CHAPTER 17

17.1 The data can be tabulated as

i	y	$(y_i - \bar{y})^2$
1	0.90	0.524755
2	1.32	0.092659
3	1.96	0.112627
4	1.85	0.050895
5	2.29	0.443023
6	1.42	0.041779
7	1.35	0.075295
8	1.47	0.023839
9	1.74	0.013363
10	1.82	0.038259
11	1.30	0.105235
12	1.47	0.023839
13	1.92	0.087379
14	1.65	0.000655
15	2.06	0.189747
16	1.55	0.005535
17	1.95	0.106015
18	1.35	0.075295
19	1.78	0.024211
20	2.14	0.265843
21	1.63	3.14E-05
22	1.66	0.001267
23	1.05	0.329935
24	1.71	0.007327
25	1.27	0.125599
Σ	**40.61**	**2.764416**

(a) $\bar{y} = \dfrac{40.61}{25} = 1.6244$

(b) $s_y = \sqrt{\dfrac{2.764416}{25-1}} = 0.339388$

(c) $s_y^2 = 0.339388^2 = 0.115184$

(d) $\text{c.v.} = \dfrac{0.339388}{1.6244} \times 100\% = 20.89\%$

(e) $t_{0.05/2,25-1} = 2.063899$

$$L = 1.6244 - \frac{0.339388}{\sqrt{25}} 2.063899 = 1.484308$$

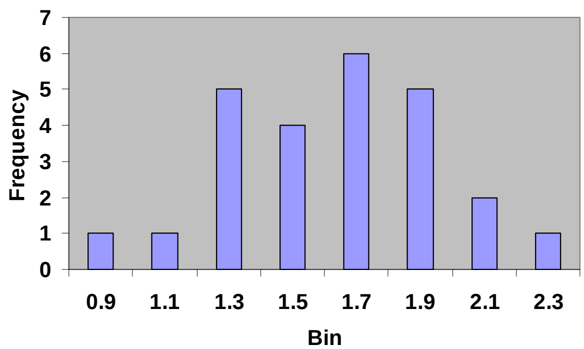
$$U = 1.6244 + \frac{0.339388}{\sqrt{25}} 2.063899 = 1.764492$$

(f) The data can be sorted and then grouped. We assume that if a number falls on the border between bins, it is placed in the lower bin.

lower	upper	Frequency
0.8	1	1
1	1.2	1

1.2	1.4	5
1.4	1.6	4
1.6	1.8	6
1.8	2	5
2	2.2	2
2.2	2.4	1

The histogram can then be constructed as



17.2 The data can be tabulated as

i	y	$(y_i - \bar{y})^2$
1	28.65	0.390625
2	28.65	0.390625
3	27.65	0.140625
4	29.25	1.500625
5	26.55	2.175625
6	29.65	2.640625
7	28.45	0.180625
8	27.65	0.140625
9	26.65	1.890625
10	27.85	0.030625
11	28.65	0.390625
12	28.65	0.390625
13	27.65	0.140625
14	27.05	0.950625
15	28.45	0.180625
16	27.65	0.140625
17	27.35	0.455625
18	28.25	0.050625
19	31.65	13.14063
20	28.55	0.275625
21	28.35	0.105625
22	28.85	0.680625
23	26.35	2.805625
24	27.65	0.140625
25	26.85	1.380625
26	26.75	1.625625
27	27.75	0.075625
28	27.25	0.600625
Σ	**784.7**	**33.0125**

(a) $\bar{y} = \dfrac{784.7}{28} = 28.025$

(b) $s_y = \sqrt{\dfrac{33.0125}{28-1}} = 1.105751$

(c) $s_y^2 = 1.105751^2 = 1.222685$

(d) $\text{c.v.} = \dfrac{1.105751}{28.025} \times 100\% = 3.95\%$

(e) $t_{0.1/2,28-1} = 1.703288$

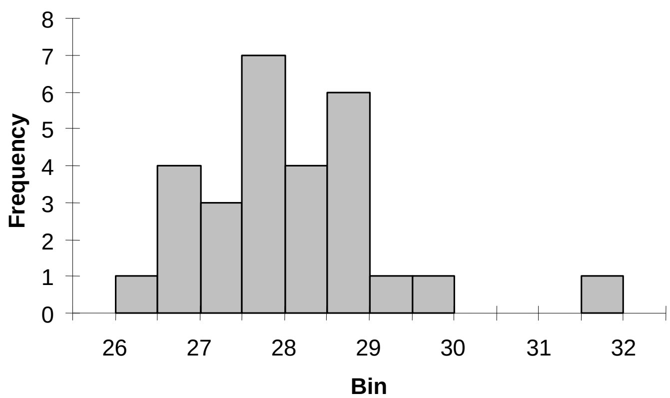
$$L = 28.025 - \frac{1.105751}{\sqrt{28}} 1.703288 = 27.66907$$

$$U = 28.025 + \frac{1.105751}{\sqrt{28}} 1.703288 = 28.38093$$

(f) The data can be sorted and grouped.

Lower	Upper	Frequency
26	26.5	1
26.5	27	4
27	27.5	3
27.5	28	7
28	28.5	4
28.5	29	6
29	29.5	1
29.5	30	1
30	30.5	0
30.5	31	0
31	31.5	0
31.5	32	1

The histogram can then be constructed as

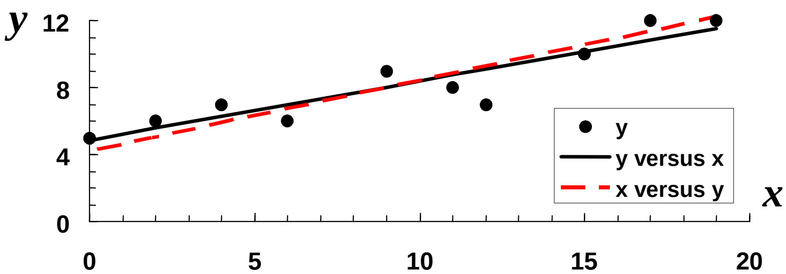


(g) 68% of the readings should fall between $\bar{y} - s_y$ and $\bar{y} + s_y$. That is, between 28.025 – 1.10575096 = 26.919249 and 28.025 + 1.10575096 = 29.130751. Twenty values fall between these bounds which is equal to 20/28 = 71.4% of the values which is not that far from 68%.

17.3 The results can be summarized as

	y versus x	x versus y
Best fit equation	$y = 4.851535 + 0.35247x$	$x = -9.96763 + 2.374101y$
Standard error	1.06501	2.764026
Correlation coefficient	0.914767	0.914767

We can also plot both lines on the same graph

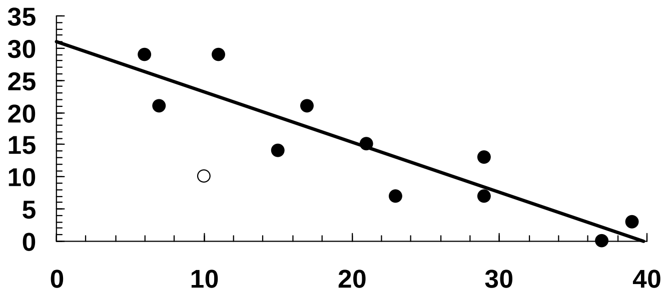


Thus, the “best” fit lines and the standard errors differ. This makes sense because different errors are being minimized depending on our choice of the dependent (ordinate) and independent (abscissa) variables. In contrast, the correlation coefficients are identical since the same amount of uncertainty is explained regardless of how the points are plotted.

17.4 The results can be summarized as

$$y = 31.0589 - 0.78055x \qquad (s_{y/x} = 4.476306; r = 0.901489)$$

At x = 10, the best fit equation gives 23.2543. The line and data can be plotted along with the point (10, 10).



The value of 10 is nearly 3 times the standard error away from the line,

23.2543 – 3(4.476306) = 9.824516

Thus, we can tentatively conclude that the value is probably erroneous. It should be noted that the field of statistics provides related but more rigorous methods to assess whether such points are “outliers.”

17.5 The sum of the squares of the residuals for this case can be written as

$$S_r = \sum_{i=1}^{n} \left(y_i - a_1 x_i\right)^2$$

The partial derivative of this function with respect to the single parameter a_1 can be determined as

$$\frac{\partial S_r}{\partial a_1} = -2\sum\left[(y_i - a_1 x_i)x_i\right]$$

Setting the derivative equal to zero and evaluating the summations gives

$$0 = \sum y_i x_i - a_1 \sum x_i^2$$

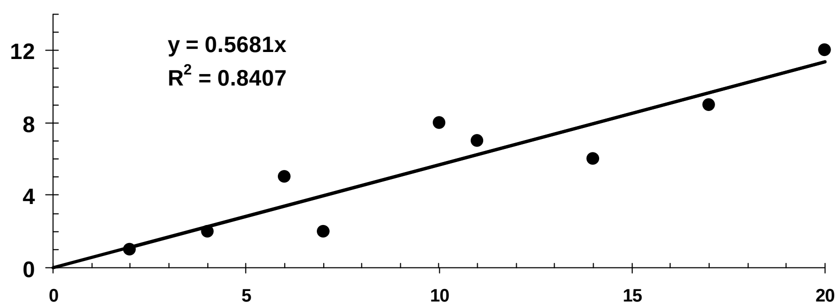
which can be solved for

$$a_1 = \frac{\sum y_i x_i}{\sum x_i^2}$$

So the slope that minimizes the sum of the squares of the residuals for a straight line with a zero intercept is merely the ratio of the sum of the dependent variables (y) times the sum of the independent variables (x) over the sum of the independent variables squared (x^2). Application to the data gives

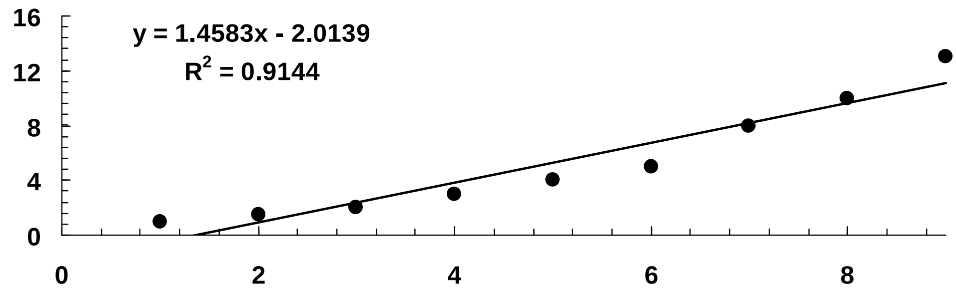
x	y	xy	x^2
2	1	2	4
4	2	8	16
6	5	30	36
7	2	14	49
10	8	80	100
11	7	77	121
14	6	84	196
17	9	153	289
20	12	240	400
		688	1211

Therefore, the slope can be computed as 688/1211 = 0.5681. The fit along with the data can be displayed as



17.6 (a) The results can be summarized as

$$y = -2.01389 + 1.458333x \qquad (s_{y/x} = 1.306653; r = 0.956222)$$

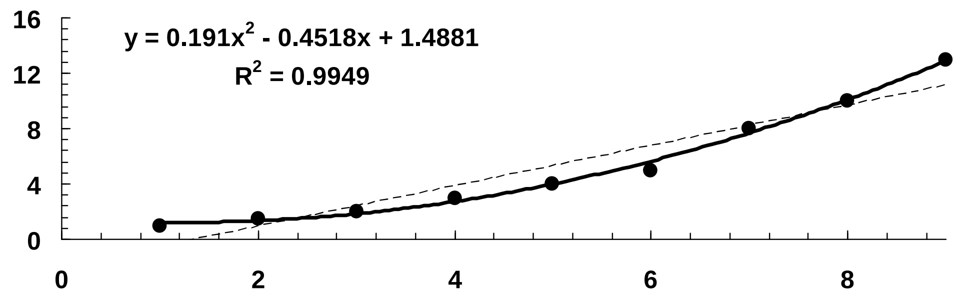


As can be seen, although the correlation coefficient appears to be close to 1, the straight line does not describe the data trend very well.

(b) The results can be summarized as

$$y = 1.488095 - 0.45184x + 0.191017x^2 \qquad (s_{y/x} = 0.344771; r = 0.997441)$$

A plot indicates that the quadratic fit does a much better job of fitting the data.



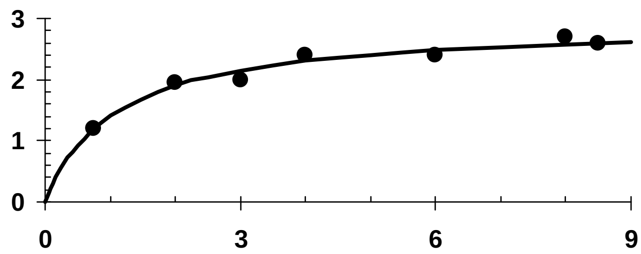
17.7 (a) We regress 1/y versus 1/x to give

$$\frac{1}{y} = 0.34154 + 0.36932\frac{1}{x}$$

Therefore, α_3 = 1/0.34154 = 2.927913 and β_3 = 0.36932(2.927913) = 1.081337, and the saturation-growth-rate model is

$$y = 2.927913\frac{x}{1.081337 + x}$$

The model and the data can be plotted as



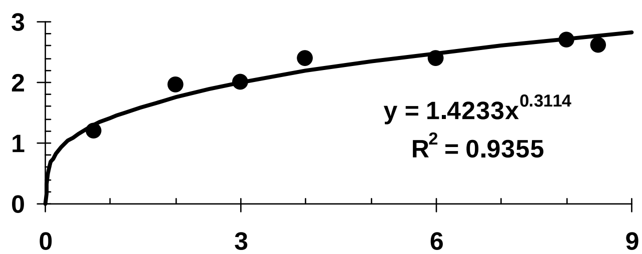
(b) We regress $\log_{10}(y)$ versus $\log_{10}(x)$ to give

$$\log_{10} y = 0.153296 + 0.311422 \log_{10} x$$

Therefore, $\alpha_2 = 10^{0.153296} = 1.423297$ and $\beta_2 = 0.311422$, and the power model is

$y = 1.423297x^{0.311422}$

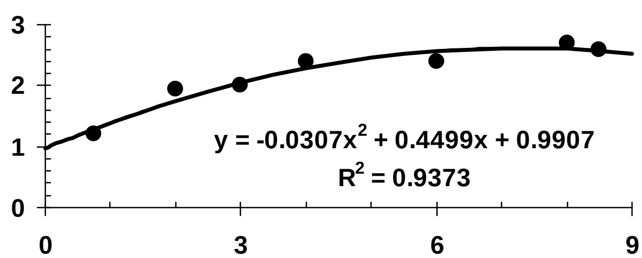
The model and the data can be plotted as



(c) Polynomial regression can be applied to develop a best-fit parabola

$y = -0.03069x^2 + 0.449901x + 0.990728$

The model and the data can be plotted as



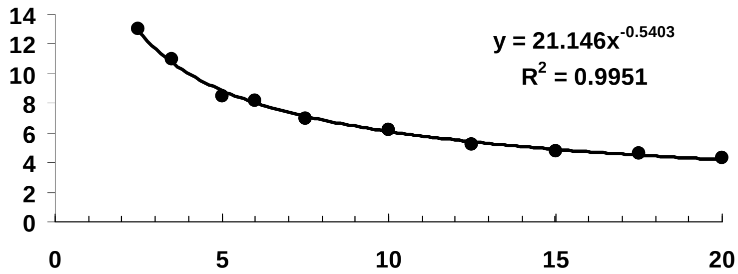
17.8 We regress $\log_{10}(y)$ versus $\log_{10}(x)$ to give

$$\log_{10} y = 1.325225 - 0.54029 \log_{10} x$$

Therefore, $\alpha_2 = 10^{1.325225} = 21.14583$ and $\beta_2 = -0.54029$, and the power model is

$y = 21.14583x^{-0.54029}$

The model and the data can be plotted as



The model can be used to predict a value of $21.14583(9)^{-0.54029} = 6.451453$.

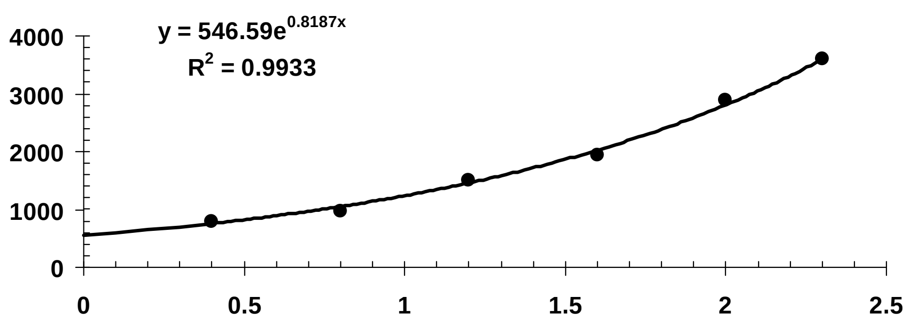
17.9 We regress ln(y) versus x to give

$\ln y = 6.303701 + 0.818651x$

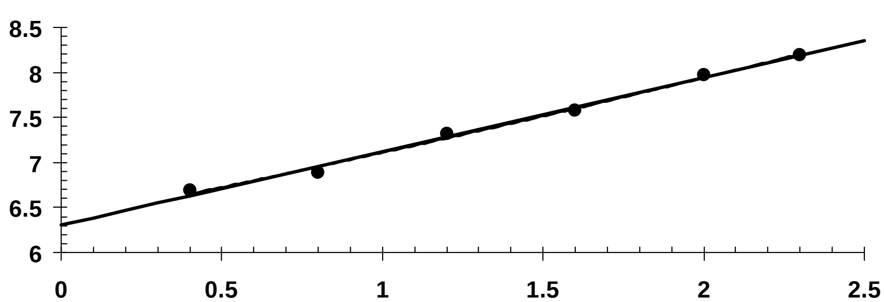
Therefore, $\alpha_1 = e^{6.303701} = 546.5909$ and $\beta_1 = 0.818651$, and the exponential model is

$y = 546.5909e^{0.818651x}$

The model and the data can be plotted as



A semi-log plot can be developed by plotting the natural log versus x. As expected, both the data and the best-fit line are linear when plotted in this way.



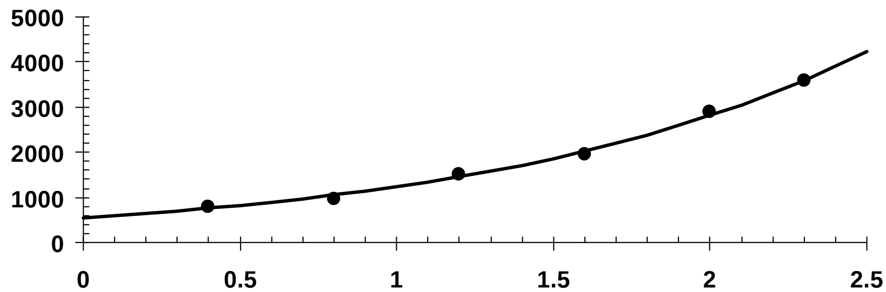
17.10 For the data from Prob. 17.9, we regress $\log_{10}(y)$ versus x to give

$$\log_{10} y = 2.737662 + 0.355536x$$

Therefore, $\alpha_5 = 10^{2.737662} = 546.5909$ and $\beta_5 = 0.355536$, and the base-10 exponential model is

$$y = 546.5909 \times 10^{0.355536x}$$

The model and the data can be plotted as



This plot is identical to the graph that was generated with the base-*e* model derived in Prob. 17.9. Thus, although the models have a different base, they yield identical results.

The relationship between β_1 and β_5 can be developed as in

$$e^{-\beta_1 t} = 10^{-\beta_5 t}$$

Take the natural log of this equation to yield

$$-\beta_1 t = -\beta_5 t \ln 10$$

or

$$\beta_1 = 2.302585\beta_5$$

This result can be verified by substituting the value of β_5 into this equation to give

$$\beta_1 = 2.302585(0.355536) = 0.818651$$

This is identical to the result derived in Prob. 17.9.

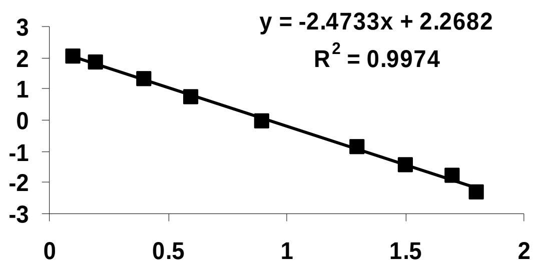
17.11 The function can be linearized by dividing it by x and taking the natural logarithm to yield

$$\ln(y / x) = \ln \alpha_4 + \beta_4 x$$

Therefore, if the model holds, a plot of $\ln(y/x)$ versus x should yield a straight line with an intercept of $\ln\alpha_4$ and a slope of β_4.

x	y	$\ln(y/x)$
0.1	0.75	2.014903

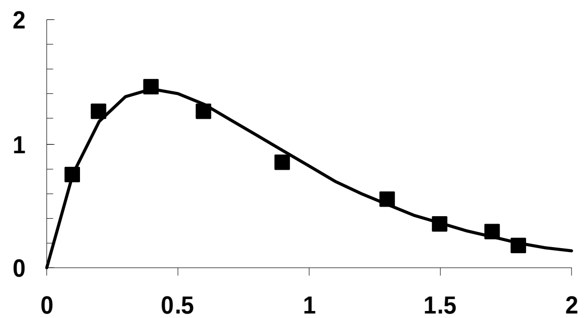
0.2	1.25	1.832581
0.4	1.45	1.287854
0.6	1.25	0.733969
0.9	0.85	-0.05716
1.3	0.55	-0.8602
1.5	0.35	-1.45529
1.7	0.28	-1.80359
1.8	0.18	-2.30259



Therefore, $\beta_4 = -2.4733$ and $\alpha_4 = e^{2.2682} = 9.661786$, and the fit is

$$y = 9.661786xe^{-2.4733x}$$

This equation can be plotted together with the data:



17.12 The equation can be linearized by inverting it to yield

$$\frac{1}{k} = \frac{c_s}{k_{\max}} \frac{1}{c^2} + \frac{1}{k_{\max}}$$

Consequently, a plot of $1/k$ versus $1/c$ should yield a straight line with an intercept of $1/k_{\max}$ and a slope of $c_s/k_{\max}$

c, mg/L	*k*, /d	$1/c^2$	$1/k$	$1/c^2 \times 1/k$	$(1/c^2)^2$
0.5	1.1	4.000000	0.909091	3.636364	16.000000
0.8	2.4	1.562500	0.416667	0.651042	2.441406
1.5	5.3	0.444444	0.188679	0.083857	0.197531
2.5	7.6	0.160000	0.131579	0.021053	0.025600
4	8.9	0.062500	0.112360	0.007022	0.003906
	Sum →	**6.229444**	**1.758375**	**4.399338**	**18.66844**

The slope and the intercept can be computed as

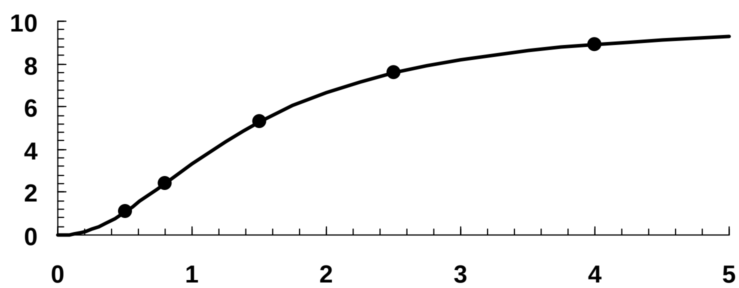
$$a_1 = \frac{5(4.399338) - 6.229444(1.758375)}{5(18.66844) - (6.229444)^2} = 0.202489$$

$$a_0 = \frac{1.758375}{5} - 0.202489\frac{6.229444}{5} = 0.099396$$

Therefore, k_{max} = 1/0.099396 = 10.06074 and c_s = 10.06074(0.202489) = 2.037189, and the fit is

$$k = \frac{10.06074c^2}{2.037189 + c^2}$$

This equation can be plotted together with the data:



The equation can be used to compute

$$k = \frac{10.06074(2)^2}{2.037189 + (2)^2} = 6.666$$

17.13 Linearization: First the natural log can be applied to give

$$\ln x = \frac{y - b}{a}$$

Multiply both sides by a

$$a \ln x = y - b$$

Rearrange to give

$$y = a \ln x + b$$

Therefore, a plot of y versus ln x should yield a straight line with a slope of a and an intercept of b.

x	y	**ln x**	**(ln x)^2**	**ln $x \times y$**
1	0.5	0	0	0
2	2	0.693147	0.480453	1.386294
3	2.9	1.098612	1.206949	3.185976
4	3.5	1.386294	1.921812	4.85203
5	4	1.609438	2.59029	6.437752
$\sum$	**12.9**	**4.787492**	**6.199504**	**15.86205**

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{5(15.86205) - 4.787492(12.9)}{5(6.199504) - (4.787492)^2} = 2.172917$$
$$b = \bar{y} - a_1\bar{x} = \frac{12.9}{5} - 2.172917\frac{4.787492}{5} = 0.499436$$
$$x = e^{(y-0.499436)/2.172917}$$
$$y = 2.172917\ln x + 0.499436 = 2.172917\ln(2.6) + 0.499436 = 2.575683$$

17.14 Linearization: Take the square root

$$\sqrt{y} = \frac{a + \sqrt{x}}{b\sqrt{x}}$$

or

$$\sqrt{y} = \frac{a}{b}\frac{1}{\sqrt{x}} + \frac{1}{b}$$

Therefore, a plot of $\sqrt{y}$ versus $1/\sqrt{x}$ should yield a straight line with a slope of a/b and an intercept of $1/b$.

x	y	$1/\sqrt{x}$	$\sqrt{y}$	$\sqrt{y}/\sqrt{x}$	$1/x$
0.5	10.4	1.414214	3.224903	4.560702	2
1	5.8	1	2.408319	2.408319	1
2	3.3	0.707107	1.81659	1.284523	0.5
3	2.4	0.57735	1.549193	0.894427	0.333333
4	2	0.5	1.414214	0.707107	0.25
		4.198671	**10.41322**	**9.855078**	**4.083333**

The slope and intercept can be computed as

$$a_1 = \frac{5(9.855078) - 4.198671(10.41322)}{5(4.083333) - 4.198671^2} = 1.992126$$
$$a_0 = \frac{10.41322}{5} - 1.992126\frac{4.198671}{5} = 0.409788$$

The constants can then be computed as

$$b = \frac{1}{0.409788} = 2.440288$$
$$a = 1.992126(2.440288) = 4.861362$$

and the prediction calculated as

$$y(1.6) = \left(\frac{4.861362 + \sqrt{1.6}}{2.440288\sqrt{1.6}}\right)^2 = 3.93904$$

17.15 MATLAB can be used to develop the solution:

```
>> format short g
>> x=[1 2 3 4 5]';
>> y=[2.2 2.8 3.6 4.5 5.5]';
>> Z=[ones(size(x)) x 1./x]
Z =
           1           1           1
           1           2         0.5
           1           3     0.33333
           1           4        0.25
           1           5         0.2

>> ZTZ=Z'*Z
ZTZ =
           5          15      2.2833
          15          55           5
      2.2833           5      1.4636

>> ZTy=Z'*y
ZTy =
        18.6
        64.1
       7.025

>> a=inv(ZTZ)*ZTy
a =
      0.3745
     0.98644
     0.84564
```

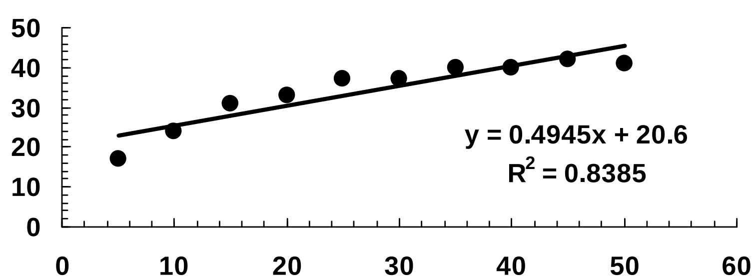
Therefore, the solution is

$$y = 0.3745 + 0.98644x + \frac{0.84564}{x}$$

17.16 (a) We regress *y* versus *x* to give

$y = 20.6 + 0.494545x$

The model and the data can be plotted as



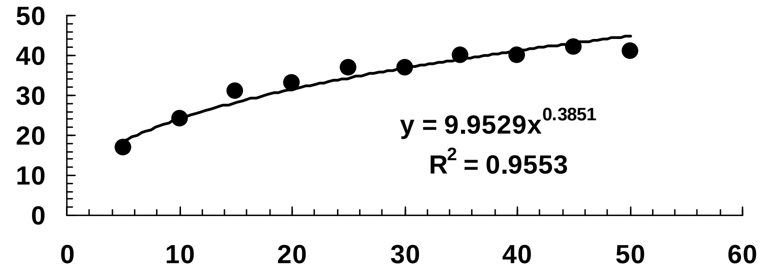
(b) We regress $\log_{10}y$ versus $\log_{10}x$ to give

$$\log_{10} y = 0.99795 + 0.385077 \log_{10} x$$

Therefore, $\alpha_2 = 10^{0.99795} = 9.952915$ and $\beta_2 = 0.385077$, and the power model is

$$y = 9.952915x^{0.385077}$$

The model and the data can be plotted as



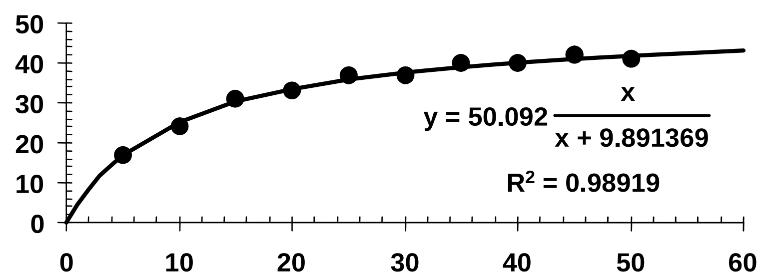
(c) We regress $1/y$ versus $1/x$ to give

$$\frac{1}{y} = 0.019963 + 0.197464\frac{1}{x}$$

Therefore, $\alpha_3 = 1/0.01996322 = 50.09212$ and $\beta_3 = 0.19746357(50.09212) = 9.89137$, and the saturation-growth-rate model is

$$y = 50.09212\frac{x}{9.89137 + x}$$

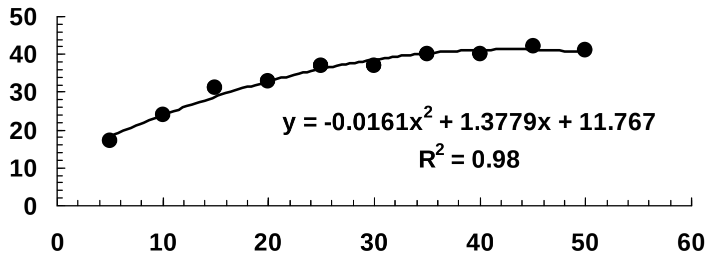
The model and the data can be plotted as



(d) We employ polynomial regression to fit a parabola

$$y = -0.01606x^2 + 1.377879x + 11.76667$$

The model and the data can be plotted as



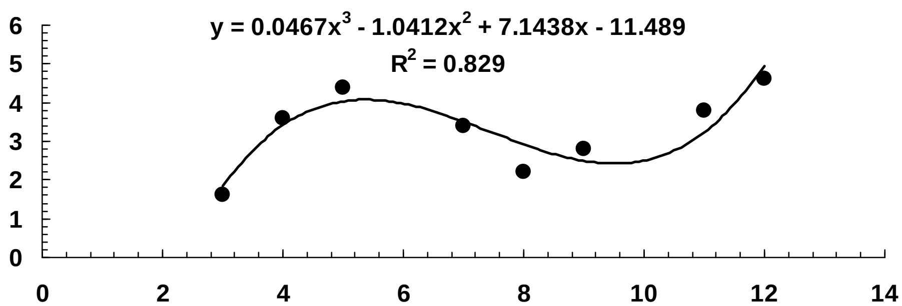
Comparison of fits: The linear fit is obviously inadequate. Although the power fit follows the general trend of the data, it is also inadequate because (1) the residuals do not appear to be randomly distributed around the best fit line and (2) it has a lower r^2 than the saturation and parabolic models.

The best fits are for the saturation-growth-rate and the parabolic models. They both have randomly distributed residuals and they have similar high coefficients of determination. The saturation model has a slightly higher r^2. Although the difference is probably not statistically significant, in the absence of additional information, we can conclude that the saturation model represents the best fit.

17.17 We employ polynomial regression to fit a cubic equation to the data

$$y = 0.046676x^3 - 1.04121x^2 + 7.143817x - 11.4887 \qquad (r^2 = 0.828981; s_{y/x} = 0.570031)$$

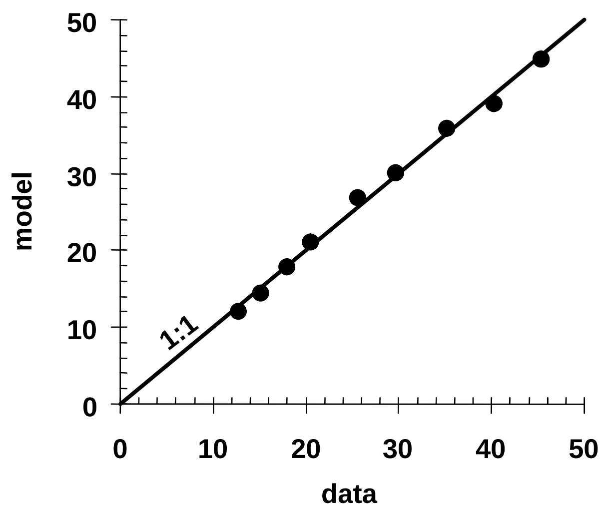
The model and the data can be plotted as



17.18 We employ multiple linear regression to fit the following equation to the data

$$y = 14.46087 + 9.025217x_1 - 5.70435x_2 \qquad (r^2 = 0.995523; s_{y/x} = 0.888787)$$

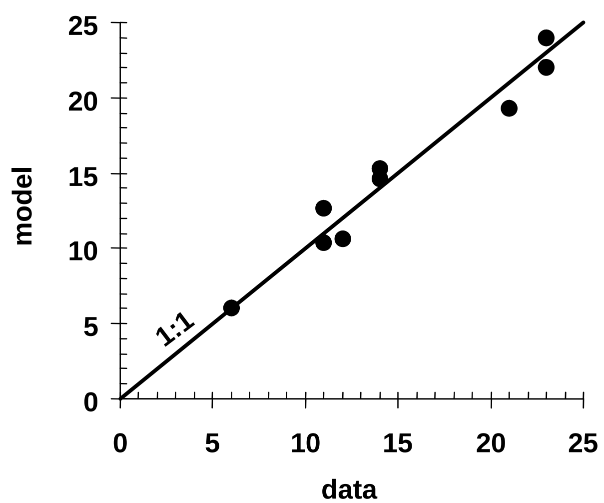
The model and the data can be compared graphically by plotting the model predictions versus the data. A 1:1 line is included to indicate a perfect fit.



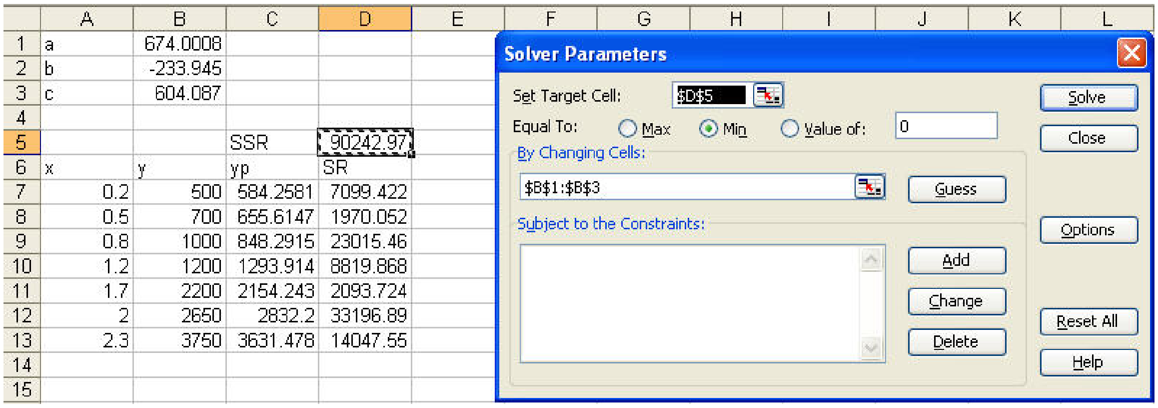
17.19 We employ multiple linear regression to fit the following equation to the data

$$y = 14.66667 - 6.66667x_1 + 2.333333x_2 \qquad (r^2 = 0.958333; s_{y/x} = 1.414214)$$

The model and the data can be compared graphically by plotting the model predictions versus the data. A 1:1 line is included to indicate a perfect fit.



17.20 We can employ nonlinear regression to fit a parabola to the data. A simple way to do this is to use the Excel Solver to minimize the sum of the squares of the residuals as in the following worksheet,



	A	B	C	D
1	a	674.0008		
2	b	-233.945		
3	c	604.087		
4				
5			SSR	90242.97
6	x	y	yp	SR
7	0.2	500	584.2581	7099.422
8	0.5	700	655.6147	1970.052
9	0.8	1000	848.2915	23015.46
10	1.2	1200	1293.914	8819.868
11	1.7	2200	2154.243	2093.724
12	2	2650	2832.2	33196.89
13	2.3	3750	3631.478	14047.55
14				
15				

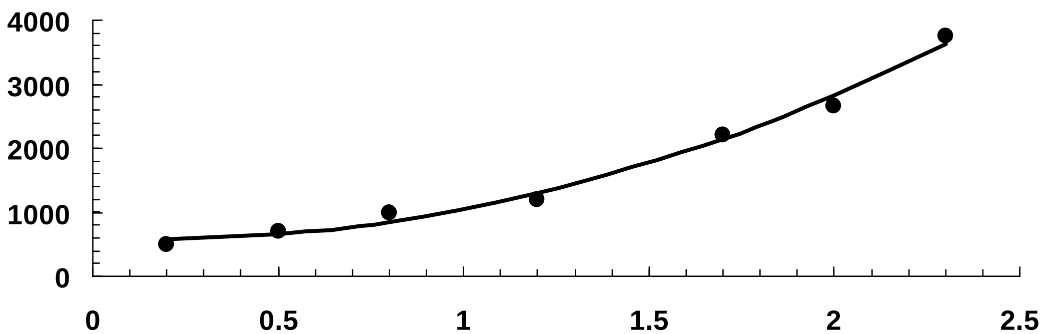
The formulas are

	A	B	C	D
1	a	674.000780425739		
2	b	-233.945011343731		
3	c	604.087039907068		
4				
5			SSR	=SUM(D7:D13)
6	x	y	yp	SR
7	0.2	500	=a*A7^2+b*A7+c_	=(B7-C7)^2
8	0.5	700	=a*A8^2+b*A8+c_	=(B8-C8)^2
9	0.8	1000	=a*A9^2+b*A9+c_	=(B9-C9)^2
10	1.2	1200	=a*A10^2+b*A10+c_	=(B10-C10)^2
11	1.7	2200	=a*A11^2+b*A11+c_	=(B11-C11)^2
12	2	2650	=a*A12^2+b*A12+c_	=(B12-C12)^2
13	2.3	3750	=a*A13^2+b*A13+c_	=(B13-C13)^2

Thus, the best-fit equation is

$$y = 674.001x^2 - 233.945x + 604.087$$

The model and the data can be displayed graphically as

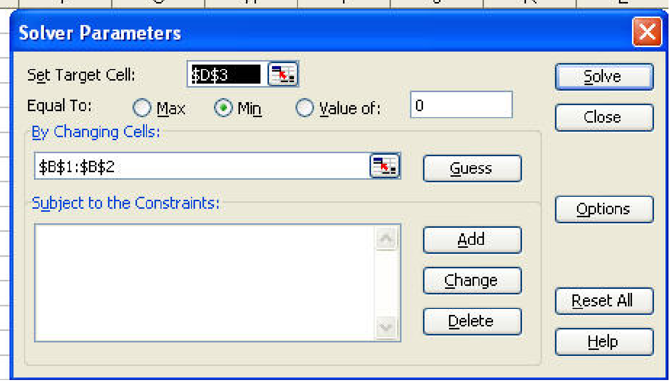


Note that if polynomial regression were used, a slightly different fit would result,

$$y = 674.007x^2 - 233.961x + 604.094$$

17.21 We can employ nonlinear regression to fit the saturation-growth-rate equation to the data from Prob. 17.16. A simple way to do this is to use the Excel Solver to minimize the sum of the squares of the residuals as in the following worksheet,

	A	B	C	D
1	a	50.52924		
2	b	10.12939		
3			SSR	6.350767
4	x	y	yp	SR
5	5	17	16.69904	0.090579
6	10	24	25.10222	1.214898
7	15	31	30.16144	0.703177
8	20	33	33.5415	0.293221
9	25	37	35.95938	1.082889
10	30	37	37.77474	0.600224
11	35	40	39.18784	0.659596
12	40	40	40.31906	0.101798
13	45	42	41.24508	0.569908
14	50	41	42.01709	1.034478
15				



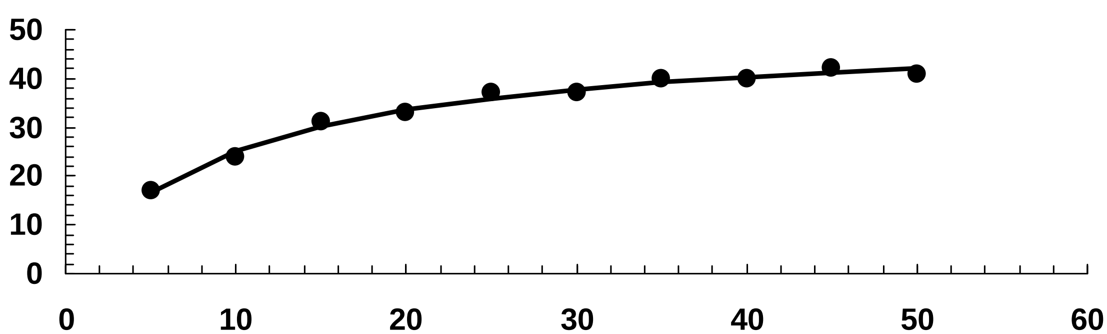
The formulas are

	A	B	C	D
1	a	50.5292424213902		
2	b	10.1293886897916		
3			SSR	=SUM(D5:D14)
4	x	y	yp	SR
5	5	17	=$B$1*A5/($B$2+A5)	=(B5-C5)^2
6	10	24	=$B$1*A6/($B$2+A6)	=(B6-C6)^2
7	15	31	=$B$1*A7/($B$2+A7)	=(B7-C7)^2
8	20	33	=$B$1*A8/($B$2+A8)	=(B8-C8)^2
9	25	37	=$B$1*A9/($B$2+A9)	=(B9-C9)^2
10	30	37	=$B$1*A10/($B$2+A10)	=(B10-C10)^2
11	35	40	=$B$1*A11/($B$2+A11)	=(B11-C11)^2
12	40	40	=$B$1*A12/($B$2+A12)	=(B12-C12)^2
13	45	42	=$B$1*A13/($B$2+A13)	=(B13-C13)^2
14	50	41	=$B$1*A14/($B$2+A14)	=(B14-C14)^2

Thus, the best-fit equation is

$$y = 50.529 \frac{x}{10.129 + x}$$

The model and the data can be displayed graphically as



Recall that for Prob. 17.16c, a slightly different fit resulted,

$$y = 50.09212 \frac{x}{9.89137 + x}$$

17.22 MATLAB provides a very nice environment for solving this problem:

(a) Prob. 17.3:

First, we can enter the data

```
>> X=[0 2 4 6 9 11 12 15 17 19]';
>> Y=[5 6 7 6 9 8 7 10 12 12]';
```

Then, we can create the *Z* matrix which consists of a column of ones and a second column of the *x*'s.

```
>> Z=[ones(size(X)) X]
Z =
     1     0
     1     2
     1     4
     1     6
     1     9
     1    11
     1    12
     1    15
     1    17
     1    19
```

Next we can develop the coefficients of the normal equations as

```
>> ZTZ=Z'*Z
ZTZ =
          10          95
          95        1277
```

We can compute the right-hand side of the normal equations with

```
>> ZTY=Z'*Y
ZTY =
    82
   911
```

We can then determine the coefficients for the linear regression as

```
>> A=inv(ZTZ)*ZTY
A =
    4.8515
    0.3525
```

This result is identical to that obtained in Prob. 17.3. Next, we can determine the r^2 and $s_{y/x}$,

```
>> Sr=sum((Y-Z*A).^2)
Sr =
    9.0740
>> r2=1-Sr/sum((Y-mean(Y)).^2)
r2 =
    0.8368
```

```
>> syx=sqrt(Sr/(length(X)-length(A)))
syx =
    1.0650
```

In order to determine the confidence intervals we can first calculate the inverse of $[Z]^T[Z]$ as

```
>> ZTZI=inv(ZTZ)
ZTZI =
    0.3410   -0.0254
   -0.0254    0.0027
```

The standard errors of the coefficients can be computed as

```
>> sa0=sqrt(ZTZI(1,1)*syx^2)
sa0 =
    0.6219
>> sa1=sqrt(ZTZI(2,2)*syx^2)
sa1 =
    0.0550
```

The *t* statistic can be determined as TINV(0.1, 10 – 2) = 1.8595. We can then compute the confidence intervals as

```
>> a0min=A(1)-1.8595*sa0;
>> a0max=A(1)+1.8595*sa0;
>> a1min=A(2)-1.8595*sa1;
>> a1max=A(2)+1.8595*sa1;
```

which yields the confidence intervals for a_0 and a_1 as [3.6951, 6.0080] and [0.2501, 0.4548], respectively.

(b) Prob. 17.17: First, we can determine the coefficients

```
>> X=[3 4 5 7 8 9 11 12]';
>> Y=[1.6 3.6 4.4 3.4 2.2 2.8 3.8 4.6]';
>> Z=[ones(size(X)) X X.^2 X.^3];
>> ZTZ=Z'*Z;
>> ZTY=Z'*Y;
>> A=inv(ZTZ)*ZTY
A =
  -11.4887
    7.1438
   -1.0412
    0.0467
```

The standard error can be computed as

```
>> Sr=sum((Y-Z*A).^2);
>> syx=sqrt(Sr/(length(Y)-length(A)))
syx =
    0.5700
```

The standard errors of the coefficients can be computed as

```
>> ZTZI=inv(ZTZ)
ZTZI =
   49.3468  -23.4771    3.2960   -0.1412
  -23.4771   11.4162   -1.6270    0.0705
    3.2960   -1.6270    0.2349   -0.0103
   -0.1412    0.0705   -0.0103    0.0005
```

```
>> sa0=sqrt(ZTZI(1,1)*syx^2)
sa0 =
    4.0043
>> sa1=sqrt(ZTZI(2,2)*syx^2)
sa1 =
    1.9260
>> sa2=sqrt(ZTZI(3,3)*syx^2)
sa2 =
    0.2763
>> sa3=sqrt(ZTZI(4,4)*syx^2)
sa3 =
    0.0121
```

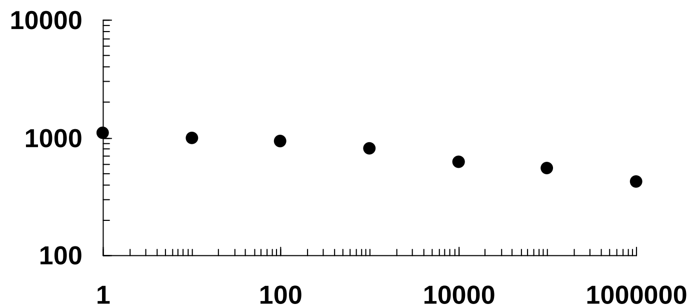
The t statistic can be determined as TINV(0.1, 8 – 4) = 2.13185. We can then compute the confidence intervals for a_0, a_1, a_2, and a_3 as [–20.0253, –2.9521], [3.0379, 11.2498], [–1.6302, –0.45219], and [0.02078, 0.072569], respectively.

17.23 Here's VBA code to implement linear regression:

```
Option Explicit
Sub Regres()
Dim n As Integer
Dim x(20) As Double, y(20) As Double, a1 As Double, a0 As Double
Dim syx As Double, r2 As Double
n = 7
x(1) = 1: x(2) = 2: x(3) = 3: x(4) = 4: x(5) = 5
x(6) = 6: x(7) = 7
y(1) = 0.5: y(2) = 2.5: y(3) = 2: y(4) = 4: y(5) = 3.5
y(6) = 6: y(7) = 5.5
Call Linreg(x, y, n, a1, a0, syx, r2)
MsgBox "slope= " & a1
MsgBox "intercept= " & a0
MsgBox "standard error= " & syx
MsgBox "coefficient of determination= " & r2
MsgBox "correlation coefficient= " & Sqr(r2)
End Sub
Sub Linreg(x, y, n, a1, a0, syx, r2)
Dim i As Integer
Dim sumx As Double, sumy As Double, sumxy As Double
Dim sumx2 As Double, st As Double, sr As Double
Dim xm As Double, ym As Double
sumx = 0
sumy = 0
sumxy = 0
sumx2 = 0
st = 0
sr = 0
'determine summations for regression
For i = 1 To n
  sumx = sumx + x(i)
  sumy = sumy + y(i)
  sumxy = sumxy + x(i) * y(i)
  sumx2 = sumx2 + x(i) ^ 2
Next i
'determine means
xm = sumx / n
ym = sumy / n
'determine coefficients
a1 = (n * sumxy - sumx * sumy) / (n * sumx2 - sumx * sumx)
a0 = ym - a1 * xm
'determine standard error and coefficient of determination
```

```
For i = 1 To n
  st = st + (y(i) - ym) ^ 2
  sr = sr + (y(i) - a1 * x(i) - a0) ^ 2
Next i
syx = (sr / (n - 2)) ^ 0.5
r2 = (st - sr) / st
End Sub
```

17.24 A log-log plot of stress versus N suggests a linear relationship.



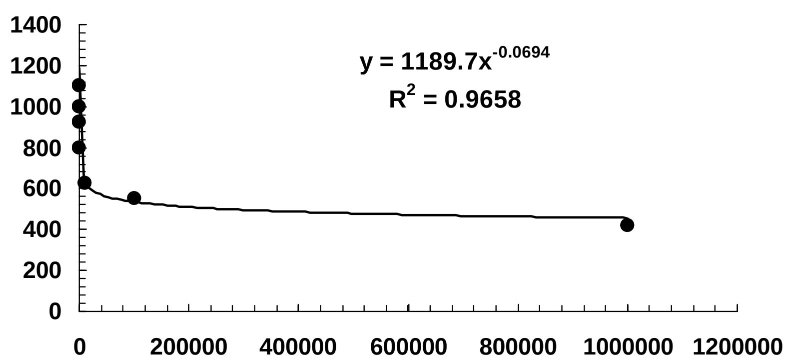
We regress $\log_{10}$(stress) versus $\log_{10}$(N) to give

$$\log_{10}(\text{stress}) = 3.075442 - 0.06943 \log_{10} N$$

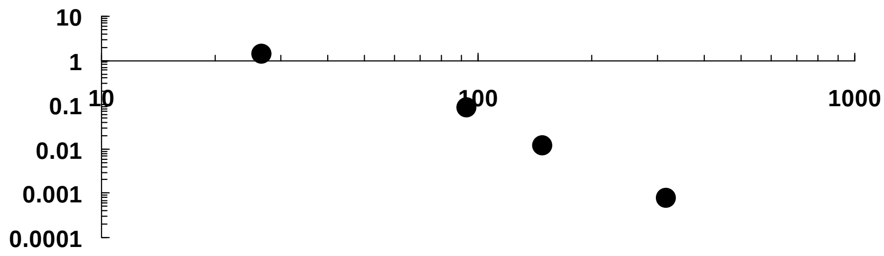
Therefore, $\alpha_2 = 10^{3.075442} = 1189.711$ and $\beta_2 = -0.06943$, and the power model is

$$\text{stress} = 1189.711 N^{-0.06943}$$

The model and the data can be plotted on untransformed scales as



17.25 A log-log plot of μ versus T suggests a linear relationship.



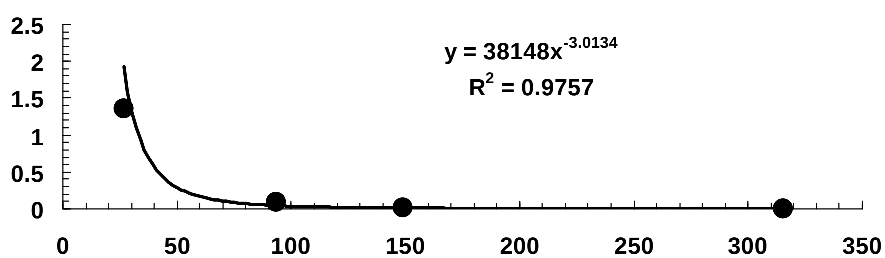
We regress $\log_{10}\mu$ versus $\log_{10}T$ to give

$$\log_{10}\mu = 4.581471 - 3.01338\log_{10}T \qquad (r^2 = 0.975703)$$

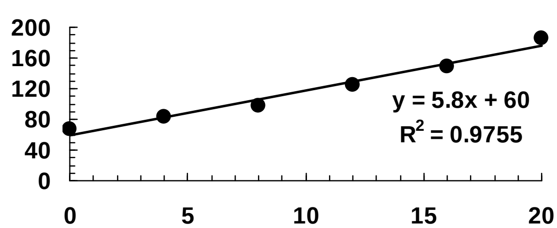
Therefore, $\alpha_2 = 10^{4.581471} = 38{,}147.94$ and $\beta_2 = -3.01338$, and the power model is

$$\mu = 38{,}147.94T^{-3.01338}$$

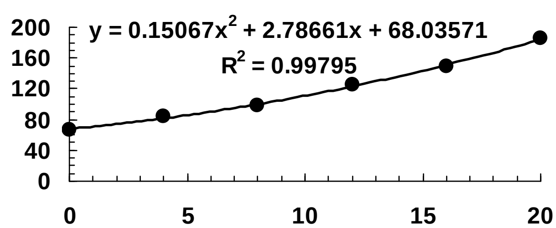
The model and the data can be plotted on untransformed scales as



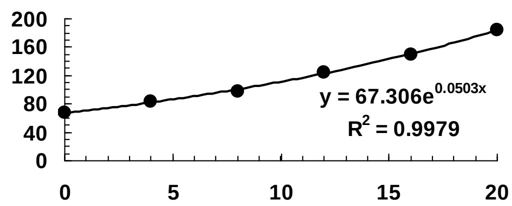
17.26 This problem was solved using an Excel spreadsheet and TrendLine. Linear regression gives



Polynomial regression yields a best-fit parabola



Exponential model:



The linear model is inadequate since it does not capture the curving trend of the data. At face value, the parabolic and exponential models appear to be equally good. However, knowledge of bacterial growth might lead you to choose the exponential model as it is commonly used to simulate the growth of microorganism populations. Interestingly, the choice matters when the models are used for prediction. If the exponential model is used, the result is

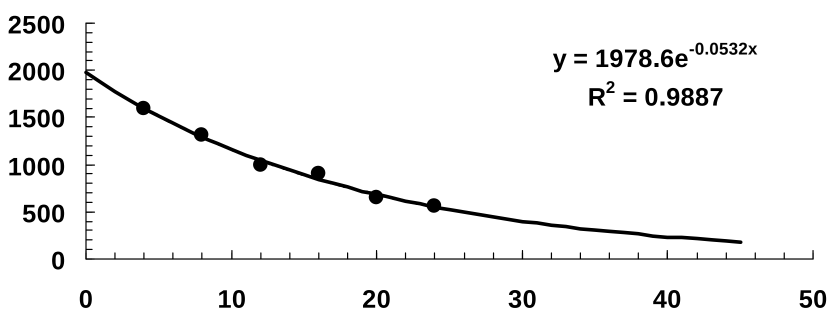
$$B = 67.306e^{0.0503(40)} = 503.3317$$

For the parabolic model, the prediction is

$$B = 0.15067t^2 + 2.78661t + 68.03571 = 420.5721$$

Thus, even though the models would yield very similar results within the data range, they yield dramatically different results for extrapolation outside the range.

17.27 The exponential model is ideal for this problem since (1) it does not yield negative results (as could be the case with a polynomial), and (2) it always decreases with time. Further, it is known that bacterial death is well approximated by the exponential model.



(a) The model says that the concentration at t = 0 was 1978.6.

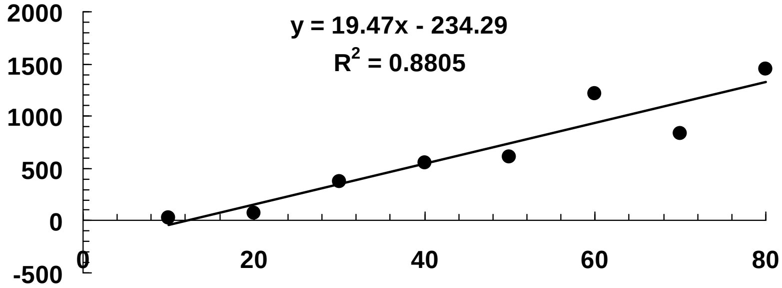
(b) The time at which the concentration reaches 200 can be computed as

$$200 = 1978.6e^{-0.0532t}$$

$$\ln\left(\frac{200}{1978.6}\right) = -0.0532t$$

$$t = \frac{\ln\left(\frac{200}{1978.6}\right)}{-0.0532} = 43.1 \text{ hr}$$

17.28 (a) Linear model



Although this model does a good job of capturing the trend of the data, it has the disadvantage that it yields a negative intercept. Since this is clearly a physically unrealistic result, another model would be preferable.

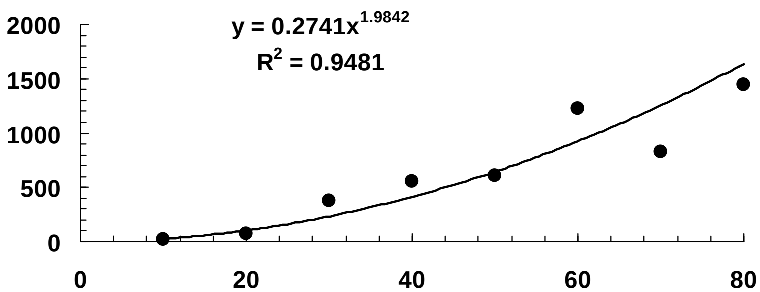
(b) Power model based on log transformations. We regress $\log_{10}(F)$ versus $\log_{10}(v)$ to give

$$\log_{10} F = -0.56203 + 1.984176 \log_{10} v$$

Therefore, $\alpha_2 = 10^{-0.56203} = 0.274137$ and $\beta_2 = 1.984176$, and the power model is

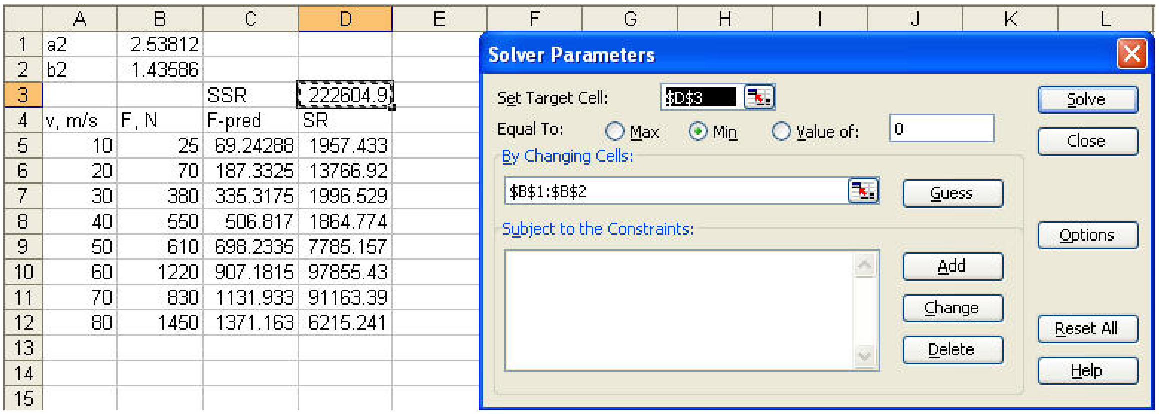
$$F = 0.274137 v^{1.984176}$$

The model and the data can be plotted as



This model represents a superior fit of the data as it fits the data nicely (the r^2 is superior to that obtained with the linear model in **(a)**) while maintaining a physically realistic zero intercept.

(c) Power model based on nonlinear regression. We can use the Excel Solver to determine the fit.



	A	B	C	D
1	a2	2.53812		
2	b2	1.43586		
3			SSR	222604.9
4	v, m/s	F, N	F-pred	SR
5	10	25	69.24288	1957.433
6	20	70	187.3325	13766.92
7	30	380	335.3175	1996.529
8	40	550	506.817	1864.774
9	50	610	698.2335	7785.157
10	60	1220	907.1815	97855.43
11	70	830	1131.933	91163.39
12	80	1450	1371.163	6215.241

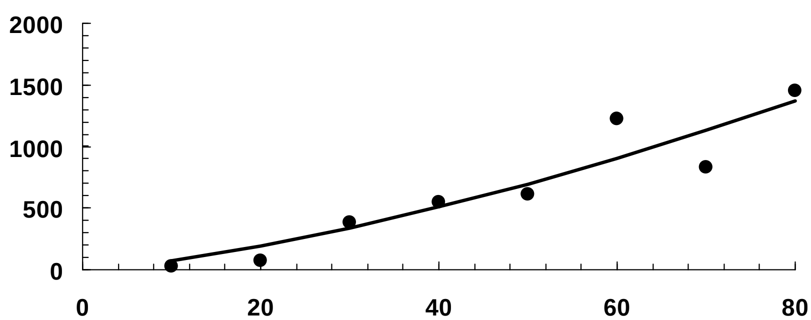
The cell formulas are

	A	B	C	D
1	a2	2.5384243920646		
2	b2	1.43585078550046		
3			SSR	=SUM(D5:D12)
4	v, m/s	F, N	F-pred	SR
5	10	25	=$B$1*A5^$B$2	=(B5-C5)^2
6	20	70	=$B$1*A6^$B$2	=(B6-C6)^2
7	30	380	=$B$1*A7^$B$2	=(B7-C7)^2
8	40	550	=$B$1*A8^$B$2	=(B8-C8)^2
9	50	610	=$B$1*A9^$B$2	=(B9-C9)^2
10	60	1220	=$B$1*A10^$B$2	=(B10-C10)^2
11	70	830	=$B$1*A11^$B$2	=(B11-C11)^2
12	80	1450	=$B$1*A12^$B$2	=(B12-C12)^2

Therefore, the best-fit model is

$$F = 2.53842 v^{1.43585}$$

The model and the data can be plotted as



This model also represents a superior fit of the data as it fits the data nicely while maintaining a physically realistic zero intercept. However, it is very interesting to note that the fit is quite different than that obtained with log transforms in **(b)**.

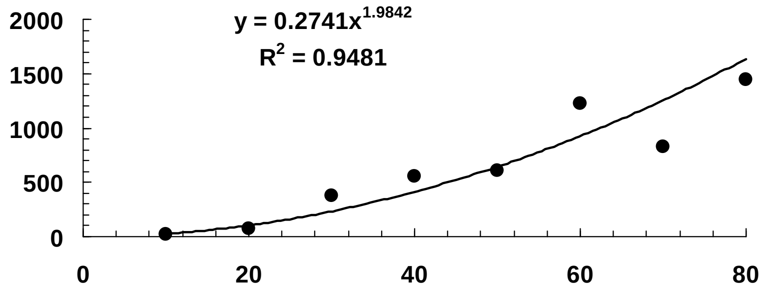
17.29 We can develop a power equation based on natural logarithms. To do this, we regress ln(F) versus ln(v) to give

$$\ln F = -1.29413 + 1.984176 \ln v$$

Therefore, $\alpha_2 = e^{-1.29413} = 0.274137$ and $\beta_2 = 1.984176$, and the power model is

$$F = 0.274137 v^{1.984176}$$

The model and the data can be plotted as



Note that this result is identical to that obtained with common logarithms in Prob. 17.28(b). Thus, we can conclude that any base logarithm would yield the same power model.

17.30 The sum of the squares of the residuals for this case can be written as

$$S_r = \sum_{i=1}^{n} \left(y_i - a_1 x_i - a_2 x_i^2 \right)^2$$

The partial derivatives of this function with respect to the unknown parameters can be determined as

$$\frac{\partial S_r}{\partial a_1} = -2 \sum \left[(y_i - a_1 x_i - a_2 x_i^2) x_i \right]$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum \left[(y_i - a_1 x_i - a_2 x_i^2) x_i^2 \right]$$

Setting the derivative equal to zero and evaluating the summations gives

$$\left(\sum x_i^2 \right) a_1 + \left(\sum x_i^3 \right) a_2 = \sum x_i y_i$$

$$\left(\sum x_i^3 \right) a_1 + \left(\sum x_i^4 \right) a_2 = \sum x_i^2 y_i$$

which can be solved for

$$a_1 = \frac{\sum x_i y_i \sum x_i^4 - \sum x_i^2 y_i \sum x_i^3}{\sum x_i^2 \sum x_i^4 - \left(\sum x_i^3 \right)^2}$$

$$a_2 = \frac{\sum x_i^2 \sum x_i^2 y_i - \sum x_i y_i \sum x_i^3}{\sum x_i^2 \sum x_i^4 - \left(\sum x_i^3\right)^2}$$

The model can be tested for the data from Table 17.28.

x	y	x^2	x^3	x^4	xy	x^2y
10	25	100	1000	10000	250	2500
20	70	400	8000	160000	1400	28000
30	380	900	27000	810000	11400	342000
40	550	1600	64000	2560000	22000	880000
50	610	2500	125000	6250000	30500	1525000
60	1220	3600	216000	12960000	73200	4392000
70	830	4900	343000	24010000	58100	4067000
80	1450	6400	512000	40960000	116000	9280000
Σ		20400	1296000	87720000	312850	20516500

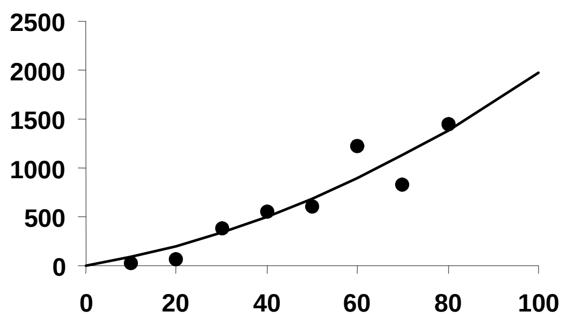
$$a_1 = \frac{312850(87720000) - 20516500(1296000)}{20400(87720000) - (1296000)^2} = 7.771024$$

$$a_2 = \frac{20400(20516500) - 312850(1296000)}{20400(87720000) - (1296000)} = 0.119075$$

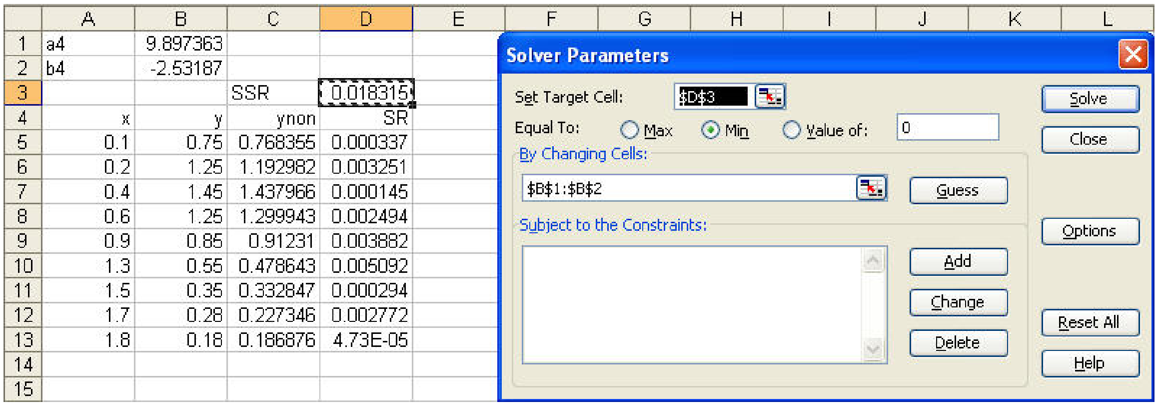
Therefore, the best-fit model is

$$y = 7.771024x + 0.119075x^2$$

The fit, along with the original data can be plotted as



17.31 We can use the Excel Solver to determine the fit.



	A	B	C	D
1	a4	9.897363		
2	b4	-2.53187		
3			SSR	0.018315
4	x	y	ynon	SR
5	0.1	0.75	0.768355	0.000337
6	0.2	1.25	1.192982	0.003251
7	0.4	1.45	1.437966	0.000145
8	0.6	1.25	1.299943	0.002494
9	0.9	0.85	0.91231	0.003882
10	1.3	0.55	0.478643	0.005092
11	1.5	0.35	0.332847	0.000294
12	1.7	0.28	0.227346	0.002772
13	1.8	0.18	0.186876	4.73E-05
14				
15				

The cell formulas are

	A	B	C	D
1	a4	9.89736315656697		
2	b4	-2.53187086003733		
3			SSR	=SUM(D5:D13)
4	x	y	ynon	SR
5	0.1	0.75	=$B$1*A5*EXP($B$2*A5)	=(B5-C5)^2
6	0.2	1.25	=$B$1*A6*EXP($B$2*A6)	=(B6-C6)^2
7	0.4	1.45	=$B$1*A7*EXP($B$2*A7)	=(B7-C7)^2
8	0.6	1.25	=$B$1*A8*EXP($B$2*A8)	=(B8-C8)^2
9	0.9	0.85	=$B$1*A9*EXP($B$2*A9)	=(B9-C9)^2
10	1.3	0.55	=$B$1*A10*EXP($B$2*A10)	=(B10-C10)^2
11	1.5	0.35	=$B$1*A11*EXP($B$2*A11)	=(B11-C11)^2
12	1.7	0.28	=$B$1*A12*EXP($B$2*A12)	=(B12-C12)^2
13	1.8	0.18	=$B$1*A13*EXP($B$2*A13)	=(B13-C13)^2

Therefore, the best-fit model is

$$y = 9.8974xe^{-2.53187x}$$

The model and the data can be plotted as