Mistaken causation? A case to teach the difference between correlation and causality

Katherine B. Hartman
Ohio University

James B. Hunt
University of North Carolina Wilmington

Carla Y. Childers
Ohio University

ABSTRACT

The case uses student ratings of marketing professors from RateMyProfessors.com to help students understand the difference between correlation and causality, and furthermore, to be able to identify the requirements for causality. Students are asked to identify causal relationships inherent in the data and to look for alternative explanations for the positive correlations presented. Teaching Notes offer relevant discussion points on causal logic fallacies (e.g., reverse cause fallacy, bidirectional cause fallacy, common cause fallacy, single cause fallacy, and coincidence), criteria for causation, and practical implications of the RateMyProfessors.com correlations.

Keywords: correlation, causality, criteria for causation, RateMyProfessors.com

Note: This is a fictitious case developed for educational use. All names used herein were created for the purposes of this case and should not be construed as factual.

INTRODUCTION FOR INSTRUCTORS

This case is designed to help students understand the difference between correlation and causality, and furthermore, to be able to identify the requirements for causality. Utilizing real student ratings of marketing professors from RateMyProfessors.com (RMP), students are asked to identify causal relationships inherent in the data and to look for alternative explanations for the positive correlations presented. This is done by asking students to critically review a hypothetical student's written interpretation of statistical analysis performed on RMP data.

Why RMP as the context? Education is a common, high involvement experience for all students. RMP is a nationally recognized website, and over four million college students use the website each month (RateMyProfessors.com, 2013). A recent survey of college students found that 71% used RMP to select among professors (Brown, Baillie, & Fraser, 2009). Thus students should be very familiar with and interested in the information presented. RMP data, and student evaluations of teaching in general, have been highly criticized, and students should be able to have lively discussion about the topic. Because students are sometimes anxious about statistics, using a case that is practical and relevant to the student audience may decrease concerns and increase engagement. In summary, the scenario should generate an interest in learning the material because the problem has depth and value.

Extensive Teaching Notes offer relevant discussion points on causal logic fallacies (e.g., reverse cause fallacy, bidirectional cause fallacy, common cause fallacy, single cause fallacy, and coincidence), criteria for causation, and practical implications of the RateMyProfessors.com correlations. Through case analysis, these somewhat mundane topics can be made more exciting, effectively bridging the gap between statistical theory and practice, and motivating students to explore the concepts of research methods and statistics.

Suggested Use of the Case and Learning Objectives

The case was written for use with undergraduate students who have a general understanding of descriptive statistics and research methods. It would ideally be used in introductory statistics or research methods courses. The proposed discussion and Teaching Notes are designed to teach the differences between correlation and causality through the RMP example. For this reason, instructors may decide to use the case either 1) before a discussion of correlation and causality, which would allow instructors to teach concepts using the example, or 2) after a discussion about correlation and causality, which would allow instructors to test students' comprehension of the concepts.

The learning objectives are for students to:

1. understand the difference between correlation and causality;
2. comprehend causal logic fallacies by exploring alternative explanations for correlation relationships; and
3. understand the requirements for proving causality.

CASE AS PRESENTED TO STUDENTS

Professor Jane Smith teaches a course focused on research and statistics. Students enrolled in her course are asked to complete an assignment designed to demonstrate an understanding of one of the statistical tests learned during the course. The assignment requires

students to collect quantitative data from any website, analyze the data using one of the statistical tests learned in the course, and interpret the results of their analysis.

Before the final assignment is due, students have the opportunity to submit a short, one-page summary of their report for review. Professor Smith has reviewed the following summary from a student and is thinking about how to help the student improve the interpretation of the results.

Student's Report Summary: "An Analysis of RateMyProfessors.Com Ratings"

RateMyProfessors.com (RMP) is a free web site that allows users to rate professors' teaching quality using an ordinal scale from 1 (worst) to 5 (best) for clarity and helpfulness. Users may also evaluate professors on easiness, rate their interest in the course, and provide written comments up to 255 characters long. For each professor, RMP aggregates user ratings and provides averages for clarity, helpfulness, overall quality (calculated by averaging clarity and helpfulness scores), and easiness. All posts are anonymous, voluntary, and publicly available.

Sample & Analysis. Data includes the average ratings of 442 marketing professors from eight colleges and 43 universities in the United States. The number of ratings for each individual professor ranged from one to 72 (Mean = 8.7, Median = 5). For the 442 professors, average ratings were calculated for Helpfulness (Mean = 3.7), Clarity (Mean = 3.6), and Easiness (Mean = 3.3).

Using professors' average scores, correlation coefficients were calculated for average easiness, helpfulness, and clarity. All relationships were significantly and positively correlated ($p < .00$). The following table provides the correlation coefficients for average professor ratings.

Table: Correlation Coefficients for Average Scores

	Average Easiness	Average Helpfulness	Average Clarity
Average Easiness	1		
Average Helpfulness	.41	1	
Average Clarity	.42	.85	1

Interpretation. The RMP data analysis identifies positive correlations between easiness and helpfulness, helpfulness and clarity, and easiness and clarity. The positive correlations suggest that professors who require less work and/or who are more lenient graders than other professors (i.e., easier) receive higher evaluations for helpfulness and clarity. The results are consistent with what is known as the grading leniency bias model (Howard & Maxwell, 1980), which predicts that actual or expected grades in a course (i.e., easiness) positively influence student evaluations of the professor. As such, if a professor wants to receive higher evaluations, then s/he should reduce the amount of work required in the course and make it easier to get better grades.

Questions for Discussion

1. What causal relationship(s) is implied in the interpretation?
2. Based upon your own experiences with professors, what are other plausible explanations for the positive correlations between easiness and helpfulness and easiness and clarity?

3. Do you believe the data has proven a causal relationship between easiness and helpfulness and easiness and clarity? Why or why not?
4. Do you agree with the statement, "If a professor wants to have higher evaluations, then s/he should reduce the amount of work required in the course and make it easier to get better grades?" Why or why not?

EXERCISE TEACHING NOTES FOR FACULTY

Discussion for Question 1: Implied Causal Relationship

The first discussion question in the case asks: "What causal relationship is implied in the interpretation?" The role of the first question is to provide students with a relatively easy question to start the discussion. The learning goal of this discussion is to provide students with a basic understanding of the variables and relationships associated with causal statements.

To answer this question, students need to review and deconstruct the following statement in the interpretation section of the case: "…professors who require less work and/or who are more lenient graders (i.e., easy) will receive higher evaluations for helpfulness and clarity." Students should identify the following from the statement:

- The variables in the implied causal relationships are ratings for easiness, ratings for helpfulness, and ratings for clarity.
- In the implied causal relationships (i.e., cause-and-effect relationships), easiness is the cause and helpfulness and clarity are the effects.
- In the implied causal relationships, the positive correlation suggests that an increase in ratings for easiness causes an increase in ratings for helpfulness and an increase in ratings for clarity.

In summary, students should understand the difference between cause and effect. In the context of the case, the interpretation section suggests A → B, where A represents easiness and B represents helpfulness and clarity.

Question 2: Alternative Explanations

The second discussion question in the case asks: "Based upon your own experiences with professors, what are other plausible explanations for the positive correlations between easiness and helpfulness and easiness and clarity?" The role of the second question is to encourage students to use their own past experiences to formulate rival explanations for the causal relationships. The learning goal of this discussion is to provide students an understanding of causal logic fallacies.

Using their own experiences, students should be able to provide several other plausible reasons for why there is a positive correlation between ratings for easiness and helpfulness and ratings for easiness and clarity. One method for executing the discussion would be to list all of the students' brainstormed ideas and then teach students to group them into similar categories based upon common causal logic fallacies. Explanations with examples of common logic fallacies are provided below.

Reverse Cause Fallacy: B causes A. The reverse cause fallacy assumes that A caused B without considering the possibility that B is actually the cause of A. In the context of the case, the student's interpretation of the correlations suggests A (easiness) causes B (helpfulness and

clarity). However, alternative explanations for the correlations are to suggest a reverse cause-and-effect where B influences A. In other words, helpfulness influences easiness, and clarity influences easiness rather than vice versa.

Otto, Sanford, and Wagner (2005) suggest and support two hypotheses about the correlations using RMP ratings data. For the clarity-easiness correlation, the authors argue: if a professor is clear [about instructions or expectations], then this makes a course easier for the student (Otto, Sanford, & Wagner, 2005, p. 27). As such, clarity influences easiness. For the helpfulness-easiness correlation, the authors argue: if a professor is willing to help a student understand the material, then this will be perceived as easier than when a professor is unwilling to be helpful (Otto, Sanford, & Wagner, 2005, p. 27). As such, helpfulness influences easiness.

An example of the reverse cause fallacy is as follows. A professor always gives assignments with clear, detailed instructions and offers extra assistance by extending his/her office hours so students can ask questions before assignments are due. Students may perceive the professor as easier than other professors because the professor consistently provides clear, detailed instructions (i.e., clarity). The professor may also be perceived as comparatively easy because he/she offers additional opportunities to answer questions (i.e., helpfulness). In the example, either perceptions of clarity or perceptions of helpfulness may cause perceptions of easiness.

Bidirectional Cause Fallacy: A causes B and B causes A. The bidirectional cause fallacy assumes that A caused B without considering the possibility that there is a reciprocal relationship between the two where A causes B and B causes A. In the context of the case, alternative explanations for the correlations are that (1) helpfulness influences easiness, which then influences perceptions of helpfulness and (2) clarity influences easiness, which then influences perceptions of clarity.

The traditional grading leniency bias model (e.g., Howard & Maxwell, 1980) suggests that actual or expected grades positively influence student evaluations of teaching. However, as stated in the reverse cause fallacy, Otto, Sanford, and Wagner (2005) hypothesize that ratings of RMP overall quality (as measured by averaging helpfulness and clarity) influences easiness ratings. Because students interact with professors over the course of an entire teaching period (i.e., session, quarter, or semester), it is plausible to conceptualize a reciprocal relationship among the variables. In short, there is a feedback loop among the variables.

An example of the bidirectional cause fallacy is as follows. A professor gives students an assignment. His/her goal is simply to ensure learning and s/he believes that the best approach is to allow students to keep revising the assignment until full credit is achieved. The professor's feedback on revisions may be perceived as helpful. This type of helpfulness may lead to perceptions of easiness in terms of grading as students are given opportunities to make revisions. At the same time, students may view the professor's easiness in grading as helpful toward their learning as well as their overall academic success. In the example, there is a reciprocal relationship between perceptions of helpfulness and perceptions of easiness.

Common Cause Fallacy: C causes both A and B. The common cause fallacy assumes that A caused B without considering the possibility that an external factor C is actually the cause for both A and B. In the context of the case, alternative explanations for the correlations are to hypothesize that ratings for helpfulness, easiness, and clarity are all influenced by some other, unknown factor. Examples include (but are not limited to) instructor likeability, instructor attractiveness, instructor competence, and students' interest in the course.

Felton, Mitchell, and Stinson (2004) found that mean RMP scores for both average quality (as measured by averaging ratings for clarity and helpfulness) and average easiness were significantly higher for professors rated as "hot" as compared to professors rated as "not." Felton, Mitchell, and Stinson (2004) conclude that instructor attractiveness has a halo effect on RMP ratings such that professors perceived as attractive will positively influence RMP ratings. In addition, Otto, Sanford, and Ross (2008) theorize that RMP ratings may be biased due to an emotional halo effect where raters rate professors with either universally high or low ratings without accurately reporting distinct aspects of faculty performance.

An example of the common cause fallacy is as follows. Two professors assign students in similar courses similar assignments. Both professors offer similar explanations for the assignment and are generally available to answer questions before class, after class, and during office hours. One professor is considered nicer, warmer, and friendlier as compared to the other professor; s/he is more welcoming during office hours and friendlier during conversations with students. In the example, differences between the professors' overall personality may cause students to perceive differences in clarity, helpfulness, and easiness.

Single Cause Fallacy (Fallacy of Reduction): A (and X and Y and Z) causes B. The single cause fallacy assumes that A caused B without considering the possibility other factors X, Y, and Z also influence B. In the context of the case, alternative explanations for the correlations are to hypothesize that easiness is caused by helpfulness and/or clarity yet also caused by other, unknown factors. The single cause fallacy is closely related to the fallacy of reduction, which suggests that the cause-and-effect relationship is part of a larger system of relationships.

Using traditional student evaluations of teaching, Howard and Maxwell (1980) test and support alternatives of the grading leniency bias model. Specifically, the study suggests teacher effectiveness positively and directly influences student motivation, which in turn positively and directly influences grades and student satisfaction with both the professor and the course. As such, the correlation between grades and evaluations is an artifact of other causal relationships. Simply stated, Howard and Maxwell (1980) suggest, "…(a) good teaching leads to better student learning, (b) greater student motivation leads to better learning, (c) greater student learning leads to higher grades, and (d) greater student learning leads to greater student satisfaction" (p. 812).

In the context of RMP data, the discussion of a single cause fallacy may involve introducing other antecedent variables into the causal relation between A → B (easiness → helpfulness, clarity) including instructor personality characteristics, student motivation to learn the material, student perceptions about the importance of the material, etc. Students could hypothesize a more complicated model where easiness mediates the relationships between the antecedents and helpfulness or clarity.

Coincidence: A and B are merely correlated. The coincidence cause fallacy assumes that A caused B without considering the possibility that the relationship between A and B is merely coincidental. In the context of the case, alternative explanations for the correlations are to hypothesize that easiness, helpfulness, and clarity are simply correlated by coincidence, i.e. there is a significant combination between the two variables that is independently explicable by joint or separate circumstance.

In the context of RMP data, coincidence may be difficult to conceptualize. Students could hypothesize that the positive correlations among the variables may be simply a matter of the limited sample of marketing professors. As such, if data were taken from a wider range of professors from other disciplines, there may be no significant correlations or even negative correlations.

Question 3: Proving causality

The third discussion question in the case asks: "Do you believe the data has proven a causal relationship between easiness and helpfulness and easiness and clarity? Why or why not?" The role of the third question is to teach students to recognize that correlation (association) is a necessary yet insufficient factor to proving causation. As such, the interpretation of the data incorrectly concludes a causal relationship when only a correlation between the variables has been demonstrated. The learning goal of this discussion is to provide students with a basic understanding of the requirements for causality.

According to Hill's (1965) criteria for causation, the minimal conditions needed to establish a causal relationship between two items in epidemiological research are as follows:

1. Temporal relationship: cause always precedes effect.
2. Strength of association: size of the association as measured by appropriate statistical tests.
3. Dose-response relationship: an increasing amount of exposure to the cause increases the effect.
4. Consistency: results are replicated in studies in different settings using different methods.
5. Plausibility: cause-and-effect relationship agrees with currently accepted understanding of processes.
6. Consideration of alternate explanations: the extent to which other possible explanations have been taken into account and have been effectively ruled out.
7. Experiment: the condition can be altered by an appropriate experimental procedure.
8. Specificity: a single cause produces a specific effect.
9. Coherence: association is compatible with existing theory and knowledge.

The following discusses the three criteria most applicable to the case.

Requirement: Correlation (Concomitant variation). One necessary yet insufficient requirement for causality is correlation. A correlation or an association refers to a relationship between two (or more) variables that change together. More specifically, concomitant variation is the extent to which two variables occur together or vary together in a way predicted by the hypothesis. Correlation relationships can be positive (direct) or negative (inverse). A positive correlation means that as one variable increases, the other variable increases. A negative correlation means that as one variable increases, the other variable decreases. However, causality refers to a relationship between two (or more) variables in which one variable causes or influences another variable. As such, "one cannot substantiate causal claims from association alone" (Pearl, 2009, p. 99).

In the context of the case, the RMP data only identifies that (1) easiness is positively correlated with helpfulness and (2) easiness is positively correlated with clarity. In other words, there is an association between easiness-helpfulness and easiness-clarity such that they vary together in the same direction. However, given the observational nature of the data, simple associations between the variables are lacking proof for a causal relationship.

Requirement: Temporal priority. A necessary yet insufficient requirement for causality is temporal priority. Hume was one of the first theorists to suggest the notion of temporal priority (see White, 1990 for a discussion). Specifically, he argued that a cause must precede an effect in

a sequence of events. In short, the cause must occur before or simultaneously with the occurrence of the effect.

In the context of the case, the RMP data has not established a temporal priority between easiness and helpfulness nor between easiness and clarity. In fact, contrary to the interpretation in the case, previous RMP ratings research actually predicts the opposite temporal relationship (e.g., Otto, Sanford, & Wagner, 2005). Specifically, students' perceptions of professor's behaviors regarding helpfulness and clarity influence students' perceptions about the difficulty of the course.

Requirement: Elimination of other possible causes. A necessary yet insufficient requirement for causality is the elimination of all other possible causes. Arguably, this is the most difficult issue to determine in understanding and proving causal relationships. Is instructor easiness alone the only source affecting helpfulness and clarity? One of the best ways of eliminating other possible causes is to make sure no confounding variables exist. Confounding variables are extraneous variables that change with levels of an independent variable (easiness). A confounding variable provides an alternative explanation of an experimental result and must be controlled to eliminate it as a plausible alternative cause.

In our RMP context, factors such as instructor likeability and attractiveness ("hotness") may be the variables actually causing student perceptions of helpfulness and/or clarity. Studies have indeed found higher ratings for professors classified as "hot" (e.g. Bonds-Raacke & Raacke, 2007; Felton, Mitchell, & Stinson, 2004; Riniolo et al., 2006). Instructor competence and communication style will certainly affect perceptions of clarity, aside from the level of difficulty of the material, assignments, and exams. If a student has previously taken a similar course, it is likely the professor and class in question will be perceived as clear because the student has been previously exposed to similar examples and material. In summary, there are many variables besides easiness that may account for perceptions of helpfulness and clarity, and these variables will need to be accounted for in any discussion or experiment regarding causality.

Another possibility is that there is a third variable (instructor likeability, instructor attractiveness, instructor competence, students' interest in the course) that drives the correlation and is likely the cause of both an increase in perceptions of easiness (students perform better when they like the professor, are interested in the material, or are motivated by effective teaching techniques) and an increase in perceptions of clarity and helpfulness (halo effect). In sum, to ascertain the cause, we must eliminate other possible alternatives.

Question 4: Practical implications

The fourth discussion question in the case asks: "Do you agree with the statement "if a professor wants to have higher evaluations, then s/he should reduce the amount of work required in the course and make it easier to get better grades"? Why or why not?" The learning goal of the fourth question is to allow students the opportunity to think about the practical implications of the correlations.

During the discussion, students should recognize that this recommendation is inappropriate. Although it is possible that reducing workloads may improve evaluations, the unknown causal relationship among the variable prohibits a definitive conclusion. It is also possible that other factors may improve evaluations. Other factors discussed might include instructor personality, teaching effectiveness, student motivation, course material, etc. As such,

making a recommendation using too little information is inappropriate, especially when a more complex system of variables may be involved.

For example, using data from more than 200,000 students, Howard and Maxwell (1980) compared three models involving expected course grades and student evaluations: (1) the grading leniency bias model, which hypothesizes actual or expected grades positively influence student evaluations, (2) the teaching effectiveness model, which hypothesizes teaching effectiveness positively influences student performance and student performance positively influences both grades and evaluations, and (3) the student characteristics model, which hypothesizes student motivation positively influences student performance and student performance positively influences both grades and evaluations. The results suggest high correlations between grades and evaluations are a by-product of a more complex system involving other variables such as teaching effectiveness and student motivation.

CONCLUSION

This case is designed to help students understand the difference between correlation and causality and associated limitations. After participating in this case, students should be able to (1) understand the difference between correlation and causality, (2) comprehend causal logic fallacies by exploring alternative explanations for correlation relationships, (3) understand the requirements for proving causality, and (4) recognize the limitations of drawing conclusions using correlations.

REFERENCES

Bonds-Raacke, J. & Raacke, J.D. (2007). The Relationship between Physical Attractiveness of Professors and Students' Ratings of Professor Quality. *Journal of Psychiatry Psychology and Mental Health, 1* (2), 1-7.

Brown, M.J., Baillie, M., & Fraser, S. (2009). Rating RateMyProfessors.com: A Comparison of Online and Official Student Evaluations of Teaching. *College Teaching, 57* (2), 89-92.

Felton, J., Mitchell, J., & Stinson, M. (2004). Web-based Student Evaluations of Professors: The Relations between Perceived Quality, Easiness, and Sexiness. *Assessment & Evaluation in Higher Education, 29* (February), 91-108.

Hill, A. B. (1965). The Environment and Disease: Association or Causation?. *Proceedings of the Royal Society of Medicine, 58*, (295-300).

Howard, G. S., & Maxwell, S.E. (1980). Correlation between Student Satisfaction and Grades: A Case of Mistaken Causation. *Journal of Educational Psychology, 72* (6), 810-820.

Otto, J., Sanford, D. & Wagner, W. (2005). Analysis of Online Student Ratings of University Faculty. *Journal of College Teaching & Learning, 2* (June), 25-30.

Otto, J., Sanford, D., Wagner, W., & Ross, D. (2008). Does Ratemyprofessor.com Really Rate My Professor?. *Assessment & Evaluation in Higher Education*, 33 (August), 355-368.

Pearl, J. (2009). Causal Inference in Statistics: An Overview. *Statistics Surveys, 3* (September), 96-146.

RateMyProfessors.com. (2013). About. Retrieved from January 10 from http://www.ratemyprofessors.com/About.jsp.

Riniolo, T. C., Johnson, K.C., Sherman, T.R., & Misso, J.A. (2006). Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?. *Journal of General Psychology, 133* (1), 19-35.

White, P.A. (1990). Ideas about Causation in Philosophy and Psychology. *Psychological Bulletin, 108* (1), 3-18.