POLS 6481, Spring 2021

Professor Scott Basinger

Reading Assignment Week 2

Distributed Friday, January 22

Due Thursday, January 28

Wooldridge (1.1 + 1.2 +) 2.1 + 2.2 + 2.3 + 2.5

Fox, *Regression Diagnostics*, pp. 21-40

1. What is the *residual* of a simple linear regression model? How is it defined? What rules must it obey?

2. Other than the fact that one pair of equations uses expectations and the other pair uses sums…

What is the difference between writing “E(*u*) = 0 and E(*xu*) = 0)” [equations 2.10 and 2.11] and writing “ = 0 and = 0” [equations 2.30 and 2.31]?

3. Explain why / how equation 2.31 { = 0} could be satisfied despite equation 2.11 { E(*xu*) = 0} being violated. What would be the consequences of this mismatch?

4. Fill in the blanks using (some of) the following terms: “discrepancy” “influential” “leverage” “DfBeta” “DfFits” “hat values” “residuals”

If a data point has an atypical value in the horizontal (*x*) dimension, then it will have high \_\_\_\_\_\_, which is measured with .

If a data point has an atypical value in the vertical (*y*) dimension, then it may have a large \_\_\_\_\_\_, which is measured with .

If a data point is unusual in both dimensions, then it may have a large effect on the regression slope, which can be measured with , and it may have a large effect on fitted values, which can be measured with .

5. In the figure below, dotted lines express the relationship between *x* and *y* using just the ○ data points; the solid line re-expresses the relationship with the data point added.

A. Which panel shows a case of high discrepancy but low leverage?

In this panel, what is the effect on the slope of adding the data point?

In this panel, what is the effect on the intercept of adding the data point?

B. The other panels show cases of high leverage; explain why the *residual* is not sufficient for diagnosing whether the data point is influential.

