict: `>> That = X*B`

Plot: `>> plot(mV, T, 'p', mV, That)`

plus labels and legend as we have done before

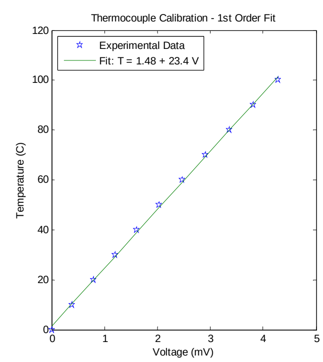
Residuals:

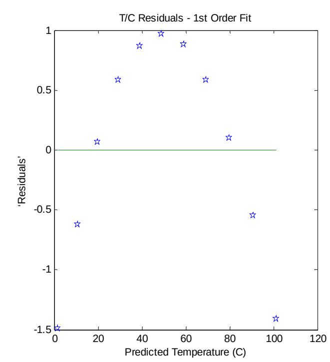
```
>> Res = T – That
>> plot (That, Res, 'p', That, Res*0)
>> ylabel 'Residuals'
>> xlabel 'Predicted Temperature (C)'
```

Below is the resulting first order fit

Residuals plot shows clear curvature

Try a second order model





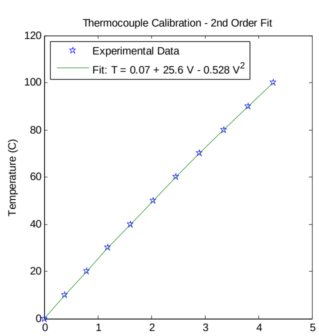
Model: $T = b + m_1(mV) + m_2(mV^2)$

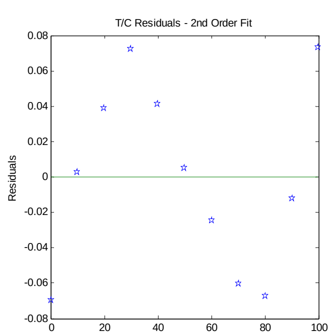
Setup : `>> X = [ones(length(mV),1), mV, mV.^2];`

Fit: `>> B = X \ T`

```
B =
   0.0696
  25.5964
  -0.5281
```

Predict & Plot: as before





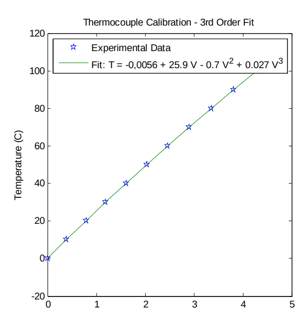
Residuals have been reduced from ± 1.5 to ± 0.08 but still show curvature at this low level – try a third order fit

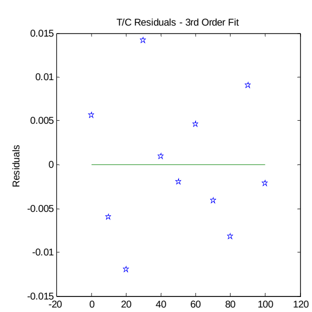
Model: $T = b + m_1(mV) + m_2(mV^2) + m_3(mV^3)$

Setup : `>> X = [ones(length(mV),1), mV, mV.^2, mV.^3];`

Fit: `>> B = X \ T`

```
B =
   -0.0056
   25.8744
   -0.6991
    0.0267
```





Residuals have been reduced from ± 0.08 to ± 0.015 and show no patterns ➔ this looks like a good model