Solution

Title: The Indoor Obstacle Course Test (IOCT)

Topics: Confounding, Causal Diagrams, Simple Linear Regression, Confidence Intervals

Background: Cadets at West Point must pass the Indoor Obstacle Course Test (IOCT) to graduate. The IOCT begins with a series of floor and climbing obstacles and ends with several laps around an indoor track. It is an exhausting test of endurance and strength. In addition to being a graduation requirement, cadets receive a letter grade that is factored into their class rank.

Shorter cadets often argue they are at a disadvantage on the obstacle course. Many obstacles appear to favor taller cadets because they are easier to reach. In this study, we will investigate the effect of height on IOCT times.

1. Watch the video of Cadet Madaline Kenyon running the IOCT. In your opinion, do some obstacles favor taller cadets? Explain.

Height seems to be an advantage on some obstacles. For example, one obstacle requires the cadets to pull themselves onto a platform over their head. In addition, rope climbing may be quicker for taller people because they require fewer locks.

The file `obstacle_course.csv` contains height (inches), IOCT times (seconds), biological sex (M/F), and whether the cadet played an intercollegiate sport for a sample of 384 cadets who ran the IOCT course in the last five years.

2. What is the explanatory variable in this study? Classify the variable as quantitative or categorical.

The explanatory variable in this study is height. It is quantitative.

3. What is the response variable in this study? Classify the variable as quantitative or categorical.

The response variable in this study is IOCT time. It is also quantitative.

4. Is this study an observational study or a randomized experiment? Explain.

This is an observational study because the researchers did not intervene on the assignment of the explanatory variable (height) to the subjects in the study.

Figure 1 depicts IOCT times in seconds versus height in inches. Table 1 contains information from the linear regression model.
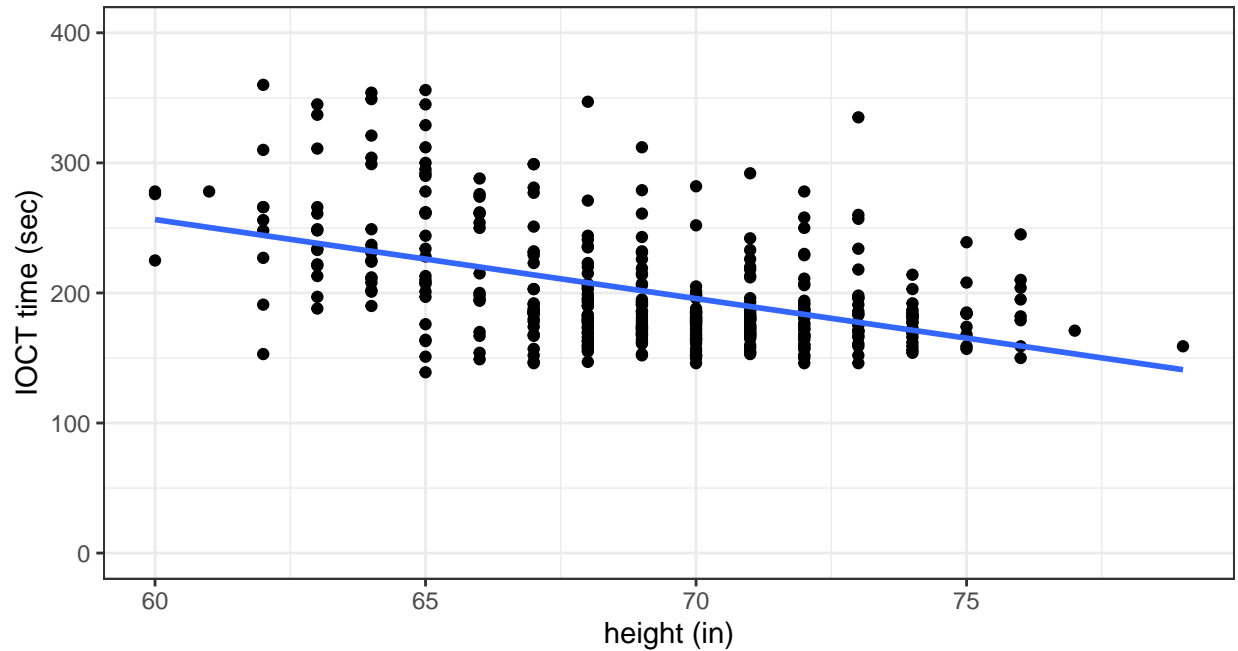

Figure 1: Indoor Obstacle Course Test (IOCT) times versus height (n = 384)

Table 1: Linear regression output for IOCT times and height.

term	estimate	std.error	statistic	p.value
(Intercept)	621.00	39.98	15.53	0
height	-6.08	0.58	-10.52	0

5. Interpret the estimate of the height coefficient in Table 1.

The average IOCT time decreases by about six seconds for each one inch increase in cadet height. In other words, taller cadets tend to do better than shorter cadets by about six seconds per inch of height.

6. Calculate and interpret a 95% confidence interval for the slope coefficient.

```
## # A tibble: 1 x 2
##    conf.low conf.high
##       <dbl>     <dbl>
## 1     -7.21     -4.94
```

7. The *p*-value for height in Table 1 indicates there is strong evidence of an association between height and IOCT time. Taller cadets tend to do better on the IOCT. Some people would say the result is *statistically significant.* However, statistical significance and practical signifigance are different. Review the grade scale for the IOCT. In your opinion, does the observed association have practical significance? Explain.

Letter grades are assigned for about 5-10 second intervals of time. For example, 2:45 to 2:49 is a B- and 2:50 to 2:54 is a C+. A six second difference in average IOCT time for each inch of height would have practical importance... that is about a half a letter grade for each inch of height.

8. A shorter cadet argues Figure 1 shows evidence the IOCT is unfair based on height. Do you agree or disagree? Explain.

While Figure 1 shows a strong association between height and IOCT time, we do not know if the taller cadets are different from the shorter cadets in ways that are important causes of faster IOCT times. Therefore, we can say there is an association between IOCT times and height but we do not know if height is the cause of the faster IOCT times.

9. Briefly explain the difference between these two conclusions.

- *Height is associated with faster IOCT times.*

- *Height causes faster IOCT times.*

Two variables are associated if they frequently change together when we observe them. In other words, association is about what we see. Association can occur when one variable causes another or when two variables share a common cause. In this case, height is associated with IOCT times if it is the cause of the faster times or if there is another variable that itself is a cause of both height and IOCT times.

One variable is a cause of a second variable if there is a change (on average) in the second variable when we intervene on the first variable. In other words, causation is about what happens when we do something (Pearl, 2018). Height is a cause of faster IOCT times if intervening on only cadet heights results in different IOCT times on average.

10. Based on the analysis presented thus far, is it possible to distinguish between these two explanations? Explain.

No, based only on Figure 1 and Table 1, we cannot determine if the relaitonship between height and IOCT times is one of correlation or causation. Both explanations are consistent with the results thus far.

11. Draw a causal diagram depicting the relationship between height, IOCT time, and sex. Explain your decisions to include/exclude arrows in the diagram.
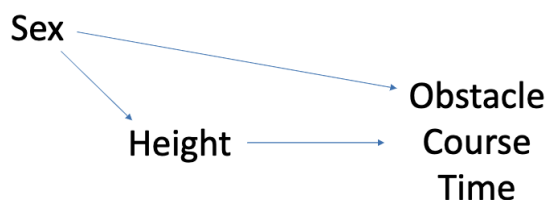

Figure 2: Causal Diagram

I included an arrow between sex and height because male cadets are taller, on average, than female cadets. I included an arrow between sex and obstacle course time because there is likely a difference in performance (on average) on the obstacle course between males and females.

12. Based on your diagram, identify the confounding variable.

The confounding variable is sex.

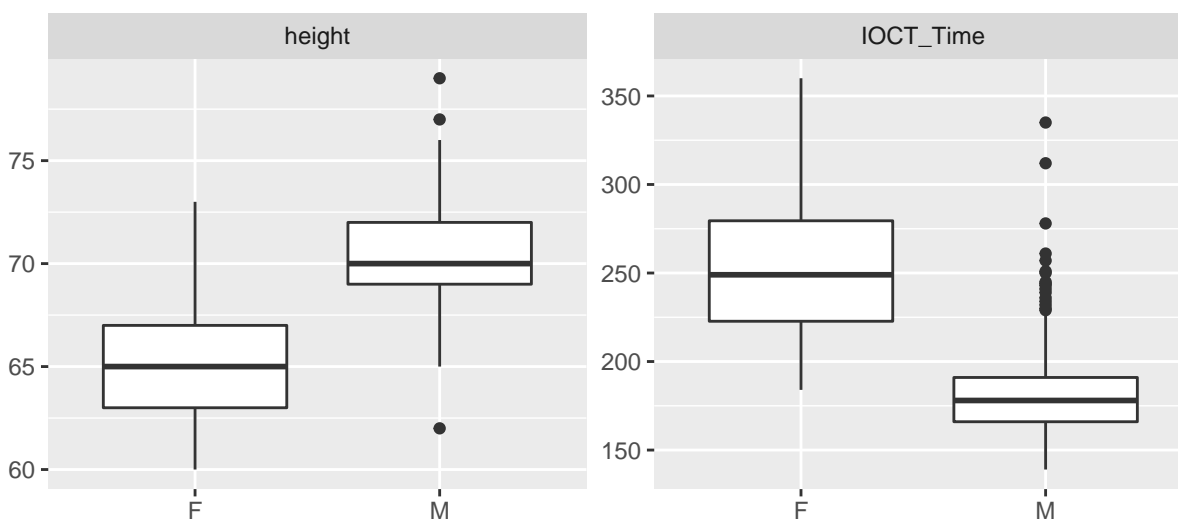Below are boxplots of height in inches and IOCT times in seconds by sex.

Figure 3: Height (inches) and IOCT time (seconds) by sex.

13. Based on Figure 2, is the estimate of the effect of height on IOCT time in Table 1 confounded by sex? If so, is the effect of height smaller or larger than that reported in Table 1? Explain.

Yes, the estimate is likely confounded. The data shows a strong relationship between height and sex and IOCT time and sex. In this case, it appears that male cadets tend to be taller and run the obstacle course faster. Therefore, the actual effect of height on IOCT time is probably smaller.

Figure 3 depicts the association between IOCT time and height by sex. Tables 2 and 3 depict regression results for female and male cadets, respectively.
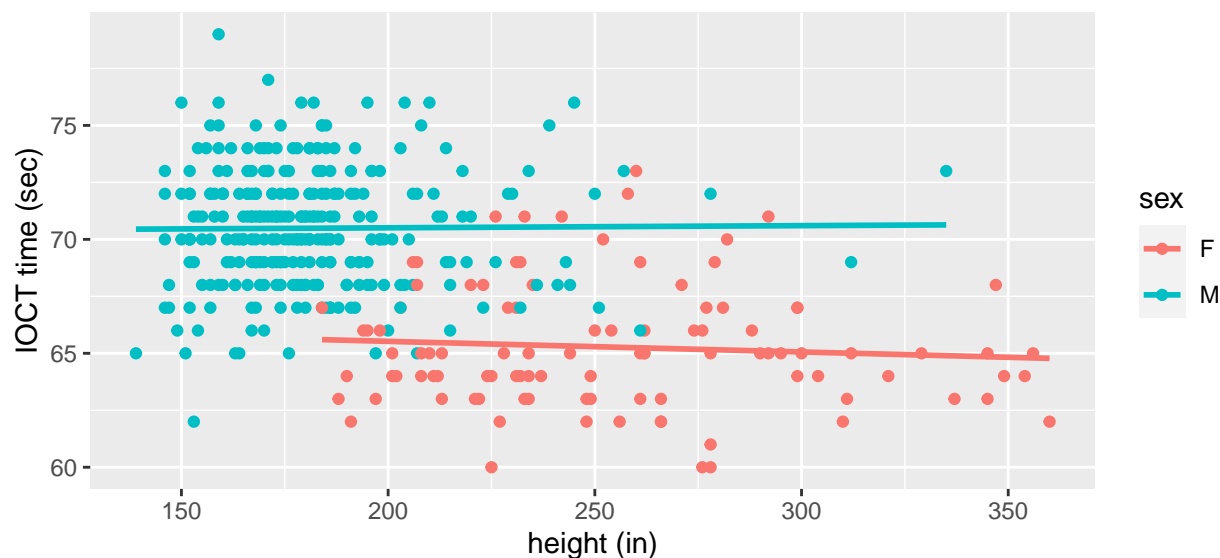
Figure 4: Indoor Obstacle Course Test (IOCT) times versus height by sex (n = 384).

Table 2: Regression results for female cadets.

term	estimate	std.error	statistic	p.value
(Intercept)	335.76	107.44	3.12	0.00
height	-1.24	1.64	-0.76	0.45

Table 3: Regression results for male cadets.

term	estimate	std.error	statistic	p.value
(Intercept)	175.61	40.92	4.29	0.00
height	0.09	0.58	0.15	0.88

14. Based on Figure 3 and Tables 2 and 3, does it appear there is an association between IOCT time and height within levels of sex? Explain.

There does not appear to be an association between height and IOCT times within levels of sex. The regression slopes in the figure and tables are very close to zero.

15. In your opinion, is there much evidence that height is an advantage on the IOCT (in other words, is height the *cause* of better IOCT times)? Explain.

Based upon the stratified analysis, it does not appear that there is much of an advantage of height. The most plausible explanation is the observed association between height and IOCT was due to confounding by sex. If there was an effect of height, we would expect to see it within levels of sex.

16. Briefly discuss two ways you could improve this study to better assess whether there is a height advantage.

- Instead of stratifying on sex, we could use multiple regression to obtain sex-adjusted estimates of the effect of height.

- We should consider other potential confounding variables. For example, intercollegiate athletes may be overrepresented in different height ranges.

- If we could recollect the data, we could match taller cadets to shorter cadets with similar levels of confounding variables.