**Quantitative Methods**

**List of Exercises N. 8**

**Selected Exercises from McClave (2014) – Chapter 12**

**12.3 Evaluating Overall Model Utility**

1. (2). Excel was used to fit the model E(y) = β0 + β1 x1+ β2 x2  to n = 20 data points, and the printout shown below was obtained:

1. What are the sample estimates of β0 , β1 and β2?
2. What is the least squares prediction equation?
3. Find SSE, MSE and s. Interpret the standard deviation in the context of the problem.
4. Test H0: β1 = 0 against H0: β1 ≠ 0. Use α = 5%.
5. Use a 95% confidence interval to estimate β2.
6. Find R2 and R2adjusted and interpret these values.
7. Find the test statistic for testing β1 = β2 = 0.
8. Find the observed significance level of the test, part g. Interpret the result.

Output: Regression equation is Y = 506.35 – 941.9 X1 – 429.1 X2

Predictor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictor | Coef | SE Coef | T | P |
| Constant | 506.346 | 45.17 | 11.21 | 0.000 |
| X1 | -941.900 | 275.08 | -3.42 | 0.003 |
| X2 | -429.060 | 379.83 | -1.13 | 0.274 |
| S = 94.251 |  |  |  |  |
| R2 =45.9% | Adjusted R2 = 39.6% |  |  |  |

Analysis of Variance:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | Sum of Squares | Mean Squares | F | Pr > F |
| Regression | 2 | 128329 | 64165 | 7.22 | 0.005 |
| Residual Error | 17 | 151016 | 8883 |  |  |
| Total | 19 | 279345 |  |  |  |

2. (17, BDYIMG). ***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 the question of interest to psychologists who 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 analyzed in the study were measured as follows: 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. 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 satisfaction with one’s own body; and 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) are saved in the file. Selected observations are listed in the next table. The psychologists used multiple regression to model desire to have cosmetic surgery (y) as a function of gender (X1), self-esteem (X2), body satisfaction (X3) and impression of reality TV (X4).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1st and last 5 observations | | | | | |
| Student | DESIRE | GENDER | SELFESTM | BODYSAT | IMPREAL |
| 1 | 11 | 0 | 24 | 3 | 4 |
| 2 | 13 | 0 | 20 | 3 | 4 |
| 3 | 11 | 0 | 25 | 4 | 5 |
| 4 | 11 | 1 | 22 | 9 | 4 |
| 5 | 18 | 0 | 8 | 1 | 6 |
| … | … | … | … | … | … |
| 166 | 18 | 0 | 25 | 3 | 5 |
| 167 | 13 | 0 | 26 | 4 | 5 |
| 168 | 9 | 1 | 13 | 5 | 6 |
| 169 | 14 | 0 | 20 | 3 | 2 |
| 170 | 6 | 1 | 27 | 8 | 3 |

1. Fit the first-order model, E(y) = β0 + β1 X1 + β2 X2+ β3 X3+ β4  X4, 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 predictiong desire to have cosmetic surgery? Test using α = 1%.
4. Which statistic R2 or R2 adjusted is the preferred measure of the 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 = 5%.
6. Find a 95% confidence interval for β4. Practically interpret the result.

**12.5 Interaction Models**

3. (43, BDYIMG). ***Reality TV and cosmetic surgery***. Refer to the Body Image: AN International Journal of Research (March 2010) study of the impact of reality TV shows on a college student’s decision to undergo cosmetic surgery. The data for the study (simulated based on statistics reported in the journal article) are saved in the file. Consider the interaction model E(y) = β0 + β1 X1 + β2 X4+ β3 X1 X4, where y = desire to have cosmetic surgery (25-point scale), X1 = 1 if male and 0 if female, X4 = impression of reality TV (7-point scale). The model was fit to the data and the resulting Excel printout appears below.

Model Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | R | R squared | Adj. R squared | St. Error of the estimate |
| 1 | 0.670 | 0.449 | 0.439 | 2.350 |

Anova

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Sum of Squares | DF | Mean Square | F | Sig |
| Regression | 747.001 | 3 | 249.000 | 45.86 | 0.000 |
| Residual | 916.787 | 166 | 5.523 |  |  |
| Total | 1663.788 | 169 |  |  |  |

Dependent Variable: DESIRE

Predictors: Constant, GENDER, IMPREAL, GENDER\_IMPREAL

Coefficients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig |
|  | B | Std. Error | Beta |
| Constant | 11.779 | 0.674 |  | 17.486 | 0.000 |
| Gender | -1.972 | 1.179 | -0.303 | -1.672 | 0.096 |
| Impreal | 0.585 | 0.162 | 0.258 | 3.617 | 0.000 |
| Gender\_Impreal | -0.553 | 0.276 | -0.378 | -2.004 | 0.047 |

1. Give the least squares prediction equation.
2. Find the predicted level of desire (y) for a male college student with an impression-of-reality-TV-sclae score of 5.
3. Coduct a test of overall model adequacy. Use α = 10%.
4. Give a practical interpretation of the adjusted R2
5. Give a practical interpretation of s.
6. Conduct a test (at α = 10%) to determine if gender (X1) and impression of reality TV show X4 interact in the prediction of the level of desire for cosmetic surgery (Y).

**12.7 Qualitative (Dummy) Variable Models**

4. (75). ***Do blondes raise more funds?*** During fundraising, does the physical appearance of the solicitor impact the level of capital raised? An economist at the University of Nevada-Reno designed an experiment to answer this question and published the results in Economic Letters (Vol. 100, 2008). Each in a sample of 955 households was contacted by a female solicitor and asked to contribute to the Center for Natural Hazards Mitigation Research. The level of contribution (in dollars) was recorded as well as their hair color of the solicitor (blond Caucasian, brunette Caucasian, or minority female).

1. Consider a model for the mean level of contribution, E(y), that allows for different means depending on the hair color of the solicitor. Create the appropriate number of dumy variables for the hair color. (use minority female as the base level).
2. Write the equation of the model, part a, incorporating the dummy variables.
3. In terms of the β’s in the model, what is the mean level of contribution for households contacted by a blond Caucasian solicitor?
4. In terms of the β’s in the model, what is the difference between the mean level of contribution for households contacted by a blond solicitor and those contacted by a minority female?
5. One theory posits that blond solicitors will achieve the highest mean contribution level, but that there will be no difference between the mean contribution levels attained by brunette Caucasian and minority females. If this theory is true, give the expected signs of the β‘s in the model.
6. The researcher found the β-estimate for the dummy variable blond Caucasian to be positive and significantly different from 0 (p-value < 0.01). The β-estimate for the dummy variable for brunette Caucasian was also positive, but not significantly different from 0 (p-value > 0.10). Do these results support the theory, part e?

5. (79, REPELL). ***Comparing mosquito repellents***. Which insect repellents protect best against mosquitoes? Consumer Reports (June 2000) tested 14 products that all claim to be an effective mosquito repellent. Each product was classified as either lotion / cream or aerosol / spray. The cost of the product (in dollars) was divided by the amount of the repellent needed to cover exposed areas of the skin (about 1/3 ounce) to obtain a cost-per-use value. Effectiveness was measured as the maximum number of hours of protection (in half-hour increments) provided when human testers exposed their arms to 200 mosquitoes. The data from the report are listed in the table below:

|  |  |  |  |
| --- | --- | --- | --- |
| Insect Repellent | Type | Cost / Use | Maximum Protection |
| Amway Hour Guard 12 | Lotion / Cream | 2.08 USD | 13.5 hours |
| Avon Skin-So-Soft | Aerosol / Spray | 0.67 | 0.5 |
| Avon BugGuard | Lotion / Cream | 1 | 2 |
| Ben’s Backyard Formula | Lotion / Cream | 0.75 | 7 |
| Bite Blocker | Lotion / Cream | 0.46 | 3 |
| BugOut | Aerosol / Spray | 0.11 | 6 |
| Cutter Skinsations | Aerosol / Spray | 0.22 | 3 |
| Cutter Unscented | Aerosol / Spray | 0.19 | 5.5 |
| Muskoll Ultra 6 Hours | Aerosol / Spray | 0.24 | 6.5 |
| Natrapel | Aerosol / Spray | 0.27 | 1 |
| Off! Deep Woods | Aerosol / Spray | 1.77 | 14 |
| Off! Skintastic | Lotion / Cream | 0.67 | 3 |
| Sawyer Deet Formula | Lotion / Cream | 0.36 | 7 |
| Repel Permanone | Aerosol / Spray | 2.75 | 24 |

1. Suppose you want to use repellent type to model the cost per use (y). Create the appropriate number of dummy variables for repellent type and write the model.
2. Fit he model, part a, to the data.
3. Give the null hypothesis for testing whether repellent type is a useful predictor of cost per use (y).
4. Conduct the test, part c, and give the appropriate conclusion. Use α = 10%.
5. Repeat parts a-d if the dependent variable is the maximum number of hours of protection (y).

**12.8 Models with Both Quantitative and Qualitative Variables**

6. (82). Consider a multiple regression model for a response y, with one quantitative independent variable X1 and one qualitative variable at 3 levels.

1. Write the first-order model that relates the mean response E(y) to the quantitative independent variable.
2. Add the main effect terms for the qualitative independent variable to the model of part a. Specify the coding scheme you use.
3. Add terms to the model of part b to allow for interaction between the quantitative and qualitative independent variables.
4. Under what circumstances will the response lines of the model in part c be parallel?
5. Under what circumstances will the model in part c have only one response line?

7. (92, WAGAP). ***Agreeableness, gender, and wages***. Do agreeable individuals get paid less, on average, than those who are less agreeable on the job? And is this gap greater for males than for females? These questions were addressed in the Journal of Personality and Social Psychology (Feb. 2012). Several variables were measured for each in a sample of individuals enrolled in the National Survey of Midlife Development in the US. Three of these variables are: (1) level of agreeableness score (where higher scores indicate a greater level of agreeableness), (2) gender (male or female), and (3) annual income (dollars). The researchers modeled mean income, E(y), as a function of both agreeableness score (X1) and a dummy variable for gender (X2 = 1 if male, 0 if female). Data for a sample of 100 individuals (simulated, based on information provided in the study) are saved in the file. The first 10 observations are listed in the accompanying table.

Data for First 10 individuals in Study:

|  |  |  |
| --- | --- | --- |
| Income | Agree Score | Gender |
| 44,770 | 3.0 | 1 |
| 51,480 | 2.9 | 1 |
| 39,600 | 3.3 | 1 |
| 24,370 | 3.3 | 0 |
| 15,460 | 3.6 | 0 |
| 43,730 | 3.8 | 1 |
| 48,330 | 3.2 | 1 |
| 25,970 | 2.5 | 0 |
| 17,120 | 3.5 | 0 |
| 20,140 | 3.2 | 0 |

Consider the model, E(y) = β0 + β1 X1 + β2 X2. The researchers theorized that for either gender, income would decrease as agreeableness score increases. If this theory is true, what is the expected sign of β1 in the model?

1. The researchers also theorized that the rate of the decrease of income with agreeableness score would be steeper for males than for females (i.e. the income gap between males and females would be greater the less agreeable the individuals are). Can this theory be tested using the model, part a? Explain.
2. Consider the interaction model, E(y) = β0 + β1 X1 + β2 X2 + β3 X1 X2. If the theory, part b, is true, give the expected sign of β1 and of β3.
3. Fit the model, part c, to the sample data. Check the signs of the estimated β coefficients. How do they compare to the expected values, part c?
4. Refer to the interaction model, part c. Give the null and the alternative hypothesis for testing whether the rate of decrease of income with agreeableness score is steeper for males than for females.
5. Conduct the test, part e. Use α = 5%. Is the researchers’ theory supported?

**12.11 + 12.12 Residual Analysis: Checking the Regression Assumptions + Some Pitfalls: Estimability, Multicollinearity, and Extrapolation**

8. (123). ***Women in top management***. The Journal of Organizational Culture, Communications and Conflict (July 2007) did a study on women in upper management positions at US firms. Monthly data (n = 252 months) were collected for several variables in an attempt to model the number of females in managerial positions (y). The independent variables included the number of females with a college degree (X1), the number of female high school graduates with no college degree (X2), the nuber of males in managerial positions (X3), the number of males with a college degree (X4), and the number of male high school graduates with no college degree (X5). The correlations are given in each part. Determine which of the correlations results in a potentia multicollinearity problem for the regression analysis.

1. The correlation relating number of females in managerial positions and number of females with a college degree: r = 0.983.
2. The correlation relating number of females in managerial positions and number of female high school graduates with no college degree: r = 0.074.
3. The correlation relating number of males in managerial positions and number of males with a college degree: r = 0.722.
4. The correlation relating number of males in managerial positions and number of male high school graduates with no college degree: r = 0.528.

9. (126, BDYIMG). ***Reality TV and cosmetic surgery.*** Refer to the Body Image: An International Journal of Research (March 2010) study of the influence of reality TV shows on one’s desire to undergo cosmetic surgery. Simulated data for the study is saved in the file. Fit the first-order model, E(y) = β0 + β1 X1 + β2 X2+ β3 X3+ β4  X4, where y = desire to have cosmetic surgery, X1 is a dummy variable for gender, X2 = level of self-esteem, X3 = level of body satisfaction, and X4 = impression of reality TV.

1. Check the data for multicollinearity. If you detect multicollinearity, what modifications to the model do you recommend?
2. 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?

10. (127, WELLS). Arsenic in groundwater. Refer to the Environmental Science & Technology (Jan. 2005) study of the reliability of a commercial kit to test for arsenic in groundwater. Fit a first-order model for arsenic level (y) as a function of latitude (X1), longitude (X2) and depth (X3) to data saved in the file. Conduct a residual analysis of the data. Based on the results, comment on each of the following:

1. assumption of mean error = 0
2. assumption of constant error variance
3. outliers
4. assumption of normally distributed errors
5. multicollinearity