Problemløsninger PS1 – BA-BMECV1031U

1 (a) The results from the OLS regression, with standard errors in parentheses, are

$$\widehat{\log}(psoda) = \underset{(0.29)}{-1.46} + \underset{(.031)}{.073}\, prpblck + \underset{(.027)}{.137} \log(income) + \underset{(.133)}{.380}\, prppov$$

$n = 401$, $R^2 = .087$

The p-value for testing H_0: $\beta_1 = 0$ against the two-sided alternative is about .018, so that we reject H_0 at the 5% level but not at the 1% level.

(b) The correlation is about –.84, indicating a strong degree of multicollinearity. Yet each coefficient is very statistically significant: the t statistic for $\hat{\beta}_{\log(income)}$ is about 5.1 and that for $\hat{\beta}_{prppov}$ is about 2.86 (two-sided p-value = .004).

(c) The OLS regression results when log(*hseval*) is added are

$$\widehat{\log}(psoda) = \underset{(.29)}{-.84} + \underset{(.029)}{.098}\, prpblck - \underset{(.038)}{.053} \log(income)$$

$$+ \underset{(.134)}{.052}\, prppov + \underset{(.018)}{.121} \log(hseval)$$

$n = 401$, $R^2 = .184$

The coefficient on log(*hseval*) is an elasticity: a one percent increase in housing value, holding the other variables fixed, increases the predicted price by about .12 percent. The two-sided p-value is zero to three decimal places.

(d) Adding log(*hseval*) makes log(*income*) and *prppov* individually insignificant (at even the 15% significance level against a two-sided alternative for log(*income*), and *prppov* is does not have a t statistic even close to one in absolute value). Nevertheless, they are jointly significant at the 5% level because the outcome of the $F_{2,396}$ statistic is about 3.52 with p-value = .030. All of the control variables – log(*income*), *prppov*, and log(*hseval*) – are highly correlated, so it is not surprising that some are individually insignificant.

(e) Because the regression in (c) contains the most controls, log(*hseval*) is individually significant, and log(*income*) and *prppov* are jointly significant, (c) seems the most reliable. It holds fixed three measure of income and affluence. Therefore, a reasonable estimate is that if the proportion of blacks increases by .10, *psoda* is estimated to increase by 1%, other factors held fixed.

2 (a) The causal (or ceteris paribus) effect of *dist* on *price* means that $\beta_1 \geq 0$: all other relevant factors equal, it is better to have a home farther away from the incinerator. The estimated equation is

$$\widehat{\log}(price) = \underset{(0.65)}{8.05} + \underset{(.066)}{.365} \log(dist)$$

$$n = 142,\ R^2 = .180,\ \bar{R}^2 = .174,$$

which means a 1% increase in distance from the incinerator is associated with a predicted price that is about .37% higher.

(b) When the variables log(*inst*), log(*area*), log(*land*), *rooms*, *baths*, and *age* are added to the regression, the coefficient on log(*dist*) becomes about .055 ($se \approx .058$). The effect is much smaller now, and statistically insignificant. This is because we have explicitly controlled for several other factors that determine the quality of a home (such as its size and number of baths) and its location (distance to the interstate). This is consistent with the hypothesis that the incinerator was located near less desirable homes to begin with.

(c) When $[\log(inst)]^2$ is added to the regression in part (b), we obtain (with the results only partially reported)

$$\widehat{\log}(price) = \underset{(2.65)}{-3.32} + \underset{(.062)}{.185} \log(dist) + \underset{(0.501)}{2.073} \log(inst) - \underset{(.0282)}{.1193} [\log(inst)]^2 + \ldots$$

$$n = 142,\ R^2 = .778,\ \bar{R}^2 = .764.$$

The coefficient on log(*dist*) is now very statistically significant, with a *t* statistic of about three. The coefficients on log(*inst*) and $[\log(inst)]^2$ are both very statistically significant, each with *t* statistics above four in absolute value. Just adding $[\log(inst)]^2$ has had a very big effect on the coefficient important for policy purposes. This means that distance from the incinerator and distance from the interstate are correlated in some nonlinear way that also affects housing price.

We can find the value of log(*inst*) where the effect on log(*price*) actually becomes negative: $2.073/[2(.1193)] \approx 8.69$. When we exponentiate this we obtain about 5,943 feet from the interstate. Therefore, it is best to have your home away from the interstate for distances less than just over a mile. After that, moving farther away from the interstate lowers predicted house price.

(d) The coefficient on $[\log(dist)]^2$, when it is added to the model estimated in part (c), is about -.0365, but its *t* statistic is only about -.33. Therefore, it is not necessary to add this complication.

3 (a) Holding other factors fixed,

$$\Delta voteA \;=\; \beta_1 \Delta \log(expendA) \;=\; (\beta_1/100)[100 \cdot \Delta \log(expendA)]$$
$$\approx\; (\beta_1/100)(\%\Delta expendA),$$

where we use the fact that $100 \cdot \Delta \log(expendA) \;\approx\; \%\Delta expendA$. So $\beta_1/100$ is the (ceteris paribus) percentage point change in *voteA* when *expendA* increases by one percent.

(b) 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 *voteA* unchanged. We can equivalently write H$_0$: $\beta_1 + \beta_2 = 0$.

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

$$\underset{}{\widehat{voteA}} = \underset{(3.93)}{45.08} + \underset{(0.382)}{6.083}\log(expendA) - \underset{(0.379)}{6.615}\log(expendB) + \underset{(.062)}{.152}\,prtystrA$$

$$n \;=\; 173, \quad R^2 \;=\; .793.$$

The coefficient on log(*expendA*) is very significant (t statistic $\approx$ 15.92), as is the coefficient on log(*expendB*) (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, $\Delta\widehat{voteA} \approx (6.083/100)\%\Delta expendA$).]
Similarly, a 10% ceteris paribus increase in spending by B reduces $\widehat{voteA}$ by about .66 percentage points. These effects certainly cannot be ignored.

While the coefficients on log(*expendA*) and log(*expendB*) 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 (b).

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

$$\widehat{voteA} \;=\; \beta_0 + \theta_1 \log(expendA) + \beta_2[\log(expendB) - \log(expendA)] + \beta_3\, prtystrA + u,$$

When we estimate this equation we obtain $\hat{\theta}_1 \approx -.532$ and se($\hat{\theta}_1$) $\approx$ .533. The t statistic for the hypothesis in part (b) is $-.532/.533 \approx -1$. Therefore, we fail to reject H$_0$: $\beta_2 = -\beta_1$.