Key

Problem Set A

The crop yield of a hectare of land (10,000 square meters) depends on the amount of fertilizer applied. The table below gives data about the amount of fertilizer applied (in hundreds of kilograms per hectare) and the crop yield (in tons per hectare).

Fertilizer applied (hundreds of kg/hectare)	Crop yield (tons per hectare)
0	1
0.5	2
1	4
1.5	6
2	7
2.5	7.5
3	7.7
3.5	7.8
4	7.85
4.5	7.88
5	7.9

A cubic regression is used to model this data.

a. Identify the independent and dependent variable. Enter the data into your calculator.

independent: amount of fertilizer applied (in hundreds of kg/ha)
dependent: crop yield (tons per ha)

b. Find the cubic regression model where x is the amount of fertilizer applied and y is the predicted crop yield.

$$\hat{y} = 0.055664x^3 - 0.92643x^2 + 4.6839x + 0.5405$$

c. What does the model predict will be the crop yield when 3500 kilograms of fertilizer are applied?

$x = 3.5$ $\hat{y} = 7.972$ tons per hectare

d. Calculate the residual for $x = 3.5$. Does the model give an overestimate or an underestimate for the true crop yield when 3500 kilograms of fertilizer are applied? Write a sentence to explain.

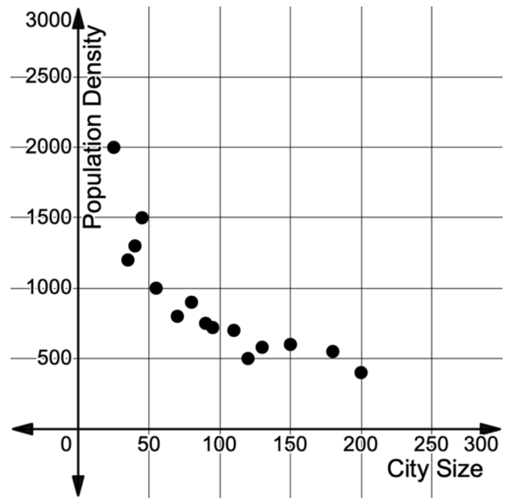
actual − predicted $= 7.8 - 7.972 = -0.172$ tons/ha

The model overestimates the true crop yield by 0.172 tons when 3500 kg of fertilizer are applied.

Problem Set B

To study the relationship between a city's size (in square miles) and a city's population density (number of people per square mile), data was collected for 15 randomly selected U.S. cities. The data is shown in the table below. A scatterplot is also given.

City Size (square miles)	Population Density (people per square mile)
45	1500
70	800
25	2000
120	500
80	900
200	400
35	1200
150	600
90	750
180	550
110	700
55	1000
40	1300
95	720
130	580



a. Explain why a logarithmic model would be appropriate to model this relationship. Enter the data into your calculator.

The population density seems to increase without bound as the city size decreases to 0. The asymptotic behavior and the general pattern of the other values resembles a reflected log graph.

b. Find the logarithmic regression model, where x is the city size, and y is the population density.

$\hat{y} = 3749.43 - 649.382 \ln x$

c. What is the population density, to the nearest person, predicted by the logarithmic model for a city of 200 square miles?

$x = 200$ square miles

$\hat{y} = 309$ people per square mile

d. Calculate the residual for the city of 200 square miles. Does the model give an overestimate or an underestimate for the true population density for a city of size 200 square miles? Write a sentence to explain.

actual − predicted $= 400 - 309 = 91$ people/mi^2

The model underestimates the true population density for a city of size 200 square miles by 91 people per square mile.

Problem Set C

After a pollutant is released in a natural environment, it undergoes a process of degradation and decomposition, causing the concentration of the pollutant to decrease over time. This decay can be modeled by an exponential function. The table provides data for the pollutant concentration (in parts per million, ppm) in a body of water in the seven days following a pollution incident.

Days	Pollutant Concentration (ppm)
0	10
1	8.5
2	7
3	5.5
4	4
5	3
6	2.5
7	2

a. Identify the independent and dependent variable. Enter the data into your calculator.

Independent: Days since pollution incident

Dependent: pollutant concentration (ppm)

b. Find an exponential regression model for this data, where x is the number of days since the pollution incident, and y is the pollution concentration.

$$\hat{y} = 10.2738(0.8111)^x$$

c. What does the model predict will be the pollutant concentration after 3 days?

$x = 3$

$\hat{y} = 5.482$ ppm

d. Calculate the residual for $x = 3$. Does the model give an overestimate or an underestimate for the true pollutant concentration after 3 days? Write a sentence to explain.

actual − predicted $= 5.5 - 5.482 = 0.018$ ppm

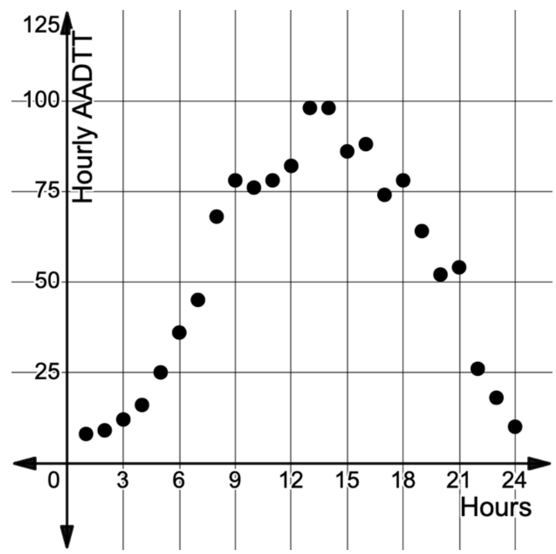
The model underestimates the true pollutant concentration after 3 days by 0.018 ppm.

Math Medic

Problem Set D

Traffic patterns are heavily influenced by the time of day. The Federal Highway Administration collects data on the Annual Average Daily Truck Traffic (AADTT) for a particular stretch of road. Selections of this data are displayed in the table below. The input variable represents the xth hour of the day ($x = 1$ represents the first hour of the day, between 12:00 AM and 1:00 AM, $x = 2$ represents the second hour of the day, between 1:00 AM and 2:00 AM, etc.). The output variable, y, is the number of trucks that cross a certain stretch of road in that hour, known as the Hourly AADTT. Truck traffic is monitored and collected over one whole year and then averaged to create the values given in the table.

x	y
1	8
3	12
4	16
5	25
8	68
10	76
11	78
13	98
15	86
18	78
20	52
21	54
24	10



a. Explain why a sinusoidal model would be appropriate to model this relationship. Then enter the data into your calculator.

The traffic pattern represents cyclical behavior that will repeat roughly every 24 hours.

b. Find the sinusoidal regression model where x is the hour of the day and y is the hourly AADTT.

$\hat{y} = 45.4532 \sin(0.23x - 1.6164) + 48.4204$

c. What is the hourly AADTT, to the nearest truck, predicted by the sinusoidal model for the time between 8 PM and 9 PM?

$x = 21$

$\hat{y} = 45$ trucks

d. Calculate the residual for the value found in part b. Does the model give an overestimate or an underestimate for the true hourly AADTT between 8 PM and 9 PM? Write a sentence to explain.

actual − predicted = 54 − 45 = 9 trucks

The model underestimates the true hourly traffic between 8 PM and 9 PM by about 9 trucks.

Math Medic