# ISyE 4031 Regression and Forecasting
## Homework 6 Solutions
## Spring 2016

1. Exercise 4.22.

a. $\hat{\beta}_5 = E(y_{d,a,B}) - E(y_{d,a,A}) = 0.21369$

$\hat{\beta}_6 = E(y_{d,a,C}) - E(y_{d,a,A}) = 0.38178$

$\hat{\beta}_6 - \hat{\beta}_5 = 0.38178 - 0.21369 = 0.16809$

A 95% C.I. for $\beta_5$ $[.21369 \pm 2.069(.06215)] = [.0851, .3423]$

A 95% C.I. for $\beta_6$ $[.38178 \pm 2.069(.06125)] = [.2551, .5085]$

Both $p$-values < .01, $\beta_5$ and $\beta_6$ are significant.

b. $\hat{y} = 25.61270 + 9.0568(.20) - 6.5767(6.50) + .58444(6.50)^2$
$- 1.15648(.20)(6.50) + .38178(1) = 8.5005$

A 95% C.I. for the mean demand: [8.4037, 8.5977]. We are 95% confident that the mean demand for all sales periods when the price difference is 0.20, the advertising expenditure is 6.50, and campaign *C* is used will be between 840,370 and 859,770 bottles.

A 95% P.I. for individual demand: [8.2132, 8.7881]. We are 95% confident that the actual demand in a particular sales period when the price difference is .20, the advertising expenditure is 6.50, and campaign *C* is used will be between 821,322 bottles and 878,813 bottles.

c. [0.0363, 0.2999]; $p$-value = 0.0147 $\beta_6$ significant at $\alpha$ = .05.

A 95% C.I. for $\beta_6$: [0.16809 $\pm$ 2.069 (0.06371)] = [0.0363, 0.2999].

2. Exercise 5.1.

a. $r(Load, Beddays) = 1.00, r(Load, Pop) = 0.936, r(Pop, Beddays) = 0.933$.

$VIF_{BedDays} = 8{,}933.1, VIF_{Load} = 9{,}597.6,\ VIF_{Pop} = 23.3$.

b. $x_2, x_4, x_5$ (*BedDays, Load, Pop*).

c. Yes, $b_4$ (says labor hours goes down as patient load increases).

d. Yes, it seems so. All $p$-values for independent variables exceed .02, and .02 is 20 times larger than 0.001.

e. Model 1 seems to be best. It has an *s* very close to the smallest *s* (equivalently, an $\overline{R}^2$ very close to largest $\overline{R}^2$) and the smallest $C_p$. It is chosen by both the stepwise and backwards procedure. It has the shorter prediction interval of 17,618 – 14,511 = 3,107 compared to 17,601 – 14,460 = 3,141. The $p$-values for the independent variables are all less than .01.

3. Exercise 5.5. Possible violation of normality because histogram appears to be skewed to the right and the normal probability plot deviates from straight line at both ends. However, according to A-D test results, $p$-value = 0.16 > 0.05 or 0.10, so we do not reject $H_0$: Errors are normal. We do not have a strong evidence, though.

Possible violation of the constant variation assumption, because the magnitudes of residuals (especially positive residuals) are larger for the fitted values that are less than 8.5.  Another evidence: the residuals below zero are more than the residuals above zero.

4. Exercise 5.9.  In Figure 5.30 (a), the points seem to fan out as the number of desktops increases.  The service time appears to vary more when more desktops are being serviced.  In Figure 5.30 (b), the points fan out.  The variation of the residuals is greater for greater number of desktops.

5. Exercise 5.10.

a. i. $\hat{y}^* = 5.0206$

A 95% P.I. for $y^*$ = [4.3402, 5.7010]

ii. $\hat{y} = e^{5.0206} = 151.5022$

A 95% $P.I.\ for\ y =\ [e^{4.3402}, \mathrm{e}^{5.7010}] = [76.7229, 299.1664]$.

b. The residual plot is curved indicating the straight-line model does not fit the data appropriately.