12.53 Forecasting movie revenues with Twitter. Refer to the *IEEE International Conference on Web Intelligence and Intelligent Agent Technology* (2010) study on using the volume of chatter on Twitter.com to forecast movie box office revenue, Exercise 12.20 (p. 671). The researchers modeled a movie's opening weekend box office revenue (y) as a function of tweet rate (x_1) and ratio of positive to negative tweets (x_2) using a first-order model.

a. Write the equation of an interaction model for $E(y)$ as a function of x_1 and x_2.

b. In terms of the β's in the model, part **a**, what is the change in revenue (y) for every 1-tweet increase in the tweet rate (x_1), holding PN-ratio (x_2) constant at a value of 2.5?

c. In terms of the β's in the model, part **a**, what is the change in revenue (y) for every 1-tweet increase in the tweet rate (x_1), holding PN-ratio (x_2) constant at a value of 5.0?

d. In terms of the β's in the model, part **a**, what is the change in revenue (y) for every 1-unit increase in the PN-ratio (x_2), holding tweet rate (x_1) constant at a value of 100?

e. Give the null hypothesis for testing whether tweet rate (x_1) and PN-ratio (x_2) interact to affect revenue (y).

12.55 Factors that affect an auditor's judgment. A study was conducted to determine the effects of linguistic delivery style and client credibility on auditors' judgments (*Advances in Accounting and Behavioral Research*, 2003). Each of 200 auditors performed an analytical review of a client's financial statement. The researchers gave the auditors different information on the client's credibility and the linguistic delivery style of the client's explanation. Each auditor then provided an assessment of the likelihood that the client's explanation accounts for the fluctuation in the financial statement. The three variables of interest—credibility (x_1), linguistic delivery style (x_2), and likelihood (y)—were all measured on a numerical scale. Regression analysis was used to fit the interaction model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$. The results are summarized in the table at the bottom of the page.

a. Interpret the phrase "client credibility and linguistic delivery style interact" in the words of the problem.

b. Give the null and alternative hypotheses for testing the overall adequacy of the model.

c. Conduct the test suggested in part **b**, using the information in the table.

d. Give the null and alternative hypotheses for testing whether client credibility and linguistic delivery style interact.

e. Conduct the test suggested in part **d**, using the information in the table.

f. The researchers estimated the slope of the likelihood–linguistic delivery style line at a low level of client credibility $(x_1 = 22)$. Obtain this estimate and interpret it in the words of the problem.

g. The researchers also estimated the slope of the likelihood–linguistic delivery style line at a high level of client credibility $(x_1 = 46)$. Obtain this estimate and interpret it in the words of the problem.

12.64 MINITAB was used to fit the complete second-order model

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$$

to $n = 39$ data points. (See the printout on p. 694)

a. Is there sufficient evidence to indicate that at least one of the parameters β_1, β_2, β_3, β_4, and β_5 is nonzero? Test, using $\alpha = .05$.


```
The regression equation is
Y = -24.56 + 1.12 X1 + 27.99 X2 - 0.54 X1X2 - 0.004 X1SQ + 0.002 X2SQ


Predictor        Coef     SE Coef       T      P
Constant       -24.563       6.531   -3.76  0.001
X1             1.19848      0.1103   10.86  0.000
X2              27.988      79.489    0.35  0.727
X1X2           -0.5397      1.0338   -0.52  0.605
X1SQ           -0.0043      0.0004  -10.74  0.000
X2SQ            0.0020      0.0033    0.60  0.550


S = 2.762      R-Sq = 79.7%    R-Sq(adj) = 76.6%

Analysis of Variance

Source           DF      SS      MS       F      P
Regression        5  989.30  197.86   25.93  0.000
Residual Error   33  251.81    7.63
Total            38 1241.11
```

MINITAB output for Exercise 12.64

b. Test H_0: $\beta_4 = 0$ against H_a: $\beta_4 \neq 0$. Use $\alpha = .01$.

c. Test H_0: $\beta_5 = 0$ against H_a: $\beta_5 \neq 0$. Use $\alpha = .01$.

d. Use graphs to explain the consequences of the tests in parts **b** and **c**.

12.73 Estimation of urban population by means of satellite images. Refer to the *Geographical Analysis* (Jan. 2007) study that demonstrated the use of satellite image maps for estimating urban population, presented in Exercise 12.18 (p. 670). A first-order model for census block population density (y) was fit as a function of the proportion of a block with low-density residential areas (x_1) and the proportion of a block with high-density residential areas (x_2). Now consider a complete second-order model for y.

a. Write the equation of the model.

b. Identify the terms in the model that allow for curvilinear relationships.

12.78 Failure times of silicon wafer microchips. Researchers at National Semiconductor experimented with tin-lead solder bumps used to manufacture silicon wafer integrated circuit chips (International Wafer Level Packaging Conference, Nov. 3–4, 2005). The failure times of the microchips (in hours) were determined at different solder temperatures (degrees Celsius). The data for one experiment are given in the following table. The researchers want to predict failure time (y) based on solder temperature (x).

WAFER

a. Construct a scatterplot for the data. What type of relationship, linear or curvilinear, appears to exist between failure time and solder temperature?

b. Fit the model, $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$, to the data. Give the least squares prediction equation.

c. Conduct a test to determine if there is upward curvature in the relationship between failure time and solder temperature. (Use $\alpha = .05$.)

Temperature (°C)	Time to Failure (hours)
165	200
162	200
164	1,200
158	500
158	600
159	750
156	1,200
157	1,500
152	500
147	500
149	1,100
149	1,150
142	3,500
142	3,600
143	3,650
133	4,200
132	4,800
132	5,000
134	5,200
134	5,400
125	8,300
123	9,700

12.195 Revenues of popular movies. The *Internet Movie Database* (www.imdb.com) monitors the gross revenues of all major motion pictures. The table on p. 763 gives both the domestic (United States and Canada) and international gross revenues for a sample of 22 popular movies.

D IMDB

a. Write a first-order model for international gross revenues y as a function of domestic gross revenues x.

b. Write a second-order model for international gross revenues y as a function of domestic gross revenues x.

c. Construct a scatterplot of these data. Which of the models appears to be a better choice for explaining variation in international gross revenues?

d. Fit the model of part **b** to the data and investigate its usefulness. Is there evidence of a curvilinear relationship between international and domestic gross revenues? Test, using $\alpha = .05$.

e. On the basis of your analysis in part **d**, which of the two models better explains the variation in international gross revenues?

12.84 MINITAB was used to fit the model

NW

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where

$$x_1 = \begin{cases} 1 & \text{if level 2} \\ 0 & \text{if not} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{if level 3} \\ 0 & \text{if not} \end{cases}$$

to $n = 15$ data points. The results are shown in the accompanying MINITAB printout.

```
The regression equation is
Y = 80.0 + 16.8 X1 + 40.4 X2

Predictor      Coef  SE Coef      T      P
Constant     80.000    4.082  19.60  0.000
X1           16.800    5.774   2.91  0.013
X2           40.400    5.774   7.00  0.000

S = 9.129     R-Sq = 80.5%    R-Sq(adj) = 77.2%

Analysis of Variance

Source           DF       SS       MS      F      P
Regression        2   4118.9   2059.5  24.72  0.000
Residual Error   12   1000.0     83.3
Total            14   5118.9
```

a. Report the least squares prediction equation.

b. Interpret the values of β_1 and β_2.

c. Interpret the following hypotheses in terms of μ_1, μ_2, and μ_3:

H_0: $\beta_1 = \beta_2 = 0$

H_a: At least one of the parameters β_1 and β_2 differs from 0

d. Conduct the hypothesis test of part **c.**

12.89 Impact of race on football card values. University of Colorado sociologists investigated the impact of race on the value of professional football players' "rookie" cards (*Electronic Journal of Sociology*, 2007). The sample consisted of 148 rookie cards of National Football League (NFL) players who were inducted into the Football Hall of Fame. The price of a card (in dollars) was modeled as a function of several qualitative independent variables: race of player (black or white), availability of the card (high or low), and position of the player (quarterback, running back, wide receiver, tight end, defensive lineman, linebacker, defensive back, or offensive lineman).

a. Create the appropriate dummy variables for each of the qualitative independent variables.

b. Write a model for price (y) as a function of race. Interpret the β's in the model.

c. Write a model for price (y) as a function of the availability of the card. Interpret the β's in the model.

d. Write a model for price (y) as a function of the player's position. Interpret the β's in the model.

12.96 Homework assistance for college students. Do college professors who provide their students with assistance on homework help improve student grades? This was the research question of interest in the *Journal of Accounting Education* (Vol. 25, 2007). A sample of 175 accounting students took a pretest on a topic not covered in class, then each was given a homework problem to solve on the same topic. The students were assigned to one of three homework assistance groups. Some students received the completed solution, some were given check figures at various steps of the solution, and some received no help at all. After finishing the homework, the students were all given a posttest on the subject. The dependent variable of interest was the knowledge gain (or test score improvement). These data are saved in the **ACCHW** file.

a. Propose a model for the knowledge gain (y) as a function of the qualitative variable, homework assistance group.

b. In terms of the β's in the model, give an expression for the difference between the mean knowledge gains of students in the "completed solution" and "no help" groups.

c. Fit the model to the data and give the least squares prediction equation.

d. Conduct the global F-test for model utility using $\alpha = .05$. Interpret the results practically.

Applying the Concepts—Advanced

12.99 Community responses to a violent crime. How communities respond to a disaster or a violent crime was the subject of research published in the *American Journal of Community Psychology* (Vol. 44, 2009). Psychologists at the University of California tracked monthly violent crime incidents in two Texas cities, Jasper and Center, before and after the murder of a Jasper citizen that had racial overtones and heavy media coverage. (Center, Texas, was selected as comparison city since it had roughly the same population and racial makeup as Jasper.) Using monthly data on violent crimes, the researchers fit the regression model:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2,$$

where y = violent crime rate (number of crimes per 1,000 population), $x_1 = \{1$ if Jasper, 0 if Center$\}$, and $x_2 = \{1$ if after the murder, 0 if before the murder$\}$.

a. In terms of the β's in the model, what is the mean violent crime rate in Center, Texas, for months following the murder?

b. In terms of the β's in the model, what is the mean violent crime rate in Jasper, Texas, for months following the murder?

c. For months following the murder, find the difference between the mean violent crime rate for Jasper and Center. (Use your answers to parts **a** and **b**.)

d. Repeat part **c** for months before the murder.

e. Note that the differences, parts **c** and **d**, are not the same. Explain why this illustrates the notion of interaction between x_1 and x_2.

f. A test for $H_0: \beta_3 = 0$ yielded a p-value $< .001$. Using $\alpha = .01$, interpret this result.

g. The regression resulted in the following β-estimates: $\hat{\beta}_1 = -429, \hat{\beta}_2 = -169, \hat{\beta}_3 = 255$. Use these estimates to illustrate that average monthly violent crime decreased in Center after the murder but increased in Jasper.

12.102 Write a model that relates $E(y)$ to two independent variables, one quantitative and one qualitative (at four levels). Construct a model that allows the associated response curves to be second order but does not allow for interaction between the two independent variables.

12.103 Consider the model

NW

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

where x_1 is a quantitative variable and x_2 and x_3 are dummy variables describing a qualitative variable at three levels, using the coding scheme

$$x_2 = \begin{cases} 1 & \text{if level 2} \\ 0 & \text{otherwise} \end{cases} \qquad x_3 = \begin{cases} 1 & \text{if level 3} \\ 0 & \text{otherwise} \end{cases}$$

The resulting least squares prediction equation is

$$\hat{y} = 44.8 + 2.2x_1 + 9.4x_2 + 15.6x_3$$

a. What is the response line (equation) for $E(y)$ when $x_2 = x_3 = 0$? When $x_2 = 1$ and $x_3 = 0$? When $x_2 = 0$ and $x_3 = 1$?

b. What is the least squares prediction equation associated with level 1? Level 2? Level 3? Plot these on the same graph.

12.105 **Reality TV and cosmetic surgery.** Refer to the *Body Image: An International Journal of Research* (Mar. 2010) study of the impact of reality TV shows on a college student's decision to undergo cosmetic surgery, Exercise 12.25 (p. 672). Consider the interaction model $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_4 + \beta_3 x_1 x_4$, where y = desire to have cosmetic surgery (25-point scale), $x_1 = \{1 \text{ if male}, 0 \text{ if female}\}$, and x_4 = impression of reality TV (7-point scale). The model was fit to the data and the resulting SAS printout appears at the bottom of the page.

IMAGE

a. Give the least squares prediction equation.

b. Find the predicted level of desire (y) for a male college student with an impression-of-reality-TV-scale score of 5.

c. Conduct a test of overall model adequacy. Use $\alpha = .10$.

d. Give a practical interpretation of R_a^2.

e. Give a practical interpretation of s.

f. Conduct a test (at $\alpha = .10$) to determine if gender (x_1) and impression of reality TV show (x_4) interact in the prediction of level of desire for cosmetic surgery (y).

g. Give an estimate of the change in desire (y) for every 1-point increase in impression of reality TV show (x_4) for female students.

h. Repeat part **g** for male students.

12.109 Whales entangled in fishing gear. Refer to the *Marine Mammal Science* (Apr. 2010) study of whales entangled in fishing gear, Exercise 12.87, (p. 703). Now consider a model for the length (y) of an entangled whale (in meters) that is a function of water depth of the entanglement (in meters) and gear type (set nets, pots, or gill nets).

a. Write a main-effects-only model for $E(y)$.

b. Sketch the relationships hypothesized by the model, part **a**. (*Hint:* Plot length on the vertical axis and water depth on the horizontal axis.)

c. Add terms to the model, part **a**, that include interaction between water depth and gear type. (*Hint:* Be sure to interact each dummy variable for gear type with water depth.)

d. Sketch the relationships hypothesized by the model, part **c**.

e. In terms of the β's in the model of part **c**, give the rate of change of whale length with water depth for set nets.

f. Repeat part **e** for pots.

g. Repeat part **e** for gill nets.

h. In terms of the β's in the model of part **c**, how would you test to determine if the rate of change of whale length with water depth is the same for all three types of fishing gear?

12.113 Workplace bullying and intention to leave. Workplace bullying is defined as harassment, persistent criticism, withholding of key information, spreading of rumors, or intimidation on the job. In *Human Resource Management Journal* (Oct. 2008), researchers employed multiple regression to examine whether perceived organizational support would moderate the relationship between workplace bullying and victims' intention to leave the firm. The dependent variable in the analysis, intention to leave (y), was measured on a quantitative scale. The two key independent variables in the study were bullying (x_1, measured on a quantitative scale) and perceived organizational support (measured qualitatively as "low," "neutral," or "high").

a. Set up the dummy variables required to represent perceived organizational support (POS) in the regression model.

b. Write a model for $E(y)$ as a function of bullying and POS that hypothesizes three parallel straight lines, one for each level of POS.

c. Write a model for $E(y)$ as a function of bullying and POS that hypothesizes three nonparallel straight lines, one for each level of POS.

d. The researchers discovered that the effect of bullying on intention to leave was greater at the low level of POS than at the high level of POS. Which of the two models, parts **b** and **c**, supports these findings?

c. For the test in part **b**, identify the complete and reduced models.
d. Portions of the MINITAB printouts for the two models are shown above. Find the values of SSE_R, SSE_C, and MSE_C on the printouts.
e. Compute the value of the test statistics for the test of part **b**.
f. Find the rejection region for the test of part **b**. Use $\alpha = .10$.
g. State the conclusion of the test in the words of the problem.

12.127 Study of supervisor-targeted aggression. "Moonlighters" are workers who hold two jobs at the same time. What are the factors that affect the likelihood of a moonlighting worker becoming aggressive toward his or her supervisor? This was the research question of interest in the *Journal of Applied Psychology* (July 2005). Completed questionnaires were obtained from $n = 105$ moonlighters, and the data were used to fit several multiple-regression models for supervisor-targeted aggression score (y). Two of the models (with R^2 values in parentheses) are shown in the table below.
a. Interpret the R^2 values for the models.
b. Give the null and alternative hypotheses for comparing the fits of Models 1 and 2.
c. Are the two models nested? Explain.
d. The nested F-test for comparing the two models resulted in $F = 42.13$ and p-value $< .001$. What can you conclude from these results?
e. A third model was fit, one that hypothesizes all possible pairs of interactions between self-esteem, history of aggression, interactional injustice at primary job, and abusive supervisor at primary job. Give the equation of this model (Model 3).
f. A nested F-test to compare Models 2 and 3 resulted in p-value $> .10$. What can you conclude from this result?

Models for Exercise 12.127

Model 1: $E(y) = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Gender}) + \beta_3(\text{Interaction injustice at secondary job}) + \beta_4(\text{Abusive supervisor at secondary job})$ $(R^2 = .101)$
Model 2: $E(y) = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Gender}) + \beta_3(\text{Interactional injustice at secondary job}) + \beta_4(\text{Abusive supervisor at secondary job}) + \beta_5(\text{Self-esteem}) + \beta_6(\text{History of aggression}) + \beta_7(\text{Interactional injustice at primary job}) + \beta_8(\text{Abusive supervisor at primary job})$ $(R^2 = .555)$

Applying the Concepts—Intermediate

12.128 Reality TV and cosmetic surgery. Refer to the *Body Image: An International Journal of Research* (Mar. 2010) study of the influence of reality TV shows on one's desire to undergo cosmetic surgery, Exercise 12.25 (p. 672). Recall that psychologists modeled desire to have cosmetic surgery (y) as a function of gender (x_1), self-esteem (x_2), body satisfaction (x_3), and impression of reality TV (x_4). The psychologists theorize that one's impression of reality TV will "moderate" the impact that each of the first three independent variables has on one's desire to have cosmetic surgery. If so, then x_4 will interact with each of the other independent variables.

IMAGE

a. Give the equation of the model for $E(y)$ that matches the theory.
b. Fit the model, part **a**, to the data saved in the **IMAGE** file. Evaluate the overall utility of the model.
c. Give the null hypothesis for testing the psychologists' theory.
d. Conduct a nested model F-test to test the theory. What do you conclude?

12.130 Glass as a waste encapsulant. The encapsulation of waste in glass is considered to be a promising solution to the problem of low-level nuclear waste. A study was 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.* 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 = Temperature (°C) of the corrosion bath

x_2 = 1 if waste type TDS-3A, 0 if not

x_3 = 1 if waste type FE, 0 if not

(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 types of waste.

12.131 Whales entangled in fishing gear. Refer to the *Marine Mammal Science* (Apr. 2010) study of whales entangled in fishing gear, Exercise 12.109 (p. 713). A first-order model for the length (y) of an entangled whale that is a function of water depth of the entanglement (x_1) and gear type (set nets, pots, or gill nets) is written as follows: $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$, where $x_2 = \{1 \text{ if set net}, 0 \text{ if not}\}$ and $x_2 = \{1 \text{ if pot}, 0 \text{ if not}\}$. Consider this model the complete model in a nested model F-test.

a. Suppose you want to determine if there are any differences in the mean lengths of entangled whales for the three gear types. Give the appropriate null hypothesis to test.

b. Refer to part **a**. Give the reduced model for the test.

c. Refer to parts **a** and **b**. If you reject the null hypothesis, what would you conclude?

d. Suppose you want to determine if the rate of change of whale length (y) with water depth (x_1) is the same for all three types of fishing gear. Give the appropriate null hypothesis to test.

e. Refer to part **d**. Give the reduced model for the test.

f. Refer to parts **d** and **e**. If you fail to reject the null hypothesis, what would you conclude?