

LM Homework 1

Joshua Ingram

2/4/2020

Problem 1

a

This result can not be taken as evidence that completing homework assignments *causes* higher grades on the final exam, however it could be evidence that there is some form of an association between the two. Not a causal relationship. This is because of the form of the study, as it is observational, and there are not necessarily reproduced “studies” that could continue to back this claim of a causal relationship. There are too many confounding variables that are not being controlled in this study.

b

It wouldn't be unreasonable to create some form of an experiment to test for this relationship, however there are a lot of variables that should be controlled. (For simplicity, we'll just look at no assignments being completed vs. students that complete all assignments. However, it could be designed in away that allows for multiple test groups.)

The experiment should have a control group, the group that does not complete assignments, and test group, the students that complete all the homework assignments. For a robust experiment we would need to control for different variables that could affect exam grades, such as study time per week, math and statistical background, and maybe even take into consideration IQ. With these variables kept consistent between the two groups (though this might be difficult in a practical setting), we would give each group the same exam. We could compare the average scores between the groups and look for a statistically significant difference in scores. This would allow us to draw conclusions about causal relationships between assignments being complete and exam scores. It would be better to have more groups with differing number of assignments completed so a model could be made to look at this relationship, rather than just a difference of proportions.

c

It seems it would be possible to do so, although these observational studies would have to be greater in number with more information collected besides just the number of homework assignments completed and the exam scores. Variables like study time, number of statistics/math classes taken before, IQ, and more could be observed. These studies and observations would have to be reproducible across many collections of data.

Problem 2

1

a

Derive that $\sum_i e_i = 0$

$$\begin{aligned}\sum_i e_i &= \sum_i \hat{y} - y_i = \sum_i \hat{\alpha} - \hat{\beta}x_i - y_i \\&= \sum_i \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x_i - y_i = \sum_i \bar{y} - y_i + \hat{\beta}x_i - \bar{\beta}x_i \\&= \sum_i \bar{y} - \sum_i y_i + \hat{\beta}(\sum_i x_i - \sum_i \bar{x}) = \sum_i y_i - \sum_i y_i + \hat{\beta}(\sum_i x_i - \sum_i x_i) \\&= 0 - \hat{\beta}(0) = 0\end{aligned}$$

Thus: $\sum_i e_i = 0$

b

Derive that $\sum_{e_i x_i} = 0$

$$\begin{aligned}\sum_{e_i x_i} &= \sum_i (\hat{y} - y_i)x_i = \sum_i (\hat{\alpha} - \hat{\beta}x_i - y_i)x_i \\&= \sum_i (\bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x_i - y_i)x_i = \sum_i \bar{y}x_i - \hat{\beta}\bar{x}x_i + \hat{\beta}x_i^2 - y_i x_i \\&= \sum_i \bar{y}x_i - \sum_i y_i x_i + \hat{\beta}(\sum_i x_i^2 - \sum_i \bar{x}x_i) \\&= \bar{y} \sum_i x_i - \sum_i y_i x_i + \hat{\beta}(\sum_i x_i^2 - \bar{x} \sum_i x_i) = \frac{1}{n} \sum_i y_i \sum_i x_i - \sum_i y_i x_i + \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2} (\sum_i x_i^2 - \frac{1}{n} (\sum_i x_i)^2) \\&= (\frac{1}{n} \sum_i y_i \sum_i x_i - \sum_i y_i x_i) + (\sum_i x_i y_i - \frac{1}{n} \sum_i x_i \sum_i y_i) = 0\end{aligned}$$

Thus: $\sum_{e_i x_i} = 0$

2

```
prestige.lm <- lm(prestige ~ education, data = Prestige)
residuals <- prestige.lm$residuals
education <- Prestige$education
sigma.e <- sum(residuals)
sigma.e.x <- sum(residuals * education)

sigma.e
```

```
## [1] 3.93019e-14
```

```
sigma.e.x
```

```
## [1] 5.044853e-13
```

Sums are not exactly zero, but very close to zero.

3

Show that the formula for the least squares estimate for the null model is $\alpha = \bar{y}$

If $\bar{y} = \hat{\alpha}$, then $\hat{\beta} = 0$

Given least squares estimate formulas:

$$\hat{\alpha} = \bar{y} - \hat{\beta}x_i \text{ where } \hat{\beta} = 0$$

$$\text{thus } \hat{\alpha} = \bar{y} - 0x_i = \bar{y}$$