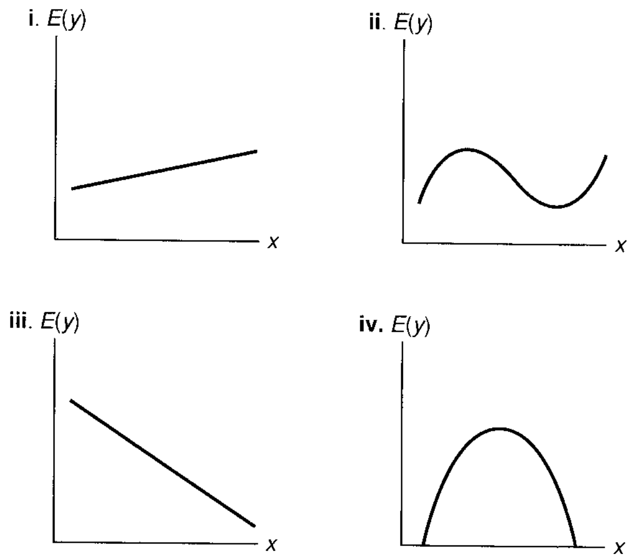
A S S I G N M E N T 3

Due date: 6:00pm, Friday, 26 November, 2021

The data files are available in Moodle.

1. **Order of polynomials.** The accompanying graphs depict pth-order polynomials for one independent variable.



   (a) For each graph, identify the order of the polynomial.

   (b) Using the parameters β_0, β_1, β_2, etc., write an appropriate model relating $E(y)$ to x for each graph.

   (c) The signs (+ or −) of many of the parameters in the models of part b can be determined by examining the graphs. Give the signs of those parameters that can be determined.

2. **Deep space survey of quasars.** Refer to the Astronomical Journal (July 1995) study of quasars detected by a deep space survey, Assignment 2, problem 6. Recall that several quantitative independent variables were used to model the quasar characteristic, rest frame equivalent width (y). The data for 25 quasars are saved in the QUASAR file.

   (a) Write a complete second-order model for y as a function of redshift (x_1), lineflux (x_2), and AB1450 (x_4).

   (b) Fit the model, part a, to the data using a sta- tistical software package. Is the overall model statistically useful for predicting y?

   (c) Conduct a test to determine if any of the curvilinear terms in the model, part a, are statistically useful predictors of y.

3. **Tire wear and pressure.** Suppose you want to use the coding system for observational data to fit a second-order model to the tire pres- sure–automobile mileage data, which are shown in the table.

TIRES2

PRESSURE	MILEAGE
x, pounds per square inch	y, thousands
30	29
31	32
32	36
33	38
34	37
35	33
36	26

(a) Give the equation relating the coded variable u to pressure, x, using the coding system for observational data.

(b) Calculate the coded values, u.

(c) Calculate the coefficient of correlation r between the variables x and x^2.

(d) Calculate the coefficient of correlation r between the variables u and u^2. Compare this value to the value computed in part c.

(e) Fit the model $E(y) = \beta - 0 + \beta_1 u + \beta_2 u_2$ using available statistical software. Interpret the results.

4. **Expert testimony in homicide trials of battered women.** Refer to the Duke Journal of Gender Law and Policy (Summer 2003) study of the impact of expert testimony on the outcome of homicide trials involving battered woman syndrome. Recall that multiple regression was employed to model the likelihood of changing a verdict from not guilty to guilty after deliberations, y, as a function of juror gender (male or female) and expert testimony given (yes or no).

(a) Write a main effects model for $E(y)$ as a function of gender and expert testimony. Interpret the β coefficients in the model.

(b) Write an interaction model for $E(y)$ as a function of gender and expert testimony. Interpret the β coefficients in the model.

(c) Based on data collected on individual juror votes from past trials, the article reported that "when expert testimony was present, women jurors were more likely than men to change a verdict from not guilty to guilty after deliberations." Assume that when no expert testimony was present, male jurors were more likely than women to change a verdict from not guilty to guilty after deliberations. Which model, part a or part b, hypothesizes the relationships reported in the article? Illustrate the model with a sketch.

5. **Cooling method for gas turbines.** Refer to the Journal of Engineering for Gas Turbines and Power (January 2005) study of a high-pressure inlet fogging method for a gas turbine

engine. Consider a model for heat rate (kilojoules per kilowatt per hour) of a gas turbine as a function of cycle speed (revolutions per minute), cycle pressure ratio, and a qualitative predictor, engine type, at three levels (traditional, advanced, and aeroderivative).

(a) Write a complete second-order model for heat rate (y) as a function of cycle speed, cycle pressure ratio, and engine type.

(b) Demonstrate that the model graphs out as three second-order response surfaces, one for each level of engine type.

(c) Fit the model to the data in the GASTUR- BINE file and give the least squares prediction equation.

(d) Conduct a global F-test for overall model adequacy.

(e) Conduct a test to determine whether the second-order response surface is identical for each level of engine type.

6. **Lead in fern moss.** A study of the atmospheric pollution on the slopes of the Blue Ridge Mountains (Tennessee) was conducted. The file LEAD- MOSS contains the levels of lead found in 70 fern moss specimens (in micrograms of lead per gram of moss tissue) collected from the mountain slopes, as well as the elevation of the moss specimen (in feet) and the direction (1 if east, 0 if west) of the slope face. The first five and last five observations of the data set are listed in the table.

LEADMOSS

SPECIMEN	LEAD LEVEL	ELEVATION	SLOPE FACE
1	3.475	2000	0
2	3.359	2000	0
3	3.877	2000	0
4	4.000	2500	0
5	3.618	2500	0
⋮	⋮	⋮	⋮
66	5.413	2500	1
67	7.181	2500	1
68	6.589	2500	1
69	6.182	2000	1
70	3.706	2000	1

Source: Schilling, J. "Bioindication of atmospheric heavy metal deposition in the Blue Ridge using the moss, *Thuidium delicatulum*," Master of Science thesis, Spring 2000.

(a) Write the equation of a first-order model relat- ing mean lead level, $E(y)$, to elevation (x_1) and slope face (x_2). Include interaction between elevation and slope face in the model.

(b) Graph the relationship between mean lead level and elevation for the different slope faces that is hypothesized by the model, part a.

(c) In terms of the β's of the model, part a, give the change in lead level for every one foot increase in elevation for moss specimens on the east slope.

(d) Fit the model, part a, to the data using an available statistical software package. Is the overall model statistically useful for predicting lead level? Test using $\alpha = .10$.

(e) Write the equation of the complete second- order model relating mean lead level, $E(y)$, to elevation (x_1) and slope face (x_2)

7. **Using glass to encapsulate waste.** Since glass is not subject to radiation damage, encapsulation of waste in glass is considered to be one of the most promising solutions to the problem of low- level nuclear waste in the environment. However, glass undergoes chemical changes when exposed to extreme environmental conditions, and certain of its constituents can leach into the surroundings. In addition, these chemical reactions may weaken the glass. These concerns led to a study undertaken jointly by the Department of Materials Science and Engineering at the University of Florida and the U.S. Department of Energy to assess the utility of glass as a waste encapsulant material.† Corrosive chemical solutions (called corrosion baths) were prepared and applied directly to glass samples containing one of three types of waste (TDS- 3A, FE, and AL); the chemical reactions were observed over time. A few of the key variables measured were

   y = Amount of silicon (in parts per million) found in solution at end of experiment. (This is both a measure of the degree of breakdown in the glass and a proxy for the amount of radioactive species released) into the environment.

$$x_1 = \text{Temperature (°C) of the corrosion bath}$$

$$x_2 = \begin{cases} 1 & \text{if waste type TDS-3A} \\ 0 & \text{if not} \end{cases} \qquad x_3 = \begin{cases} 1 & \text{if waste type FE} \\ 0 & \text{if not} \end{cases}$$

   Waste type AL is the base level. Suppose we want to model amount y of silicon as a function of temperature (x_1) and type of waste (x_2, x_3).

   (a) Write a model that proposes parallel straight-line relationships between amount of silicon and temperature, one line for each of the three waste types.

   (b) Add terms for the interaction between temperature and waste type to the model of part a.

   (c) Refer to the model of part b. For each waste type, give the slope of the line relating amount of silicon to temperature.

   (d) Explain how you could test for the presence of temperature–waste type interaction.

8. **Modeling product sales.** A company wants to model the total weekly sales, y, of its product as a function of the variables packaging and location. Two types of packaging, P_1 and P_2, are used in each of four locations, L_1, L_2, L_3, and L_4.

   (a) Write a main effects model to relate $E(y)$ to packaging and location. What implicit assump- tion are we making about the interrelationships between sales, packaging, and location when we use this model?

   (b) Now write a model for $E(y)$ that includes inter- action between packaging and location. How many parameters are in this model (remember to include β_0)? Compare this number to the number of packaging–location combinations being modeled.

   (c) Suppose the main effects and interaction mod- els are fit for 40 observations on weekly sales. The values of SSE are

$$\text{SSE for main effects model} = 422.36$$

$$\text{SSE for interaction model} = 346.65$$

Determine whether the data indicate that the interaction between location and packaging is important in estimating mean weekly sales. Use $\alpha = .05$. What implications does your conclusion have for the company's marketing strategy?