HW5

Homework 5

Please answer the following questions about empirical modeling.

Question 1:

In 1976, Marc and Helen Bornstein studied the pace of life.[2] To see if life becomes more hectic as the size of the city becomes larger, they systematically observed the mean time required for pedestrians to walk 50 feet on the main streets of their cities and towns. In Table 4.5, we present some of the data they collected. The variable P represents the population of the town or city, and the variable V represents the mean velocity of pedestrians walking the 50 feet. Problems 1–5 are based on the data in Table 4.5.

1. Fit the model $V = CP^a$ to the "pace of life" data in Table 4.5. Use the transformation $\log V = a \log P + \log C$. Plot $\log V$ versus $\log P$. Does the relationship seem reasonable?
   a. Make a table of $\log P$ versus $\log V$.
   b. Construct a scatterplot of your log–log data.
   c. Eyeball a line l onto your scatterplot.
   d. Estimate the slope and the intercept.
   e. Find the linear equation that relates $\log V$ and $\log P$.
   f. Find the equation of the form $V = CP^a$ that expresses V in terms of P.
2. Graph the equation you found in Problem 1f superimposed on the original scatterplot.

Table 4.5 Population and mean velocity over a 50-foot course, for 15 locations*

Location	Population P	Mean velocity V (ft/sec)
(1) Brno, Czechoslovakia	341,948	4.81
(2) Prague, Czechoslovakia	1,092,759	5.88
(3) Corte, Corsica	5,491	3.31
(4) Bastia, France	49,375	4.90
(5) Munich, Germany	1,340,000	5.62
(6) Psychro, Crete	365	2.76
(7) Itea, Greece	2,500	2.27
(8) Iraklion, Greece	78,200	3.85
(9) Athens, Greece	867,023	5.21
(10) Safed, Israel	14,000	3.70
(11) Dimona, Israel	23,700	3.27
(12) Netanya, Israel	70,700	4.31
(13) Jerusalem, Israel	304,500	4.42
(14) New Haven, U.S.A.	138,000	4.39
(15) Brooklyn, U.S.A.	2,602,000	5.05

*Bornstein data.

3. Using the data, a calculator, and the model you determined for V (Problem 1f), complete Table 4.6.
4. From the data in Table 4.6, calculate the mean (i.e., the average) of the Bornstein errors $|V_{observed} - V_{predicted}|$. What do the results suggest about the merit of the model?

Table 4.6 Observed mean velocity for 15 locations

Location*	Observed velocity V	Predicted velocities
1	4.81	
2	5.88	
3	3.31	
4	4.90	
5	5.62	
6	2.76	
7	2.27	
8	3.85	
9	5.21	
10	3.70	
11	3.27	
12	4.31	
13	4.42	
14	4.39	
15	5.05	

*For location names, see Table 4.5.

Question 2:

(a)

Consider the "pace of life" data from Problem 1, Section 4.1. Consider fitting a 14th-order polynomial to the data. Discuss the disadvantages of using the polynomial to make predictions. If a computer is available, determine and graph the polynomial.

(b)

In the following data, X represents the diameter of a ponderosa pine measured at breast height, and Y is a measure of volume—number of board feet divided by 10. Make a scatterplot of the data. Discuss the appropriateness of using a 13th-degree polynomial that passes through the data points as an empirical model. If you have a computer available, fit a polynomial to the data and graph the results.

X	17	19	20	22	23	25	31	32	33	36	37	38	39	41
Y	19	25	32	51	57	71	141	123	187	192	205	252	248	294

Question 3:

For the data sets in Problems 1–4, construct a divided difference table. What conclusions can you make about the data? Would you use a low-order polynomial as an empirical model? If so, what order?

1.

x	0	1	2	3	4	5	6	7
y	2	8	24	56	110	192	308	464

2.

x	0	1	2	3	4	5	6	7
y	23	48	73	98	123	148	173	198

3.

x	0	1	2	3	4	5	6	7
y	7	15	33	61	99	147	205	273

4.

x	0	1	2	3	4	5	6	7
y	1	4.5	20	90	403	1808	8103	36,316

Due: 10:00am 5. April, **. Please email your homework to TA.**