1. Use dataset AED_KNEEREPLACE.DTA.

(a) Regress medcharge against medcost.

(b) Is the coefficient for medcost statistically significant at significance level 0.05? Explain.

(c) Test the hypothesis that the coefficient for medcost equals one against the alternative that it does not equal one at significance level 0.05. What do you conclude? (Hint: This requires some extra computation).

(d) Test the claim that a one dollar increase in cost is associated with a more than one dollar increases in the charge, at significance level 0:05. What do you conclude?

(e) Which of assumptions 1-4 in the notes are necessary for this analysis to be valid?

2. Use data on health expenditures and outcomes in various OECD countries. The data used is in file oecdhealth2008.dta. Use describe for variable descriptions (Similar analysis of 2009 data is given in chapter 11 of the coursepack).

(a) If per capita GDP rises by $1,000, by how much does the % of GDP spent on health change?

(b) If the slope coefficient equals zero what is the implied income elasticity of health expenditures? Explain.

(c) Test at 5% whether or not there is a relationship between healthgdp and gdppc.

(d) Are there any outliers?

(e) Which countries are outliers?

3. Use data in file ADVERTISING.DTA at the course website. The data are for 200 regional markets in the U.S.

(a) Regress sales on tv.

(b) Predict population conditional mean sales for TV advertising expenditure of $100,000.

(c) Provide a 95% confidence interval for this population conditional mean with TV advertising expenditure of $100,000. [Hint: Use the formula for se($\widehat{y_{CM}}$) given in the course pack. You will need output $s_e, n, \bar{x}, and\ s_x^2$ and can compute $\sum_{i=1}^{n}(x_i - \bar{x})^2$as s_x^2multiplied by $n-1$, since $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.]

(d) Provide a 95% confidence interval for actual sales with TV advertising expenditure of $100,000. [Hint:
Use $se(\widehat{y_f}) = s_e \times \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$.]

(e) What graph would best fit the data? What do you learn from the combined graph?

4.

(a) Create a two way scatter plot between sales and tv.

(b) Given your graph in (a), do the errors appear to be homoskedastic or heteroskedastic?

(c) How have the results changed by using heteroskedastic-robust standard errors rather than default standard errors?