Extra Practice Problems for Midterm Exam

1. Suppose you claim that X causes Y and you find the expected correlation between measure of X and Y. There are four ways to obtain a correlation between X and Y. What are they? What role does randomization play in eliminating two of the possibilities?

2. How do we typically choose the line that the "best fits" the points in a scatterplot? What are the implications for big, medium, and small errors?