

Causality and Multiple Regression Supplement -
MA206 Probability and Statistics

Authors TBD

2021-03-05

Contents

1 (Kevin) Motivation	5
1.1 Goals of quantitative research (describe, predict, cause-and-effect)	5
1.2 Why multivariable methods?	5
1.3 Why multiple regression?	5
2 (Kevin) Causality	7
2.1 What does it mean for one thing to cause another	7
2.2 Randomized controlled experiments	7
2.3 Observational Studies	7
2.4 Confounding	7
2.5 Causal Diagrams	7
3 (COL Watts) Quantitative explanatory variable with categorical confounding variable	9
3.1 Unadjusted effect of X on Y	9
3.2 Effect of X and Y adjusting for C .	9
3.3 Assessing model adequacy	9
4 (Kevin) Interactions	11
5 Activity	13
6 (COL Watts) Quantitative explanatory variable with quantitative confounding variable	21
6.1 Effect of X and Y adjusting for C .	21
6.2 Assessing model adequacy	21

Chapter 1

(Kevin) Motivation

1.1 Goals of quantitative research (describe, predict, cause-and-effect)

Researchers and practitioners in nearly every discipline employ quantitative methods to answer important questions and make decisions. Regardless if it is finance, medicine, psychology, etc., all researchers typically have one of the following three goals when they conduct quantitative research:

- **describe** - (define here)
- **predict** - (define here)
- **cause-and-effect** - (define here)

For example, let's say a medical researcher investigates a newly discovered variant of the coronavirus. Initially, researchers will probably focus on describe studies. *What is the prevalence of the variant in the population? What groups of individuals are most vulnerable to the variant?* Follow-on studies will have the p (continue example here)

1.2 Why multivariable methods?

1.3 Why multiple regression?

Chapter 2

(Kevin) Causality

- 2.1 What does it mean for one thing to cause another
- 2.2 Randomized controlled experiments
- 2.3 Observational Studies
- 2.4 Confounding
- 2.5 Causal Diagrams

Chapter 3

(COL Watts) Quantitative explanatory variable with categorical confounding variable

Introduce multiple regression by adding a categorical confounding variable

Explanatory (X) - quantitative

Response (Y) - quantitative

Confounding (C) - categorical

3.1 Unadjusted effect of X on Y

short review of simple regression

3.2 Effect of X and Y adjusting for C .

3.3 Assessing model adequacy

Chapter 4

(Kevin) Interactions

Extend the example in the last chapter...introduce in terms of effect modification.
We have a new research question, “does the affect of X on Y change based on another variable of interest?”

Chapter 5

Activity

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.

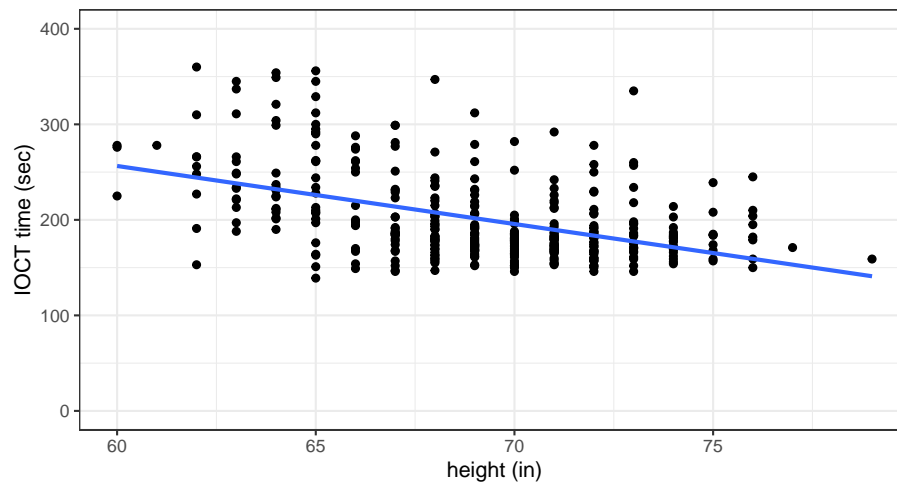
The file `obstacle_course.csv` contains height (inches), IOCT times (seconds), biological sex (M/F), and whether the cadet played an intercollegiate sport for a sample of 384 cadets who ran the IOCT course in the last five years.

2. What is the explanatory variable in this study? Classify the variable as quantitative or categorical.
3. What is the response variable in this study? Classify the variable as quantitative or categorical.
4. Is this study an observational study or a randomized experiment? Explain.

Table 5.1: 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

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



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

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

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 significance are different. Review the grade scale for the IOCT. In your opinion, does the observed association have practical significance? Explain.

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

9. Briefly explain the difference between these two conclusions.
 - *Height is associated with faster IOCT times.*
 - *Height causes faster IOCT times.*

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

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

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

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

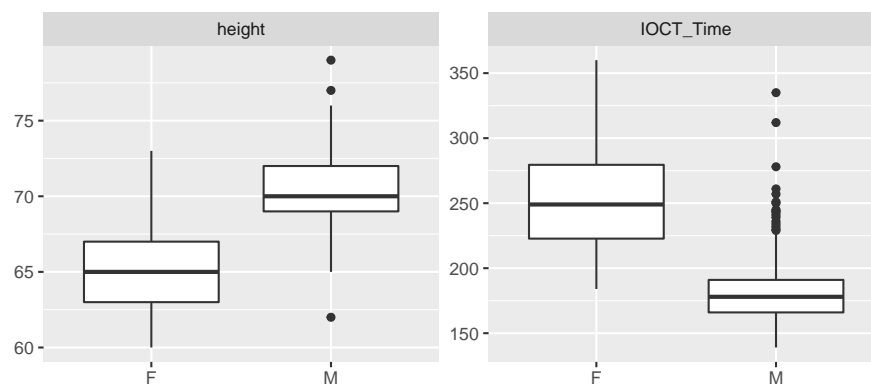


Figure 5.1: Height (inches) and IOCT time (seconds) by sex.

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.

Table 5.2: 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 5.3: 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

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.

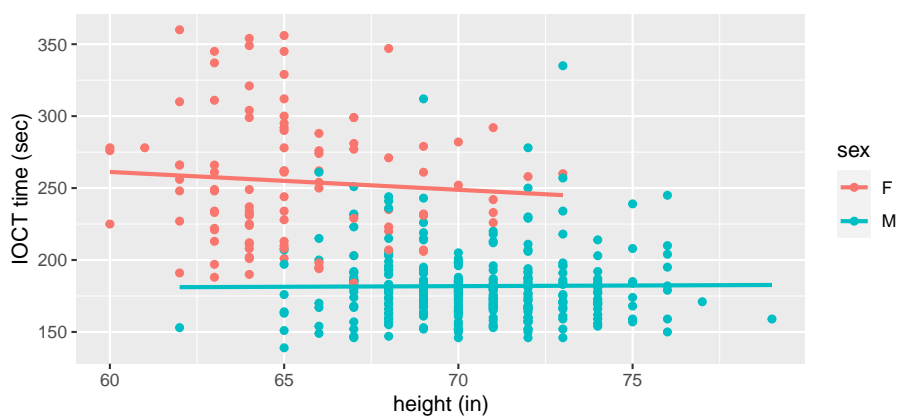


Figure 5.2: Indoor Obstacle Course Test (IOCT) times versus height by sex ($n = 384$).

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.

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

Chapter 6

(COL Watts) Quantitative explanatory variable with quantitative confounding variable

Explanatory (X) - quantitative

Response (Y) - quantitative

Confounding (C) - quantitative

6.1 Effect of X and Y adjusting for C .

6.2 Assessing model adequacy