Psychometrics: An introduction

Overview

- A brief history of psychometrics
- The main types of tests
- The 10 most common tests
- Why psychometrics?: Clinical versus actuarial judgment

A brief history

- Testing for proficiency dates back to 2200 B.C., when the Chinese emperor used grueling tests to assess fitness for office

Francis Galton

- Modern psychometrics dates to Sir Francis Galton (1822-1911), Charles Darwin's cousin

• Interested in (in fact, obsessed with) individual differences and their distribution
• 1884-1890: Tested 17,000 individuals on height, weight, sizes of accessible body parts, + behavior: hand strength, visual acuity, RT etc
• Demonstrated that objective tests could provide meaningful scores

James Cattell

- James Cattell (studied with Wundt & Galton) first used the term 'mental test'in 1890

• His tests were in the 'brass instruments' tradition of Galton
  • mostly motor and acuity tests
• Founded 'Psychological Review'(1897)

Clark Wissler

- Clark Wissler (Cattell's student) did the first basic validational research, examining the relation between the old 'mental test' scores and academic achievement

• His results were largely discouraging
• He had only bright college students in his sample
  • Why is this a problem?
• Wissler became an anthropologist with a strong environmentalist bias.

Alfred Binet

- Goodenough (1949): The Galtonian approach was like "inferring the nature of genius from the the nature of stupidity or the qualities of water from those of….hydrogen and oxygen".

- Alfred Binet (1905) introduced the first modern intelligence test, which directly tested higher psychological processes (real abilities & practical judgments)
  - i.e. picture naming, rhyme production, weight ordering, question answering, word definition.
- Also motivated IQ (Stern, 1914): mental 'age' divided by chronological age

The rise of psychometrics

- Lewis Terman (1916) produced a major revision of Binet's scale
- Robert Yerkes (1919) convinced the US government to test 1.75 million army recruits
- Post WWI: Factor analysis emerged, making other aptitude and personality tests possible

What is a psychometric test?

- A test is a *standardized* procedure for *sampling* behavior and *describing* it using *scores or categories*
  - Most tests are predictive of of some non-test behavior of interest
  - Most tests are *norm-referenced* = they describe the behavior in terms of *norms*, test results gathered from a large group of subjects (the *standardization sample*)
  - Some tests are *criterion-referenced* = the objective is to see if the subject can attain some pre-specified criterion.

The main types of tests

- Intelligence tests: Assess intelligence
- Aptitude tests: Assess capability
- Achievement tests: Assess degree of accomplishment
- Creativity tests: Assess capacity for novelty
- Personality tests: Assess traits
- Interest inventories: Assess preferences for activities
- Behavioral tests: Measure behaviors and their antecedents/consequences
- Neuropsychological tests: Measure cognitive, sensory, perceptual, or motor functions

The 10 most commonly used tests

1.) Wechsler Intelligence Scale for Children (WISC)
2.) Bender Visual-Motor Gestalt Test
3.) Wechsler Adult Intelligence Scale (WAIS)
4.) Minnesota Multiphasic Personality Inventory (MMPI)
5.) Rorschach Ink Blot Test
6.) Thematic Apperception Test (TAT)
7.) Sentence Completion
8.) Goodenough Draw-A-Person Test
9.) House-Tree-Person Test
10.) Stanford-Binet Intelligence Scale

From Brown & McGuire, 1976

Clinical versus actuarial judgment

- Clinical judgment = reaching a decision by processing information in ones head
- Actuarial judgment = reaching a decision without employing human judgment, using empirically-established relations between data and the event of interest
  - 'Actuarial': *ad. L. actu [{amac}] ri-us*, a keeper of accounts
  - Note that some of the data in an actuarial judgment may be qualitative clinical observations, allowing a mixture of methods

Clinical versus actuarial judgment

- Paul Meehl (1954) first addressed the question: Which is better?

- His ground rules for comparison:
  - Both methods should draw from the same data set (this was relaxed by others, with no changes in results)
  - Cross-validation should be required, to avoid using variation specific to the data set
  - There should be explicit prediction of success, recidivism, or recovery

Meehl (1954): Results

- He looked at between 16 and 20 studies (depending on inclusion criteria)
  - "…it is clear that the dogmatic, complacent assertion sometimes heard from clinicians that 'naturally' clinical prediction, being based on 'real understanding' is superior, is simply not justified by the facts to date".
- In all but one case, predictions made by actuarial means were equal to or better than clinical methods
  - In a later paper, he changed his mind about the one.

Thirty years later...

"There is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly as this one."

Paul Meehl, 1986

Where are clinician's strengths? I

i.) Theory-mediated judgments

- If the predictor knows the relevant causal influences, can measure them, and has a model specific enough to take him/her from theory to fact
- However, are there any reasons to doubt this potential advantage?

Where are clinician's strengths? II

ii.) Ability to use rare events

- If the predictor knows that the current case is an exception to the statistical trend, s/he can use that information to over-ride the trend
- it is in theory possible to build these into actuarial methods
  - Why is it very difficult in practice?
  - Why might we worry about clinicians ability to incorporate rare events into prediction?

Where are clinician's strengths? III

iii.) Able to detect complex predictive cures

- Humans beings still (for now) are masters at recognizing some complex configurations, such as facial expressions etc.

Where are clinician's strengths? IV

iv.) Able to re-weight utilities in real-time

- For ethical, legal, or humanitarian reasons, we might decide to do things differently than usual in particular cases.

Where are actuarial strengths? I

i.) Immunity from fatigue, forgetfulness, hang-overs, hostility, prejudice, ignorance, false association, over-confidence, bias, and random fluctuations in judgment.

Where are actuarial strengths? II

ii.) Consistency & proper weighting

- variables are weighted the same way every time, according to their actual demonstrable contributions to the criterion of interest
- irrelevant variables are properly weighted to zero

Where are actuarial strengths? III

iii.) Feedback & base-rates 'built-in' to the system

- Clinicians rarely know how they are doing because they don't get immediate feedback and because they have imperfect memory
- actuarial records constitute perfect memories of how things came out in similar cases and can include a larger and wider sample than a human can ever hope to see

Where are actuarial strengths? IV

iv.) Not overly sensitive to optimal weightings

- Even simplistic actuarial judgments often beat human judgments