Midterm

Economics 435: Quantitative Methods

Fall 2009

1 Properties of expected values

a) Prove that $cov(x, y) = E(xy) - E(x)E(y)$.

b) Prove that if $E(u|x) = 0$, then $cov(x, u) = 0$.

c) Let y be a random variable and let x be a binary random variable. Show that:

$$\frac{E(xy)}{E(x)} = E(y|x = 1)$$

This question is a little harder than (a) or (b), so if you're not sure how to do it, skip ahead to question 2.

2 The consequences of skipping class

A recent paper[1] by Dobkin, Gil, and Marion (DGM) measures the effect of class attendance on final exam grades in a large economics lecture class. In the class analyzed by DGM, attendance is mandatory for all students who score below the class median on the midterm. DGM use this institutional detail to devise what's known as a regression discontinuity design for measuring both the effect of attending class and the effect of being required to attend class. We will stick to answering the second question, simply because it's easier. I have also simplified their empirical setting somewhat.

You have a data set $D_n = \{(a_i, m_i, r_i, c_i, f_i)\}_{i=1}^n$ where:

$$\begin{aligned}
a_i &= \begin{cases} 1 & \text{if student } i \text{ attended class regularly after the midterm} \\ 0 & \text{if student } i \text{ did not attended class regularly after the midterm} \end{cases} \\
m_i &= \text{midterm grade of student } i \\
r_i &= \begin{cases} 1 & \text{if student } i\text{'s midterm grade is below the class median} \\ 0 & \text{if student } i\text{'s midterm grade is above the class median} \end{cases} \\
c_i &= \begin{cases} 1 & \text{if student } i\text{'s midterm grade is close to (within } \epsilon \text{ points of) the class median} \\ 0 & \text{if student } i\text{'s midterm grade is not close to the class median} \end{cases} \\
f_i &= \text{final exam grade of student } i \\
f_i(1) &= \text{final exam grade student } i \text{ would receive if required to attend class regularly} \\
f_i(0) &= \text{final exam grade student } i \text{ would receive if not required to attend class regularly}
\end{aligned}$$

You should interpret your data set as a random sample from a larger population, and should assume that there is no exact linear dependence in the sample among any of the variables.

[1] The paper is "Causes and Consequences of Skipping Class in College" (*Economics of Education Review*, forthcoming) and is available at `http://people.ucsc.edu/~marion/`.

a) Would an OLS regression of f_i on r_i provide a good measure of the effect of mandatory attendance? Briefly (in a sentence or two) explain why.

b) Let TE_i be the effect of being required to attend class on the final exam grade of student i. Define TE_i in terms of the variables above.

c) One possible identifying assumption is that among students whose miderm score is close to the median, students that are just above the median are no different in potential outcomes from students who are just below the median. In other words:

$$E(f_i(1)|r_i = 1, c_i = 1) = E(f_i(1)|r_i = 0, c_i = 1) = E(f_i(1)|c_i = 1) \quad (1)$$

$$E(f_i(0)|r_i = 1, c_i = 1) = E(f_i(0)|r_i = 0, c_i = 1) = E(f_i(0)|c_i = 1) \quad (2)$$

Given assumptions (1) and (2), prove that:

$$A\hat{T}E_c = \frac{\frac{1}{n}\sum_{i=1}^{n} f_i r_i c_i}{\frac{1}{n}\sum_{i=1}^{n} r_i c_i} - \frac{\frac{1}{n}\sum_{i=1}^{n} f_i(1-r_i)c_i}{\frac{1}{n}\sum_{i=1}^{n}(1-r_i)c_i}$$

is a consistent estimator for $E(TE_i|c_i = 1)$. Hints:

- You should find the result in part (c) of quesiton 1 of this exam useful.
- You should also find it useful to know that if r and c are both binary
  - rc and $(1-r)c$ are also binary.
  - $rc = 1$ if and only if $r = 1$ and $c = 1$.
  - $(1-r)c = 1$ if and only if $r = 0$ and $c = 1$.

d) As a researcher, you get to decide on how close counts as "close", i.e. the value of ϵ. What are the advantages and disadvantages of setting a very small value of ϵ?

e) Suppose that:

$$E(f_i(0)|m_i, r_i) = \beta_0 + \beta_2 m_i \quad (3)$$

$$E(f_i(1)|m_i, r_i) = \beta_1 + \beta_2 m_i \quad (4)$$

Find $E(TE_i)$ in terms of the model parameters $(\beta_0, \beta_1, \beta_2)$.

f) Let $\hat{\beta}_r$ be the regression coefficient on r_i from an OLS regression of f_i on (m_i, r_i). Given assumptions (3) and (4), prove that $\hat{\beta}_r$ is a consistent estimator for $E(TE_i)$. You can use the theorem that describes conditions (MLR1-MLR4) for consistency of OLS.