

**Homework 1 Solutions:**

**Chapter 1: #12, 13**

**Chapter 1: #2, 4, 5, 8, 18, 29, 30**

**Assigned Mon, Sept 28:**

- 1.12**    **a.** Observational study. Participants were not randomly assigned to exercise varying amounts, they were simply observed to see how much they exercised.  
**b.** The conclusion reached by the investigator is not justified because the study was an observational study and not a randomized experiment. For instance, it's possible that the cause and effect relationship is in the other direction: people who are often sick are not able to exercise as much. There are many possible confounding variables, that lead people to both exercise different amounts and have different numbers of colds, that have nothing to do with the exercise *causing* fewer colds. (See part c for examples.)  
**c.** Possible explanatory variables are general health, diet, climate where they live, pollution where they live, occupation. All of these might lead some people to both exercise more and have fewer colds, and others to exercise less and have more colds.  
**d.** The investigator would have to randomly assign people to different amounts of exercise. This may not be ethical, and it may be difficult to get cooperation, but it is the only way to be able to conclude a cause and effect relationship.
- 1.13**    **a.** Observational study. Participants were not randomly assigned to spend different amounts of time in class preparation, they were simply asked how much time they spent.  
**b.** The conclusion reached by the investigator is not justified because the study was an observational study and not a randomized experiment. A cause and effect conclusion cannot be made.  
**c.** Possible confounding variables are general conscientiousness, other obligations that take time away from both preparing for the seminar and being productive on the job, commitment to the company, and so on. A correct answer should be something that is related to amount of time spent preparing for class, and that affects productivity.  
**d.** The investigator would have to randomly assign people to different amounts of class preparation time. For instance, the employer could designate specific work hours for class preparation, that differed for different employees, randomly assigned.

**Assigned Wed, Sept 30:**

- 1.2**     $Y = 300 + 2X$ ; this is a functional relation, not a statistical one. There is no variability in  $Y$  for a given  $X$ .
- 1.4**    This illustrates the error term,  $\epsilon_i$ . If there were no error terms, the  $Y$  value would be the same for all units with the same  $X$  value.
- 1.5**    The student is incorrect. The  $E\{Y\}$  does not include the error term,  $\epsilon_i$ . So either the student should remove the  $E$  on the left hand side, or remove the error term on the right hand side.
- 1.8**     $E\{Y\}$  would still be 104 but the  $Y$  value for a new case wouldn't necessarily be 108.  $E\{Y\}$  is a fixed constant, but each individual value of  $Y$  is a random variable.
- 1.18**    It is not necessarily true that  $\sum \epsilon = 0$  when summed over the  $n$  cases. It is true that it would be 0 if summed over the entire population, but not for any particular set of  $n$  cases.

- 1.29** The implication is that when  $X = 0$ ,  $Y = 0$  as well. The regression function plot would be a line going through the origin ( $X = 0$ ,  $Y = 0$ ) with a slope of  $\beta_1$ .
- 1.30** The implication is that  $X$  is no help in trying to explain or predict  $Y$ . The regression function would plot as a horizontal line at  $Y = \beta_0$ , the same across the entire range of  $X$ .