Statistics One

Lecture 6
Measurement

1

Three segments

- Reliability
- Validity
- Sampling

2

Lecture 6 ~ Segment 1

Reliability

3

Reliability

- Important concepts & topics
  - Classical test theory
  - Reliability estimates

4

Reliability

- Classical test theory
  - Raw scores (X) are not perfect
  - They are influenced by bias and chance error
    - For example, measurement of body temperature

5

Reliability

- Classical test theory
  - In a perfect world, it would be possible to obtain a "true score" rather than a "raw score" (X)
    - X = true score + bias + error
    - This is also known as "true score theory"

6

Reliability

- A measure (X) is considered to be reliable as it approaches the true score
  - The problem is we don't know the true score
  - So, we estimate reliability

7

Reliability

- Methods to estimate reliability
  - Test / re-test
  - Parallel tests
  - Inter-item estimates

8

Reliability

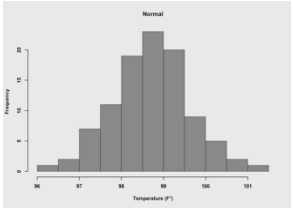
- Example: Body temperature
  - Orally
  - Internally
  - Infrared thermometer: “The wand”

9

Body temperature

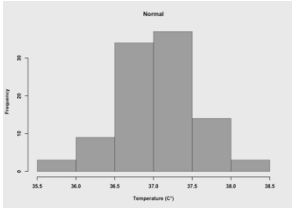
10

Body temperature F°



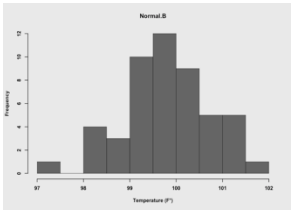
11

Body temperature C°



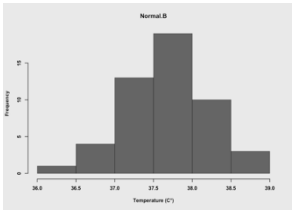
12

Body temperature F°: Biased



13

Body temperature C°: Biased



14

Reliability

- Test / re-test
  - Measure everyone twice
    - X1
    - X2

15

Reliability

- Test / re-test
  - The correlation between X1 and X2 is an estimate of reliability
    - However, if the bias is uniform then we won't detect it with the test / re-test method

16

Reliability

- Parallel tests
  - Measure body temperature with the wand (X1) and with an oral thermometer (X2)
  - The correlation between X1 and X2 is an estimate of reliability
    - AND, now the bias of the wand will be revealed

17

Reliability

- Inter-item
  - Inter-item is the most commonly used method in the social sciences
    - Test / re-test and parallel tests are time consuming
    - Inter-item is therefore more cost efficient

18

Reliability

- Inter-item
  - For example, suppose a 20-item survey is designed to measure extraversion
    - Randomly select 10 items to get sub-set A (X1)
    - The other 10 items become sub-set B (X2)
    - The correlation between X1 and X2 is an estimate of reliability

19

Segment summary

- Classical test theory (true score theory)
- Reliability estimates
  - Test / re-test
  - Parallel tests
  - Inter-item estimates

20

END SEGMENT

21

Lecture 6 ~ Segment 2

Validity

22

Validity

- What is a construct?
  - How to operationalize a construct
  - Construct validity
    - Content validity
    - Convergent validity
    - Divergent validity
    - Nomological validity

23

Validity

- What is a construct?
  - An ideal "object" that is not directly observable
    - As opposed to "real" observable objects
    - For example, "intelligence" is a construct

24

Validity

- How do we operationalize a construct?
  - The process of defining a construct to make it observable and quantifiable
    - For example, intelligence tests

25

Validity

- Construct validity
  - Content validity
  - Convergent validity
  - Divergent validity
  - Nomological validity

26

Validity

- An example:
  - Construct: Verbal ability in children

27

Validity

- How to operationalize?
  - A vocabulary test

28

Validity

- Construct validity
  - Content validity
    - Does the test consist of words that children in the population and sample should know?

29

Validity

- Construct validity
  - Convergent validity
    - Does the test correlate with other, established measures of verbal ability?
      - For example, reading comprehension

30

Validity

- Construct validity
  - Divergent validity
    - Does the test correlate less well with measures designed to test a different type of ability?
      - For example, spatial ability

31

Validity

- Construct validity
  - Nomological validity
    - Are scores on the test consistent with more general theories, for example, of child development and neuroscience
      - For example, a child with neural damage or disease to brain regions associated with language development should score lower on the test

32

Reliability & Validity: Review

- Important concepts & topics
  - Classical test theory
  - Reliability estimates
  - Construct validity

33

END SEGMENT

34

Lecture 6 ~ Segment 3

Sampling

35

Sampling

- Important concepts & topics
  - Random and representative sampling
  - Sampling error
  - Standard error

36

Sampling

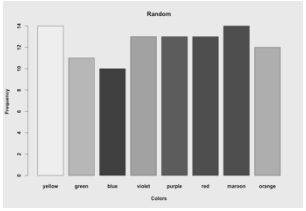
- Random and representative
- Recall the color wheel from Lecture 1

37

Illustration

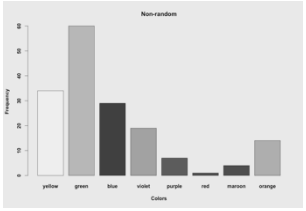
38

Random



39

Not random



40

Sampling error

- *Sampling error*: The difference between the population and the sample
  - Notice that even the "random" histogram is not "perfectly" random
  - There is some fluctuation due to *sampling error*

41

Sampling error

- PROBLEM!
  - We typically don't know the population parameters
  - So, how do we estimate sampling error?

42

Sampling error

- Sampling error mainly depends on the size of the sample, relative to the size of the population
  - As sample size increases, sampling error decreases
- It also depends on the variance in the population

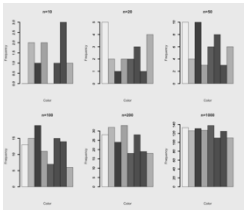
43

Sampling error

- Assume 6 samples from a normal population
  - N = 10
  - N = 20
  - N = 50
  - N = 100
  - N = 200
  - N = 1000

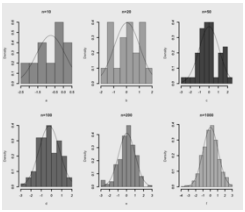
44

Sampling error



45

Sampling error



46

Sampling error

- Sampling error is estimated from the size of the sample and the variance in the sample
  - Under the assumption that the sample is random and representative of the population

47

Standard error

- Standard error is an estimate of amount of sampling error
  - SE = SD / SQRT(N)
  - SE: Standard error
  - SD: Standard deviation of the sample
  - N: Size of the sample

48

Segment Summary

- Important concepts & topics
  - Random and representative sampling
  - Sampling error
  - Standard error

49

END SEGMENT

50

END LECTURE 6

51