Week 1
Course Overview

What psychometrics is...

Test theory: How tests work and how to make them work.

- How to construct a good test.
- How to evaluate a test.
- How best to score a test.
- How to avoid over-interpreting a test.
- How to justify your use and interpretation of a test.

...and is not...

Not individual assessment:

- It is an essential skill for those who do assessment.
- But psychometrics is not how to do assessment.
- Psychometrics and individual assessment can be thought of as overlapping topics.
- Note that the use of terms like “assessment” and “evaluation” vary with context.

...and is also not.

Not statistics:

- Statistical research benefits psychometrics.
- However, good psychometrics requires substantive knowledge, not just statistics.
- Understanding psychometric theory & psychological theory go hand in hand in test construction.
- E.g., Test statistics can tell you that an item does not work.
- Substantive knowledge tells you why, or how to make a better item.

Why psychometrics?

- Every experiment relies on effective measures of experimental outcomes.
- Every evaluation study relies on effective measures of program outcomes.
- Every treatment study relies on effective measures of treatment outcomes.

- Developing a measure makes a great dissertation.
  - Use it for research after your dissertation.
  - Less risky than trying to develop a measure and use it in your dissertation.

- Understand the theory behind the measures that you use.
  - Avoid uninformed mistakes.
  - Make better choices between measures.

- Research into a test also sheds light on the construct that it measures.
  - Likewise, research into a construct sheds light on the test.
  - Ideally, theory and measurement of a construct should advance together.

Syllabus

- Further note on software:
  - This course is not about learning software.
  - This course is not a substitute for a full course on SEM or IRT.
  - Nonetheless, toying with data offers a good opportunity to get more from the course.
  - The course project involves using R (with provided scripts).

- Learn by doing:
  - Like riding a bicycle: you can read about it, but...
  - Use what you know: Excel, SPSS, a calculator.
  - Play with formulas
  - Try to make them work or not work,
  - Take them apart to see how they work.
  - Test out ideas with simulated data.

What is a test?

"A test is a device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process" (2014 Standards, p. 2)

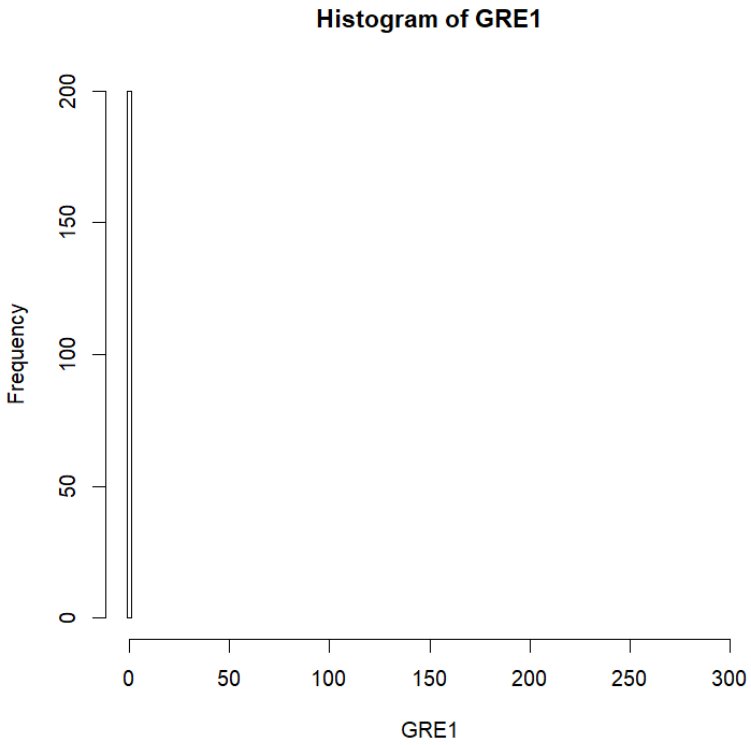
- A test does not have to be called a test to fit this definition (e.g., "assessment", "scale" or "inventory").
- A test does not have to be paper and pencil.
- A test does not have to be scored correct/incorrect.
- Standardization is a matter of degree.
- A test does not have to be self-report (e.g., behavioral observation)

Example of Psychometric Thinking

- Model GRE scores
- Mean & SD: very simple example.
- Plots illustrate the results of each model.
- Note GRE Psychology Subject Test ~N(600, 100)

Model 1: GRE = 0

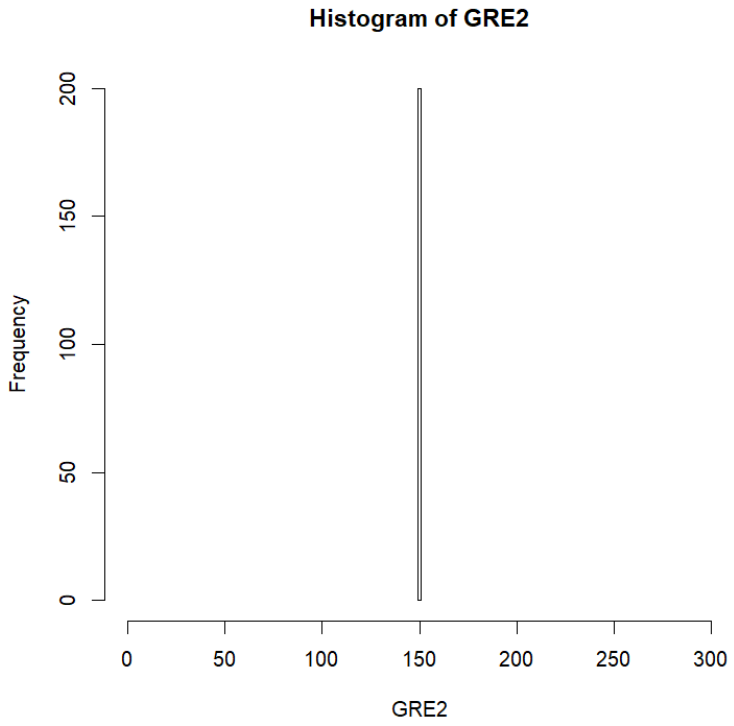
- Strengths: You can write an equation, make predictions, and test them.
- Weaknesses: The model does not fit empirical observation.



Model 2: GRE = mean = 150

- Strengths: If you had to guess somebody's GRE with no further info, this is the guess that gets you closest to their GRE on average.

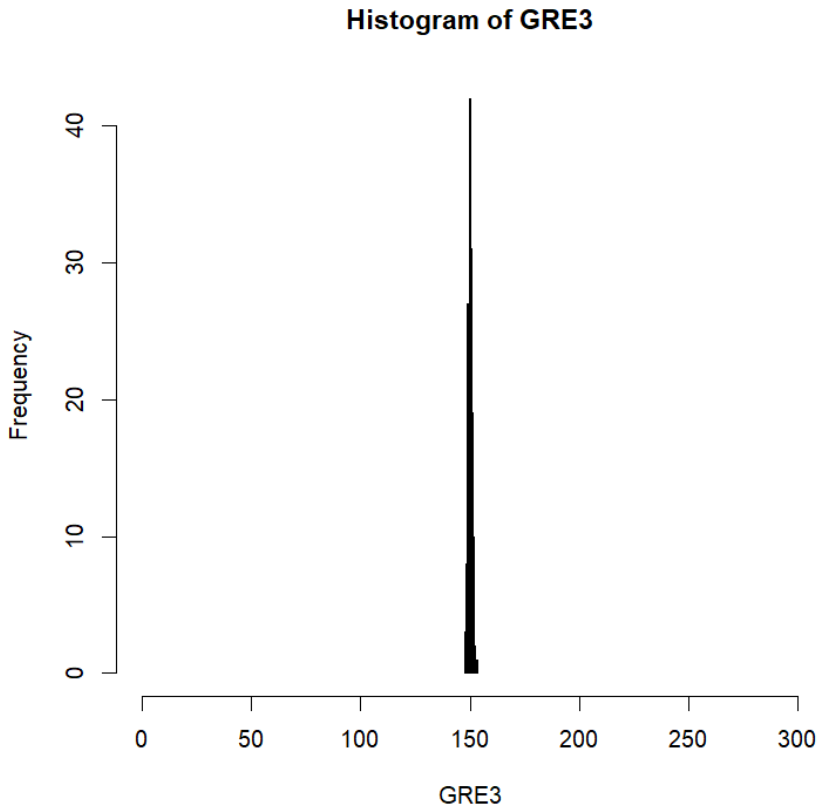
- Weaknesses: No variability, empirical GRE scores vary.



Model 3:
GRE = mean + error = 150 + e, e ~N(0, 1)

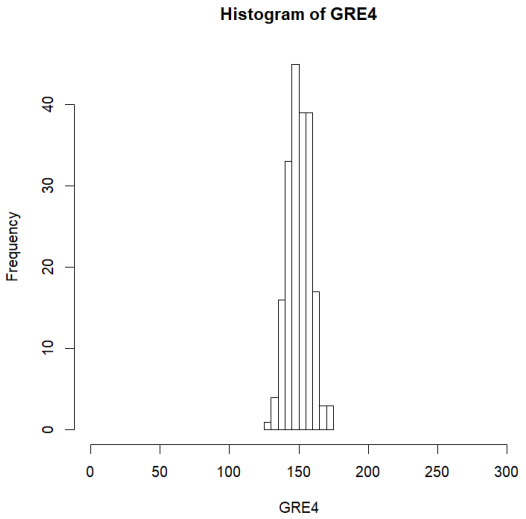
- Strengths:
  Correct mean,
  variability
  around mean.

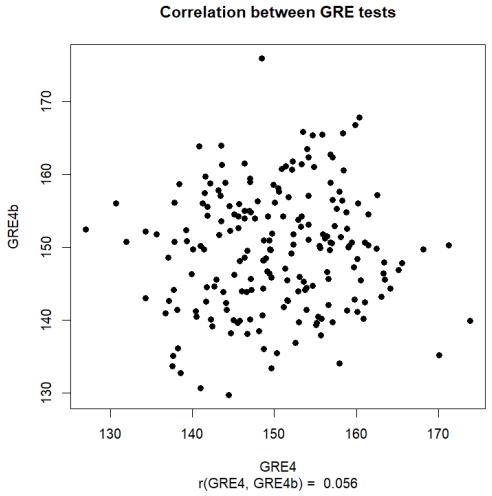
- Weaknesses:
  Not enough
  variability.



Model 4:
GRE = mean + error = 150 + e, e ~N(0, 8.5)

- Strengths: Correct mean and SD.
- Weaknesses: Treats GRE as completely random. Predicts retest correlation = zero.
- Note that the GRE models so far are like regressions with no predictors.

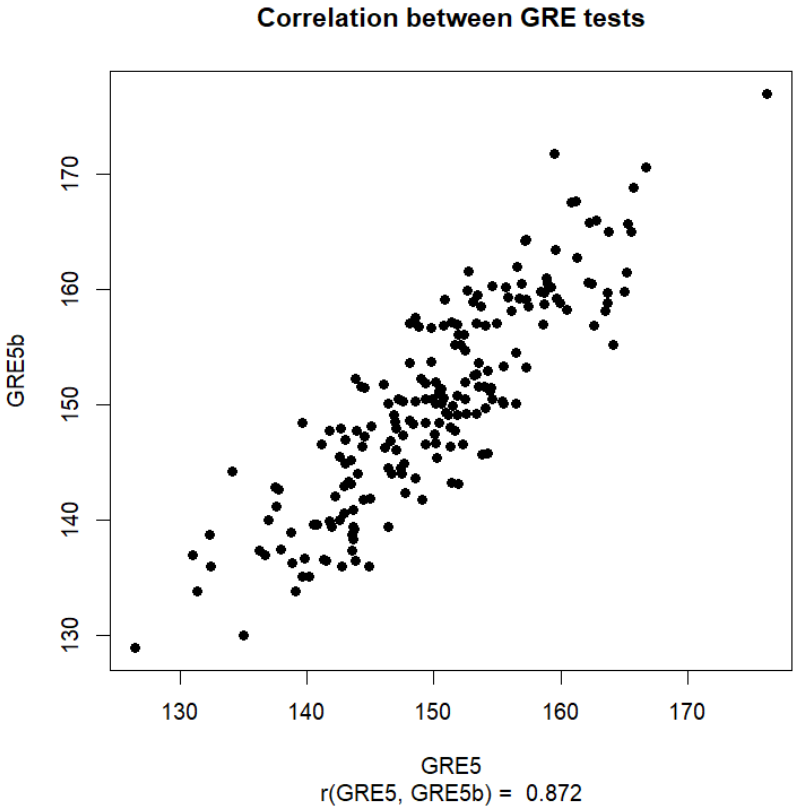




# Model 5:
GRE = mean + error = 150 + e, e~N(0,8.5)
GRE retake = 150 + (.9 * GRE mean deviation) + residual
residual ~N(0, 4)

- Strengths: Correct correlation (or at least a reasonable one)

- Weaknesses: It seems implausible that retest scores are just random deviations from previous scores. It is more likely that both are random deviations from a common cause. Also, this model assumes no change in participants between tests.



Psychometric Models

- The rest of the semester will present models that can predict the correlation more accurately.
- These include true score models, factor models, and IRT models.
- The latter two model item scores rather than test scores.

Software Notes

- The above graphs were produced by the R script <Week 1 v5.R>. Results will vary for stochastic models.
- See accompanying R script and Excel file for illustrations of how to implement the above simulation.