PS 236: Causal Inference
Problem Set 1

UC Berkeley, Fall 2008

Due: Thursday, September 25

Your solutions must be submitted in hard copy to my mailbox in the Political Science main office (Barrows Hall 210) prior to the office's closure on the due date (prior to 5pm). Except under extraordinary circumstances, no late assignments will be accepted.

Question 1

A fair coin is flipped twice. Let $H\cdot$ and $\cdot H$ represent heads on the first and second flip respectively, while the other flip may be heads or tails. For parts a–d, define the sample space and the set of desired outcomes and calculate the probabilities. Explain your responses to parts e and f.

a. $\Pr(\cdot H)$

b. $\Pr(\cdot H \cap H\cdot)$

c. $\Pr(\cdot H \cup H\cdot)$

d. $\Pr(\cdot H | H\cdot)$

e. Provide a pair of independent events in this problem.

f. Provide a pair of events that are not independent.

Question 2

Assume that the true relationship between the random variables X and Y is $Y = X^2$. Let X be distributed symmetrically around 0; *i.e.*, $f(x) = f(-x)\ \forall\ x \in \mathbb{R}$. A researcher, unaware of the true model relating X and Y, estimates $Y_i = \alpha + \beta X_i + \epsilon_i$ using a random sample of X. The latter point posits that the empirical distribution of X is identical to its theoretical distribution.

a. What is the covariance between Y and X? *Hint*: The integration by substitution rule implies that

$$\int_a^b g(x)f(x)\,dx = \int_{-b}^{-a} g(-x)f(-x)\,dx$$

b. Calculate $\hat{\beta}$.

c. Calculate $\hat{\alpha}$.

d. What might the researcher conclude about the relationship between X and Y?

Question 3

Using the home ownership data file that was used in section, perform the following analyses in `R`. Please provide the answers in two separate parts. In the first, simply give the answer to the question. In the second, provide "clean" code that generates the solutions that you have given, but without extraneous code or `R` printouts (carrots, calculations, *etc.*). The code should be printed as a single file, with comments used to define the code relevant to each question.

a. Do any variables have missing values? If so, how many?

Create a new variable that is equal to 1 if the `poverty` variable, the ratio of an individual's income to the state poverty line, is less than 100 and 0 otherwise. This converts the poverty ratio to an indicator of poverty status.

People own their homes if the `ownershd` variable is equal to `Owned with mortgage or loan` or `Owned free and clear`. They rent their homes if this variable is equal to `With cash rent` or `No cash rent`.

b. How many observations do not fit into the homeowner *v.* renter dichotomy?

c. What fraction of homeowners is in poverty? What fraction of renters?

d. Are a larger fraction of homeowners in poverty in California or in Oregon?

e. What are the income means, standard deviations, and quartiles for homeowners and for renters?

f. What are the income means, standard deviations, and quartiles for homeowners and for renters in poverty?

g. Give the median income by state.