Homework #5 (Solutions)
STAT-UB.0003: Regression and Forecasting Models

Solutions adapted from N.S. Boudreau's *Instructor's Solution Manual* (2011).

1. MBS Ex. 12.72.

Solution:

(a)

$$\text{Race}_{\text{black}} = \begin{cases} 1 & \text{if race is black} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Race}_{\text{white}} = \begin{cases} 1 & \text{if race is white} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Avail}_{\text{high}} = \begin{cases} 1 & \text{if availability is high} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Avail}_{\text{low}} = \begin{cases} 1 & \text{if availability is low} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{QB}} = \begin{cases} 1 & \text{if position is quarterback} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{RB}} = \begin{cases} 1 & \text{if position is running back} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{WR}} = \begin{cases} 1 & \text{if position is wide receiver} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{TE}} = \begin{cases} 1 & \text{if position is tight end} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{DL}} = \begin{cases} 1 & \text{if position is defensive lineman} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{LB}} = \begin{cases} 1 & \text{if position is linebacker} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{DB}} = \begin{cases} 1 & \text{if position is defensive back} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Pos}_{\text{OL}} = \begin{cases} 1 & \text{if position is offensive lineman} \\ 0 & \text{otherwise} \end{cases}$$

(b) One potential model is

$$\text{Price} = \beta_0 + \beta_1 \text{Race}_{\text{black}} + \varepsilon.$$

In this model,

β_0 = mean price for race white,
β_1 = difference in mean price between races white and black.

(c) One potential model is

$$\text{Price} = \beta_0 + \beta_1 \text{Avail}_{\text{high}} + \varepsilon.$$

In this model,

β_0 = mean price for availability low,
β_1 = difference in mean price between availability high and low.

(d) One potential model is

$$\text{Price} = \beta_0 + \beta_1 \text{Pos}_{\text{QB}} + \beta_2 \text{Pos}_{\text{RB}} + \beta_3 \text{Pos}_{\text{WR}} + \beta_4 \text{Pos}_{\text{TE}} + \beta_5 \text{Pos}_{\text{DL}} + \beta_6 \text{Pos}_{\text{LB}} + \beta_7 \text{Pos}_{\text{DB}} + \varepsilon$$

In this model,

β_0 = mean price for position offensive lineman (OL),
β_1 = difference in mean price between position QB and position OL,
β_2 = difference in mean price between position RB and position OL,
β_3 = difference in mean price between position WR and position OL,
β_4 = difference in mean price between position TE and position OL,
β_5 = difference in mean price between position DL and position OL,
β_6 = difference in mean price between position LB and position OL,
β_7 = difference in mean price between position DB and position OL.

2. MBS Ex. 12.74.

Solution:

(a) $E(y \mid x) = \beta_0 + \beta_1 x$.

(b) β_0 =mean relative optimism for analysts who worked for sell-side firms.

(c) Yes.

(d) Yes. If the buy-side analysts are less optimistic, then their estimates will be smaller than the sell-side estimates. Thus, the estimate of β_1 will be negative.

3. MBS Ex. 12.78, parts (a)–(d).

Solution:

(a) One possible model is

$$\text{Improve} = \beta_0 + \beta_1 \text{Assist}_{\text{Check}} + \beta_2 \text{Assist}_{\text{Full}} + \varepsilon$$

(b) In this model, the difference between "completed solution" and "no help" would be β_2.

(c) Here is the fit:

```
Analysis of Variance

Source          DF   Adj SS  Adj MS  F-Value  P-Value
Regression       2    6.643   3.322     0.45    0.637
  ASSIST_CHECK   1    1.121   1.121     0.15    0.697
  ASSIST_FULL    1    2.803   2.803     0.38    0.538
Error           72  527.357   7.324
Total           74  534.000


Model Summary

      S   R-sq  R-sq(adj)  R-sq(pred)
2.70636  1.24%      0.00%       0.00%


Coefficients

Term             Coef  SE Coef  T-Value  P-Value   VIF
Constant        2.433    0.494     4.92    0.000
ASSIST_CHECK    0.287    0.733     0.39    0.697  1.22
ASSIST_FULL    -0.483    0.781    -0.62    0.538  1.22


Regression Equation

IMPROVE = 2.433 + 0.287 ASSIST_CHECK - 0.483 ASSIST_FULL
```

The least squares prediction equation is

$$\widehat{\text{Improve}} = 2.433 + 0.287 \text{Assist}_{\text{Check}} - 0.483 \text{Assist}_{\text{Full}}.$$

(d) The p-value for the F test is $p = 0.637$. There is no evidence that the model is useful.

4. MBS Ex. 12.81.

Solution:

(a) One possible model is

$$\text{Rating} = \beta_0 + \beta_1 \text{Rating}_S + \beta_2 \text{Rating}_V + \varepsilon$$

(b) Here is the Minitab output:

```
    Analysis of Variance

Source       DF   Adj SS   Adj MS  F-Value  P-Value
Regression    2   123.27   61.633    20.45    0.000
  RATING_S    1   114.12  114.116    37.87    0.000
  RATING_V    1    63.37   63.375    21.03    0.000
Error       321   967.35    3.014
Total       323  1090.62


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.73596  11.30%     10.75%       9.64%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   3.167    0.167    18.96    0.000
RATING_S  -1.454    0.236    -6.15    0.000  1.33
RATING_V  -1.083    0.236    -4.59    0.000  1.33


Regression Equation

RECALL = 3.167 - 1.454 RATING_S - 1.083 RATING_V
```

The least squares prediction equation is

$$\widehat{\text{Recall}} = 3.167 - 1.454\text{Rating}_S - 1.083\text{Rating}_V$$

(c) To test overall model utility, we look at the p-value for the F test. Minitab reports this as $p = 0.000$, which we can interpret as $p < 0.001$ (the p-value is never exactly equal to 0). Since $p < \alpha$, we reject the null hypothesis that all slope coefficients are 0; we have very strong evidence that the model is useful.

(d) The means for the three groups are

$$\bar{y}_N = \hat{\beta}_0 = 3.167,$$
$$\bar{y}_S = \hat{\beta}_0 + \hat{\beta}_1 = 3.167 - 1.454 = 1.713,$$
$$\bar{y}_V = \hat{\beta}_0 + \hat{\beta}_2 = 3.167 - 1.083 = 2.084.$$

5. MBS Ex. 12.90, parts (c)–(d).

Solution: (Not graded.)