Chapter 3 Model Fitting (TD3)

3.1 Fitting Models to Data Graphically
3.2 Analytic Methods of Model Fitting

Exercise in Class

1. The following table gives the elongation e in inches per inch (in./in.) for a given stress S on a steel wire measured in pounds per square inch (1b/in.^{2}). Test the model $e = c_1 S$ by plotting the data. Estimate c_1 graphically.

$S \times 10^{-3}$	5	10	20	30	40	50	60	70	80	90	100
$e \times 10^5$)	0	19	57	94	134	173	216	256	297	343	390

2. In the following data, x is the diameter of a ponderosa pine in inches measured at breast height and y is a measure of volume number of board feet divided by 10. Test the model $y = ax^b$ by plotting the transformed data. If the model seem reasonable, estimate the parameters a and b of the model graphically.

x	17	19	20	22	23	25	28	31	32	33	36	37	38	39	41
y	19	25	32	51	57	71	113	141	123	187	192	205	252	259	294

3. The following data represent (hypothetical) energy consumption normalized to the year 1900. Plot the data. Test the model $Q = ae^{bx}$ by plotting the transformed data. Estimate the parameters of the model graphically.

x	Year	Consumption Q
0	1900	1.00
10	1910	2.01
20	1920	4.06
30	1930	8.17
40	1940	16.44
50	1950	33.12
60	1960	66.69
70	1970	134.29
80	1980	270.43
90	1990	544.57
100	2000	1096.63

4. Using elementary calculus, show that the minimum and maximum points for $y = f(x)$ occur among the minimum and maximum point for $y = f^2(x)$. Assuming $f(x) \geq 0$. why can we minimize $f(x)$ by minimizing $f^2(x)$?

5. For each of the following data sets, formulate the mathematical model that minimizes the largest deviation between the data and the line $y = ax + b$. If a computer is available, solve for the estimates of a and b.

a.

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

b.

x	29.1	48.2	72.7	92.0	118	140	165	190
y	0.0493	0.0821	0.123	0.154	0.197	0.234	0.274	0.328

c.

x	2.5	3.0	3.5	4.0	4.5	5.0	5.5
y	4.32	4.83	5.27	5.74	6.26	6.79	7.23

6. For the folowing data, formulate the mathematical model that minimizes the largeast deviation between the data and the model $y = c_1x^2 + c_2x + c_3$. If a computer is available , solve for the estimates of c_1, c_2 and c_3.

x	0.1	0.2	0.3	0.4	0.5
y	0.06	0.12	0.36	0.65	0.95

Assignment

Deadline: 01 December, 2022.

1. The following data represent the growth of a population of fruit flies over a 6-week period. Test the following models by plotting an appropriate set of data. Estimate the parameters of the following model.

a. $P = c_1t$

b. $P = ae^{bt}$

t (days)	7	14	21	28	35	42
P (number of observed flies)	8	41	133	250	280	297

2. In 1610 the german astronomer Johannes Kepler became direction of the Prague Observatory. Kepler had been helping Tycho Brahe in collecting 13years of observation on the relative motion of the planet Mars. By 1609 Kepler had formulated his first two laws:

i Each planet moves on an ellipse with the sun at one focus.

ii For each planet, the line from the sun to the planet sweeps out equal areas in equal times. Kepler spent many years verifying these laws and formulating a third law, which relates the planets' orbital periods and mean distances from the sun.

a. Plot the period time T versus the mean distance r using the following updated observational data.

planet	Period(day)	Mean distance from the sun (millions of kilometers)
Mercury	88	57.9
Venus	225	108.2
Earth	365	149.6
Mars	687	227.9
Jupiter	4,329	778.1
Saturn	10,753	1428.2
Uranus	30,660	2837.9
Neptune	60,150	4488.9

b. Assuming a relationship of the form

$$T = cr^a$$

determine the parameters C and a by plotting $\ln T$versus $\ln r$. Does the model seem reasonable? Try to formulate Kepler's third law.

3. For the following data, formulate the mathematical model that minimizes the largest deviation bewteen the data and the model $P = ae^{bx}$. If a computer is available, solve for the estimates of a and b.

t	7	14	21	28	35	42
p	8	41	133	250	280	297

4. Suppose the variable x_1 can assume any real value. Show that the following substitution using nonnegative variable x_2 and x_3 permits x_1 to assume any real value.

$$x_1 = x_2 - x_3, \quad \text{where } x_1 \text{ is unconstrained}$$

and

$$x_2 \geq 0 \quad \text{and} \quad x_3 \geq 0.$$

Thus, if a computer code allows only nonnegative variable, the substitution allows for solving the linear program in the variable x_2 and x_3 and then recovering the value of the variable x_1.

3.3 Applying the Least-Squares Criterion
3.4 Choosing a Best Model

1. Use Equations (3.5) and (3.6) to estimate the coefficients of the line $y = -ax + b$ such that the sum of the squared deviations between the line and the following data points is minimized.

a.

x	1.0	2.3	3.7	4.2	6.1	7.0
y	3.6	3.0	3.2	5.1	5.3	6.8

b.

x	29.1	48.2	72.2	92.0	118	140	165	199
y	0.0493	0.0821	0.123	0.154	0.197	0.234	0.274	0.328

c.

x	2.5	3.0	3.5	4.0	4.5	5.0	5.5
y	4.32	4.83	5.27	5.74	6.26	6.79	7.23

For each problem, compute D and d_{max} to bound c_{max}. Compare the results to your solutions to Problem 2 in Section 3.2.

2. Make an appropriate transformation to fit the model $P = ae^{bt}$ using Equation (3.4). Estimate a and b.

t	7	14	21	28	35	42
P	8	41	133	250	280	297

3. Find a model using the least-squares criterion either on the data or on the transformed data (as appropriate). Compare your results with the graphical fits obtained in the problem set 3.1 by computing the deviations, the maximum absolute deviation, and the sum of the squared deviations for each model. Find a bound on c_{max} if the model was fit using the least-squares criterion.

a. Problem $4a$ in Section 3.1   **b. Problem $4b$ in Section 3.1**

4. a. In the following data, W represents the weight of a fish (bass) and l represents its length. Fit the model $W = kl^3$ to the data using the least-squares criterion.

Length, l (in.)	**14.5**	**12.5**	**17.25**	**14.5**	**12.625**	**17.75**	**14.125**	**12.625**
Weight, W(oz)	**27**	**17**	**41**	**26**	**17**	**49**	**23**	**16**

b. In the following data, g represents the girth of a fish. Fit the model $W = klg^2$ to the data using the least-squares criterion.

Length, l (in.)	**14.5**	**12.5**	**17.25**	**14.5**	**12.625**	**17.75**	**14.125**	**12.625**
Girth, g (in.)	**9.75**	**8.375**	**11.0**	**9.75**	**8.5**	**12.5**	**9.0**	**8.5**
Weight, W (oz)	**27**	**17**	**41**	**26**	**17**	**49**	**23**	**16**

Assignment

Deadline: 12 December, 2022

1. Derive the equations that minimize the sum of the squared deviations between a set of data points and the quadratic model $y = c_1x + c_2x + c_3x$. Use the equations to find estimates of c1, c2, and c3 for the following set of data.

x	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
y	**0.06**	**0.12**	**0.36**	**0.65**	**0.96**

Compute D and d_{max} to bound C_{max}. Compare the results with your solution to problem 3 in Section 3.2.

2. A general rule for computing a person's weight is as follows: For a female, multiply the height in inches by 3.5 and subtract 108; for a male, multiply the height in inches by 4.0 and subtract 128. If the person is small bone-structured, adjust this computation by subtracting 10%; for a large bone-structured person, add 10%. No adjustment is made for an average-size person. Gather data on the weight versus height of people of differing age, size, and gender. Using Equation (3.4), fit a straight line to your data for males and another straight line to your data for females. What are the slopes and intercepts of those lines? How do the results compare with the general rule?

3. Write a computer program that finds the least-squares estimates of the coefficients in the following models.

a.$y = ax^2 + bx + c$

b. $y = ax^n$

4. Write a computer program that computes the deviation from the data points and any model that the user enters. Assuming that the model was fitted using the least-squares criterion, compute D and d_{max}. Output each data point, the deviation from each data point, D, d_{max}, and the sum of the squared deviations.