SOLUTIONS TO PROBLEMS

4.1 (i) and (iii) generally cause the $t$ statistics not to have a $t$ distribution under $H_0$. Homoskedasticity is one of the CLM assumptions. An important omitted variable violates Assumption MLR.3. The CLM assumptions contain no mention of the sample correlations among independent variables, except to rule out the case where the correlation is one.

4.5 (i) $0.412 \pm 1.96(0.94)$, or about 0.228 to 0.596.

(ii) No, because the value 0.4 is well inside the 95% CI.

(iii) Yes, because 1 is well outside the 95% CI.

4.8 (i) We use Property VAR.3 from Appendix B: $\text{Var}(\mathbf{\beta}_1 - 3 \mathbf{\beta}_2) = \text{Var}(\mathbf{\beta}_1) + 9 \text{Var}(\mathbf{\beta}_2) - 6 \text{Cov}(\mathbf{\beta}_1, \mathbf{\beta}_2)$.

(ii) $t = (\mathbf{\beta}_1 - 3 \mathbf{\beta}_2 - 1)/\text{SE}(\mathbf{\beta}_1 - 3 \mathbf{\beta}_2)$, so we need the standard error of $\mathbf{\beta}_1 - 3 \mathbf{\beta}_2$.

(iii) Because $\theta_1 = \beta_1 - 3\beta_2$, we can write $\beta_1 = \theta_1 + 3\beta_2$. Plugging this into the population model gives

$$y = \beta_0 + (\theta_1 + 3\beta_2)x_1 + \beta_2x_2 + \beta_3x_3 + u$$

$$= \beta_0 + \theta_1x_1 + \beta_2(3x_1 + x_2) + \beta_3x_3 + u.$$  

This last equation is what we would estimate by regressing $y$ on $x_1$, $3x_1 + x_2$, and $x_3$. The coefficient and standard error on $x_1$ are what we want.
Holding other factors fixed

\[ \Delta \text{vote}_A = \beta_1 \Delta \log(\text{expend}_A) = (\beta_1 / 100) [100 \cdot \Delta \log(\text{expend}_A)] \approx (\beta_1 / 100) (\%\Delta \text{expend}_A), \]

where we use the fact that \(100 \cdot \Delta \log(\text{expend}_A) \approx \%\Delta \text{expend}_A\). So \(\beta_1 / 100\) is the (ceteris paribus) percentage point change in \(\text{vote}_A\) when \(\text{expend}_A\) increases by one percent.

(ii) The null hypothesis is \(H_0: \beta_2 = -\beta_1\), which means a \(z\)% increase in expenditure by A and a \(z\)% increase in expenditure by B leaves \(\text{vote}_A\) unchanged. We can equivalently write \(H_0: \beta_1 + \beta_2 = 0\).

(iii) The estimated equation (with standard errors in parentheses below estimates) is

\[ \text{vote}_A = 45.08 + 6.083 \log(\text{expend}_A) - 6.615 \log(\text{expend}_B) + .152 \text{prtyst}_A \]

\[ (3.93) \quad (0.382) \quad (0.379) \quad (.062) \]

\[ n = 173, \quad R^2 = .793. \]

The coefficient on \(\log(\text{expend}_A)\) is very significant (\(t\) statistic \(\approx 15.92\)), as is the coefficient on \(\log(\text{expend}_B)\) (\(t\) statistic \(\approx -17.45\)). The estimates imply that a 10% ceteris paribus increase in spending by candidate A increases the predicted share of the vote going to A by about .61 percentage points. [Recall that, holding other factors fixed, \(\Delta \text{vote}_A \approx (6.083/100) \%\Delta \text{expend}_A\).]

Similarly, a 10% ceteris paribus increase in spending by B reduces \(\text{vote}_A\) by about .66 percentage points. These effects certainly cannot be ignored.

While the coefficients on \(\log(\text{expend}_A)\) and \(\log(\text{expend}_B)\) are of similar magnitudes (and opposite in sign, as we expect), we do not have the standard error of \(\hat{\beta}_1 + \hat{\beta}_2\), which is what we would need to test the hypothesis from part (ii).

(iv) Write \(\theta_1 = \beta_1 + \beta_2\), or \(\beta_1 = \theta_1 - \beta_2\). Plugging this into the original equation, and rearranging, gives

\[ \hat{\text{vote}}_A = \beta_0 + \theta_1 \log(\text{expend}_A) + \beta_2 [\log(\text{expend}_B) - \log(\text{expend}_A)] + \beta_3 \text{prtyst}_A + u, \]

When we estimate this equation we obtain \(\hat{\theta}_1 \approx -.532\) and \(\text{se}(\hat{\theta}_1) \approx .533\). The \(t\) statistic for the hypothesis in part (ii) is \(-.532/.533 \approx -1\). Therefore, we fail to reject \(H_0: \beta_2 = -\beta_1\).
4.16 (i) If we drop \( rbisyr \) the estimated equation becomes

\[
\log(\text{salary}) = 11.02 + 0.0677 \text{ years} + 0.0158 \text{ gamesyr} \\
(0.27) \quad (0.0121) \quad (0.0016) \\
+ 0.0014 \text{ bavg} + 0.0359 \text{ hrunsyr} \\
(0.0011) \quad (0.0072)
\]

\( n = 353, \quad R^2 = 0.625. \)

Now \( hrunsyr \) is very statistically significant (\( t \) statistic \( \approx 4.99 \)), and its coefficient has increased by about two and one-half times.

(ii) The equation with \textit{runsysr, fldperc, and sbasesyr} added is

\[
\log(\text{salary}) = 10.41 + 0.0700 \text{ years} + 0.0079 \text{ gamesyr} \\
(2.00) \quad (0.0120) \quad (0.0027) \\
+ 0.0053 \text{ bavg} + 0.0232 \text{ hrunsyr} \\
(0.00110) \quad (0.0086) \\
+ 0.0174 \text{ runsysr} + 0.0010 \text{ fldperc} - 0.0064 \text{ sbasesyr} \\
(0.0051) \quad (0.0020) \quad (0.0052)
\]

\( n = 353, \quad R^2 = 0.639. \)

Of the three additional independent variables, only \textit{runsysr} is statistically significant (\( t \) statistic \( = 0.0174/0.0051 \approx 3.41 \)). The estimate implies that one more run per year, other factors fixed, increases predicted salary by about 1.74%, a substantial increase. The stolen bases variable even has the "wrong" sign with a \( t \) statistic of about \(-1.23 \), while \textit{fldperc} has a \( t \) statistic of only \(.5. \)

Most major league baseball players are pretty good fielders; in fact, the smallest \textit{fldperc} is 800

4.16 cont

(which means \(.800). \) With relatively little variation in \textit{fldperc}, it is perhaps not surprising that its effect is hard to estimate.

(iii) From their \( t \) statistics, \textit{bavg, fldperc, and sbasesyr} are individually insignificant. The \( F \) statistic for their joint significance (with 3 and 345 df) is about \(.69 \) with \( p \)-value \( \approx .56. \)

Therefore, these variables are jointly very insignificant.