Substantive Significance of Multivariate Regression Coefficients

Jane E. Miller
Institute for Health, Health Care Policy and Aging Research
Rutgers University

Aims of this talk

- Provide concrete approaches to presenting substantive significance
  - Many statistics textbooks show how to assess and present statistical significance.
  - Few if any show how to assess and present substantive significance.
- Show how to implement a balanced presentation of
  - Inferential statistics for formal hypothesis testing.
  - Interpretation of substantive significance of findings in the context of the specific research question.

What is substantive significance?

- Real-world relevance to topic.
- In various disciplines, substantive significance =
  - "clinically…
  - "economically…
  - "educationally…
  - …meaningful" variation.
- Substantive significance of an association between two variables.
  - "So what?"
  - "How much does it matter?"

Example: Effects of video gaming

"For every hour a boy played a video game, he read *just two minutes less* than a boy who didn't play video games. Notably, non-gaming boys didn't read much at all either, spending *only eight minutes a day* with a book."

- Time spent playing video games shows a statistically significant negative association with time spent reading among adolescents.
- But is that gradient substantively significant?
  - Should video gaming be banned as a way of increasing reading time?

Cummings and Vandewater, 2007. "Relation of Adolescent Video Game Play to Time Spent in Other Activities," Archives of Pediatrics and Adolescent Medicine.

What ? does inferential statistics answer?

- "How likely would it be to obtain a coefficient at least as large as that estimated based on the sample, if in fact there is no effect of gaming on reading in the population from which the sample was drawn?"
- The *p*-value tells us the probability of falsely rejecting the null hypothesis.
  - Want *p* to be as small as possible.
  - Conventional levels of "statistical significance" : $p<.05$
- Strictly speaking, $p<.05$ tells us that for a large sample such as that used in the gaming study, the estimated coefficient on time spent gaming is at least 1.96 times its standard error.

What questions DOESN'T it answer?

- Whether the relationship is
  - Causal
    - Association ≠ causation
    - 2 variables can be correlated w/o one causing the other.
  - In the expected direction
    - The estimated coefficient could be stat sig but in the opposite of the hypothesized direction.
  - Big enough to matter in the real-world context
    - Each hour spent gaming reduced reading time by 2 minutes. Is that enough to induce genuine concern on the part of parents or teachers?

Conclusion: Don’t stop at “$p<.05$”!

- “$p<.05$” answers only part of what we want to know about our research question.
  - It is a necessary but not sufficient part of statistical analysis.
- Also need to consider questions about
  - Substantive significance
    - Direction
    - Size
  - Causality
    - Non-causal associations should not be used to inform policy or program changes.
    - Confounding or spurious associations should be ruled out.
      - Often why we estimate a multivariate model.

Principles for presenting results

- Name the specific variables.
  - Avoid writing about “my dependent variable” or “β.”
  - Avoid using acronyms from your database ☹
- Incorporate units into prose description.
- Report numbers in tables.
  - Complete set of coefficients, standard errors, GOF statistics.
- Interpret numbers in text.
  - Use prose to ask and answer research question, using #s as evidence.

What to report for coefficients

- Direction (AKA "sign")
  - For categorical variables, which category has higher value?
    - E.g., which gender has higher morbidity?
  - For continuous variables, is the trend up, down, or level?
    - E.g., as age increases, does income rise, fall, or remain constant?
- Magnitude
  - How much bigger?
- Statistical significance
  - Is the association statistically significant?

Interpreting βs: continuous predictors

- The unstandardized coefficient on a continuous predictor in an OLS model measures
  - The absolute difference in the dependent variable (Y) for a <u>one-unit increase</u> in the independent variable (X_i).
    - The slope of the relationship between X_i and Y.
  - Effect size is in original units of Y.

- Example topic: Birth weight by mother's age.
  - Dependent variable = birth weight in grams
  - Independent variable = mother's age in years

  - Both are continuous variables.

Mother's age as a predictor of birth weight

- Poor: "Mother's age and child's birth weight are correlated ($p<0.01$)."
  - Names the concepts involved and conveys statistical significance.
  - Omits direction or magnitude of the association.
- Better: "As mother's age increases, her child's birth weight also increases ($p<0.01$)."
  - Conveys concepts (dependent and independent variables), direction and statistical significance.
  - Doesn't report magnitude of the age/birth weight relationship.
- Best: "For each additional year of mother's age at the time of her child's birth, the child's birth weight increases by 10.7 grams ($p<0.01$)."
  - Concepts, units, direction, magnitude, and statistical significance.

Interpreting βs: categorical predictors

- The coefficient on a categorical predictor in an OLS model measures
  - Concept of a "1-unit increase" in X_i does **NOT** make sense.
  - The absolute difference in Y for the category of interest compared to the reference category.
  - Effect size is in original units of Y.

- Example: gender as a predictor of birth weight
  - Dummy variable BBBOY for "baby boy"
  - Coded
    - 1= boy
    - 0= girl = omitted (reference) category.

Gender as a predictor of birth weight

- Poor: "The β for 'BBBOY' is 116.1 with a s.e. of 12.3 (Table 14.3)."
  - Uses a cryptic acronym rather than naming the concept.
  - Reports the same information as the table, but does not interpret it.
  - Readers must calculate test statistic & compare it against critical value.
- Poor [version #2]: "Boys weigh significantly more at birth than girls."
  - Concepts and direction but not magnitude.
  - Statistical significance is ambiguous: Is the term "significant" intended in the **statistical** sense or to describe a **large difference**?
- Slightly better: "Gender is associated with a difference of 116.1 grams in birth weight ($p<.01$)."
  - Concepts, magnitude, and statistical significance but not direction: Was birth weight higher for boys or for girls?
- Best: "At birth, boys weigh on average 116 grams more than girls ($p<.01$)."
  - Concepts, reference category, direction, magnitude, and statistical significance.

The Goldilocks principle

- No one size numeric contrast is universally suited to interpretation of βs on all continuous X_i.
- When a 1-unit contrast is
  - Too big
    - Variables with low mean values and small range
      - E.g., variables measured in proportions
      - Range= 0-1, so 1 unit change is entire possible range!
  - Too small
    - Variables with high mean values and/or a wide range
      - E.g., a 1 mm Hg increase in blood pressure isn't clinically meaningful, nor can it be measured that precisely
  - Just right

Parsing the Goldilocks problem

- Identifying appropriate sized contrasts for interpreting βs on continuous variables depends on the
  - Topic of the research question
    - Read the literature to understand the substantive context of the analysis.
      - Can help identify pertinent contrasts in X_i
  - Context (when, where, who)
    - E.g., a $1 increase in weekly income might suit some developing countries today, or U.S. in the early 20th century, but not the U.S. today.
  - Distribution of variables
    - Examine level and range empirically.

Consider the dependent variable

- Topic and distribution of Y matters, too!
  - Birth weight
    - Plausible range is ~500 grams to ~5,000 grams
    - A β of 1.0 would be trivially small.
  - Grade point average (GPA)
    - Plausible range is 0.0 (F) to 4.0 (A)
    - A β of 1.0 is ¼ of the definitionally possible range!

Concluding points

- Emphasize the substantive issues behind the statistical analyses.
- Maintain a focus on your research question.
  - Avoid generic wording about “coefficients” or “variables.”
- Aim for a balanced presentation of statistical significance and substantive importance.
  - Use prose to ask and answer research question.
  - Use tables to report comprehensive, detailed statistics.

Resolving the Goldilocks problem

- Modify variable(s) so that a 1-unit increase fits the data and topic.
  - Change level of aggregation
    - Hourly income instead of weekly income.
  - Change the scale of measurement
    - Income in $100s or $1,000s instead of $1
- Take logs of X and/or Y
  - Commonly done in economics.
- Consider using categorical versions of continuous variables.
  - 20-unit ranges of blood pressure instead of continuous mm Hg.