

## Moving Towards Multivariate Analysis

In this assignment, you are looking at three variables at a time – what is called multivariate analysis.

Multivariate analysis helps us explore why a relationship exists.

For example, let's say I take a random selection of 1100 people. If I measure each person's height in inches and I give each person the same vocabulary test (number of words out of 100 correctly defined), I would find a strong correlation between the two. Taller people, on average, know more words than shorter people. Why should this be so?

Answering that question invokes theory. Theory is an interpretation of data. In this case, an interpretation of why taller people know more words than shorter people. Our theory might be:

1. directly causal, in the sense of height causing someone to know more or fewer words,
2. indirectly causal, in the sense that height is causally linked to some social process that creates vocabulary knowledge, or
3. spurious, in the sense that height and vocabulary have nothing to do with one another but are coincident because an omitted variable(s) is causing both outcomes

**An example of 1** would probably have to be physiological. You might theorize that the same biological processes that account for growth account for brain development, thus leading taller people to be capable of learning language qualitatively better than shorter people. If this theory is correct, then we should see the same relationship whether we look at men or women, baby boomers or millennials, democrats or republicans, and so on.

**An example of 2** might suggest that height opens up opportunities for education that then create imbalances in vocabulary in the population. Maybe we, as a society, tend to give taller people more chances at a lot of different educational opportunities – scholarships, honors classes, suggestions that you pursue a graduate degree, etc. Since it is reasonable to assume increased formal education increases one's vocabulary, we would be theorizing that education has a direct effect on vocabulary and that height has an indirect effect on vocabulary through education. If this is correct, then when we look just at people who got graduate degrees, we would expect height and vocabulary to be unrelated. When we look just at people who didn't get any formal education, we would expect height and vocabulary to be unrelated. In other words, controlling for education, I would expect that height and vocabulary to be unrelated. I am theorizing that height causes vocabulary knowledge only indirectly through education.

**An example of 3** would be thinking that the relationship between height and vocabulary is misleading and purely (or mostly) the product of another cause of both of these variables. For example, age is definitely related to both height and vocabulary. People tend to grow taller from birth to age 20 or so. And they tend to acquire more and more vocabulary during this time as well. Maybe shorter people know fewer words because shorter people are kids and kids know relatively fewer words. If we divide our sample into ages, do short 40 year olds know fewer words than tall 40 year olds? Taking age into account probably eliminates the relationship between height and vocabulary knowledge. The bivariate relationship is misleading, what many people call “spurious”, because there is in no way in which height causes vocabulary. It’s an illusion created by the omission of an important factor related to both height (x) and vocabulary score (y).

**You all found a relationship between opinions about race in America and evaluations of how Trump has influenced race relations. But why does this relationship exist?**

**Did pre-existing opinions about race in America determine how people were going to react to Trump?**

**Or are both sets of opinions a product of general political ideology?**

**Or maybe both are a product of political party loyalty?**

**These are the sorts of questions I’d like you to explore in this paper.**

**The dataset contains the demographic variables listed on the following pages.**

The list is long. Moreover, in your final write-up, you don’t need to discuss more than one of these demographic variables depending on what you find.

**• F\_ACSWEB**

0. Does not access the Internet by paying a cell phone company or Internet service provider
1. Accesses Internet by paying a cell phone company or Internet service provider

**• F\_AGE CAT**

1. 18-29
2. 30-49
3. 50-64
4. 65+
99. DK/REF

• **F\_ATTEND**

1. More than once a week
2. Once a week
3. Once or twice a month
4. A few times a year
5. Seldom
6. Never
99. Refused

• **F\_BORN**

1. Yes, born-again or evangelical Christian
2. No, not born-again or evangelical Christian
99. Refused

• **F\_CITIZEN**

1. Yes
2. No
99. Refused

• **F\_CREGION**

1. Northeast
2. Midwest
3. South
4. West

• **F\_EDUCCAT**

1. College graduate+
2. Some College
3. H.S. graduate or less
99. Don't know/Refused

• **F\_EDUCCAT2**

1. Less than high school
2. High school graduate
3. Some college, no degree

4. Associate's degree
5. College graduate/some post grad
6. Postgraduate
99. Don't know/Refused

• **F\_HISP**

1. Yes
2. No
99. Refused

• **F\_IDEO**

1. Very conservative
2. Conservative
3. Moderate
4. Liberal
5. Very liberal
99. Refused

• **F\_INCOME**

1. Less than \$10,000
2. \$10,000 to less than \$20,000
3. \$20,000 to less than \$30,000
4. \$30,000 to less than \$40,000
5. \$40,000 to less than \$50,000
6. \$50,000 to less than \$75,000
7. \$75,000 to less than \$100,000
8. \$100,000 to less than \$150,000
9. \$150,000 or more
99. Refused

• **F\_INCOME RECODE**

1. \$75,000+
2. \$30-\$74,999
3. <\$30,000
99. Don't know/Refused

• **F\_MARITAL**

1. Married
2. Living with a partner
3. Divorced
4. Separated
5. Widowed
6. Never been married
99. Refused

• **F\_METRO**

1. Metropolitan
2. Non-metropolitan

• **F\_NATIVITY**

1. U.S.
2. Puerto Rico
3. Other U.S. territory
4. Another country
99. Refused

• **F\_PARTY FINAL**

1. Republican
2. Democrat
3. Independent
4. Something else
99. Refused

• **F\_PARTYLN FINAL**

1. The Republican Party
2. The Democratic Party
99. Refused

• **F\_PARTYSUM FINAL**

1. Rep/Lean Rep

2. Dem/Lean Dem
9. DK/Refused/No lean

• **F\_RACETHNMOD**

1. White
2. Black
3. Hispanic
4. Other
5. Asian
9. Refused

• **F\_REG**

1. You are ABSOLUTELY CERTAIN that you are registered to vote at your current address
2. You are PROBABLY registered, but there is a chance your registration has lapsed
3. You are NOT registered to vote at your current address
99. Refused

• **F\_RELIG**

1. Protestant (for example, Baptist, Methodist, Non-denominational, Lutheran, Presbyterian, Pentecostal, Episcopalian, Refo)
2. Roman Catholic
3. Mormon (Church of Jesus Christ of Latter-day Saints or LDS)
4. Orthodox (such as Greek, Russian, or some other Orthodox church)
5. Jewish
6. Muslim
7. Buddhist
8. Hindu
9. Atheist
10. Agnostic
11. Something else
12. Nothing in particular
99. Refused

• **F\_SEX**

1. Male
2. Female

99. Refused

• **F\_VOLSUM**

1. Yes

2. No

99. Refused