

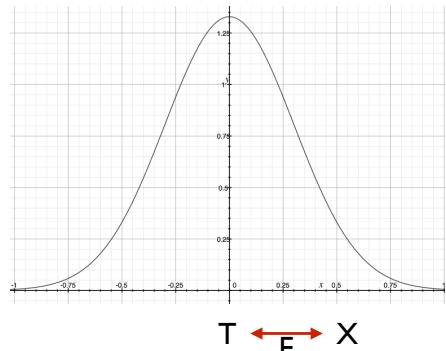
# Review : Reliability

- History
- Classical Test Score Theory
  - Models of reliability
  - Sources of error
- Estimating Reliability
- Increasing Reliability
- SEM and Confidence Intervals

## Classical Test-Score Theory

- $T$  = True Score
- $X$  = Observed
- $E$  = Error

- $X = T+E$
- $E = X-T$



Psychology 402 - Fall 2014 - Dr. Michael Diehr

## Reliability: errors & methods

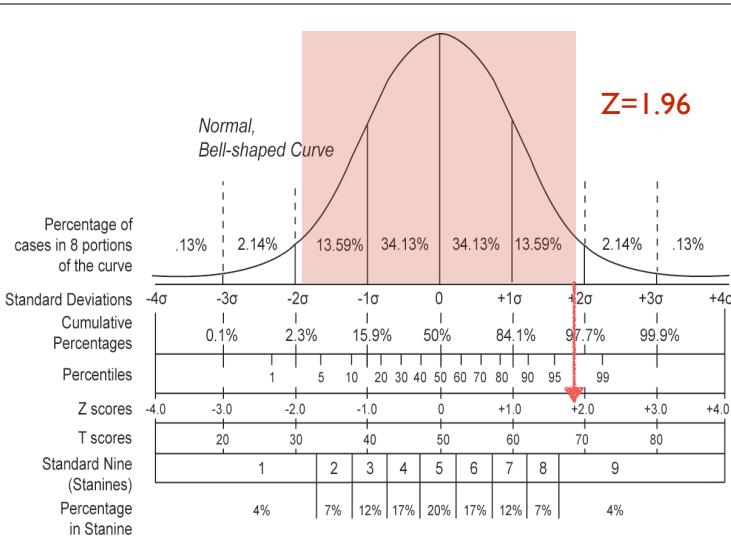
	Description	Name	Statistic
Time Sampling	1 test given two times	test-retest reliability	correlation between scores
Item Sampling	2 different tests given once	Alternate or Parallel forms	correlation between forms
Internal Consistency	One test, multiple items	Split Half or internal reliability	Cronbach's Alpha
Observer Differences	One test w/ 2+ observers	inter-observer reliability	Kappa

Psychology 402 - Fall 2014 - Dr. Michael Diehr

## Standard Error of Measurement

- $SEM = S\sqrt{1-r}$
- $S$  = std dev of measured scores
- $r$  = reliability coefficient of test
- SEM : 68% confidence interval (1 standard deviation, e.g.  $Z=\pm 1.0$ )

Psychology 402 - Fall 2014 - Dr. Michael Diehr



Psychology 402 - Fall 2014 - Dr. Michael Diehr

## Confidence Interval

- “How likely is a true score to fall within a range”
- $Z$  = z-score associated with % range
- Confidence interval =  $Z * SEM$
- Example:
  - 95% confidence interval :  $Z = 1.96$
  - $SEM = 4.64$
  - $1.96 * 4.64 = 9.1$
  - 95% CI =  $\pm 9.1$  points
  - Range =  $X \pm CI$
  - $106 \pm 9.1 = \text{range from } 96.9 \dots 115.1$

Psychology 402 - Fall 2014 - Dr. Michael Diehr

## Increasing Reliability

- **Increase N** (number of questions, items or tests)
- **Focus on common characteristic**
  - tests are more reliable if all items measure a single characteristic
- **Use Item Analysis** (“discriminability analysis”) to find items that best measure a single characteristic

## Reliability Summary

- Reliability: consistency of measurement
- Source of error → how to measure reliability
- Reliability coefficients ~ correlation
- Reliability is NOT Validity
- Reliability is a foundation