Modeling

Just knowing that there is an effect or having some idea as to its source may not be very useful.

We want to be able to:
- make predictions
- estimate the effects of levels not studied

We must develop a model.

Types of models

Theoretical (parametric) model

- Data follows from a theory.
- Uses known theoretical laws or principles.
- Example - Calibration curve of absorbance vs. concentration.
- While the model may not be linear, we typically attempt to transform it to make it linear.

Types of models

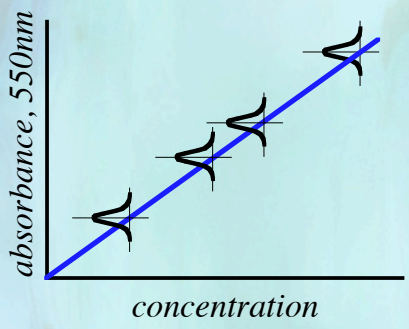
Empirical (nonparametric) model.
- No theoretical basis for the model.
- An equation is simply fit to the data.

You must assume that the trend continues outside the experimental ranges.

Example - polynomial fit of a GC trace.

Example

Linear fit of a theoretical relationship.



Linear models, $Y = b_1X$

If X is known and we control it, then $\sigma^2_X = 0$ and a b_1 (slope) is found that minimizes:

$$\sum \frac{\left(Y_i - \hat{Y}\right)^2}{\sigma_y^2} \qquad \hat{Y} = predicted$$

We simply want to minimize the difference the measured and predicted values.

Linear model

We're just trying to minimize the standard error (SE) for our model.

$$SE = \sum \frac{\left(Y_i - \hat{Y}\right)^2}{\sigma_y^2} = \sum \frac{1}{\sigma_y^2}\left(Y_i - b_i X_i\right)^2$$

The optimum value for b1 is the one that minimizes the standard error.

We also must assume that the values are homoscedastic: (unweighted least squares)

$$\sigma_{y1}^2 = \sigma_{y2}^2 = \ldots = \sigma_{yn}^2$$

Homoscedatic data

homoscedastic

not homoscedastic

Linear model

To find the value of b_1 that minimizes the standard error, we set:

$$\frac{\partial SE}{\partial b_1} = 0$$

$$\Rightarrow b_1 \sum x_i^2 = \sum x_i y_i$$

Solving for b_1 gives $b_1 = \frac{\sum x_i y_i}{\sum x_i^2}$

Linear model

Weighted least squares

Used when the data is not homoscedastic.

$$\left(\sigma_{y1}^2 \neq \sigma_{y2}^2 \neq \ldots. \neq \sigma_{yn}^2 \neq\right)$$

$$b_1 = \frac{\sum \frac{x_i y_i}{\sigma_{y1}^2}}{\sum \frac{x_i^2}{\sigma_{y1}^2}}$$

Each measurement is weighted by the reciprocal of its variance.

Example

ppm	abs	xy	x^2
1	0.020	0.020	1
2	0.038	0.076	4
3	0.064	0.192	9
4	0.077	0.308	16
5	0.105	0.525	25
	sums	1.121	55

$$b_1 = \frac{\sum x_i y_i}{\sum x_i^2} = 0.0204 \; so \; Y = 0.0204 \; X$$

Y-intercept

- What if your data does not go through the origin?
- Y-intercept = $\bar{Y}$ - b_1 $\bar{X}$
- You simply need to calculate the mean for both X and Y and use the above equation.

How good is the fit?

An ANOVA can can be conducted to determine the 'goodness of fit' for the model.

Source of Variance	Sum of Squares	DF
Model	$\sum \left(\bar{y} - \hat{y}\right)^2$	p
Residual	$\sum \left(y_i - \hat{y}\right)^2$	n-p-1
Total	$\sum \left(y_i - \bar{y}\right)^2$	n-1

p = number of parameters.
n = number of data points.

ANOVA of our earlier example

Calculate absorbances based on $Y = b_1X$

$b_1 = 0.0204 \quad \bar{Y} = 0.0608$

ppm	abs_{exp}	abs_{calc}	model	residual
1	0.020	0.020	0.0402	0.0004
2	0.038	0.0408	0.0198	0.0028
3	0.064	0.0612	-0.0006	-0.0028
4	0.077	0.0816	-0.021	0.0046
5	0.105	0.102	-0.003	-0.0030
		$\sum(\bar{y} - \hat{y})^2$	0.00416	0.000046

ANOVA of our earlier example

Source	Sum of Squares	DF	Mean Square (SS/DF)
Model	0.00416	1	0.00416
Residual	0.000046	5-1-1	0.0000153
Total	0.00441	5-1	0.001103

F test will show that the model accounts for a significant amount of the variance compared to the residual. F = 272

Correlation coefficient

The ratio of model/total variance is commonly referred to as the correlation coefficient.

$$r = \sqrt{\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}} = b_1\frac{\sigma_x}{\sigma_y}$$

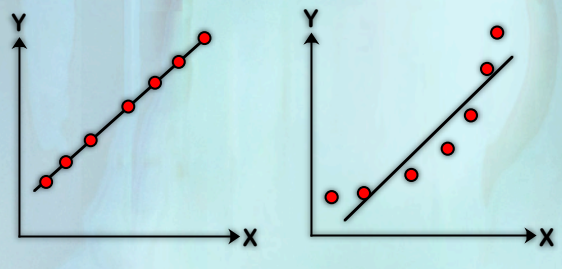
This value is commonly reported by programs and calculators that can conduct linear regression fits.

Goodness of fit.

The ratio of model/total variance tells us how much variance the model accounts for (correlation coefficient,r). The ratio of model to residual mean squares would be the 'F' value.

- In the last example, r was 0.943 (88.9%) which is a pretty good fit.
- If the model/total ratio is < 0.8, then you should consider looking at a different model.
- Graphing the data is an excellent way to see the nature of your relationship.

Goodness of fit

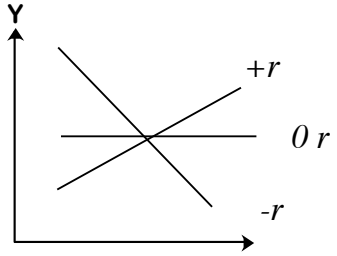


Both examples show a linear regression fit of a data set. The one on the right indicates that the 'best' model may not be a linear one.

Correlation coefficient

Values range between -1 to +1 where

- -1 indicates perfect correlation with a negative slope.
- +1 indicates perfect correlation with a positive slope.
- 0 no correlation - this is rare, even a 'bad' fit will have at least some correlation.



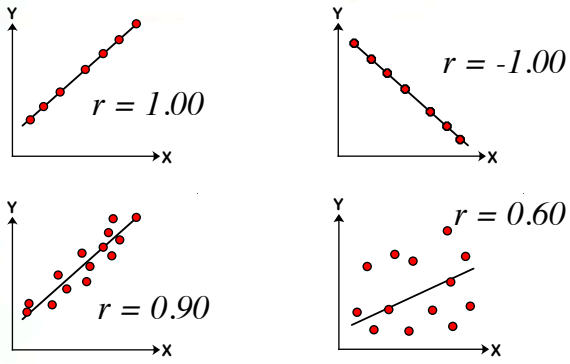
Correlation coefficient

Relationship between correlation coefficient (r) and proportion of variance (r^2).

Correlation, |r|	100 r^2
0.10	1
0.20	4
0.50	25
0.80	64
0.90	81
0.95	90
0.99	99
1.00	100

|r| values below 0.9 indicate a poor relationship.

Correlation coefficient

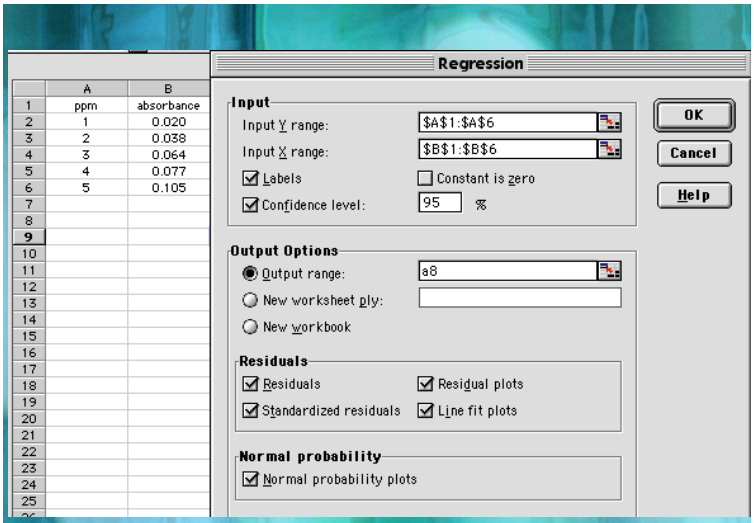


The same thing with Excel

It should come as no surprise that we can do the same calculations using Excel.

Two approaches

- Data analysis add-on
  Detailed analysis
- Trendlines
  Quick and dirty regression lines when producing a graph.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.995147836
R Square	0.990319216
Adjusted R Squa	0.987092289
Standard Error	0.179636518
Observations	5

ANOVA

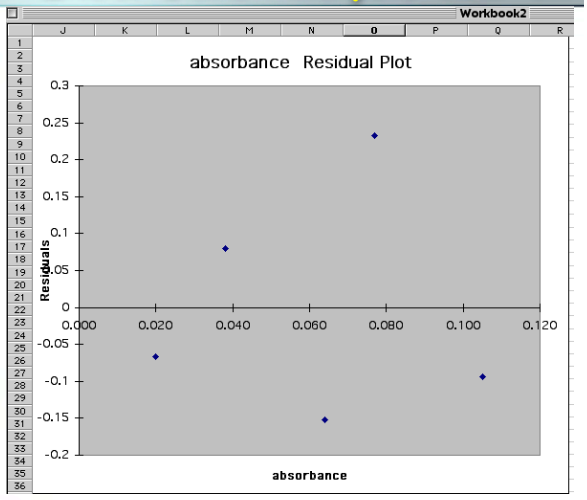
	df	SS	MS	F	Significance F
Regression	1	9.903192165	9.903192165	306.8922717	0.000405434
Residual	3	0.096807835	0.032269278		
Total	4	10			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.11907137	0.183025576	0.650572303	0.561704977	-0.46339824	0.701540984	-0.46339824	0.701540984
absorbance	47.38369457	2.704804898	17.51834101	0.000405434	38.77579014	55.991599	38.77579014	55.991599

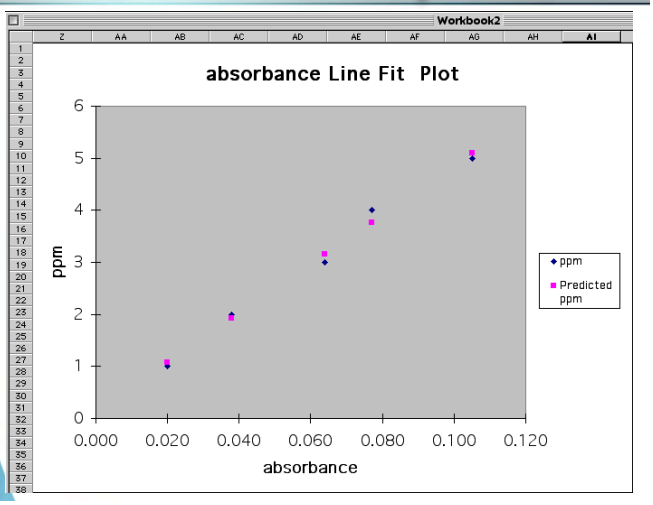
RESIDUAL OUTPUT

Observation	Predicted ppm	Residuals	tandard Residuals
1	1.066745262	-0.06674526	-0.42903743
2	1.919651764	0.080348236	0.516477122
3	3.151627823	-0.15162782	-0.97466111
4	3.767615852	0.232384148	1.493761427
5	5.0943593	-0.0943593	-0.60654001

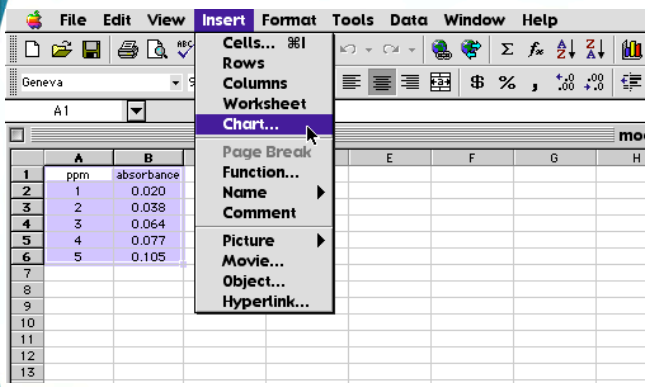
Excel Data Analysis add-on



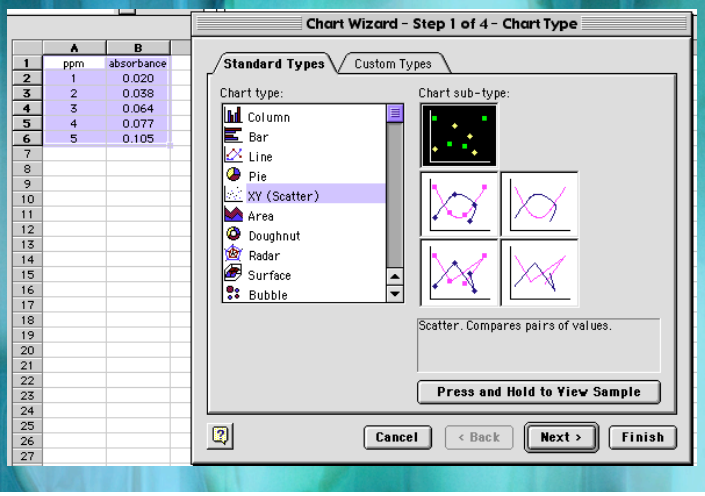
Excel Data Analysis add-on



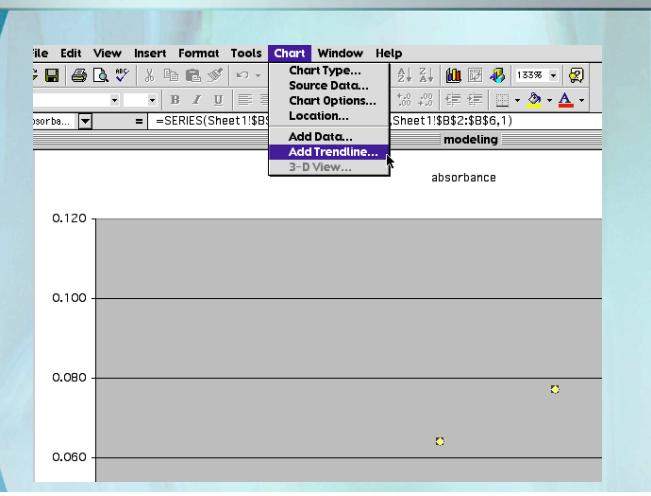
Excel trendline



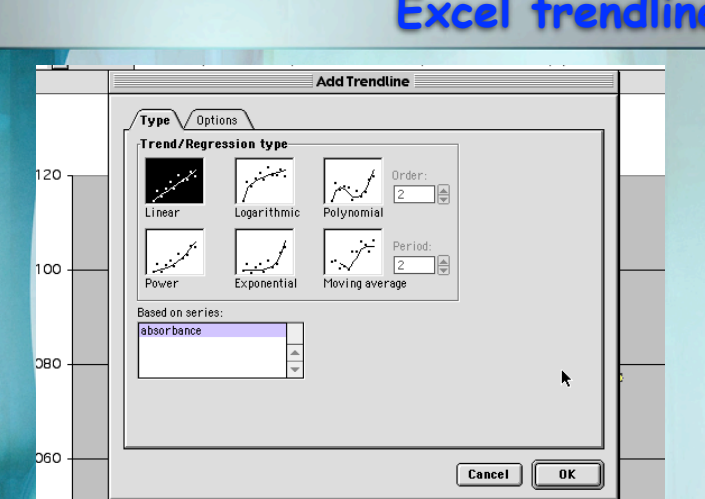
Excel trendline



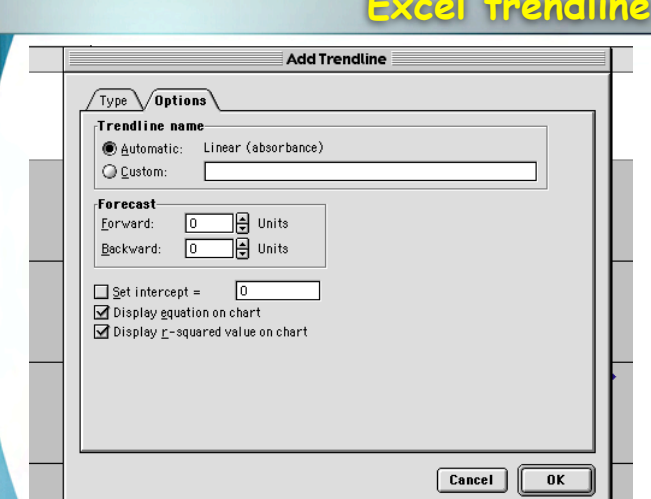
Excel trendline



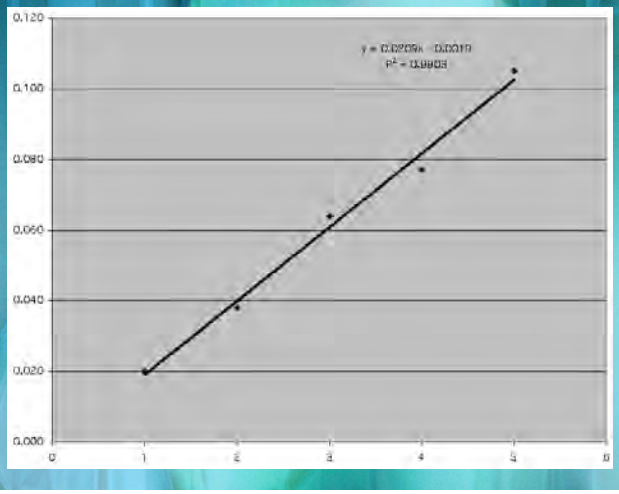
Excel trendline



Excel trendline



Excel trendline



Using XLStat

- As one would expect, XLStat is also able to do linear regression.
- It provides the same information and then some.

ppm	absorbance
1	0.020
2	0.038
3	0.064
4	0.077
5	0.105

Analysis of variance:

Source	DF	SS	Mean sqr	F	Pr > F
Model	1	0.004	0.004	306.892	0.000
Error	3	0.000	0.000		
Corrected Tot	4	0.004			

Computed against model Y=Mean(Y)

Type I Sum of Squares analysis:

Source	DF	SS	Mean sqr	F	Pr > F
ppm	1	0.004	0.004	306.892	0.000

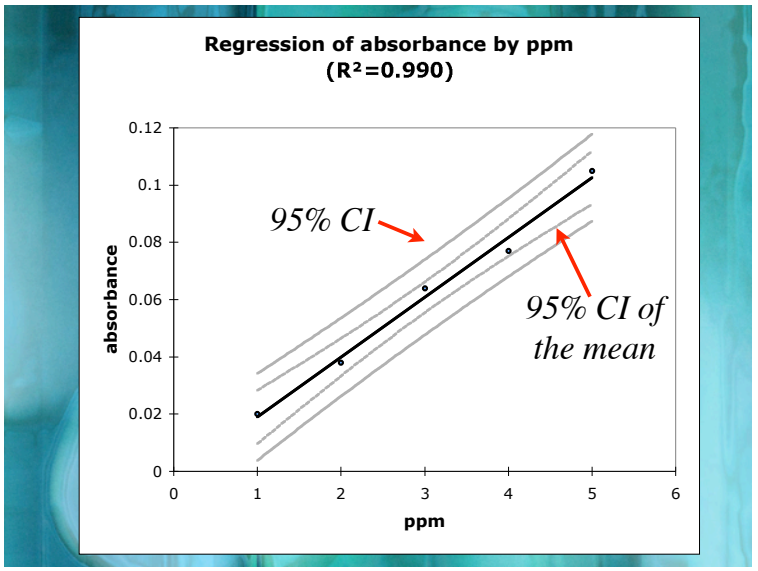
Type III Sum of Squares analysis:

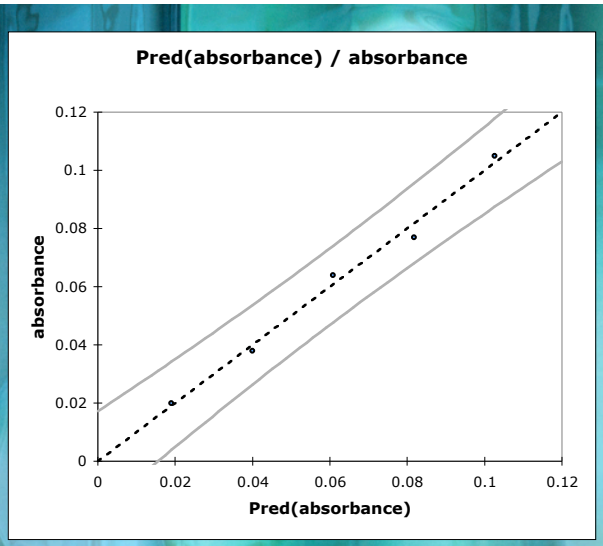
Source	DF	SS	Mean sqr	F	Pr > F
ppm	1	0.004	0.004	306.892	0.000

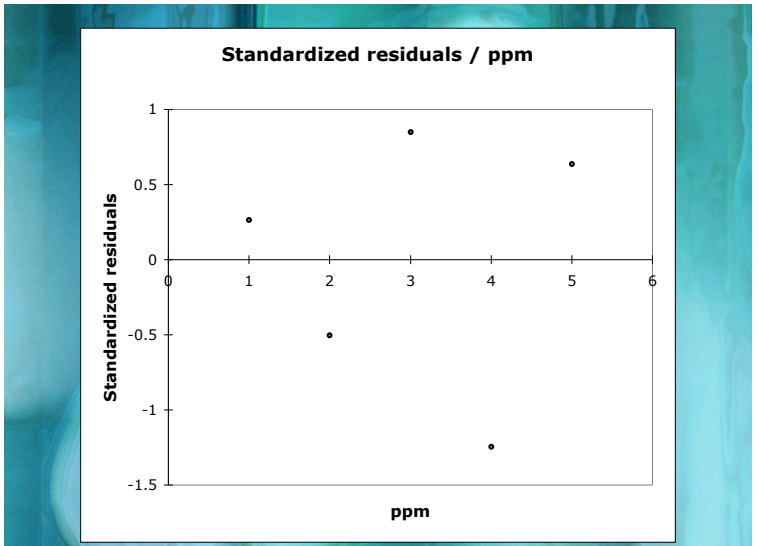
Model parameters:

Source	Value	Std error	t	Pr > |t|	Lower (95%)	Upper (95%)
Intercept	-0.002	0.004	-0.480	0.664	-0.014	0.011
ppm	0.021	0.001	17.518	0.000	0.017	0.025

Equation of the model:
absorbance = -0.0019+0.0209*ppm







Data transformations

- In some cases, it is best to do a simple transformation of your data prior to attempting a linear regression fit.
- The goal of the transformation is to make the resulting relationship linear.
- An ANOVA analysis can be conducted on the transformed data.

Data transformations

Transform		Equation of the line
Y	X	$Y = b X + a$
Y	1/X	$Y = a + b / X$
1/Y	X	$Y = 1 / (a + b X)$
X/Y	X	$Y = X / (a + b X)$
log Y	X	$Y = a b^X$
log Y	log X	$Y = a X^b$
Y	X^n	$Y = a + b X^n$

b = slope, a = intercept.

Multiple linear regression

MLR assumes a linear relationship between X_i and y, with superimposed noise (e). It also assumes that there are no interactions between $X_1, X_2, \ldots X_n$.

$Y = b_0 + b_1X_1 + b_2X_2 + \ldots . b_nX_n + e$

We then 'fit' a regression equation. The bn values are considered to be estimates of the true population parameters, βn.

Multiple linear regression

For a two variable model:

$$Y = b_1 X_1 + b_2 X_2$$

As in the one variable model, we end up with:

$$b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum Y X_1$$
$$b_2 \sum X_2^2 + b_1 \sum X_1 X_2 = \sum Y X_2$$

We can then solve these two 'normal equations.' It becomes more difficult as additional variables are added in.

Multiple linear regression

- Results must be interpreted more carefully that with simple regression.
- All of the b values are now tied together and must be interpreted as a group.
- R^2 values will increase as you add additional X values to the model - even random numbers.
- Use adjusted R^2 values to see if new predictors improved model.

Multiple linear regression

$$R^2 = \frac{SS_{Total} - SS_{residual}}{SS_{Total}}$$

$$adjusted\ R^2 = \frac{MS_{Total} - MS_{residual}}{MS_{Total}}$$

R^2 simply looks a how well all of the X account for Y. The adjusted R^2 is weighted by the number of X used.

Multiple linear regression

One advantage of an MLR approach is that you can obtain multiple measurements (X) for a single response. This can help eliminate noise.

We'll look at one example using MLR - Determination of Octane Number by NIR.

We'll revisit this example in a later unit when we compare it to other multivariate calibration methods.

Octane number

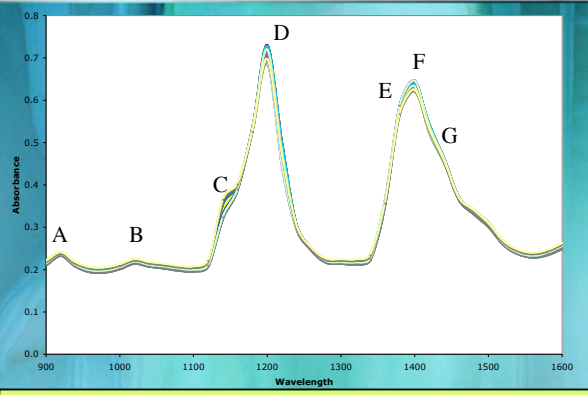
Rating Octane of Gasoline using near-IR.
ASTM method is complex and expensive.
A simple spectral method would be more desirable.

Experimental
A series of unleaded gasoline samples were assayed by the ASTM method.

NIR spectra (900-1600 nm) were also obtained.

A matrix was constructed from the spectra (20 nm intervals) were the X matrix and the ASTM octane number was the Y matrix.

Octane Number, NIR spectra

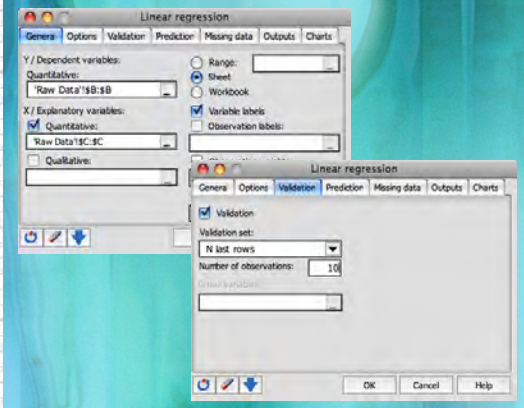


A 915 nm, CH_2 stretch
B 1021 nm, CH_2/CH_3 combination band
C 1151 nm, aromatic and CH_3 stretch
D 1194 nm, CH_3 stretch
E 1394 nm, CH_2 combination bands
F 1412 nm aromatic & CH_2 combination bands
G 1435 nm aromatic & CH2 combination bands

Octane Number, XLStat MLR

Sample #	Octane #	900
1	87.3	0.2146
2	87.0	0.2198
3	87.1	0.2164
4	89.7	0.2201
5	84.9	0.2138
6	84.7	0.2143
7	89.3	0.2203
8	87.6	0.2167
9	84.5	0.2139
10	91.7	0.2218
11	87.1	0.2151
12	87.9	0.2214
13	83.7	0.2161
14	83.7	0.2149
15	87.5	0.2189
16	91.7	0.2230
17	91.2	0.2209
18	89.4	0.2205
19	91.7	0.2320
20	84.0	0.2159
21	87.1	0.2220
22	87.3	0.2183
23	83.6	0.2155
24	87.1	0.2207
25	91.7	0.2316
26	84.2	0.2157
27	92.5	0.2267
28	87.5	0.2198
29	91.4	0.2316
30	87.1	0.2202

Single variable - 900 - just for comparison.



Octane Number, MLR

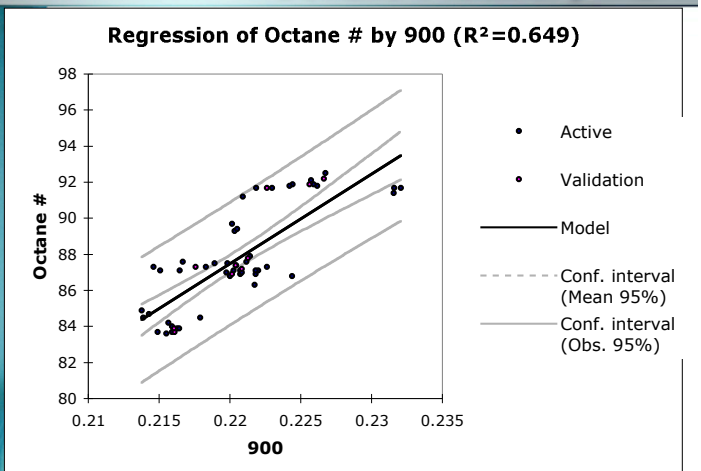
Regression of variable Octane #:

Goodness of fit statistics:	
Observations	49.000
Sum of weight	49.000
DF	47.000
R^2	0.649
Adjusted R^2	0.642

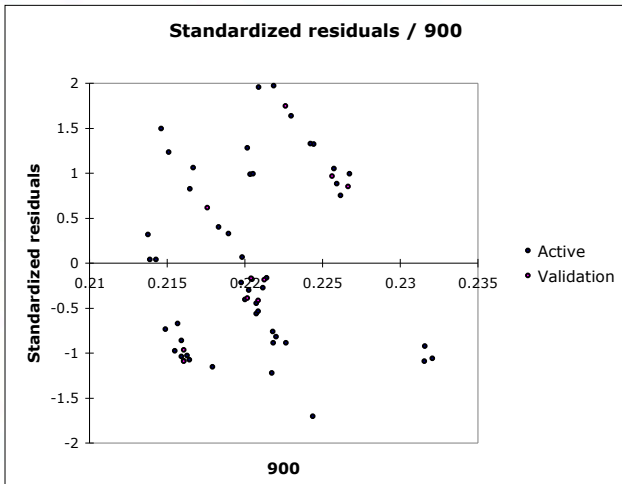
Analysis of variance:

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	1	244.297	244.297	86.953	< 0.0001
Error	47	132.048	2.810		
Corrected Tot	48	376.345			

Octane Number, MLR



Octane Number, MLR



Octane Number, MLR

Using all values results in a significant improvement in the fit..

Regression of variable Octane #:

Goodness of fit statistics:

Observations	49.000
Sum of weight	49.000
DF	39.000
R^2	0.995
Adjusted R^2	0.994

Analysis of variance:

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	9	374.482	41.609	871.305	< 0.0001
Error	39	1.862	0.048		
Corrected Tot	48	376.345			

Octane Number, MLR

Sum of squares analysis shows which lines are the most significant for the fit.

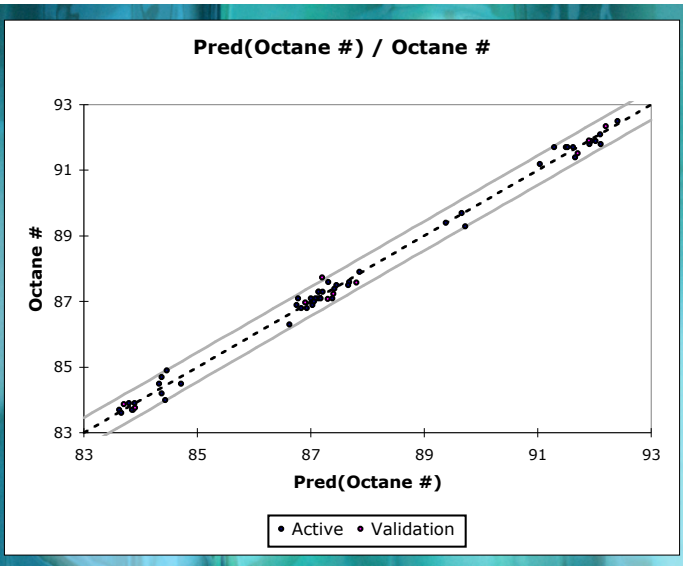
Type III Sum of Squares analysis:

Source	DF	Sum of squares	Mean squares	F	Pr > F
1080	1	0.002	0.002	0.034	0.855
960	1	0.002	0.002	0.037	0.849
1320	1	0.011	0.011	0.227	0.636
920	1	0.018	0.018	0.386	0.538
980	1	0.092	0.092	1.931	0.173
900	1	0.309	0.309	6.470	0.015
1060	1	0.445	0.445	9.317	0.004
940	1	0.540	0.540	11.316	0.002
1000	1	0.747	0.747	15.649	0.000

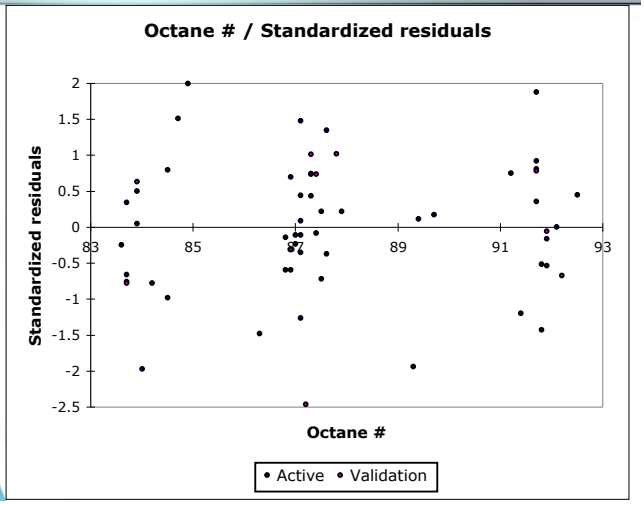
What happened to the other variables?

Multicollinearity

- A common problem with multiple linear regression.
- Example - reporting both the pH and pOH of a system.
- You need to eliminate redundant information or it will skew your results.
- A common approach is to eliminate all but one variable with similar correlations.
- XLStat automatically does this.



Octane Number, MLR



Using the 'best' 4 lines

Goodness of fit statistics:

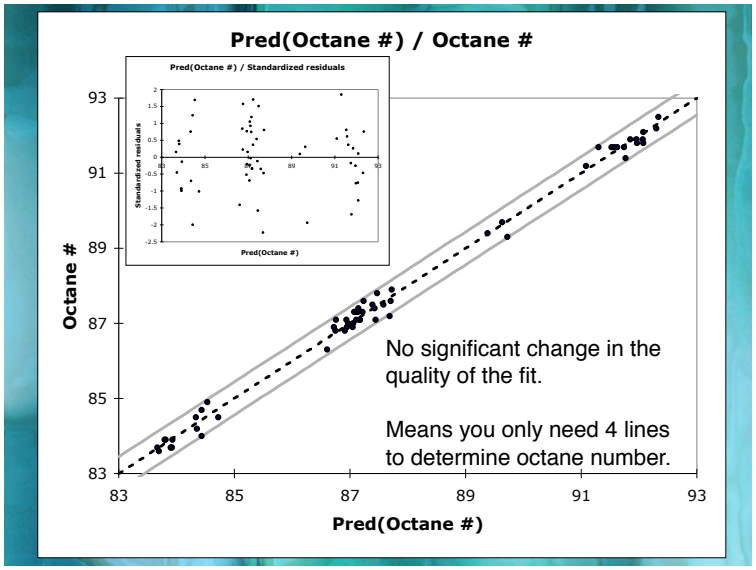
Observations	59.000
Sum of weight	59.000
R^2	0.994
Adjusted R^2	0.994

Analysis of variance:

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	4	459.183	114.796	2439.792	< 0.0001
Error	54	2.541	0.047		
Corrected Tot	58	461.724			

Type III Sum of Squares analysis:

Source	DF	Sum of squares	Mean squares	F	Pr > F
900	1	0.794	0.794	16.877	0.000
940	1	3.532	3.532	75.064	< 0.0001
1000	1	2.072	2.072	44.045	< 0.0001
1060	1	0.673	0.673	14.304	0.000



ANCOVA

XLStat does a good job of switching to the proper type of model based on the type of data.

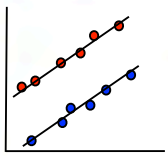
Number of X variables	Qual	Quant	Mixed
1	Simple ANOVA	LR	-
2 or more	ANOCA	MLR	ANCOVA

ANCOVA

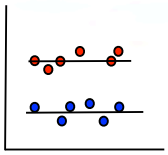
You can use the method to tell:

- If the qualitative variables are significant.
- If one gets the same basic model (slope) for the quantitative variables.
- If you can build a model that can account for both types of factors (when significant.)

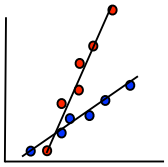
ANCOVA examples



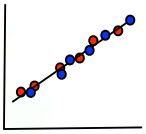
Both treatment and covariate variables are significant with the same model.



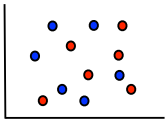
Treatment is significant but covariate isn't.



Both treatment and covariate variables are significant but the models are different.



Covariate variable is significant, treatment is not.



Neither are significant.

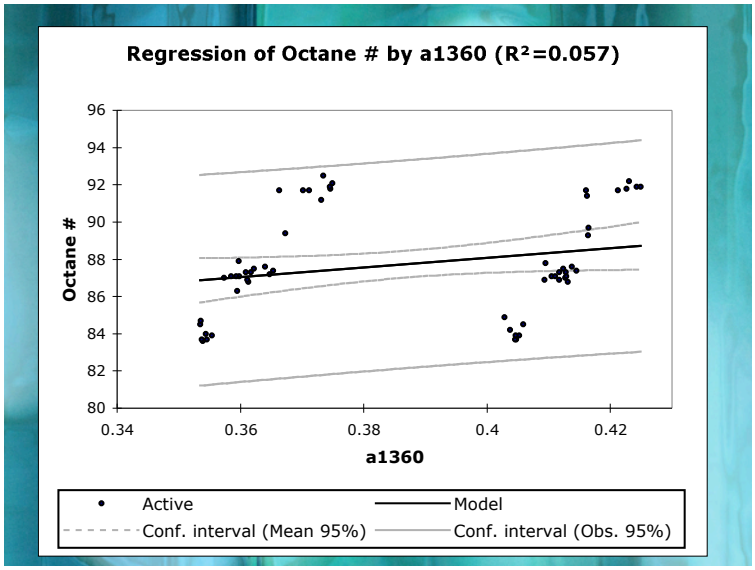
Modified Octane example

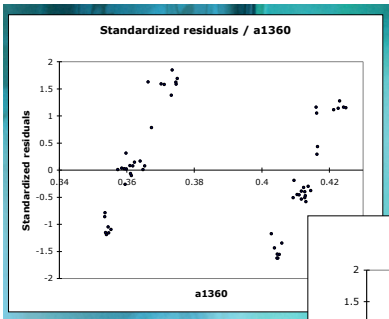
- The near IR spectra can vary based not only on octane number.
- The presence of oxygenates can cause changes.
- Seasonal blends can also cause changes.
- ANCOVA can be used to deal with this type of situation.

Blend	Octane #	a1360
Summer	83.6	0.35
Summer	83.7	0.35
Summer	83.7	0.35
Winter	83.7	0.40
Winter	83.7	0.40
Summer	83.9	0.36
Winter	83.9	0.41
Winter	83.9	0.40
Summer	84.0	0.35
Winter	84.2	0.40
Summer	84.5	0.35
Winter	84.5	0.41
Summer	84.7	0.35
Winter	84.9	0.40
Summer	86.3	0.36
Summer	86.8	0.36
Winter	86.8	0.41
Summer	86.9	0.36
Winter	86.9	0.41
Winter	86.9	0.41
Summer	87.0	0.36
Winter	87.0	0.41
Summer	87.1	0.36
Summer	87.1	0.36
Summer	87.1	0.36
Winter	87.1	0.41
Winter	87.1	0.41
Winter	87.1	0.41
Summer	87.2	0.36
Summer	87.3	0.36
Summer	87.3	0.36
Winter	87.3	0.41
Winter	87.3	0.41
Summer	87.4	0.37

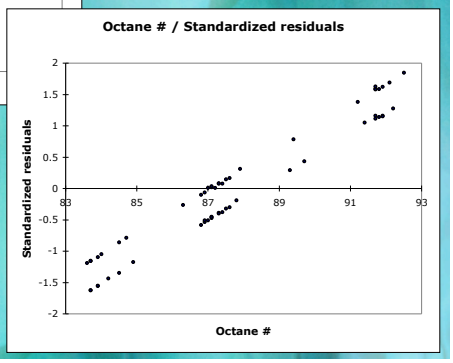
For simplicity, we'll look at a single near IR region. It has one of the highest correlations with octane number of those evaluated earlier.

Start with a simple linear regression analysis, ignoring the fact there are both summer and winter blends included.

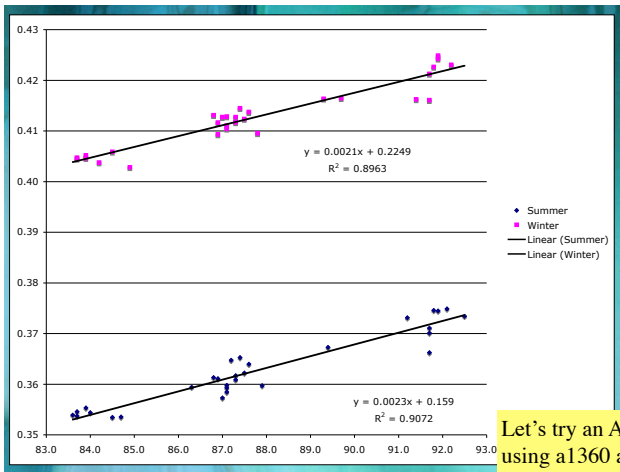




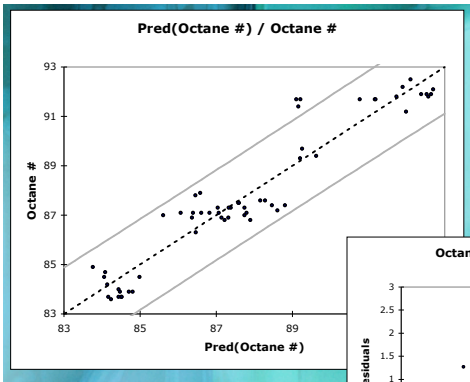
It seems pretty clear that a simple linear regression has a problem. It is also obvious that there are two types of samples.

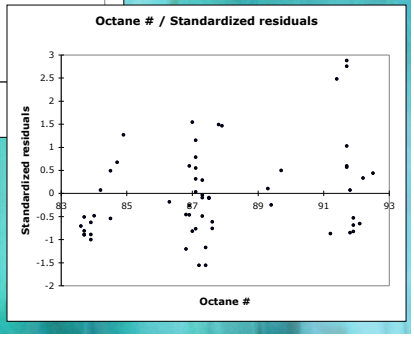


Evaluating a simple scatterplot may help.



Let's try an ANCOVA - using a1360 as the covariate and the 'blend' type as an additional factor.





Non-linear regression

- With many problems, it is not always possible to set up a simple linear regression solution, even with data transformation.
- Non-linear regression methods permit the user to fit a set of parameters to a model which can involve many variables.
- This type of problem is typically solved using a computer program.

Non-linear regression

The approach used to fit a model will vary based on the program used and the options chosen.

We'll give an overview as to the
general goal
typical user options
potential problems

Non-linear regression

The goal of any non-linear least squares regression fit is to minimize the error between experimental and modeled values.

$$\min \Sigma \left(y_i - f(x) \right)^2$$

where f(x) is the function to be fit (Y = f(x))

It is assumed that the function contains one or more adjustable parameters.

Non-linear regression

Example function. $y_i = e^{X_i Z}$

The goal is to find. $\min \sum \left(y_i - e^{X_i Z} \right)^2$

The adjustable parameter is Z.

x	y
1	2
2	4
3	8

Non-linear regression

Brute force approach.

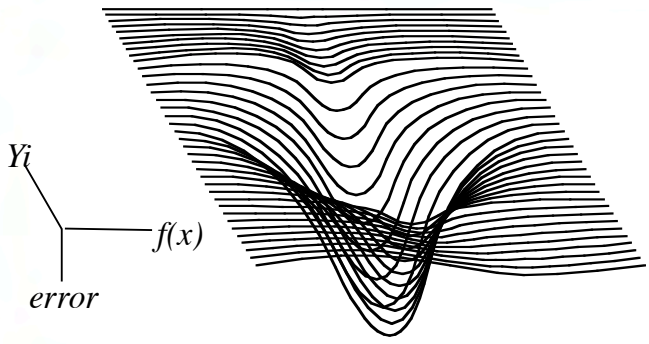
- Since this is a simple example, you could just set up a simple program to test a range of Z values.
- Given enough time, such a search would determine that the optimum solution for Z.
- With more complex models (more parameters to fit), this approach becomes difficult to do.

Non-linear regression

Response surface.

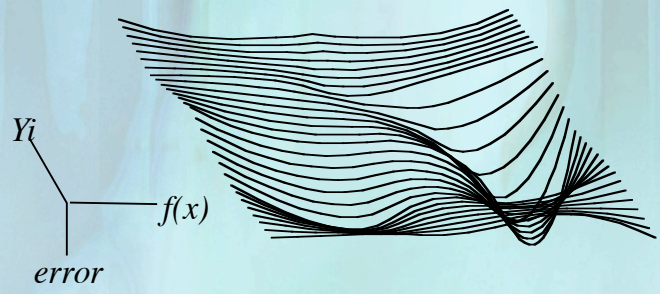
- As a program attempts to find an optimum solution, it is evaluating potential solutions, looking for an error minimum. This results in an N dimensional 'surface' being produced of possible solutions.
- Non-linear regression fitting programs evaluate the best direction for subsequent estimates based on changes in this surface.

Response surface



For our simple model, the program will attempt to find this minimum value.

Response surface



This can be more difficult as the number of adjustable parameters increases. There can be several false minima that must be avoided.

Non-linear regression options

Most programs include a range of options you can select or modify.

This not only permits you to

- speed up processing time
- avoid false convergence.
- control the tolerance of your fit
- . . .

Non-linear regression options

Typical options

Initial estimate of parameters

An initial guess of your values can not only save processing time but avoid false convergence as well.

Scaling of parameters

The magnitude of your parameters can vary greatly. Scaling will give them comparable 'weight' during the fit.

Non-linear regression options

First Derivative Function

Some programs require that you provide the first derivative (Jacobian) for your function. It is used to determine the best direction to try it's next estimate.

Tolerance

Various types including how big a jump it can make to the next estimate, how good the numbers are, ...

Non-linear regression options

Maximum iterations

How many 'guesses' it should attempt before quitting.

Limits

Do you want to set any upper/lower limits for the parameters?

Convergence

How good does the estimate need to be.

Non-linear regression

Each program has its own approach(s) as it attempts to find the optimum solution.

Your best bet is to:

Try several different programs if possible.
Test each program using the various options available.
Try different rearrangements of your model to see what effect it has.

Things are usually OK if the changes above still result is the same basic solution

Actual model

For this model, the goal is to determine the minimum error solution for the following:

$$\left(\frac{V_M}{RT}h+\chi_s\right)\phi_f^2+\phi_f+\ln(1-\phi_f)=0$$

$RT = 2479.05$

$\chi_s = 0.34$

V_M = *molar volume of solvent*

ϕ_f = *polymer volume fraction*

$$h = (D_p-D_s)^2+0.25\left[(P_p-P_s)^2+(H_p-H_s)^2\right]$$

D_s, P_s, H_s – known constants

D_p, P_p, H_p – *parameters to be fit*

Using Excel's 'Solver'

- The Solver is a built in non-linear least squares function for Excel.
- Parameters are held in specific cells that it will alter to find a solution.
- It tests by seeking a minimize or maximize value held in another 'cell'
- All intermediates must be also be cell functions.
- Several options are available.

Experimental data

	Solvent	MW	Vm	D	P	H	Gain
7	Acetonitrile	41.1	52.6	15.3	18.0	6.1	0.463
8	Acrylonitrile	53.1	67.1	16.4	17.4	6.8	0.617
9	Benzaldehyde	106.1	101.5	19.4	7.4	5.3	0.222
10	Benzene	78.1	89.4	18.4	0.0	2.0	0.117
11	Butanol, 1-	74.1	91.5	16.0	5.7	15.7	0.027
12	Butanone, 2-	72.1	90.1	16.0	9.0	5.1	3.312
13	Butyl Acetate	116.1	133.5	15.1	3.7	6.3	1.571
14	CCl4	153.8	97.1	17.8	0.0	0.6	0.158
15	Chloroform	119.4	80.7	17.8	3.1	5.7	0.207
16	Cyclohexane	84.2	108.7	16.8	0.0	0.2	0.031
17	Dichloroethane, 1,	99.0	79.4	19.0	7.4	4.1	0.116
18	Dichloromethane	84.9	63.9	18.2	6.3	6.1	0.230
19	Dimethylformamide	73.1	77	17.4	13.7	11.3	1.152
20	Dimethylsulfoxide	78.1	71.3	18.4	16.4	10.2	0.540
21	Dioxane, 1,4-	88.1	85.7	19.0	1.8	7.4	0.788
22	Ethanol	46.1	58.5	15.8	8.8	19.4	0.032
23	Ethyl Acetate	88.1	98.5	15.8	5.3	7.2	2.949
24	Ethyl Ether	74.1	104.8	14.5	2.9	5.1	0.279
25	Furfural	96.1	83.2	18.6	14.9	5.1	0.207
26	Isooctane	114.2	166.1	14.3	0.0	0.0	0.015
27	Pentane	72.2	117.4	14.5	0.0	0.0	0.034
28	Propanol, 2-	60.1	76.8	15.8	6.1	16.4	0.027
29	Pyridine	79.1	80.9	19.0	8.8	5.9	0.487
30	Toluene	92.1	106.8	18.0	1.4	2.0	0.114
31	TCE	131.4	90.2	18.0	3.1	5.3	0.150
32	Xylene, o-	106.2	121.2	17.8	1.0	3.1	0.107

Vs	Vf	Hansen	A	A1	delta^2
0.593	0.471	41.9	0.273	0.166	0.011
0.780	0.404	29.2	0.184	0.113	0.005
0.212	0.714	7.6	0.331	0.537	0.042
0.134	0.797	43.6	1.215	0.799	0.173
0.033	0.941	28.7	1.239	1.882	0.413
4.138	0.113	22.3	0.015	0.007	0.000
1.806	0.226	37.4	0.121	0.030	0.008
0.100	0.841	51.9	1.681	1.000	0.463
0.140	0.791	21.4	0.648	0.774	0.016
0.040	0.930	59.3	2.543	1.735	0.653
0.093	0.850	11.4	0.509	1.046	0.289
0.173	0.753	10.5	0.346	0.646	0.090
1.214	0.303	10.6	0.062	0.058	0.000
0.492	0.518	12.8	0.190	0.211	0.000
0.766	0.408	19.5	0.169	0.116	0.003
0.040	0.930	40.1	1.111	1.723	0.375
3.297	0.138	25.1	0.026	0.011	0.000
0.394	0.573	48.9	0.789	0.278	0.262
0.179	0.747	14.0	0.451	0.627	0.031
0.022	0.960	82.5	5.412	2.270	9.874
0.056	0.905	80.3	3.388	1.445	3.775
0.034	0.939	31.7	1.165	1.854	0.476
0.498	0.515	5.6	0.138	0.209	0.005
0.133	0.800	38.5	1.277	0.808	0.221
0.103	0.837	21.4	0.785	0.978	0.037
0.122	0.813	37.0	1.419	0.862	0.310

Calculated values

$$\underbrace{\left(\frac{V_M}{RT}h+\chi_s\right)\phi_f^2+\phi_f}_{A}+\underbrace{\ln(1-\phi_f)}_{A1}=0$$

We will be minimizing by tracking the sum of squares - that is what is held in the delta2 column

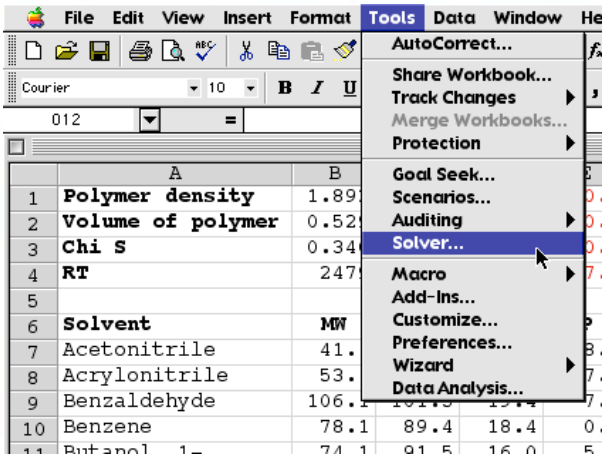
Parameters and constants

	A	B	C	D	E
1	Polymer density	1.892		Dp	20.0
2	Volume of polymer	0.529		Pp	10.0
3	Chi S	0.340		Hp	10.0
4	RT	2479		SSE	17.5
5					

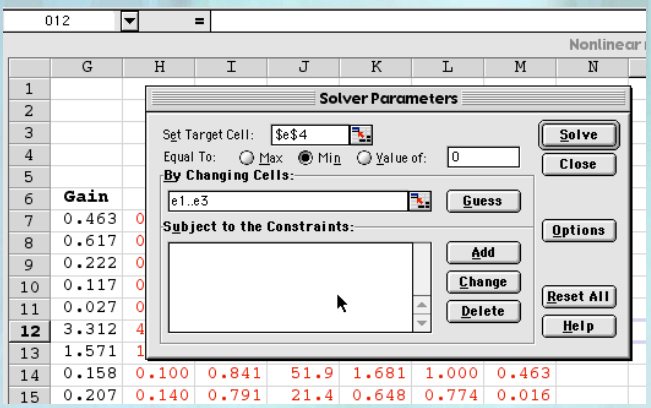
Column B contains known constants for the experiment

Column E contains the values to be adjusted - Dp, Pp and Hp. SSE is the sum of squares that will be tested.

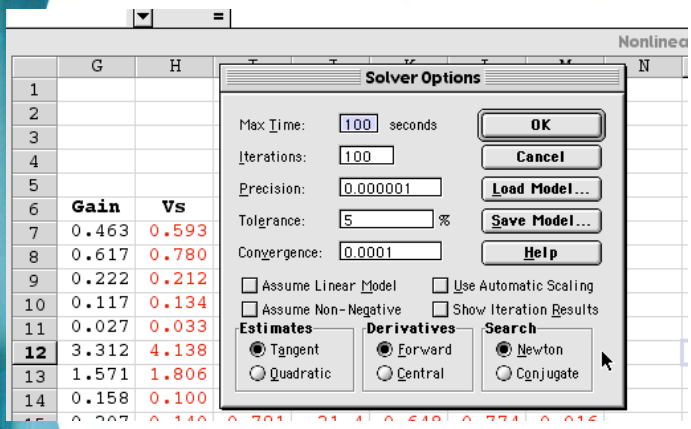
Using Excel's 'Solver'



Using Excel's 'Solver'



Using Excel's 'Solver'



Final results

	A	B	C	D	E	F
1	**Polymer density**	1.892		**Dp**	15.4	
2	**Volume of polymer**	0.529		**Pp**	10.5	
3	**Chi S**	0.340		**Hp**	2.5	
4	**RT**	2479		**SSE**	0.6	
5						
6	**Solvent**	**MW**	**Vm**	**D**	**P**	**H**
7	Acetonitrile	41.1	52.6	15.3	18.0	6
8	Acrylonitrile	53.1	67.1	16.4	17.4	6

Solver summary of results

Microsoft Excel 8.0 Answer Report
Worksheet: [Nonlinear model]Model
Report Created: 4/19/99 6:15:19 AM

Target Cell (Min)

Cell	Name	Original Value	Final Value
E4	SSE	17.5	0.6

Adjustable Cells

Cell	Name	Original Value	Final Value
E1	Dp	20.0	15.4
E2	Pp	10.0	10.5
E3	Hp	10.0	2.5

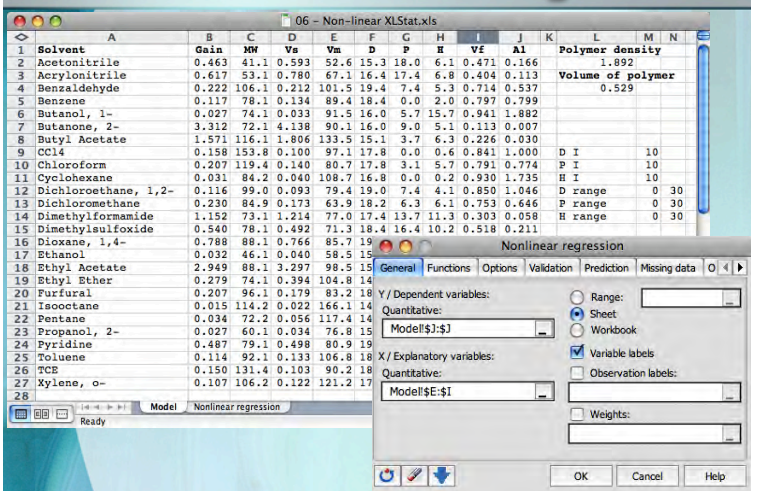
Constraints

NONE

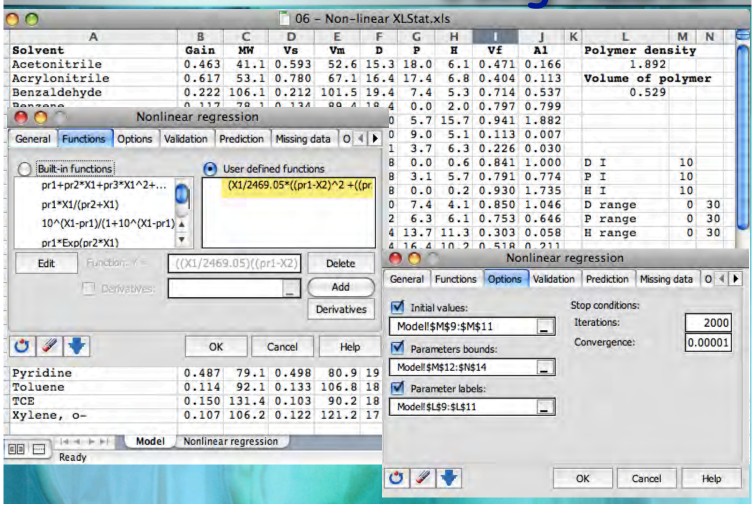
Using XLStat

- XLStat takes a different approach.
- You still have X and Y variables but you build the model on a separate menu.
- Parameter to be fit are included in the model equation as pr1, pr2, pr3.....

Using XLStat



Using XLStat



Using XLStat