Introduction to the Principles and Practice of Clinical Research

Measures

David Black, Ph.D.

Pediatric and Developmental Neuropsychiatry, NIMH

With thanks to Audrey Thurm, Daniel Pine, Erin McClure, F. Xavier Castellanos & David Rubinow

The Uncertainty Principle

"The more precisely the POSITION is determined,
the less precisely the MOMENTUM is known."

Werner Heisenberg

Outline

1. What is a measure?
2. Types of measures?
3. Selection of measures?
4. Is your measure any good?
5. Additional factors influencing measurement?

What is a measure?

- To answer, need to know ….
  - Construct: concept or idea of interest in research protocol [depression]
  - Variable: operational definition of that construct [symptoms: flat affect, suicidal ideation, etc.]
  - Measure: Instrument used to assess level or change of the variable [Hamilton Rating Scale for Depression – HAM-D]
- Assumptions of variables:
  - Changes over time/across subjects or groups
  - Varies in an observable, quantifiable fashion
- Assumptions of measures
  - Measure = true value + error
  - Error = random error + non-random (biased) error

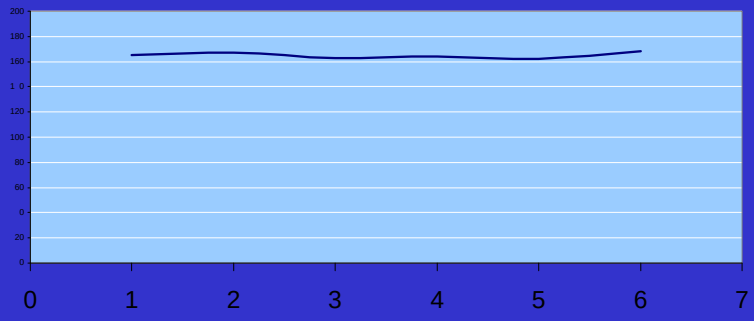
Types of Error

- Non-random error = Systematic Error = Bias
  - Reduce by good measurement and study design
- Random error = non-systematic error = noise in the measure
  - Minimize with use of clean measures (only measure the variable/construct of interest)
  - Larger sample: Central limit theorem
    http://www.stat.sc.edu/~west/javahtml/CLT.html

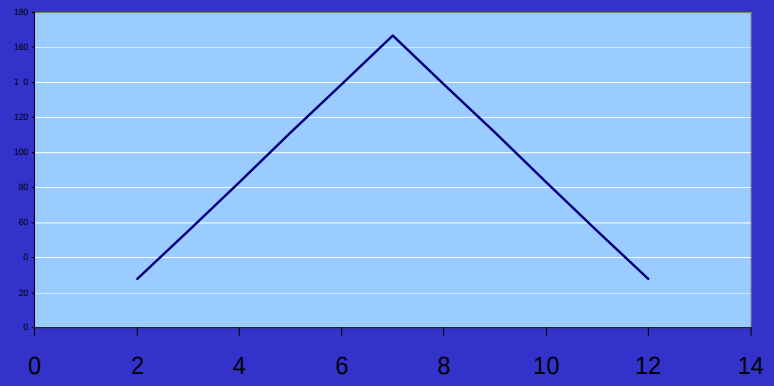
Central Limit Theorem in Action

Sum of All Rolled Dice

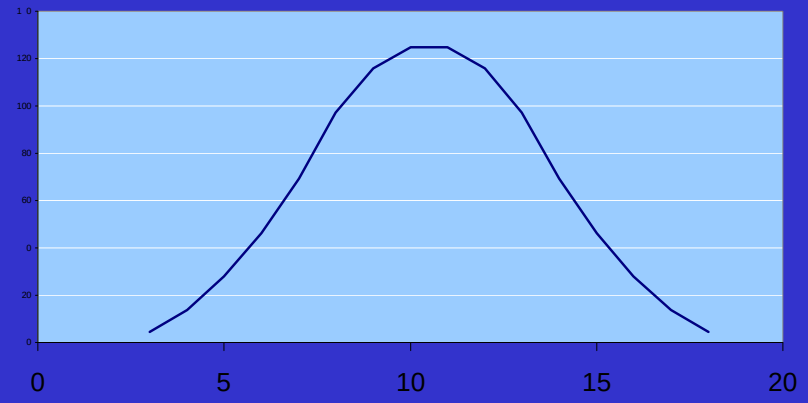
One Die



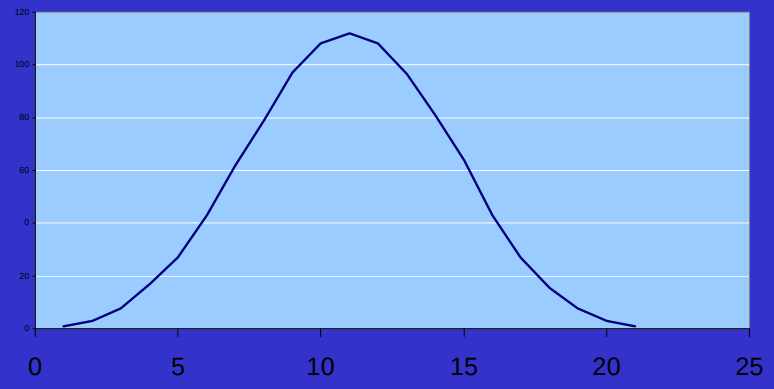
Two Dice



Three Dice



Four Dice



Types of Measures

- Qualitative Measures
  - Narratives
  - Provide rich detailed information often in narrative format
- Quantitative Measures
  - Questionnaires
  - Structured Interviews

Qualitative vs. Quantitative

Qualitative

- Richly detailed, nuanced
- Difficult to analyze
- Codifying may be subjective and time consuming

Quantitative

- Reduces complex concepts to numbers
- Lends to analysis
- Differences between groups can be readily tested
- Subtle differences may be missed

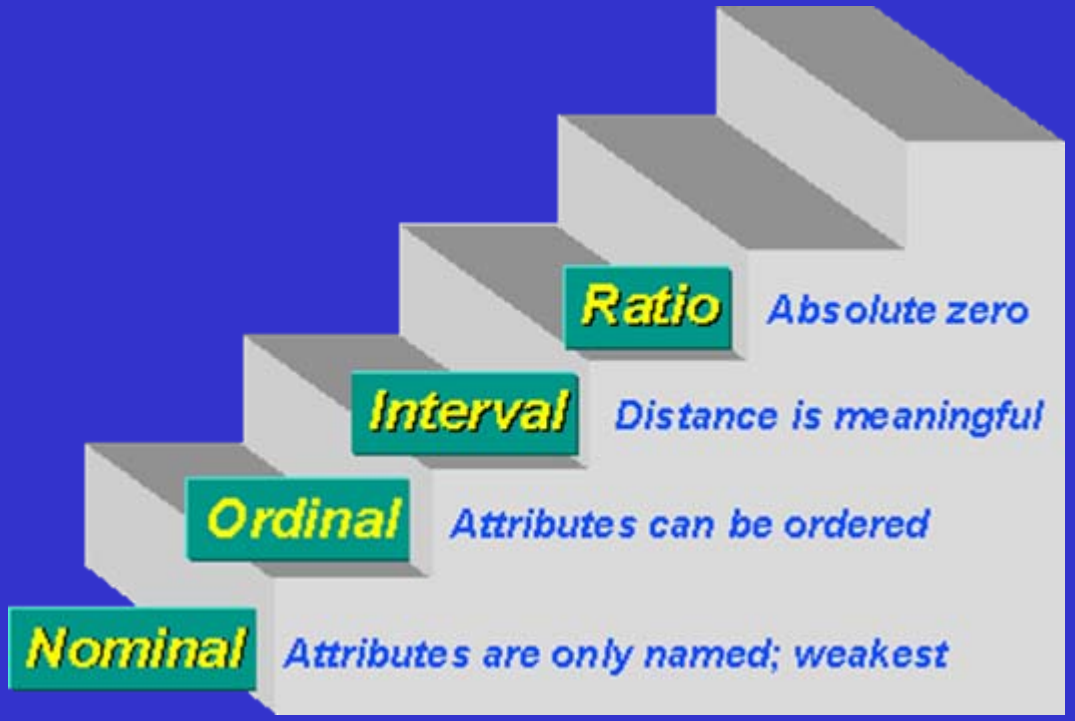
Measures must be reduced to numbers to be analyzed

- How measures are reduced to numbers have implications for:
  - Analysis of data: Statistical strategies
  - Interpretation of data
    - When are differences meaningful
      - Clinical significance
    - Interpretation of relative differences
      - Threshold
      - Continuous change

Types of Scales

- Nominal or Categorical
  - Classification or set of categories; mutually exclusive and collectively exhaustive – e.g., gender; sick vs. healthy
- Ordinal
  - Mutually exclusive classes that form an ordered series; rank order – e.g., Class standing; seriousness of a tumor
- Interval
  - Ordered series of ranks with equal intervals between any two pairs of adjacent classes – e.g., temperature
- Ratio
  - An interval scale with a true zero point origin – e.g., weight

Types of Scales



Selecting the right Measure?

Considerations

- Study design
  - Pre-post (measure change within group over time)
  - Group design (group differences)
  - Cross-sectional (sensitive in different populations)

Considerations

- Variable Characteristics & Content domain
  - Narrow vs. Broad
    - Specific phobia vs. general negative affect
  - Discrete vs. general
    - visual construction vs. general intelligence
  - Continuous vs. Categorical
    - symptom improvement vs. survival
  - Symptom vs. Syndrome
    - Pain vs. Depression

Considerations

- Variable Characteristics & Content domain
  - Syndrome specific or general construct
    - Symptom, biomarker vs. quality of life
  - Severity vs. frequency (or both)
    - How depressed are you vs. how many drinks last week
  - Absolute value vs. change from baseline
    - Compared to reference group vs. improvement
  - Subjective vs. objective
    - Pain level vs. change in laboratory value

Considerations

- Purpose of the measure
  - Initial eligibility criteria
    - screening instruments often have lower thresholds for inclusion
  - Measure improvement
    - capture subtle treatment changes
    - State measures vs. trait measures
  - Assignment of groups
    - Depressed vs. not depressed
    - Do you eliminate the middle ground to ensure distinct groups?

Considerations

- Characteristics of the measure
  - Range: Sensitive in range you are measuring
  - Specificity: Discriminates between groups
  - Sensitivity: Will identify case if present
  - Self-report/Clinician administered

Considerations

- Practical administration issues
  - Monetary cost
  - Burden to patient (time, intrusiveness)
  - Time investment to administer
  - Training needs to obtain valid and reliable administration

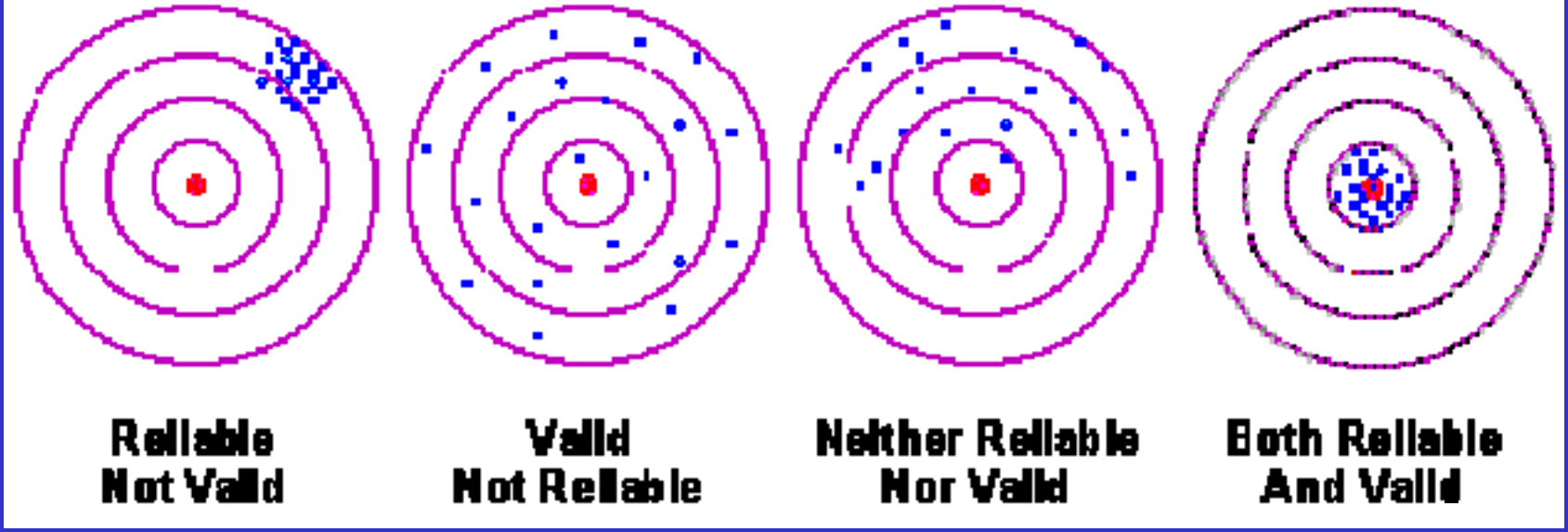
What makes a good measure?

Reliability, Validity, &
Demand characteristics

Good psychometric properties

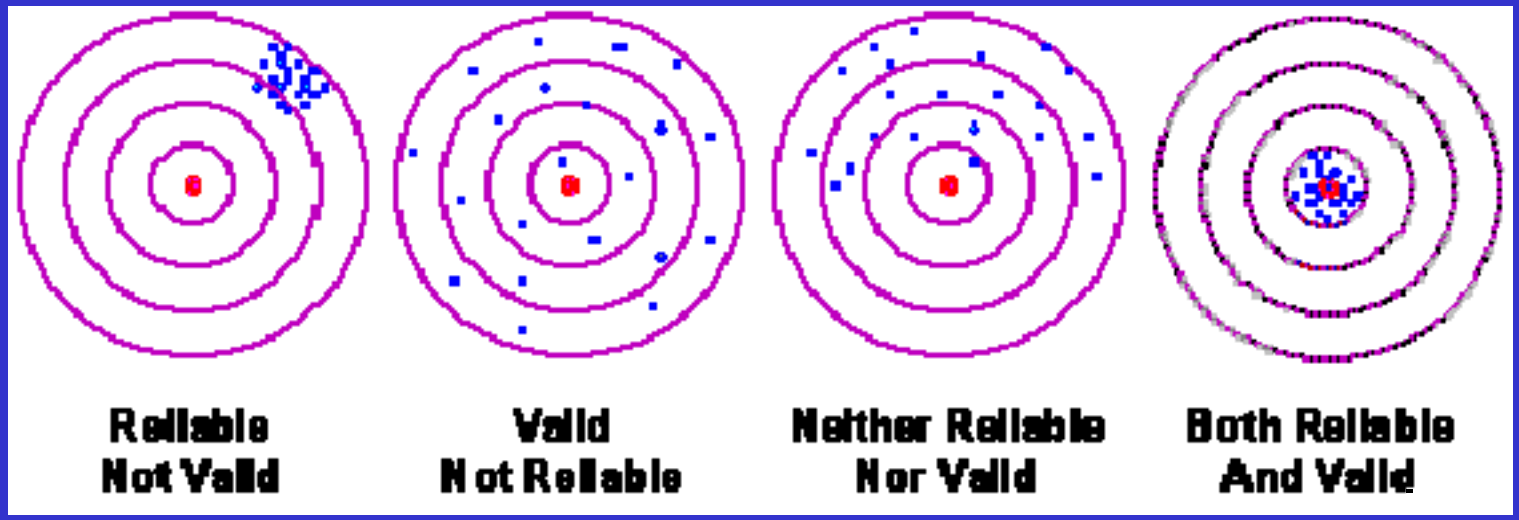
- Reliability
  - The consistency with which a measure assesses a given trait; i.e., agreement between two measures obtained by the same or **maximally similar** methods
- Validity
  - The extent to which a measure actually measures a trait; i.e., agreement between two measures obtained by **maximally different** methods

Reliability and Validity



Reliability and Statistical Power

- Increasing reliability reduces error



•Decreased error improves sensitivity to detect change, which improves sample power

Types of Reliability

- Test-retest
  - Temporal stability of a test from one measurement session to another
- Internal-consistency reliability
  - Also known as reliability of components – average of the intercorrelations of single test items; these coefficients go up as the number of test items increase

Inter-rater Reliability

- Extent to which two raters rate the same behavior similarly

- Measuring reliability:

Suitable Methods

- Kappa
- Intraclass correlation coefficient (ICC)

Unsuitable Methods

- Percent agreement
- Chi-square
- Correlation

Bartko & W T Carpenter, **J Nerv Ment Dis** 1976, 163:307-314

Diagnostic Reliability - % Agreement

B	Rater A			
	S	MD	Nr	Total
S	**40**	20	0	60
MD	0	**10**	10	20
Nr	10	0	**10**	20
Total	50	30	20	100

Bartko & Carpenter
J Nerv Ment Dis 1976, 163:307-314

Expected Frequencies:

B	Rater A			
	S	MD	Nr	Total
S	**30**	18	12	60
MD	10	**6**	4	20
Nr	10	6	**4**	20
Total	50	30	20	100

Types of Validity

- Face (Content) Validity
  - Do items appears to measure the construct of interest
- Criterion-related Validity
  - Comparison with independent, direct measures
- Construct Validity
  - Measurement of the theoretical construct

Sensitivity & Specificity

- Sensitivity: Probability of testing positive, given you have the disease
- Specificity: Probability of testing negative given you don't have the disease

**** Insensitive to population base rates***

- Positive Predictive Power (PPP): Probability of disease given positive test
- Negative Predictive Power (NPP): Probability of no disease given negative test

****Sensitive to population base rates***

Sensitivity & Specificity

	Disease Present	Disease Absent	
Positive Result	Group A True Positive	Group B False Positive	A/A+B Positive Predictive Power
Negative Result	Group C False Negative	Group D True Negative	D/C+D Negative Predictive Power
	A/A+C Sensitivity	D/B+D Specificity	

Sensitivity & Specificity:

Role of population base rates in test selection

- Assumptions for demonstration
  - Sensitivity = 99.9%
  - Specificity = 99.9%
  - Population sampled = 1 million subjects
- Base rates (frequency of occurrence in population):
  - 1%
  - 0.1%
  - 10%

http://www.musc.edu/dc/icrebm/sensitivity.html

1% Base Rate of Disease

	Disease Present	Disease Absent	
Positive Result	9,990	990	A/A+B PPP = 91%
Negative Result	10	989,010	D/C+D NPP = 99.9%
Total	10,000	990,000	
	A/A+C Sensitivity = 99.9%	D/B+D Specificity 99.9%	

http://www.musc.edu/dc/icrebm/sensitivity.html

.1% Base Rate of Disease

	Disease Present	Disease Absent	
Positive Result	999	999	A/A+B PPP = 50%
Negative Result	1	998,001	D/C+D NPP = 99.99%
	A/A+C Sensitivity = 99.9%	D/B+D Specificity 99.9%	

http://www.musc.edu/dc/icrebm/sensitivity.html

10% Base Rate of Disease

	Disease Present	Disease Absent	
Positive Result	99,900	900	A/A+B PPP = 99%
Negative Result	100	899,100	D/C+D NPP = 99.99%
	A/A+C Sensitivity = 99.9%	D/B+D Specificity 99.9%	

http://www.musc.edu/dc/icrebm/sensitivity.html

Demand Characteristics

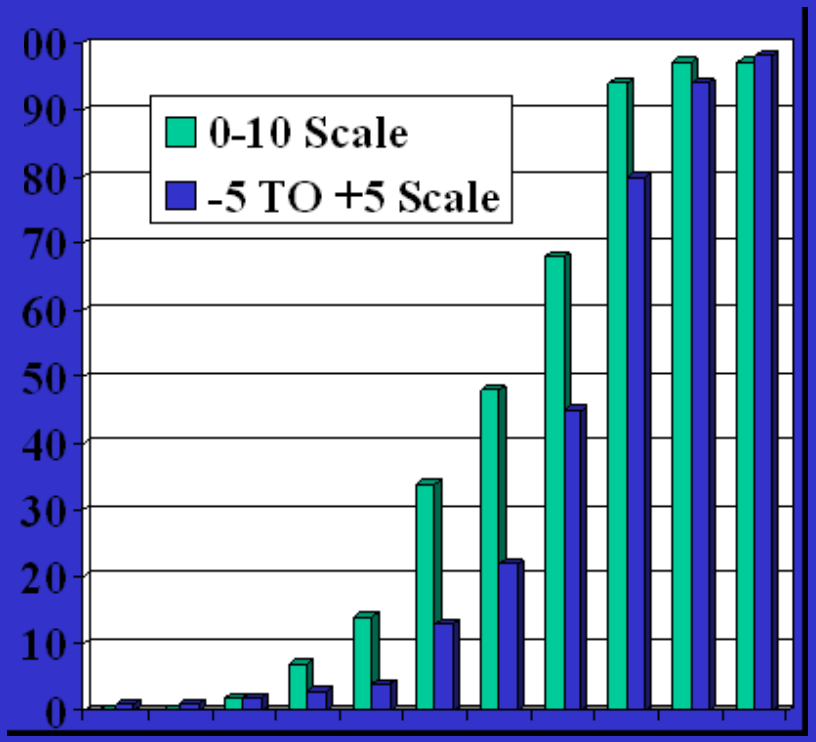
- Factors that impact how respondents rate themselves
  - Scale anchors
  - Response sets
  - Number of points on rating scale
  - Length of questionnaire
- Participant Reactivity
- Investigator Factors

Effect of Endpoint Labels

How successful would you say you have been in your life?

-5	0	+5
Not at all successful		Extremely successful

0	5	10



Schwartz et al., 1991 Public Op Q 570-582

Effects of Response Alternatives

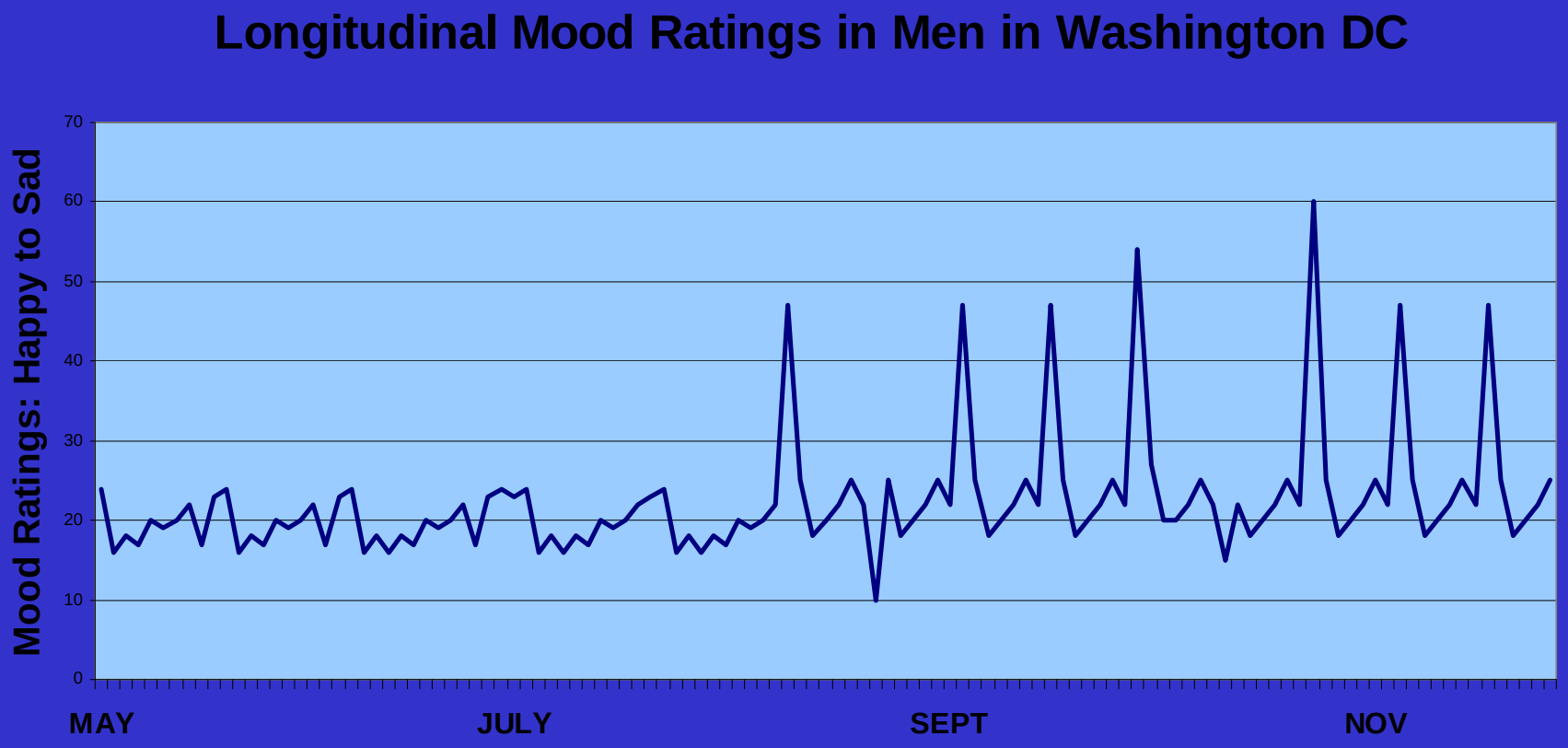
Reported Daily TV Viewing (%)				
Low Frequency Alternatives			High Frequency Alternatives	
Up to ½ h	7.4		Up to 2 ½ h	62.5
½ to 1 h	17.7		2 ½ to 3 h	23.4
1 to 1 ½ h	26.5		3 to 3 ½ h	7.8
1 ½ to 2 h	14.7		3 ½ to 4 h	4.7
2 to 2 ½ h	17.7		4 to 4 ½ h	1.6
Over 2 ½ h	16.2		Over 4 ½ h	0

Schwartz et al., 1985, Public Opinion Q 49: 388-395

Other Factors Influence a Measure

- Performance variables
  - Skill and care of administration
  - Practice, floor & ceiling effects
  - Typographical errors
- Test conditions: temperature, noise, privacy
- Insensitivity in range of interest
- Measurement Interval and duration
- Infinite unknown factors = RANDOMIZE

How long and how often should the measure be applied?



Subject Reactivity

- Social facilitation on easy tasks
- Anxiety on harder tasks
- Impression management
  - Want to be socially desirable
  - Want to please the researcher
- May lead to unrepresentative behavior

Expectancy Effects

- Usually in the predicted direction of hypotheses
- May occur at multiple levels research
  - Development of measure
  - Observation of behavior
  - Recording of behavior
  - Data analysis
  - Interpretation
- A primary reason for double-blind, placebo controlled trials & Peer review process

Measures - Summary

- Choose your measures carefully
  - Know their weaknesses
- Consider psychometric properties carefully
- Beware of biased error above all
  - Central Limit Theorem

Useful Resources

- http://www.socialresearchmethods.net/kb/index.php
  - Good discussion of research methods
- http://www.musc.edu/dc/icrebm/sensitivity.html
  - Sensitivity and Specificity
- http://www.intmed.mcw.edu/clincalc/bayes.html
  - Helpful statistics calculator (sensitivity and specificity)
- http://www.stat.sc.edu/~west/javahtml/CLT.html
  - Central limit theorem
- **http://www.uq.oz.au/~hmrburge/stats/index.html**
  - Primer on statistics
- **(book)** Rosenthal & Rosnow (1991). Essentials of Behavioral Research.