MA 542 REGRESSION ANALYSIS
SPRING 2018
HW - 7 - Solution Key

1. (Chapter 7 question 7)
a)

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 14.819  14.819 11.4649  0.001125 **
X2         1 72.802  72.802 56.3262 9.699e-11 ***
X4         1 50.287  50.287 38.9062 2.306e-08 ***
X3         1  0.420   0.420  0.3248  0.570446
Residuals 76 98.231   1.293
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
```

b)

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0$$

Full model : $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \epsilon_i$
Reduced model : $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_4 X_{i4} + \epsilon_i$

$$F^* = \frac{MSR(X_3|X_1, X_2, X_4)}{MSE(X_1, X_2, X_3, X_4)} = \frac{.420}{1.293} = 0.3248$$

But,

$$F_{0.99,1,79} = 6.98057 \quad \text{OR} \quad P - value = 0.5704457$$

Since $F^* = 0.3248 < F_{0.99,1,79} = 6.98057$ or $P - value = 0.5704457 > 0.01$, H_0 is not rejected. So the predictor X_3 can be dropped from the model given that X_1, X_2 and X_4 are in the model.

2. (Chapter 7 question 8)

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_1 : not H_0$$

Full model : $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \epsilon_i$
Reduced model : $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_4 X_{i4} + \epsilon_i$

$$F^* = \frac{(SSR(X_2|X_1, X_4) + SSR(X_3|X_1, X_2, X_4))/2}{MSE(X_1, X_2, X_3, X_4)} = \frac{27.857 + 0.420}{1.293}/2 = 10.9389$$

But,

$$F_{0.99,2,76} = 4.89584 \quad \text{OR} \quad P - value = 6.682136 \times 10^{-05}$$

Since $F^* = 19.6155 > F_{0.99,2,79} = 4.89584$ or $P - value = 6.682136 \times 10^{-05} < 0.01$, H_0 is rejected. So the predictor X_2 and X_3 can not be dropped from the model with X_1 and X_2.

3. (Chapter 7 question 10)

$$H_0 : \beta_1 = -0.1, \quad \beta_2 = 0.4 \quad \text{vs} \quad H_1 : not \;\; H_0$$

Full model : $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \epsilon_i$
Reduced model : $Y_i + 0.1X_{i1} - 0.4X_{i2} = \beta_0 + \beta_3 X_{i3} + \beta_4 X_{i4} + \epsilon_i$

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} / \frac{SSE(F)}{df_f} = \frac{110.141 - 98.231}{78 - 76} / 1,293 = 4.6056$$

But,

$$F_{0.99,2,79} = 4.89584 \quad \text{OR} \quad P - value = 0.01294$$

Since $F^* = 4.6056 < F_{0.99,2,79} = 4.89584$ or $P - value = 0.01294 > 0.01$, H_0 is not rejected. So $\beta_1 = -0.1$, and $\beta_2 = 0.4$.

4. (Chapter 7 question 15)

$$R^2_{Y4} = \frac{SSR(X_4)}{SSTO} = \frac{67.775}{236.557} = 0.2865$$

So 28.65% of the total variation of Y (Rental rates) is explained by X_4 (total square footage).

$$R^2_{Y1} = \frac{SSR(X_1)}{SSTO} = \frac{14.819}{236.557} = 0.0626$$

So 6.26% of the total variation of Y (Rental rates) is explained by X_1 (the age).

$$R^2_{Y1|4} = \frac{SSR(X_1|X_4)}{SSE(X_4)} = \frac{42.275}{168.782} = 0.2505$$

25.05% left over variation of Y (Rental rates) after regressing on X_4 (total square footage) is explained by X_1 (the age).

$$R^2_{Y14} = \frac{SSR(X_1, X_4)}{SSTO} = \frac{67.775 + 42.275}{236.557} = 0.4652$$

46.52% variation of Y (Rental rates) X_4 (total square footage) is explained by X_1 (the age) and X_4 (total square footage).

$$R^2_{Y2|14} = \frac{SSR(X_2|X_1, X_4)}{SSE(X_1, X_4)} = \frac{27.857}{126.508} = 0.2202$$

22.02% left over variation of Y (Rental rates) after regressing on X_1 (the age) and X_4 (total square footage) is explained by X_2 (operating expenses and taxes).

$$R^2_{Y3|124} = \frac{SSR(X_3|X_1, X_2, X_4)}{SSE(X_1, X_2, X_4)} = \frac{0.42}{98.65} = 0.004257$$

0.4257% left over variation of Y (Rental rates) after regressing on X_1, (the age), X_2 (operating expenses and taxes) and X_4 (total square footage) is explained by X_3 (vacancy rates).

$$R^2 = 0.5847$$

So 58.47% of the total variation of Y (Rental rates) is explained by the predictors X_1, (the age), X_2 (operating expenses and taxes), X_3 (vacancy rates) and X_4 (total square footage).