Assignment Name: Homework 8

Student Name: Mo Pei

8. Market research was conducted for a national retail company to compare the relationship between sales and advertising during the warm spring and summer seasons as compared with the cool fall and winter seasons. The data were collected over a period of several years. They tested the model below, and found that it was significant.

$Y= \beta_0 + \beta_1 Season + \beta_2 Advertising + \varepsilon$

Where Season is 0 for warm months and 1 for cool months.

They wondered if they could do a better job of predicting the average sales if they used higher order terms.

1. If they wanted to test the complete second order model what would it be?

   Y= β0+ β1Season + β2Advertising + β3Advertising^2 + β4Advertising* Season +ε

   No beta for season square because the value of season ranges from 0 to 1 the same that of season^2.

2. Use the dataset [Market Research] to conduct a nested f-test to see whether higher order terms, as found in the complete second order model, would help to significantly improve the model above.

Nested Model

$Y = \beta_0 + \beta_1 Season + \beta_2 Advertising + \varepsilon$

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .977[a] | .954 | .948 | 2.3556 |

a. Predictors: (Constant), ADV, SEASON

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 1721.817 | 2 | 860.909 | 155.152 | .000[b] |
| | Residual | 83.232 | 15 | 5.549 | | |
| | Total | 1805.049 | 17 | | | |

a. Dependent Variable: SALES

b. Predictors: (Constant), ADV, SEASON

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 104.528 | 3.229 | | 32.367 | .000 |
| | SEASON | -7.768 | 1.135 | -.388 | -6.846 | .000 |
| | ADV | 2.979 | .206 | .820 | 14.477 | .000 |

a. Dependent Variable: SALES

$Y = \beta0 + \beta1Season + \beta2Advertising + \beta3Advertising^2 + \beta4Advertising * Season + \varepsilon$

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .987[a] | .973 | .965 | 1.9277 |

a. Predictors: (Constant), AdvAdv, AdvSeason, SEASON, ADV

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 1756.739 | 4 | 439.185 | 118.183 | .000[b] |
| | Residual | 48.310 | 13 | 3.716 | | |
| | Total | 1805.049 | 17 | | | |

a. Dependent Variable: SALES

b. Predictors: (Constant), AdvAdv, AdvSeason, SEASON, ADV

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 90.306 | 14.054 | | 6.426 | .000 |
| | SEASON | 7.944 | 5.271 | .397 | 1.507 | .156 |
| | AdvSeason | -1.070 | .353 | -.780 | -3.034 | .010 |
| | ADV | 4.338 | 1.791 | 1.194 | 2.423 | .031 |
| | AdvAdv | -.027 | .056 | -.229 | -.481 | .639 |

a. Dependent Variable: SALES

Nested F test

H0: $ß3 = ß4 = 0$

H1: at least one coefficient square of advertisement or interaction of advertisement and season is different from zero

F= (SSModel_full-SSModel_nested)/#predictors tested / SSE_full/ (n-k-1)

$F= ((1756.739 - 1721.817)/2)/(48.310/ (18-4-1)) = 4.698675$

Numerator is 2 and denominator is 13, critical value is 3.8056 for P value of 0.05

$F=4.698675 > 3.8056$ thus $p < 0.05$

So we can reject null, and conclude that one of the tested betas square of advertisement, interaction between advertisement and season is\are useful in predicting average sales, after accounting for Season and ADV (advertising).

8b. A statistician who owns a pedometer was interested in predicting the number of miles (Miles) they walked in a day. To do this, they recorded several variables for 42 different days. This included, the number of steps taken at a moderate pace that day (Moderate), and whether or not there was rain that day (Rain). The dataset is called Walking. The researcher used the following model to predict the number of miles walked in a day:

$Y= \beta_0+ \beta_1 \text{Moderate} + \varepsilon$

The wondered if they could do a better job if they let the slope and/or the intercept vary depending on whether it rained or not. Use a nested-F to test this theory.

$Y = \beta_0 + \beta_1 Moderate + \varepsilon$

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .893[a] | .797 | .792 | .4525 |

a. Predictors: (Constant), Moderate

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 32.194 | 1 | 32.194 | 157.244 | .000[b] |
| | Residual | 8.190 | 40 | .205 | | |
| | Total | 40.384 | 41 | | | |

a. Dependent Variable: Miles

b. Predictors: (Constant), Moderate

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.912 | .118 | | 16.140 | .000 |
| | Moderate | .001 | .000 | .893 | 12.540 | .000 |

a. Dependent Variable: Miles

$Y = \beta_0 + \beta_1 Moderate + \beta_2 Rain + \beta_3 ModerateRain + \varepsilon$

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .909[a] | .827 | .813 | .4289 |

a. Predictors: (Constant), ModerateRain, Rain, Moderate

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 33.392 | 3 | 11.131 | 60.494 | .000[b] |
| | Residual | 6.992 | 38 | .184 | | |
| | Total | 40.384 | 41 | | | |

a. Dependent Variable: Miles

b. Predictors: (Constant), ModerateRain, Rain, Moderate

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.674 | .199 | | 8.413 | .000 |
| | Moderate | .001 | .000 | 1.424 | 5.471 | .000 |
| | Rain | .076 | .274 | .039 | .279 | .782 |
| | ModerateRain | .000 | .000 | -.588 | -1.776 | .084 |

a. Dependent Variable: Miles

Nested F test

H0: ß2 = ß3 = 0

H1: at least one predictor of coefficient of rain or interaction between rain and moderate is different from zero

F= (SSModel_full-SSModel_nested)/#predictors tested / SSE_full/ (n-k-1)

F= ((33.392- 32.194)/2)/( 6.992/ (42-3-1)) = 3.255435

Numerator is 2 and denominator is 38, critical value is 3.24481 for P value of 0.05

F=3.255435> 3.24481 thus $p < 0.05$

So we can reject null, and conclude that one of the tested betas rain or not rain and interaction between rain indicator and moderate is\are useful in predicting average miles, after accounting for moderate.

**3.36 *Caterpillar metabolic rates.*** In **Exercise 1.31** on **page 66**, we learned that the transformed body sizes (*LogBodySize*) are a good predictor of transformed metabolic rates (*LogMrate*) for a sample of caterpillars with data in **MetabolicRate**. We also notice that this linear trend appeared to hold for all five stages of the caterpillar's life. Create five different indicator variables, one for each level of *Instar*, and fit a multiple regression model to estimate five different regression lines. Only four of your five indicator variables are needed to fit the multiple regression model. Can you explain why? You should also create interaction variables to allow for the possibility of different slopes for each regression line. Does the multiple regression model with five lines provide a better fit than the simple linear regression model?

1)

Since it has five levels and so it omits instar level 5. Thereby, only four indictors variables are useful.

2)

Original model:

LogMrate = β0+ β1 LogBodySize + ε

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .974[a] | .948 | .948 | .175218534808 |

a. Predictors: (Constant), LogBodySize

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 169.020 | 1 | 169.020 | 5505.255 | .000[b] |
| | Residual | 9.303 | 303 | .031 | | |
| | Total | 178.322 | 304 | | | |

a. Dependent Variable: LogMrate

b. Predictors: (Constant), LogBodySize

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | 1.307 | .014 | | 96.331 | .000 |
| | LogBodySize | .916 | .012 | .974 | 74.197 | .000 |

a. Dependent Variable: LogMrate

New model:

LogMrate = β0+ β1 LogBodySize + + β2Instar1 + β3Instar2 + β4Instar3 + β5Instar4 + ß6 LogBodySize Instar1 + ß7 LogBodySize Instar2 + ß8 LogBodySize Instar3+ ß9 LogBodySize Instar4 + error

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .975[a] | .950 | .948 | .173854578449 |

a. Predictors: (Constant), instar_4_LogBodySize, LogBodySize, instar_3, instar_2, instar_4, instar_3_LogBodySize, instar_1, instar_2_LogBodySize, instar_1_LogBodySize

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 169.406 | 9 | 18.823 | 622.750 | .000[b] |
| | Residual | 8.916 | 295 | .030 | | |
| | Total | 178.322 | 304 | | | |

a. Dependent Variable: LogMrate

b. Predictors: (Constant), instar_4_LogBodySize, LogBodySize, instar_3, instar_2, instar_4, instar_3_LogBodySize, instar_1, instar_2_LogBodySize, instar_1_LogBodySize

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.319 | .047 | | 27.823 | .000 |
| | LogBodySize | .980 | .114 | 1.041 | 8.632 | .000 |
| | instar_1 | -.066 | .209 | -.028 | -.314 | .753 |
| | instar_2 | .016 | .120 | .008 | .134 | .894 |
| | instar_3 | -.026 | .061 | -.015 | -.420 | .675 |
| | instar_4 | -.061 | .053 | -.035 | -1.142 | .254 |
| | instar_1_LogBodySize | -.101 | .151 | -.087 | -.668 | .505 |
| | instar_2_LogBodySize | -.029 | .137 | -.022 | -.214 | .831 |
| | instar_3_LogBodySize | -.068 | .120 | -.040 | -.564 | .573 |
| | instar_4_LogBodySize | -.227 | .127 | -.062 | -1.792 | .074 |

a. Dependent Variable: LogMrate

Nested F test

H0: ß2 = ß3 = ß4 = ß5 = ß6 = ß7 = ß8 = ß9 = 0

H1: at least one of the coefficients of the indicators or interactions with LosBodySize is different from zero

F= (SSModel_full-SSModel_nested)/#predictors tested / SSE_full/ (n-k-1)

F= ((169.406- 169.020)/8)/( 8.916/ (305-9-1)) = 1.596428

Numerator is 8 and denominator is 295, critical value is 3.24481 for P value of 0.05

F=1.596428< 1.96985032 thus p > 0.05

So we fail to reject null, and conclude that the tested betas (the predictors we added to form full model) indicator of instar and interaction between different level of instar and logbodysize are not useful in predicting average miles, after accounting for moderate.