Solution to Homework 6

Chapter 6

6.3 (i) The turnaround point is given by $\hat{\beta}_1/(2|\hat{\beta}_2|)$, or .0003/(.000000014) ≈ 21,428.57; remember, this is sales in millions of dollars.

(ii) Probably. Its *t* statistic is about –1.89, which is significant against the one-sided alternative H_0: $\beta_1 < 0$ at the 5% level (*cv* ≈ –1.70 with *df* = 29). In fact, the *p*-value is about .036.

(iii) Because *sales* gets divided by 1,000 to obtain *salesbil*, the corresponding coefficient gets multiplied by 1,000: (1,000)(.00030) = .30. The standard error gets multiplied by the same factor. As stated in the hint, $salesbil^2 = sales/1{,}000{,}000$, and so the coefficient on the quadratic gets multiplied by one million: (1,000,000)(.0000000070) = .0070; its standard error also gets multiplied by one million. Nothing happens to the intercept (because *rdintens* has not been rescaled) or to the R^2.

$$\widehat{rdintens} = \underset{(0.429)}{2.613} + \underset{(.14)}{.30}\, salesbil - \underset{(.0037)}{.0070}\, salesbil^2$$

$n = 32, \quad R^2 = .1484.$

(iv) The equation in part (iii) is easier to read because it contains fewer zeros to the right of the decimal. Of course the interpretation of the two equations is identical once the different scales are accounted for.

Chapter 7

7.1 (i) The coefficient on *male* is 87.75, so a man is estimated to sleep almost one and one-half hours more per week than a comparable woman. Further, $t_{male} = 87.75/34.33 \approx 2.56$, which is close to the 1% critical value against a two-sided alternative (about 2.58). Thus, the evidence for a gender differential is fairly strong.

(ii) The *t* statistic on *totwrk* is –.163/.018 ≈ –9.06, which is very statistically significant. The coefficient implies that one more hour of work (60 minutes) is associated with .163(60) ≈ 9.8 minutes less sleep.

(iii) To obtain R_r^2, the *R*-squared from the restricted regression, we need to estimate the model without *age* and age^2. When *age* and age^2 are both in the model, *age* has no effect only if the parameters on both terms are zero.

7.8 (i) We want to have a constant semi-elasticity model, so a standard wage equation with marijuana usage included would be

$$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u.$$

Then $100 \cdot \beta_1$ is the approximate percentage change in *wage* when marijuana usage increases by one time per month.

(ii) We would add an interaction term in *female* and *usage*:

$$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + \beta_6 female \cdot usage + u.$$

The null hypothesis that the effect of marijuana usage does not differ by gender is H_0: $\beta_6 = 0$.

(iii) We take the base group to be nonuser. Then we need dummy variables for the other three groups: *lghtuser*, *moduser*, and *hvyuser*. Assuming no interactive effect with gender, the model would be

$$\log(wage) = \beta_0 + \delta_1 lghtuser + \delta_2 moduser + \delta_3 hvyuser + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u.$$

(iv) The null hypothesis is H_0: $\delta_1 = 0$, $\delta_2 = 0$, $\delta_3 = 0$, for a total of $q = 3$ restrictions. If n is the sample size, the *df* in the unrestricted model – the denominator *df* in the F distribution – is $n - 8$. So we would obtain the critical value from the $F_{q,n\text{-}8}$ distribution.

(v) The error term could contain factors, such as family background (including parental history of drug abuse), that could directly affect wages and also be correlated with marijuana usage. We are interested in the effects of a person's drug usage on his or her wage, so we would like to hold other confounding factors fixed. We could try to collect data on relevant background information.

7.11 (i) The coefficient on *male* in the second equation means that in the sample of 856 students, if we consider males and females with the same level of grade point average at the start of the term, the male scores on average 3.83, which is more than the female. The 95% confidence interval is $3.83 \pm (1.96)0.74$ or about 2.38 to 5.28. Clearly, this interval excludes zero.

(ii) The F test for joint significance of *male* and *colgpa*, with 1 and 852 *df*, is about .12 with *p*-value ≈ .729; these variables are jointly very insignificant. Therefore, we can conclude that there is no gender difference in the model.

(iii) The last equation is obtained by replacing *male.colgpa* with *male*.(*colgpa* – 2.81); the coefficient on *male* is now gender difference when *colgpa* = 2.81. From part (ii), we observe that there is no gender difference in the model with the interaction and so the coefficient on male is closer to that in the second equation.

SOLUTIONS TO COMPUTER EXERCISES

C7.4 (i) The two signs that are pretty clear are $\beta_3 < 0$ (because *hsperc* is defined so that the smaller the number the better the student) and $\beta_4 > 0$. The effect of size of graduating class is not clear. It is also unclear whether males and females have systematically different GPAs. We may think that $\beta_6 < 0$, that is, athletes do worse than other students with comparable characteristics. But remember, we are controlling for ability to some degree with *hsperc* and *sat*.

(ii) The estimated equation is

$$\widehat{\text{colgpa}} = \underset{(0.079)}{1.241} - \underset{(.0164)}{.0569}\, hsize + \underset{(.00225)}{.00468}\, hsize^2 - \underset{(.0006)}{.0132}\, hsperc$$

$$+ \underset{(.00007)}{.00165}\, sat + \underset{(.018)}{.155}\, female + \underset{(.042)}{.169}\, athlete$$

$$n = 4{,}137, \quad R^2 = .293.$$

Holding other factors fixed, an athlete is predicted to have a GPA about .169 points *higher* than a nonathlete. The t statistic $.169/.042 \approx 4.02$, which is very significant.

(iii) With *sat* dropped from the model, the coefficient on *athlete* becomes about .0054 (se$\approx$ .0448), which is practically and statistically not different from zero. This happens because we do not control for SAT scores, and athletes score lower on average than nonathletes. Part (ii) shows that, once we account for SAT differences, athletes do better than nonathletes. Even if we do not control for SAT score, there is no difference.

(iv) To facilitate testing the hypothesis that there is no difference between women athletes and women nonathletes, we should choose one of these as the base group. We choose female nonathletes. The estimated equation is

$$\widehat{\text{colgpa}} = \underset{(0.076)}{1.396} - \underset{(.0164)}{.0568}\, hsize + \underset{(.00225)}{.00467}\, hsize^2 - \underset{(.0006)}{.0132}\, hsperc$$

$$+ \underset{(.00007)}{.00165}\, sat + \underset{(.084)}{.175}\, femath + \underset{(.049)}{.013}\, maleath - \underset{(.018)}{.155}\, malenonath$$

$$n = 4{,}137, \quad \text{R}^2 = .293.$$

The coefficient on *femath* = *female·athlete* shows that *colgpa* is predicted to be about .175 points higher for a female athlete than a female nonathlete, other variables in the equation fixed. The hypothesis that there is no difference between female athletes and female nonathletes is tested by using the t statistic on *femath*. In this case, $t = 2.08$, which is statistically significant at the 5% level against a two-sided alternative.

(v) Whether we add the interaction *female·sat* to the equation in part (ii) or part (iv), the outcome is practically the same. For example, when *female·sat* is added to the equation in part (ii), its coefficient is about .000051 and its t statistic is about .40. There is very little evidence that the effect of *sat* differs by gender.