Solutions to Homework 7

David M. Rocke

BIM 105, Fall 2019

Problem 1

1. The article "Application of Analysis of Variance to Wet Clutch Engagement" (M. Mansouri, M. Khonsari, et al., Proceedings of the Institution of Mechanical Engineers, 2002:117–125) presents the following fitted model for predicting clutch engagement time in seconds (y) from engagement starting speed in m/s (x_1), maximum drive torque in Nm (x_2), system inertia in kg m^2 (x_3), and applied force rate in kN/s (x_4):

$$
\begin{aligned}
y &= -0.83 + 0.017x_1 + 0.0895x_2 + 42.771x_3 \\
&\quad +0.027x_4 - 0.0043x_2x_4
\end{aligned}
$$

The sum of squares for regression was SSR $=$ 1.08613 and the sum of squares for error was SSE $=$ 0.036310. There were 44 degrees of freedom for error.

(a) Predict the clutch engagement time when the starting speed is 20 m/s, the maximum drive torque is 17 Nm, the system inertia is 0.006 kg m^2, and the applied force rate is 10 kN/s.

$$\begin{aligned} y &= -0.83 + 0.017(20) + 0.0895(17) \\ &\quad +42.771(0.006) + 0.027(10) - 0.0043(17)(10) \\ &= 0.827126 \end{aligned}$$

(b) Is it possible to predict the change in engagement time associated with an increase of 2 m/s in starting speed? If so, find the predicted change. If not, explain why not.

It is possible, as is true for any predictor that is not involved in an interaction:

$$0.017 \times 2 = 0.034$$

(c) Is it possible to predict the change in engagement time associated with an increase of 2 Nm in maximum drive torque? If so, find the predicted change. If not, explain why not.

The change depends on the value of $x4$ because of the $x2 : x4$ interaction, so we cannot find the value of the change without knowing the value of $x4$. The change is

$$0.0895 \times 2 - 0.0043 \times 2 \times x4$$

(d) Compute the coefficient of determination R^2.

(e) Compute the F statistic for testing the null hypothesis that all the coefficients are equal to 0. Can this hypothesis be rejected?

$$
\begin{aligned}
\text{SSR} &= 1.08613 \\
\text{SSE} &= 0.036310 \\
\text{SST} &= 1.08613 + 0.03613 = 1.12244 \\
R^2 &= \text{SSR}/\text{SST} = 0.968 \\
F(5,44) &= \frac{\text{SSR}/5}{\text{SSE}/44} = 263.2317 \\
p &\approx 0
\end{aligned}
$$

2. The file `HW7-2.csv` presents measurements of mean noise levels in dBA (y), roadway width in m (x_1), and mean speed in km/h (x_2), for 10 locations in Bangkok, Thailand, as reported in the article "Modeling of Urban Area Stop-and-Go Traffic Noise" (P. Pamanikabud and C. Tharasawatipipat, Journal of Transportation Engineering, 1999:152–159). Construct a good linear model to predict mean noise levels using roadway width, mean speed, or both, as predictors. Provide the standard deviations of the coefficient estimates and the P-values for testing that they are different from 0. Explain how you chose your model.

For this problem, transformations are not needed, which can be shown by histograms, or just by the narrow range of the variables. One way would be to start with the model $y \sim x1 * x2$, eliminate the interaction as not significant, then eliminate $x1$ as not significant. The standard deviations of the intercept and $x2$ coefficient are 2.0671 and 0.063783 from the output. The p-values are 4.4×10^{-10} and 0.0252, so both null hypotheses are rejected.

```
>> summary(HW72)

   y: 10x1 double

       Values:

           Min         78.1
           Median     78.55
           Max           81
```

```
x1: 10x1 double

    Values:

        Min          6
        Median     6.5
        Max         12
```

```
x2: 10x1 double

    Values:

        Min           28.26
        Median       30.445
        Max           38.73
```

```
>> fitlm(HW72,'y~x2')


Linear regression model:
    y ~ 1 + x2

Estimated Coefficients:
                   Estimate       SE        tStat       pValue
                   ________    ________    ______    __________

    (Intercept)     73.228       2.0671    35.425    4.4134e-10
    x2             0.17511     0.063783    2.7454      0.025237


Number of observations: 10, Error degrees of freedom: 8
Root Mean Squared Error: 0.71
R-squared: 0.485,  Adjusted R-Squared: 0.421
F-statistic vs. constant model: 7.54, p-value = 0.0252
```

3. The file `HW7-3.csv` consist of yield measurements from many runs of a chemical reaction. The quantities varied were the temperature in $^\circ$C ($x1$), the concentration of the primary reactant in % ($x2$), and the duration of the reaction in hours ($x3$). The dependent variable (y) is the fraction converted to the desired product.

(a) Fit the linear model predicting y from the three variables without interactions or quadratic terms.

(b) Two of the variables in this model have coefficients significantly different from 0 at the 15% level. Fit a linear regression model containing these two variables.

(c) Compute the product (interaction) of the two variables referred to in part (b). Fit the model that contains the two variables along with the interaction term.

(d) Based on the results in parts (a) through (c), specify a model that appears to be good for predicting y from x_1, x_2, and x_3.

(e) Might it be possible to construct an equally good or better model in another way?

The first model suggests that x1 is not informative. The second shows that neither x2 nor x3 can be removed, while the third shows that the interaction is not significant. I would pick the second model. It is always possible that there is a better model, using transformations, or combinations that have not been tried.

```
>> fitlm(HW73)

Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate        SE         tStat        pValue
                   _________    ________    ________    __________

    (Intercept)       10.552       3.539      2.9818     0.0037274
    x1             -0.033435    0.097034    -0.34457       0.73126
    x2               0.54648     0.20168      2.7097     0.0081281
    x3                1.8305     0.24774      7.3887    8.9608e-11


Number of observations: 90, Error degrees of freedom: 86
Root Mean Squared Error: 4.06
R-squared: 0.493,  Adjusted R-Squared: 0.476
F-statistic vs. constant model: 27.9, p-value = 1.07e-12
```

```
>> fitlm(HW73,'y~x2+x3')

Linear regression model:
    y ~ 1 + x2 + x3

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    _______    ______    __________

    (Intercept)     9.9458      3.0544    3.2562     0.0016102
    x2             0.48736     0.10546    4.6212    1.3143e-05
    x3              1.8329     0.24639    7.4392    6.7278e-11


Number of observations: 90, Error degrees of freedom: 87
Root Mean Squared Error: 4.04
R-squared: 0.493,  Adjusted R-Squared: 0.481
F-statistic vs. constant model: 42.2, p-value = 1.52e-13
```

```
>> fitlm(HW73,'y~x2*x3')

Linear regression model:
    y ~ 1 + x2*x3

Estimated Coefficients:
                   Estimate       SE        tStat       pValue  
                   ________    ________    _______    ________

    (Intercept)      17.956      9.1033     1.9725    0.051766
    x2              0.16285     0.36306    0.44854     0.65489
    x3              0.54885      1.3965    0.39302     0.69528
    x2:x3          0.051789    0.055437    0.93418     0.35282


Number of observations: 90, Error degrees of freedom: 86
Root Mean Squared Error: 4.04
R-squared: 0.498,  Adjusted R-Squared: 0.48
F-statistic vs. constant model: 28.4, p-value = 7.38e-13
```