1. **Reality TV and cosmetic surgery**

**How much influence does the media, especially reality television programs, have on one’s decision to undergo cosmetic surgery?** This was a question of interest to psychologists. They went on and published an article in **Body Image**: An International Journal of Research (March 2010). In the study, 170 college students answered questions about their impressions of reality TV shows featuring cosmetic surgery, level of self-esteem, satisfaction with one’s own body, and desire to have cosmetic surgery to alter one’s body.

**The variables measured and analyzed in the study were follows:**

|  |  |  |
| --- | --- | --- |
| Variable | Scale | Description of Scale |
| DESIRE | scale ranging from 5 to 25 | Where the higher the value, the greater the interest in having cosmetic surgery. |
| GENDER | 1 if male, 0 if female | Gender |
| SELFESTM | scale ranging from 4 to 40 | Where the higher the value, the greater the level of self-esteem. |
| BODYSAT | scale ranging from 1 to 9 | Where the higher the value, the greater the level of self-esteem |
| BODYSAT | scale ranging from 1 to 9 | Where the higher the value, the greater the satisfaction with one’s own body. |
| IMPREAL | scale ranging from 1 to 7 | Where the higher the value, the more one believes reality television shows featuring cosmetic surgery are realistic. |

The data for the study (simulated based on statistics reported in the journal article) saved in the BDYIMG file (included). The psychologists used multiple regression to model desire to have cosmetic surgery as a function of gender, self-esteem, body satisfaction, and impression of reality TV.

1. Fit the first order model to the data in the file. Give the least squares prediction equation.
2. Interpret the β estimates in the words of the problem.
3. Is the overall model statistically useful for predicting desire to have cosmetic surgery? Test using α = 0.01.
4. Which statistic, R2 or Ra2, is the preferred measure of model fit? Practically interpret the value of this statistic.
5. Conduct a test to determine whether desire to have cosmetic surgery decreases linearly as level of body satisfaction increases.   
   Use α = 0.05.
6. Find a 95% confidence interval for the coefficient of the variable IMPREAL. Practically interpret the result.
7. Give the least squares prediction equation for a model that includes only the gender and impression of reality TV and the interaction between them.
8. Find the predicted level of desire for a male college student with an impression-of-reality-TV-scale score of 5.
9. Give a practical interpretation of Ra2.
10. Give a practical interpretation of s.
11. Conduct a test (α = 0.10) to determine if gender and impression of reality TV show interact in the prediction of level of desire for cosmetic surgery.
12. Give an estimate of the change in desire for every 1-point increase in impression of reality TV show for female students.
13. Repeat part l for male students.

Suppose that psychologists theorized 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 impression of reality TV will interact with each of the other independent variables.

1. Give the equation for the model for E(y) that matches the theory.
2. Fit the model, part n, to the simulated data saved in the file. Evaluate the overall utility of the model.
3. Give the null hypothesis for testing the psychologists’ theory.
4. Conduct a nested model F-test to test the theory. What do you conclude?
5. Check the data for multicollinearity for the first model (part a). If you detect multicollinearity, what modifications to the model do you recommend?
6. Conduct a complete residual analysis for the model. Do you detect any violations of the assumptions? If so, what modifications to the model do you recommend.

2. **IQs and the Bell Curve**

The Bell Curve (Free Press, 1994), written by Richard Herrnstein and Charles Murray (H&M), is a controversial book about race, genes, IQ, and economic mobility. The book heavily employs statistics and statistical methodology in an attempt to support the authors’ positions on the relationships among these variables and their social consequences. The main theme of The Bell Curve can be summarized as follows:

1. Measured intelligence (IQ) is largely genetically inherited.
2. IQ is correlated positively with a variety of socio-economic status success measures, such as prestigious job, high annual income, and high educational attainment.
3. From 1 and 2, it follows that socioeconomic successes are largely genetically caused and therefore resistant to educational and environmental interventions (such as affirmative action).

The statistical methodology (regression) employed by the authors and the inferences derived from the statistics were critiqued in Chance (Summer 1995) and the Journal of the American Statistical Association (Dec. 1995). The following are just a few of the problems with H&M’s use of regression that are identified:

**Problem 1**

H & M consistently use a trio of independent variables – IQ, socioeconomic status, and age – in a series of first-order models designed to predict dependent social outcome variables such as income and unemployment. (Only on a single occasion are interaction terms incorporated.) Consider, for example, the model

**E(y) = β0 + β1x1 + β2x2 + β3x3**

Where y = income, x1 = IQ, x2 = socioeconomic status, and x3 = age. H & M employ t-tests on the individual β parameters to assess the importance of the independent variables. As with most of the models considered in The Bell Curve, the estimate of β1 in the income model is positive and statistically significant at α = 0.05, and the associated t-value is larger (in absolute value) than the t-values associated with the other independent variables. Consequently, H&M claim that IQ is a better predictor of income than the other two independent variables. No attempt was made to determine whether the model was properly specified or whether the model provides an adequate fit to the data.

**Problem 2**

In an appendix, the authors describe multiple regression as a “mathematical procedure that yields coefficients for each of the [independent variables], indicating how much of the change in [the dependent variable] can be anticipated for a given change in any particular [independent] variable, with all the others held constant.” Armed with this information and the fact that the estimate of β1 in the model above is positive, H&M infer that a high IQ necessarily implies (or causes) a high income, and a low IQ inevitably leads to a low income. (Cause-and –effect inferences like this are made repeatedly throughout the book.)

**Problem 3**

The title of the book refers to the normal distribution and its well-known “bell shaped” curve. There is a misconception among the general public that scores on intelligence tests (IQ) are normally distributed. In fact, most IQ scores have distributions that are decidedly skewed. Traditionally, psychologists and psychometricians have transformed these scores so that the resulting numbers have a precise normal distribution. H&M make a special point to do this. Consequently, the measure of IQ used in all the regression models is normalized (i.e., transformed so that the resulting distribution is normal), despite the fact that regression methodology does not require predictor (independent) variables to be normally distributed.

**Problem 4**

A variable that is not used as a predictor of social outcome in any of the models in The Bell Curve is level of education. H&M purposely omit education from the models, arguing that IQ causes education, not the other way around. Other researchers who have examined H&M’s data report that when education is included as an independent variable in the model, the effect of IQ on the dependent variable (say, income) is diminished.

* 1. Comment on each of the problems identified. Why do each of these problems cast a shadow on the inferences made by the authors?
  2. Using the variables specified in the model above, describe how you would conduct the multiple regression analysis. (Propose a more complex model and describe the appropriate model tests, including a residual analysis.)

3. **The Condo Sales**

This case involves investigation of the factors that affect the sale price of Oceanside condominium units. It represents an extension of an analysis of the same data by Herman Kelting. Although condo sale prices have increased dramatically over the past 20 years, the relationship between these factors and sale price remains about the same. Consequently, the data provide valuable insight into today’s condominium sales market.

The sales data were obtained for a mew oceanside condominium complex consisting of two adjacent and connected eight floor buildings. The complex contains 200 units of equal size (approximately 500 square feet each). The locations of the buildings relative to the ocean, the swimming pool, the parking lot, etc., are shown in the accompanying figure. There are several features of the complex that you should note:

1. The units facing south, called ocean view, face the beach and ocean. In addition, units in building 1 have a good view of the pool. Units to the rear of the building, called bay-view, face the parking lot and an area of land that ultimately borders a bay. The view from the upper floors of these units is primarily of woody, sandy terrain. The bay is very distant and barely visible.
2. The only elevator in the complex is located at the end of building 1, as are the office and the game room. People moving to or from the higher floor units in building 2 would likely use the elevator and move through the passages to their units. Thus, units on the higher floors an at a greater distance from the elevator would be less convenient; they would require greater effort in moving baggage, groceries, and so on and would be farther away from the game room, the office, and the swimming pool. These units also possess an advantage: there would be the least amount of traffic through the hallways in the area and hence they are the most private.
3. Lower-floor Oceanside units are most suited to active people; they open onto the beach, ocean, and pool. They are within easy reach of the game room, and they are easily reached from the parking area.
4. Checking the layout of the condominium complex, you discover that some of the units in the center of the complex, units ending in number 11 and 14, have part of their view blocked.
5. The condominium complex was completed at the time of the 1975 recession: sales were slow, and the developer was forced to sell most of the unit at auction approximately 18 months after opening. Consequently, the auction data was completely buyer specified and hence consumer oriented in contrast to most other real estate sales data that are, to a high degree, seller and broker specified.
6. Many unsold units in the complex were furnished by the developer and rented prior to the auction. Consequently, some of the units bid on and sold at auction had furniture, others did not.

This condominium complex is obviously unique. For example, the single elevator located at one end of the complex produces a remarkably high level of both inconvenience and privacy for the people occupying units on the top floors in building 2. Consequently, the developer is unsure of how the height of the unit (floor number), distance of the unit from the elevator, presence or absence of an ocean view, etc., affect the prices of the units sold at auction. To investigate these relationships, the following data (saved in the data file) were recorded for each of the 106 units sold at the auction:

* 1. ***Sale price*.** Measured in hundreds of dollars (adjusted for inflation).
  2. ***Floor height*.** The floor location of the unit; the variable levels are 1, 2,…., 8.
  3. ***Distance from elevator.*** This distance, measured along the length of the complex, is expressed in number of condominium units. An additional two units of distance was added to the units in building 2 to account for the walking distance in the connecting area between the two buildings. Thus, the distance of unit 105 from the elevator would be 3, and the distance between unit 113 and the elevator would be 9. The variable levels are 1, 2, …….., 15.
  4. ***View of ocean****.* The presence or absence of an ocean view is recorded for each unit and specified with a dummy variable (1 if the unit possesses an ocean view and 0 if not). Note that units not possessing an ocean view face the parking lot.
  5. *End unit.* We expect the partial reduction of view of end units on the ocean side (numbers ending in 11) to reduce their sale price. The ocean view of these end unit is partially blocked by building 2. This qualitative variable is also specified with a dummy variable (1 if the unit has a unit number ending in 11 and 0 if not).
  6. ***Furniture****.* The presence or absence of furniture is recorded for each unit and is represented with a single dummy variable (1 if the unit was furnished and o if not).

Your objective for this case is to build a regression model that accurately predicts the sale price of a condominium unit sold at auction. Prepare a professional document that presents the results of your analysis. Include graphs that demonstrate how each of the independent variables in your model affects auction price. The data set is included under the name CONDO.