The (empirical) case for analyzing Likert data with parametric tests

Heather Wojton, PhD

Institute for Defense Analyses

04/xx/2017

Roadmap

- Evaluating human-system interactions in OT
- The great debate
- Let the data decide
- Recommendations for analysts
- I stop talking

Human-system interaction affects mission success

Likert data is commonly used to measure HSI in OT

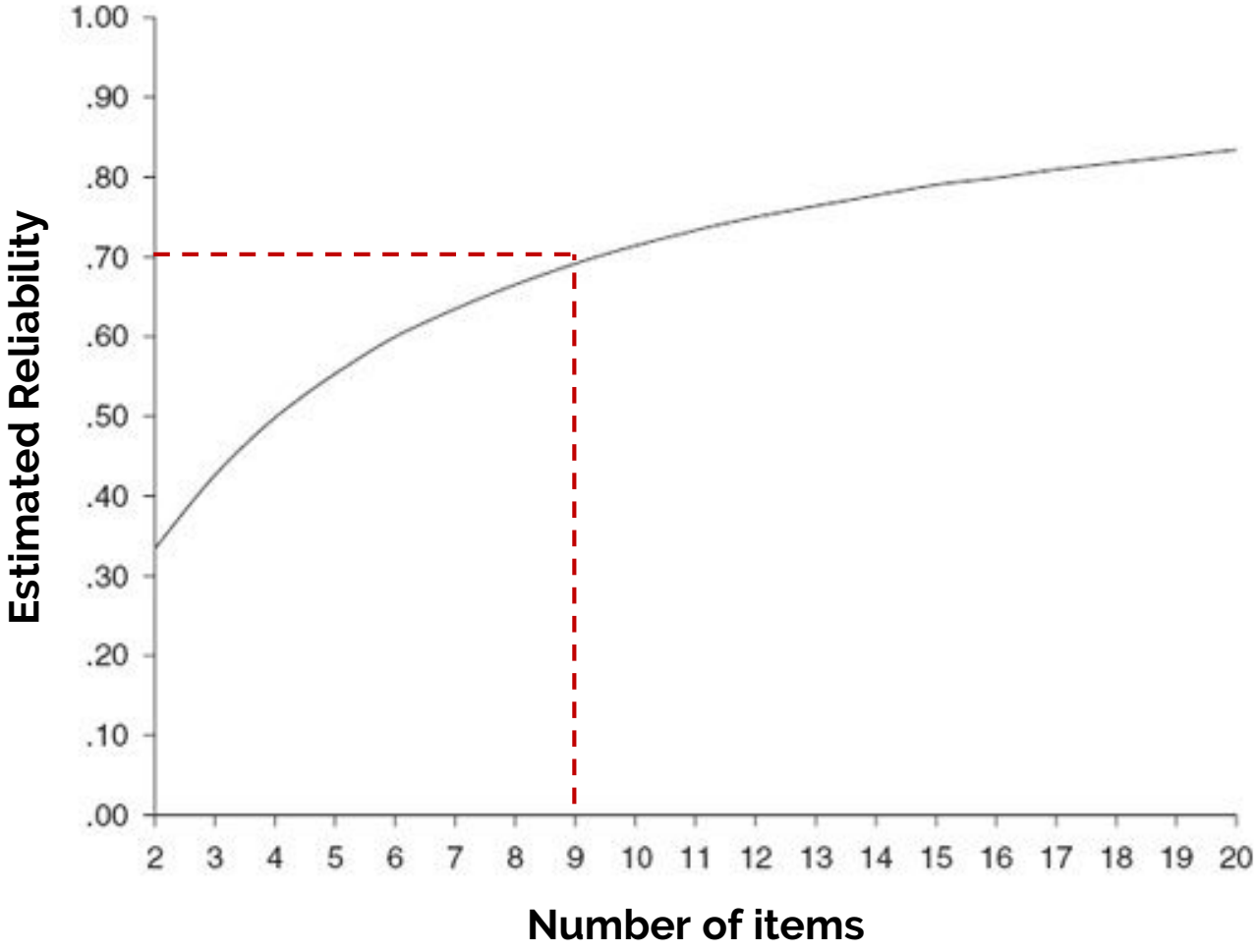
Data can be collected using a single item in the Likert response format…

How easy was it to navigate the interface?

Very Difficult						Very Easy
1	2	3	4	5	6	7

…or using a **Likert Scale** (Likert, 1932; Likert & Hayes, 1957)

Likert Scales *typically* include 8 or more items



Furr & Bacharach (2014)

Testers **disagree** on appropriate analysis methods for Likert data

Roadmap

- Evaluating human-system interactions in OT
- The great debate
- Let the data decide
- Recommendations for analysts
- I stop talking

Not all numbers are created equal

Stevens (1946) proposed 4 levels of measurement

Levels of Measurement

Property of Numbers

	Nominal	Ordinal	Interval	Ratio
Identity	X	X	X	X
Order		X	X	X
Quantity			X	X
Rational Zero				X
Example	*Sex*	*Education*	*Memory*	*Behavior*

Researchers have criticized this classification system

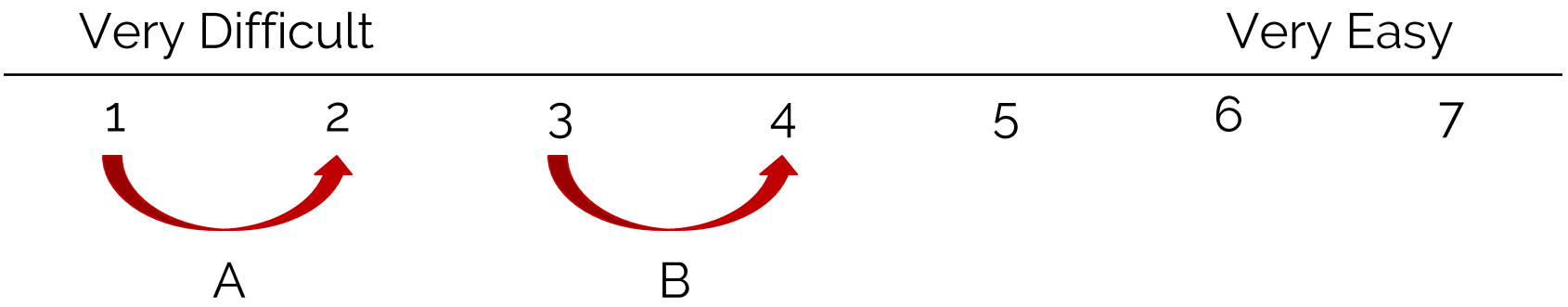
(Mitchell, 1986; Velleman & Wilkinson, 1993)

More nuanced classification systems exist

(Chrisman, 1998; Mosteller & Tukey, 1977; van den Berg, 1991)

The *ordinal-ist* argument is grounded in Stevens' levels of measurement

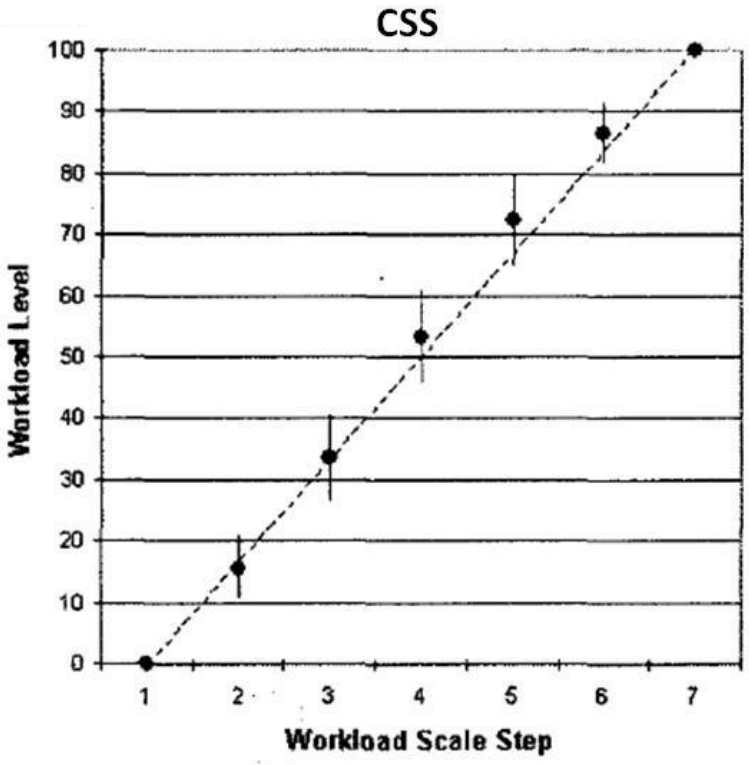
How easy was it to navigate the interface?

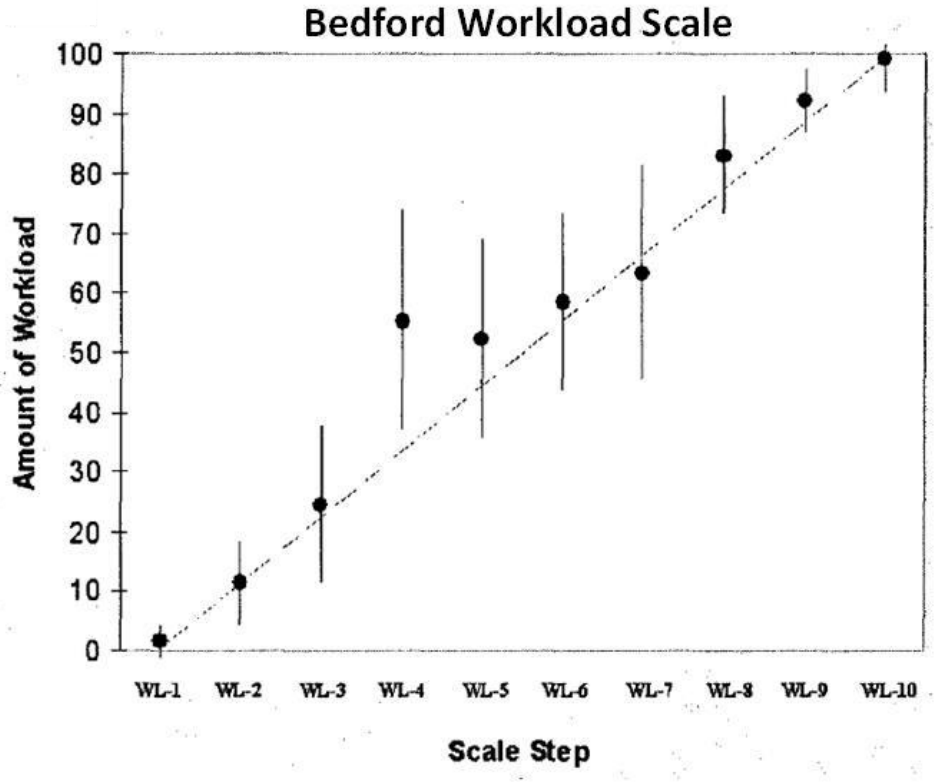


How do you know the distance between points is equal?

A = B **OR** A ≠ B

Ordinal-ists argue you can't guarantee distances are equal and thus, Likert data is ordinal

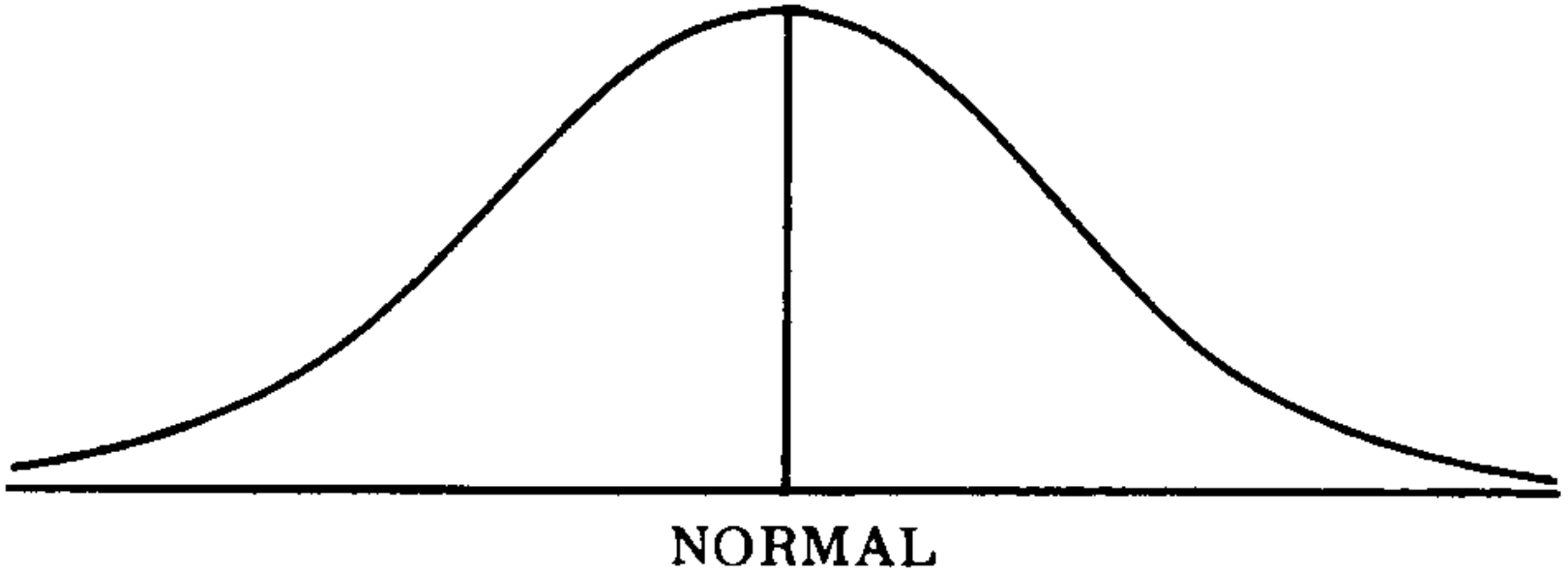




George, Edward J. (2004)

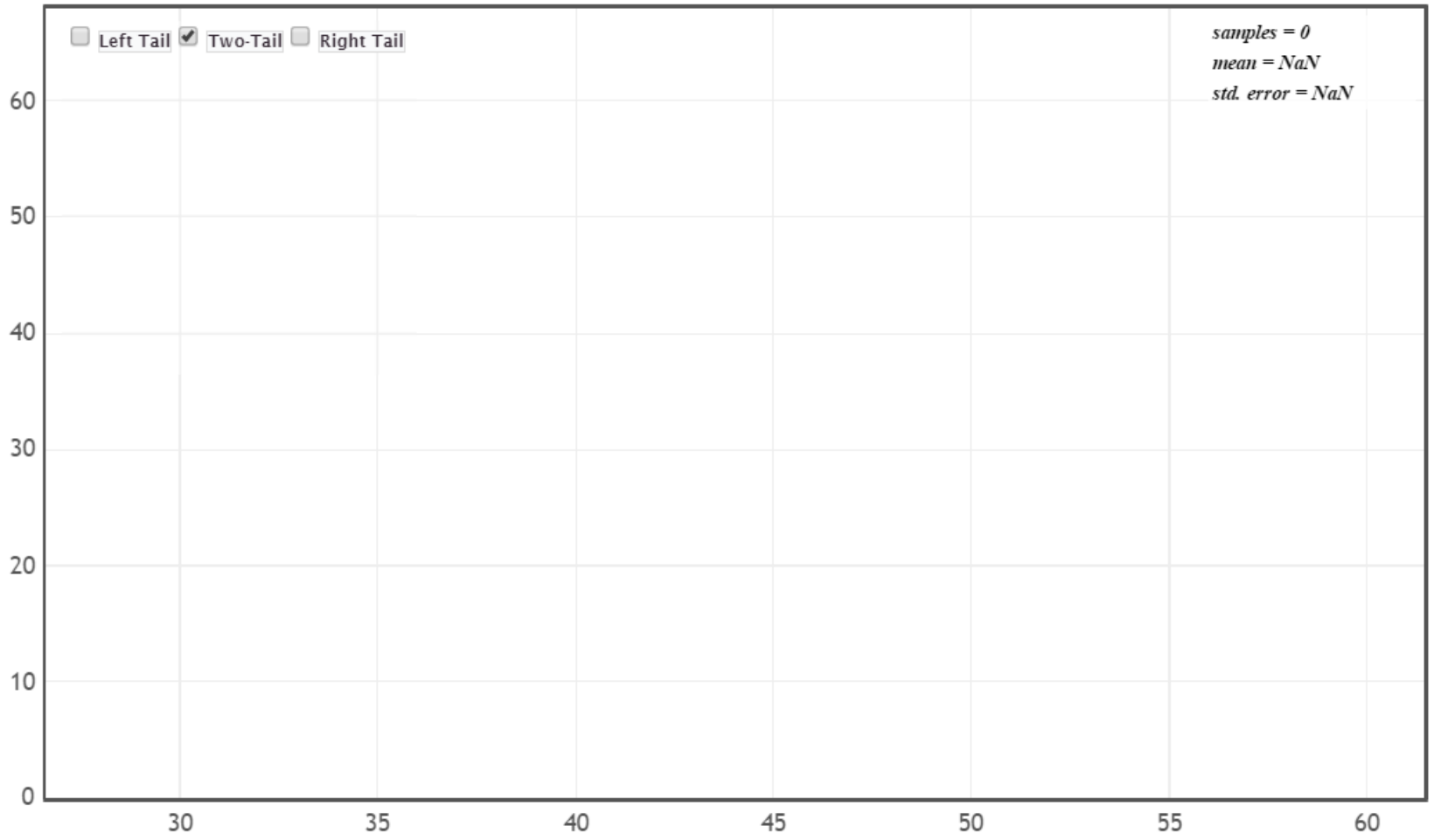
Ordinal-ists argue that Likert data violate the assumptions of parametric tests

They argue Likert data is not **continuous** or **normally distributed**



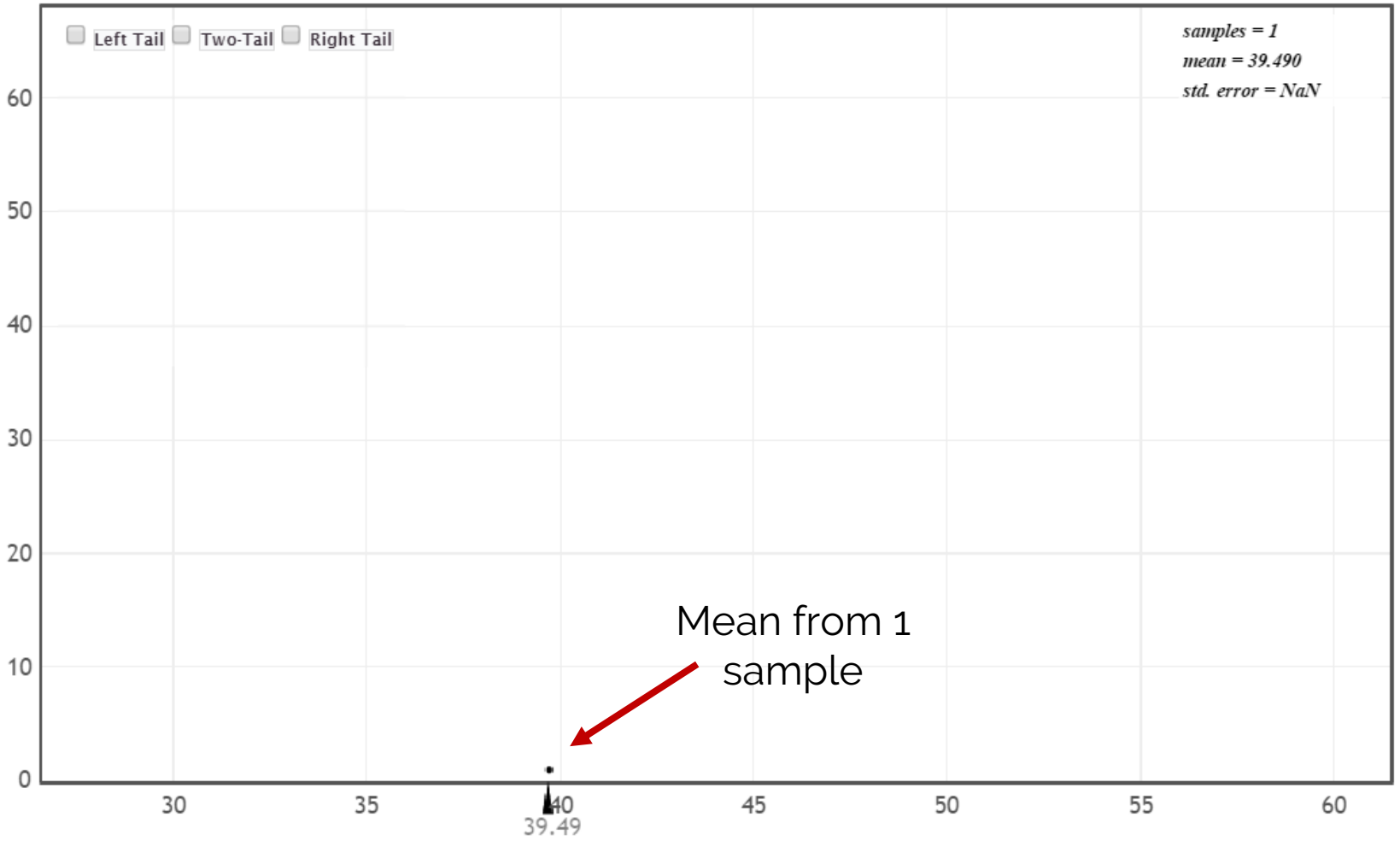
Argument is easier to apply to single Likert items than Likert scales

Normality assumption is often misunderstood



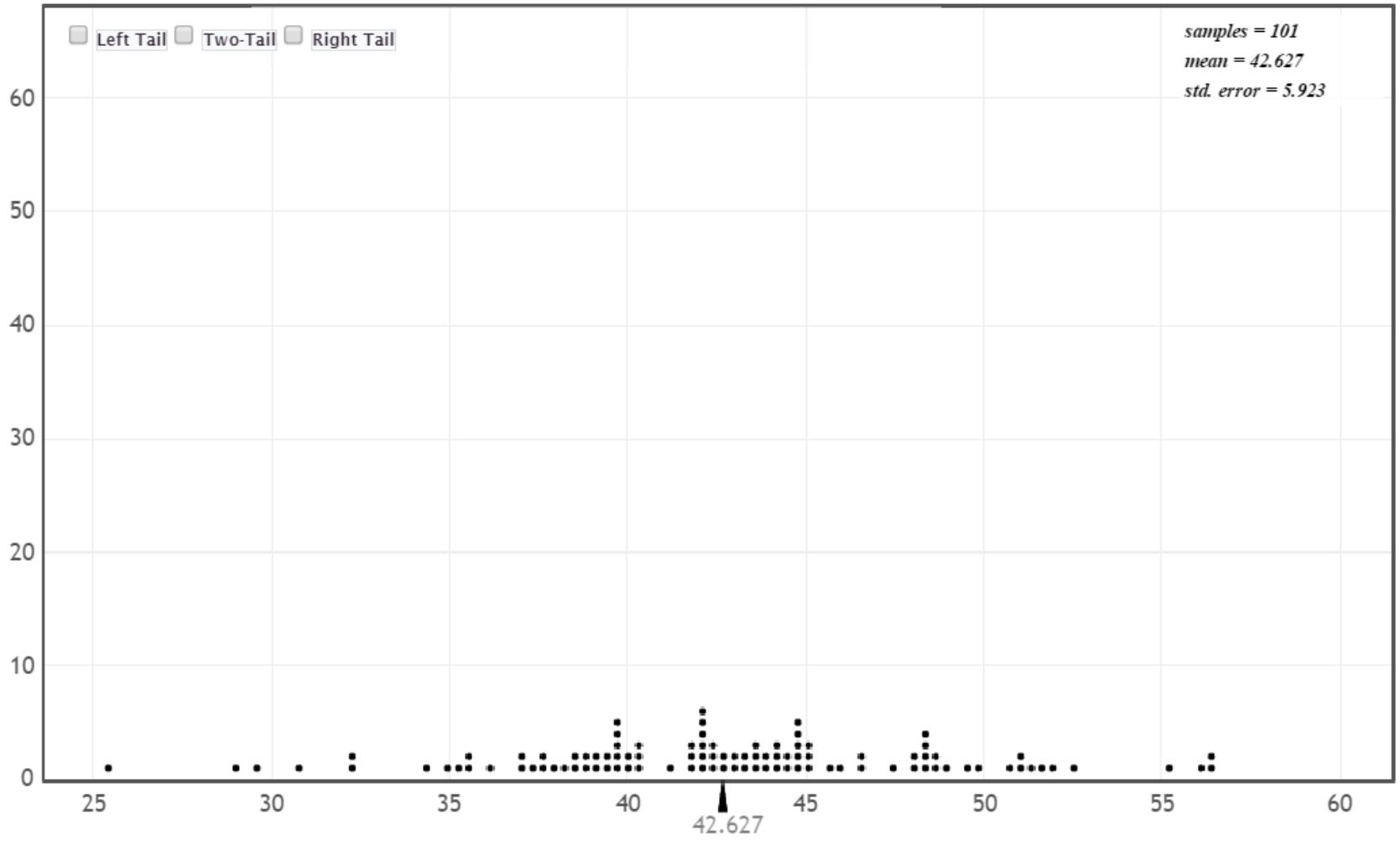
www.lock5stat.com

Normality assumption is often misunderstood



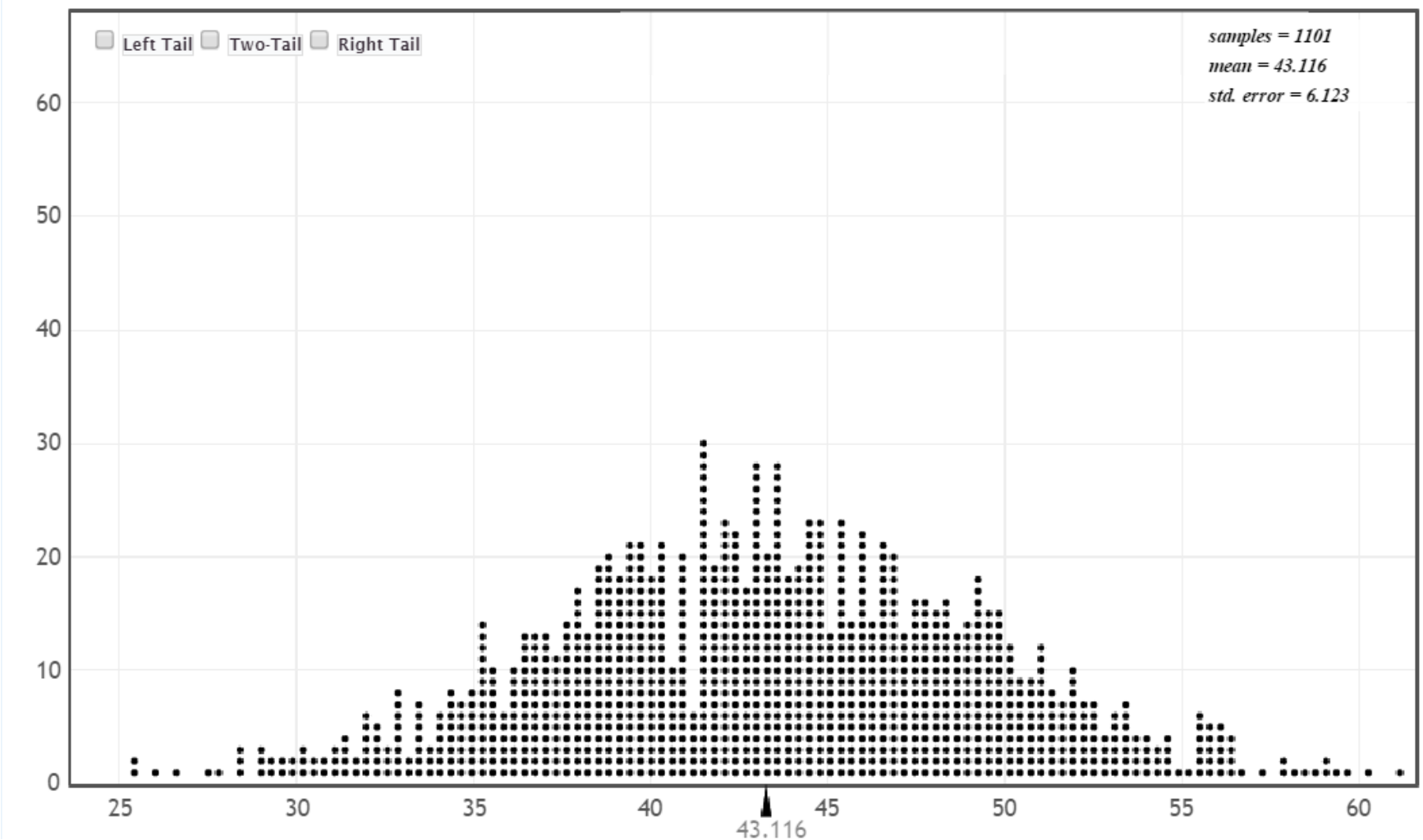
www.lock5stat.com

Normality assumption is often misunderstood



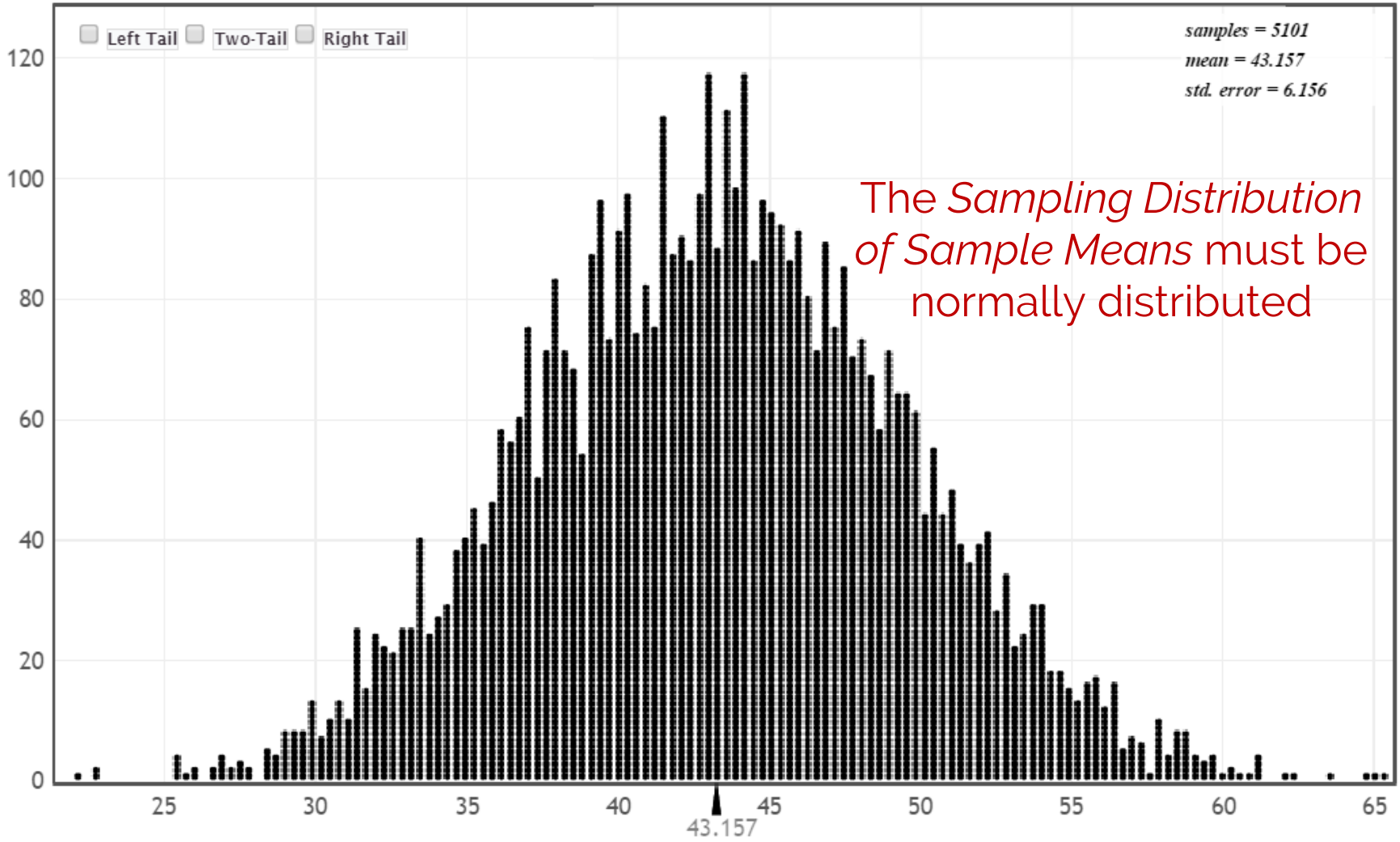
www.lock5stat.com

Normality assumption is often misunderstood



www.lock5stat.com

Normality assumption is often misunderstood



The *Sampling Distribution of Sample Means* must be normally distributed

www.lock5stat.com

Ordinal-ists argue that we will experience higher error rates using parametric statistics on Likert data because of these violations

(Nunnally, 1967; Jaimeson, 2004)

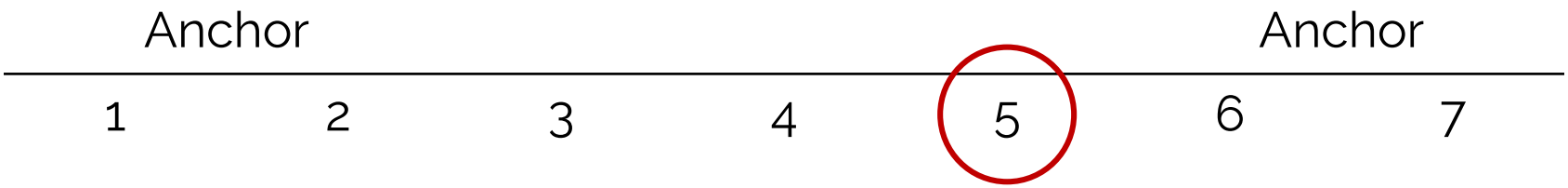
The Problem: Ordinal-ists are all theory and no data

Roadmap

- Evaluating human-system interactions in OT
- The great debate
- Let the data decide
- Recommendations for analysts
- I stop talking

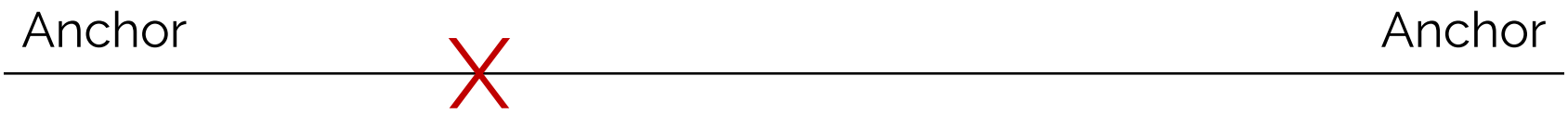
The distance between scale points is an empirical question

Question



vs.

Question



Visual Analogue Scale (VAS) scores correspond to actual changes in intensity of stimuli and are widely recognized as interval data

(Bolognese et al., 2003; Joyce et al., 1975; Myles et al., 1999; Price et al., 1983)

Evidence indicates the distance between scale points is roughly equal

Multiple studies have replicated this effect

(Baggaley & Hull, 1983; Carifio, 1976; Carifio 1978; Davey et al., 2007; Mauret & Pierce, 1998; Parker et al., 2002)

The effect of non-normality on error rates is an empirical question

The F-test is **robust** to violations of normality (Bartlett, 1937; Boneau, 1960; Box & Anderson, 1955; David & Johnson, 1951; Glass et al., 1972; Gombolay & Shah, 2016; Lindquist, 1953; Norton, 1952; Pearson, 1931)

Glass (1972) examined the effect of scale length on type I error rate in F-tests

The F-test controlled type I error rates for scales with at least 5 points

Skewness, kurtosis, and moderate heterogeneity of variance had little impact

Five-Point Scale

Populations Sampled	n	*Nominal Significance Levels* .10	.05	.01
A,A,A,A	11	.1014	.0516	.0104
A,A,A,A	51	.0986	.0518	.0096
B,B	11	.1074	.0518	.0130
B,B	51	.1024	.0516	.0118
B,B,B,B	11	.0976	.0516	.0104
B,B,B,B	51	.1004	.0492	.0108
C,C	11	.0992	.0470	.0088
C,C	51	.1000	.0502	.0094
C,C,C,C	11	.1016	.0480	.0096
C,C,C,C	51	.0974	.0522	.0108
A,B*	11	.1128	.0556	.0100
A,B	51	.0996	.0504	.0106
A,B,B,B	11	.1040	.0550	.0132
A,B,B,B	51	.1016	.0494	.0128

We **do not** risk higher error rates when analyzing Likert data with the F-test or t-test

The effect of non-normality on error rates is an empirical question

The Pearson correlation is **robust** to violations of normality

(Pearson, 1931, 1932a, 1932b; Dunlap, 1931; Havelick & Peterson, 1976; Murray, 2013)

Norman (2010) asked participants to complete 8, 10-point Likert format questions on 2 occasions

Computed Pearson and Spearman correlations for responses across the 2 occasions

Predicted the Spearman correlation from the Pearson correlation

	Original 10 point scales	Collapsed 5 point scales	Transformed 4 point scales
Slope	1.001	1.018	0.995
Intercept	−0.007	−0.013	−0.0003
Correlation	0.990	0.992	0.987
Mean Pearson	0.529	0.521	0.485
Mean Spearman	0.523	0.517	0.488

Pearson performed equivalent to Spearman even when data was severely non-normal

We **do not** risk higher error rates when analyzing Likert data with the Pearson correlation or Regression

Roadmap

- Evaluating human-system interactions in OT
- The great debate
- Let the data decide
- Recommendations for analysts
- I stop talking

Use **parametric statistics** to analyze your Likert data

Likert data approximates interval data

Greater power to detect an effect

Error rates are not higher

Roadmap

- Evaluating human-system interactions in OT
- The great debate
- Let the data decide
- Recommendations for analysts
- I stop talking

In science, data trumps theory

Questions?