Introduction to Statistics: Homework 1 (Multivariate Regression and Causality)

1.

a. Write two survey questions that you could use to measure people's *generosity*. Include details about what the response options would be and their numerical ranges.

   1) How many dollars did you donate to charity last month? Responses would be entered as dollar amounts. They would range from 0 to whatever the highest dollar amount people entered was.

   2) Some people are very generous while others prefer to focus on taking care of themselves. How would you rate yourself on a scale from 0 to 10 where 0 means you prefer to focus on taking care of yourself and 10 means you are very generous. This variable would range from 0 to 10.

b.

	Coefficient	Standard Error	T
Dollars Donated	0.020	.005	4.0
Constant	1.2	.020	6.0

c. Support for Government Spending = 1.2 + 0.020*(Dollars Donate to Charity Last Month) + u

d. The coefficient on the constant is the estimated support for government aid to the poor among respondents who donated zero dollars last month.

e. The coefficient on Dollars Donated is an estimate of the relationship between this variable and support for government aid to the poor. For every additional dollar donated last month we expect support for aid to the poor to increase by 0.020 units.

f. Political ideology is one variable that might be expected to be associated with support for government aid to the poor. Because liberals tend to be more supportive of government spending we would expect liberals to be more supportive of government aid to the poor and conservatives to be less supportive. Whether including this variable as an independent variable would affect the coefficient on the measure of generosity would depend on whether ideology is associated with generosity. If liberals tend to be more generous than conservatives, then including a measure of ideology in the model would probably reduce the size of the coefficient on generosity. This is because including ideology as a control is a way of avoiding giving generosity explanatory "credit" for variation in support for government aid to the poor that can also be explained by ideology.

2. For this set of questions use the “mortality” dataset. This data is from 1992. We are interested in what affects the level of infant mortality in a country.
   The variables in the dataset are:
   MORTINFT: Number of infant deaths per 1,000 live births.
   TVPERCAP: Number of television sets per capita
   PHYSTOT: Number of physicians per 1,000 people
   CTRYNAME: Name of the country

   a. Units of analysis: Countries
   b. Ranges
      MORTINFT: 3.9 to 144 (also acceptable: 140.1)
      TVPERCAP: 0 to 0.75 (also acceptable: 0.75)
      PHYSTOT: 0.02 to 4.58 (also acceptable: 4.56)
   c. MORTINFT: 43.96
      TVPERCAP: 0.197
      PHYSTOT: 1.46
   d.
      i. Infant Mortality = 74.81 + (-156.44)(TVs per capita) + u
      ii. Yes, TVs per capita is statistically significant. The absolute value of T is greater than 2 and the p-value is less than .05.
      iii. -19.054 [74.81 + (-156.44)(.6)]
      iv. The relationship between the number of TVs per capita and the infant mortality rate is probably not causal. Unless we think that TVs are an important source of information related to public health, this is likely to be a spurious relationship. In other words, this is likely due to the fact that the measure of TVs per capita is associated with other factors that cause lower infant mortality rates.
   e.
      i. 0.6841
      ii. The variable physicians per 1,000 people is positively associated with TVs per capita. The more physicians per 1,000 people there are in a country, more TVs per capita there are.
   f. Next, run a multivariate regression predicting infant mortality using the number of TVs per capita and the number of physicians per 1,000 people as independent variables.

i. Infant Mortality = 78.62 + (-108.42)(TVs per capita) + (-9.08)(Physicians per 1,000 people) + u

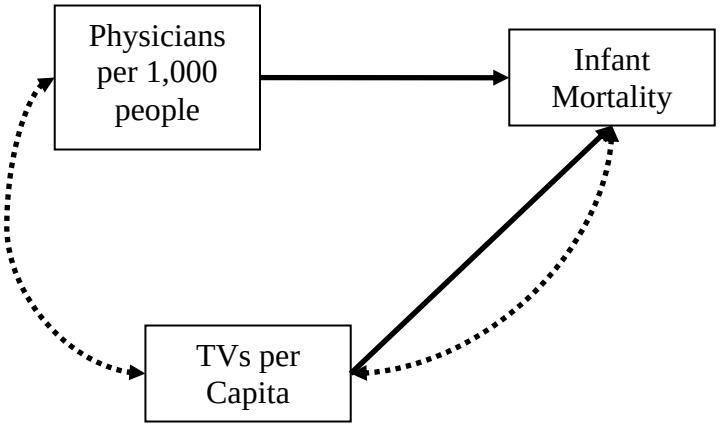
ii. Yes, number of televisions per capita is a statistically significant predictor of infant mortality. We know because although the absolute value of T statistic is somewhat smaller than in the bivariate regression it is still greater than 2. Also the p-value is less than .05. Physicians per 1,000 people is also a statistically significant predictor. The absolute value of the t-statistic is greater than 2 and the p-value is less than .05.

iii. 8.01 [78.62 -108.42*.4 - 9.08*3]

iv. Yes, it is smaller in this model. This is probably due to the fact that TVs per capita is positively associated with physicians per 1,000 people. Controlling for the number of physicians provides a way to account for the fact that the association between TVs per capita and infant mortality can, in part, be explained by the fact that countries with more TVs per capita tend to have more physicians.

g. My interpretation is that there is a causal relationship between physicians per 1,000 people and infant mortality. Having more doctors available reduces the likelihood of infant mortality. I would guess that the relationship between TVs per capita and physicians per 1,000 people is not causal at all. TVs per capita may have a small causal effect on infant mortality (because TVs can provide access to health information). This is indicated by the solid arrow connecting these two variables. However, not all of the relationship between these two variables is causal – most can probably be explained by other factors that affect both TVs per capita and Infant mortality.

h.



i. Income per capita might also confound the relationship between TVs per capita and infant mortality. This could be measured as average monthly income in each country (in US dollars).

ii. Including income per capita in the model would probably reduce the size of the coefficient on TVs per capita. This is because I would expect this variable to be positively associated with both TVs per capita and infant mortality. We would expect the relationship between TVs per capita and infant mortality to be weaker after accounting for variation in income per capita.

iii. Including per capita income would probably also reduce the size of the coefficient on physicians per capita. Income is likely to be positively associated with physicians per 1,000 people and with infant mortality. Without controlling for income per capita, physicians per capita could be getting "credit" for the associations between other economic factors that are likely to affect infant mortality (e.g., nutrition and access to clean water).

iv. I would draw causal arrows from income per capita to each of the three variables because I would expect that income causally affects each of these other measures.