Homework 5

Solution

Question 1. Reality TV and cosmetic surgery (Page 201 Questions 4.30 and Page 236 Data set: BDYIMG)

4.30(a) (5 pts) Give the least squares prediction equation.

```
### Import data and fit the interaction model
bdyimg = read.csv("STAT 3113 Data Sets/BDYIMG.csv")
fit_bdyimg_int = lm(DESIRE ~ GENDER + IMPREAL + GENDER:IMPREAL, data=bdyimg)

summary(fit_bdyimg_int)
```

```
##
## Call:
## lm(formula = DESIRE ~ GENDER + IMPREAL + GENDER:IMPREAL, data = bdyimg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1174 -1.6597 -0.1174  1.6518  5.8826
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     11.7789     0.6736  17.486  < 2e-16 ***
## GENDER          -1.9722     1.1792  -1.672 0.096311 .
## IMPREAL          0.5846     0.1616   3.617 0.000395 ***
## GENDER:IMPREAL  -0.5533     0.2761  -2.004 0.046705 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.35 on 166 degrees of freedom
## Multiple R-squared:  0.449,  Adjusted R-squared:  0.439
## F-statistic: 45.09 on 3 and 166 DF,  p-value: < 2.2e-16
```

Answer: The least squares prediction equation is

$$\hat{y} = 11.779 - 1.972x_1 + 0.585x_4 - 0.553x_1x_4.$$

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

```
new.data = data.frame(GENDER = 1, IMPREAL = 5)
predict(fit_bdyimg_int, newdata=new.data)
```

```
##        1
## 9.963267
```

Answer: For $x_1 = 1$ and $x_4 = 5$, the predicted value is

$$\hat{y} = 9.96.$$

4.30(c) (10 pts) Conduct a test of overall model adequacy. Use $\alpha = .10$.

```
source("anova_alt.R")
anova_alt(fit_bdyimg_int)
```

```
## Analysis of Variance Table
##
##          Df      SS      MS      F           P
## Source    3  747.00 249.000 45.086 2.2926e-21
## Error   166  916.79   5.523
## Total   169 1663.79   9.845
```

Answer:

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

H_a : At least one of the $\beta s \neq 0$.

As the p-value is 2.293e-21 $< \alpha = 0.1$, we reject H_0. We conclude there is sufficient evidence to indicate the model is adequate in predicting desire to have cosmetic surgery at $\alpha = 0.1$.

4.30(d) (5 pts) Give a practical interpretation of R_a^2.

Answer: $R_a^2 = 0.439$.

43.9% of the sample variation in the desire to have cosmetic surgery around its mean is explained by the model containing gender, impression of reality TV and the interaction of the two variables, adjusted for the number of terms in the model and sample size.

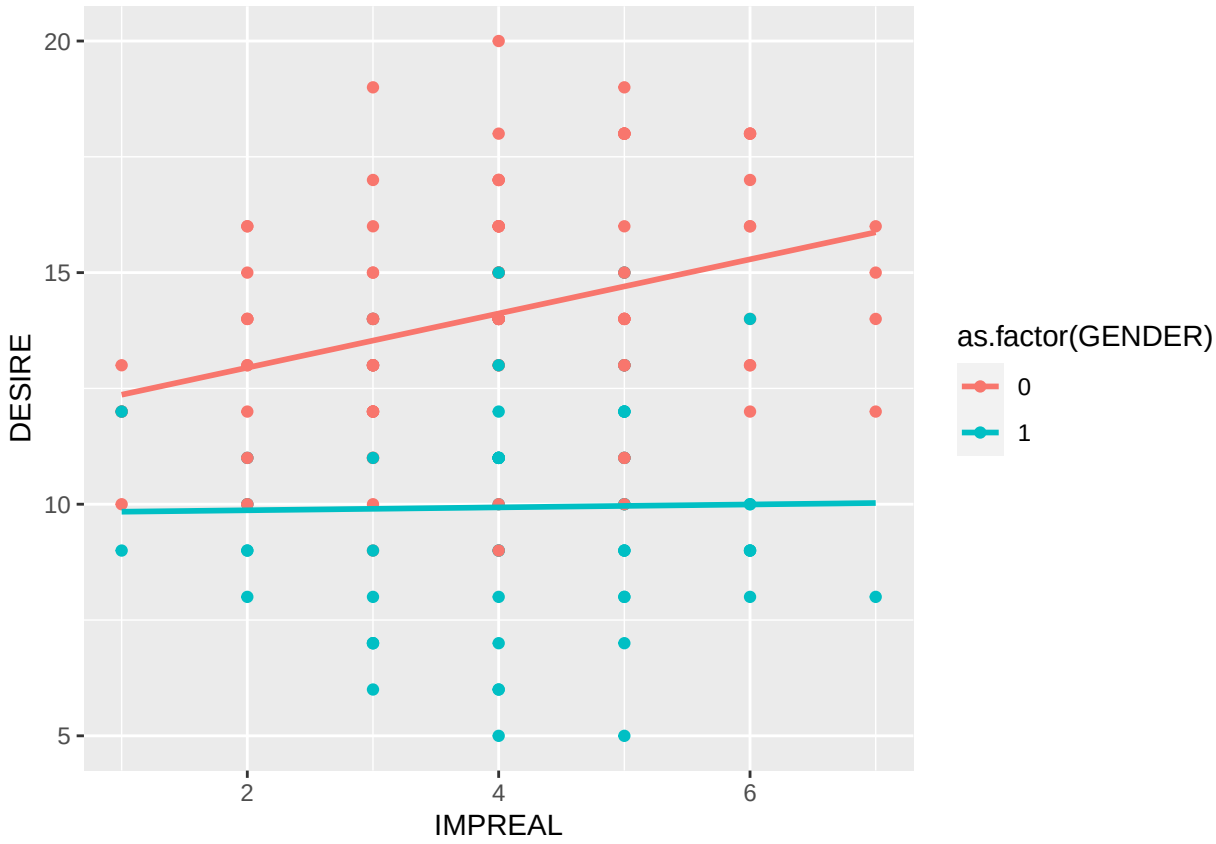
4.30(e) (5 pts) Give a practical interpretation of s.

Answer: $s = 2.35$. Most (more than 95%) of the observed values of desire will fall within $2s = 2(2.35) = 4.7$ units of their predicted values.

4.30(f) (10 pts) 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).

```
### Not required in the homework.
### Just would like to show the evidence of the interaction
library(ggplot2)
ggplot(bdyimg, aes(x=IMPREAL, y=DESIRE,
                   color = as.factor(GENDER)))+
  geom_point()+
  geom_smooth(se=FALSE, method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Answer: To determine if gender and impression of reality TV interact, we test:

$H_0 : \beta_3 = 0$

$H_a : \beta_3 \neq 0$

```
summary(fit_bdyimg_int)
```

```
##
## Call:
## lm(formula = DESIRE ~ GENDER + IMPREAL + GENDER:IMPREAL, data = bdyimg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1174 -1.6597 -0.1174  1.6518  5.8826
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     11.7789     0.6736  17.486  < 2e-16 ***
## GENDER          -1.9722     1.1792  -1.672 0.096311 .
## IMPREAL          0.5846     0.1616   3.617 0.000395 ***
## GENDER:IMPREAL  -0.5533     0.2761  -2.004 0.046705 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.35 on 166 degrees of freedom
## Multiple R-squared:  0.449,  Adjusted R-squared:  0.439
## F-statistic: 45.09 on 3 and 166 DF,  p-value: < 2.2e-16
```

The test statistic is $t = -2.004$ and the p-value is $p = 0.047$.

Since the p-value is less than $\alpha = 0.10$, H_0 is rejected. There is sufficient evidence to indicate gender and impression of reality TV interact to affect desire to have cosmetic surgery at $\alpha = 0.10$.

4.68(a) (5 pts) Give the equation of the model for $E(y)$ that matches the theory.

Answer:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_4 + \beta_6 x_2 x_4 + \beta_7 x_3 x_4,$$

where y = desire to have cosmetic surgery, x_1 = gender, x_2 = self_esteem, x_3 = body satisfaction, x_4 = impression of reality.

4.68(b) (5 pts) Fit the model, part(a), to the simulated data saved in the file.

```
fit_bdyimg_full = lm(DESIRE ~ GENDER + SELFESTM +
                               BODYSAT + IMPREAL +
                               GENDER:IMPREAL +
                               SELFESTM:IMPREAL +
                               BODYSAT:IMPREAL, data = bdyimg)
summary(fit_bdyimg_full)
```

```
##
## Call:
## lm(formula = DESIRE ~ GENDER + SELFESTM + BODYSAT + IMPREAL +
##     GENDER:IMPREAL + SELFESTM:IMPREAL + BODYSAT:IMPREAL, data = bdyimg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4295 -1.5039 -0.1417  1.4827  6.1053
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.092057   2.012896   6.504 9.27e-10 ***
## GENDER           -1.889953   2.073809  -0.911    0.363
## SELFESTM         -0.090775   0.117596  -0.772    0.441
## BODYSAT           0.134995   0.474874   0.284    0.777
## IMPREAL           0.745972   0.491797   1.517    0.131
## GENDER:IMPREAL   -0.064728   0.511041  -0.127    0.899
## SELFESTM:IMPREAL  0.009771   0.028078   0.348    0.728
## BODYSAT:IMPREAL  -0.112132   0.116010  -0.967    0.335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.236 on 162 degrees of freedom
## Multiple R-squared:  0.5132, Adjusted R-squared:  0.4922
## F-statistic:  24.4 on 7 and 162 DF,  p-value: < 2.2e-16
```

```
anova_alt(fit_bdyimg_full)
```

```
## Analysis of Variance Table
##
##         Df      SS      MS    F          P
## Source   7  853.89 121.984 24.4 1.7181e-22
## Error  162  809.90   4.999
## Total  169 1663.79   9.845
```

Answer:

- The fitted regression equation is:

$$\hat{y} = 13.09 - 1.89x_1 - 0.091x_2 + 0.135x_3 + 0.746x_4 - 0.065x_1x_4 + 0.0098x_2x_4 - 0.112x_3x_4.$$

- The overall utility of the model – *Not required in the homework*

Answer: To test the overall utility of the model, we test:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

H_a : At least 1 $\beta_i \neq 0$.

The test statistic is $F = 24.40$ and the p-value is $p = 0.000$. Since the p-value is so small, we reject H_0. There is sufficient evidence to indicate the model is useful for predicting desire to have cosmetic surgery.

4.68(c) (5 pts) Give the null hypothesis for testing the psychologists' theory.

Answer: To determine if impression of reality TV interacts with each of the other independent variables, the null hypothesis is:

$$H_0 : \beta_5 = \beta_6 = \beta_7 = 0$$

4.68(d)(10 pts) Conducted a nested model F-test to test the theory. What do you conclude?

```
fit_bdyimg_red = lm(DESIRE ~ GENDER + SELFESTM +
                                BODYSAT + IMPREAL, data=bdyimg)

anova(fit_bdyimg_red, fit_bdyimg_full)
```

```
## Analysis of Variance Table
##
## Model 1: DESIRE ~ GENDER + SELFESTM + BODYSAT + IMPREAL
## Model 2: DESIRE ~ GENDER + SELFESTM + BODYSAT + IMPREAL + GENDER:IMPREAL +
##     SELFESTM:IMPREAL + BODYSAT:IMPREAL
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    165 835.95
## 2    162 809.90  3    26.058 1.7374 0.1614
```

Answer: The reduced model is

$$E(y) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$$

The hypotheses are

$H_0 : \beta_5 = \beta_6 = \beta_7 = 0$

H_a : At least one of the $\beta s \neq 0$.

From the R output, the p-value is $0.1614 > \alpha = 0.05$. Therefore, we fail to reject H_0. There is insufficient evidence to indicate impression of reality TV interacts with at least one of the other independent variables at $\alpha = 0.05$. In other words, we prefer the reduced model.

Question 2. Commercial refrigeration systems (Page 209 Question 4.40)

What is the hypothesized sign (positive or negative) of the β_2 parameter in the model?

Answer: (5 pts) Because the graph is curved, we would hypothesize that the model should be $E(y) = \beta_0 + \beta_1x + \beta_2x^2$. Since the curve opens downward, β_2 will be negative.

Question 3. Shopping on Black Friday (Page 210 Question 4.44 Data set: BLKFRIDAY)

(a) (5 pts) Fit the quadratic model, $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$, to the data using statistical software. Give the estimated regression equation.

```
blackfriday = read.csv("STAT 3113 Data Sets/BLKFRIDAY.csv")

fit_bf = lm(YEARS ~ AGE + I(AGE^2), data=blackfriday)
summary(fit_bf)
```

```
##
## Call:
## lm(formula = YEARS ~ AGE + I(AGE^2), data = blackfriday)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.015  -4.769  -0.937   4.048  13.985
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.933164  10.238764  -0.872    0.389
## AGE          0.703857   0.551842   1.275    0.211
## I(AGE^2)    -0.003412   0.006725  -0.507    0.615
##
## Residual standard error: 7.427 on 35 degrees of freedom
## Multiple R-squared:  0.4291, Adjusted R-squared:  0.3965
## F-statistic: 13.15 on 2 and 35 DF,  p-value: 5.493e-05
```

```
anova_alt(fit_bf)
```

```
## Analysis of Variance Table
##
##        Df     SS     MS      F          P
## Source  2 1450.9 725.47 13.153 5.4932e-05
## Error  35 1930.4  55.16
## Total  37 3381.4  91.39
```

Answer: the estimated regression equation is

$$\hat{y} = -8.9 + 0.704x - 0.00341x^2$$

(c) (10 pts) Conduct a test to determine if the relationship between age (x) and number of years shopping on Black Friday (y) is best represented by a linear or quadratic function. Use $\alpha = .01$.

Answer: To determine if the relationship between age and the number of years shopping on Black Friday is quadratic, we test:

$H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

The test statistic is $t = -0.51$ and the p-value is 0.615, which is greater than $\alpha = .01$, we fail to reject H_0. There is insufficient evidence to indicate the relationship between age and the number of years shopping on Black Friday is quadratic at $\alpha = 0.01$.

Question 4. Homework assistance for accounting students

(Refer to Page 228 Question 4.58, Data set: ACCHW)

(a) (5 pts) Propose a model for the knowledge gain (y) as a function of the qualitative variable, homework assistance group. (Suppose we use "NO" as the base level).

Answer:

We use the "NO" help level as the base level. As there are three levels (completed, check figures, no help), we need two dummy variables as follows:

$x_1 = 1$, if completed solution; $= 0$ if not.

$x_2 = 1$, if check figures; $= 0$ if not.

The model for knowledge gain as a function of homework assistant group is

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

(b) (5 pts) 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''.

Answer: In terms of the model, the difference between the mean knowledge gain between students in the complete solution group and the no help group is β_1

(c) (5 pts) Fit the model to the data and give the least squares prediction equation.

```
### Read the data
acchw = read.csv("STAT 3113 Data Sets/ACCHW.csv")

### Set 'NO' as the base level.
acchw$ASSIST=relevel(as.factor(acchw$ASSIST), ref="NO")

### Please fit the model below.

fit_acchw = lm(IMPROVE ~ ASSIST, data=acchw)
summary(fit_acchw)
```

```
##
## Call:
## lm(formula = IMPROVE ~ ASSIST, data = acchw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.433 -2.433  0.050  1.567  6.567
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.4333     0.4941   4.925  5.2e-06 ***
## ASSISTCHECK   0.2867     0.7329   0.391    0.697
## ASSISTFULL   -0.4833     0.7813  -0.619    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.706 on 72 degrees of freedom
## Multiple R-squared:  0.01244,    Adjusted R-squared:  -0.01499
## F-statistic: 0.4535 on 2 and 72 DF,  p-value: 0.6372
```

Answer: The least regression prediction equation is

$$\hat{y} = 2.433 - 0.483x_1 + 0.287x_2$$