An introduction to Psychometric Theory
Issues in Scaling

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2019
Outline

Types of scales and how to describe data
   Describing data graphically
      Central Tendency

More scaling examples
   Shape

Types of scales
   Examples

Measures of dispersion
   What is the fundamental scale?
**Four types of scales and their associated statistics**

**Table:** Four types of scales and their associated statistics (Rossi, 2007; Stevens, 1946) The statistics listed for a scale are invariant for that type of transformation.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Basic operations</th>
<th>Transformations</th>
<th>Invariant statistic</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>equality</td>
<td>Permutations</td>
<td>Counts</td>
<td>Detection</td>
</tr>
<tr>
<td></td>
<td>$x_i = x_j$</td>
<td></td>
<td>Mode</td>
<td>Species classification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\chi^2$ and $(\phi)$ correlation</td>
<td>Taxons</td>
</tr>
<tr>
<td>Ordinal</td>
<td>order</td>
<td>Monotonic</td>
<td>Median</td>
<td>Mhos Hardness scale</td>
</tr>
<tr>
<td></td>
<td>$x_i &gt; x_j$</td>
<td>(homeomorphic)</td>
<td>Percentiles</td>
<td>Beaufort Wind (intensity)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x' = f(x)$</td>
<td>Spearman correlations*</td>
<td>Richter earthquake scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>f is monotonic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval</td>
<td>differences</td>
<td>Linear</td>
<td>Mean ($\mu$)</td>
<td>Temperature ($^\circ$F, $^\circ$C)</td>
</tr>
<tr>
<td></td>
<td>($x_i - x_j &gt; (x_k - x_l$)</td>
<td>(Affine)</td>
<td>Standard Deviation ($\sigma$)</td>
<td>Beaufort Wind (velocity)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x' = a + bx$</td>
<td>Pearson correlation ($r$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Regression ($\beta$)</td>
<td></td>
</tr>
<tr>
<td>Ratio</td>
<td>ratios</td>
<td>Multiplication</td>
<td>Coefficient of variation ($\frac{\sigma}{\mu}$)</td>
<td>Length, mass, time</td>
</tr>
<tr>
<td></td>
<td>($x_i &gt; x_j \frac{x_k}{x_l}$)</td>
<td>(Similarity)</td>
<td></td>
<td>Temperature ($^\circ$K)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$x' = bx$</td>
<td></td>
<td>Heating degree days</td>
</tr>
</tbody>
</table>

The Beaufort wind speed scale is interval with respect to the velocity of the wind, but only ordinal with respect to the effect of the wind. The Richter scale of earthquake intensity is a logarithmic scale of the energy released but
**Graphical and tabular summaries of data**

1. The Tukey 5 number summary shows the important characteristics of a set of numbers
   - Maximum
   - 75th percentile
   - Median (50th percentile)
   - 25th percentile
   - Minimum

2. Graphically, this is the box plot
   - Variations on the box plot include confidence intervals for the median
The summary command gives the Tukey 5 numbers

```r
> summary(sat.act)

gender    education    age    ACT    SATV
Min. :1.000 Min. :0.000 Min. :13.00 Min. : 3.00 Min. :200.0
1st Qu.:1.000 1st Qu.:3.000 1st Qu.:19.00 1st Qu.:25.00 1st Qu.:550.0
Median :2.000 Median :3.000 Median :22.00 Median :29.00 Median :620.0
Mean :1.647  Mean :3.164  Mean :25.59  Mean :28.55  Mean :612.2
3rd Qu.:2.000 3rd Qu.:4.000 3rd Qu.:29.00 3rd Qu.:32.00 3rd Qu.:700.0
Max. :2.000  Max. :5.000  Max. :65.00  Max. :36.00  Max. :800.0
```
A box plot of the first 4 sat.act variables

A Tukey Boxplot

```
boxplot(sat.act[1:4], main="A Tukey Boxplot")
```
A violin or density plot of the first 5 epi.bfi variables

Density plot

- epiE
- epiS
- epilmp
- epilie
- epiNeur
The `describe` function gives more descriptive statistics

```r
> describe(sat.act)

vars  n  mean   sd median trimmed  mad min max range  skew  kurtosis   se
gender 1 700 1.65 0.48  2  1.68  0.00  1  2   1 -0.61 -1.62  0.02
education 2 700 3.16 1.43  3  3.31  1.48  0  5   5 -0.68 -0.07  0.05
age 3 700 25.59 9.50  22 23.86  5.93 13  65  52  1.64  2.42  0.36
ACT 4 700 28.55 4.82  29 28.84  4.45  3  36  33 -0.66  0.53  0.18
SATV 5 700 612.23 112.90  620 619.45 118.61 200 800 600 -0.64  0.33  4.27
SATQ 6 687 610.22 115.64  620 617.25 118.61 200 800 600 -0.59 -0.02  4.41
```
Multiple measures of central tendency

**mode**  The most frequent observation. Not a very stable measure, depends upon grouping. Can be used for categorical data.

**median**  The number with 50% above and 50% below. A powerful, if underused, measure. Not sensitive to transforms of the shape of the distribution, nor outliers. Appropriate for ordinal data, and useful for interval data.

**mean**  One of at least seven measures that assume interval properties of the data.
Multiple ways to estimate the mean

Arithmetic mean \( \bar{X} = X. = \left( \sum_{i=1}^{N} X_i \right) / N \) \( \text{mean}(x) \)

Trimmed mean throws away the top and bottom t% of observations. This follows the principle that all data are normal at the middle. \( \text{mean}(x, \text{trim}=.1) \)

Winsorized mean Find the arithmetic mean after replacing the n lowest observations with the nth value, and the N largest values with the Nth largest. \( \text{winsor}(x, \text{trim}=.2) \)

Geometric Mean \( \bar{X}_{\text{geometric}} = N \sqrt{\prod_{i=1}^{N} X_i} = e^{\Sigma(ln(x))/N} \) (The anti-log of the mean log score). \( \text{geometric.mean}(x) \)

Harmonic Mean \( \bar{X}_{\text{harmonic}} = \frac{N}{\sum_{i=1}^{N} 1/X_i} \) (The reciprocal of the mean reciprocal). \( \text{harmonic.mean}(x) \)

Circular Mean \( \bar{X}_{\text{circular}} = tan^{-1} \left( \frac{\sum cos(x)}{\sum sin(x)} \right) \) \( \text{circular.mean}(x) \)

(\( x \) is in radians)

\( \text{circadian.mean} \) \( \text{circular.mean}(x) \) (\( x \) is in hours)
Circular statistics

Table: Hypothetical mood data from six subjects for four mood variables. The values reflect the time of day that each scale achieves its maximum value for each subject. Each mood variable is just the previous one shifted by 5 hours. Note how this structure is preserved for the *circular mean* but not for the arithmetic mean.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Energetic Arousal</th>
<th>Positive Affect</th>
<th>Tense Arousal</th>
<th>Negative Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>14</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>16</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>18</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>20</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>22</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>24</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Arithmetic Mean</td>
<td>14</td>
<td>19</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Circular Mean</td>
<td>14</td>
<td>19</td>
<td>24</td>
<td>5</td>
</tr>
</tbody>
</table>
Some hypothetical data stored in a data.frame

<table>
<thead>
<tr>
<th>Participant</th>
<th>Name</th>
<th>Gender</th>
<th>$\theta$</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bob</td>
<td>Male</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Debby</td>
<td>Female</td>
<td>3</td>
<td>14</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Alice</td>
<td>Female</td>
<td>7</td>
<td>18</td>
<td>14</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>Gina</td>
<td>Female</td>
<td>6</td>
<td>17</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>Eric</td>
<td>Male</td>
<td>4</td>
<td>15</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>Fred</td>
<td>Male</td>
<td>5</td>
<td>16</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>Chuck</td>
<td>Male</td>
<td>2</td>
<td>13</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

> s.df <- read.clipboard()
> dim(s.df)  #how many elements are in each dimension
[1] 7 7
> str(s.df)  #show the structure
'data.frame': 7 obs. of 7 variables:
$ Participant: int 1 2 3 4 5 6 7
$ Name        : Factor w/ 7 levels "Alice","Bob",..: 2 4 1 7 5 6 3
$ Gender      : Factor w/ 2 levels "Female","Male": 2 1 1 1 2 2 2
$ theta       : int 1 3 7 6 4 5 2
$ X           : int 12 14 18 17 15 16 13
$ Y           : num 2 6 14 12 8 10 4
$ Z           : int 1 4 64 32 8 16 2

g a
Saving the data.frame in a readable form

The previous slide is readable by humans, but harder to read by computer. PDFs are formatted in a rather weird way. We can share data on slides by using the `dput` function. Copy this output to your clipboard from the slide, and then get it into R directly.

```r
> dput(sf.df)

structure(list(ID = 1:7, Name = structure(c(2L, 4L, 1L, 7L, 5L, 6L, 3L), .Label = c("Alice", "Bob", "Chuck", "Debby", "Eric", "Fred", "Gina"), class = "factor"), gender = structure(c(2L, 1L, 1L, 1L, 2L, 2L, 2L), .Label = c("Female", "Male"), class = "factor"),
          theta = c(1L, 3L, 7L, 6L, 4L, 5L, 2L), X = c(12L, 14L, 18L, 17L, 15L, 16L, 13L), Y = c(2L, 6L, 14L, 12L, 8L, 10L, 4L),
          Z = c(1L, 4L, 64L, 32L, 8L, 16L, 2L)), .Names = c("ID", "Name", "gender", "theta", "X", "Y", "Z"), class = "data.frame", row.names = c(NA, -7L))

my.data <- structure(list(ID = 1:7, Name = structure(c(2L, 4L, 1L, 7L, 5L, 6L, 3L), .Label = c("Alice", "Bob", "Chuck", "Debby", "Eric", "Fred", "Gina"), class = "factor"), gender = structure(c(2L, 1L, 1L, 1L, 2L, 2L, 2L), .Label = c("Female", "Male"), class = "factor"),
          theta = c(1L, 3L, 7L, 6L, 4L, 5L, 2L), X = c(12L, 14L, 18L, 17L, 15L, 16L, 13L), Y = c(2L, 6L, 14L, 12L, 8L, 10L, 4L),
          Z = c(1L, 4L, 64L, 32L, 8L, 16L, 2L)), .Names = c("ID", "Name", "gender", "theta", "X", "Y", "Z"), class = "data.frame", row.names = c(NA, -7L))
```
Sorting the data can display certain features

We use the order function applied to the "Names" column and then to the 4th column.

```r
> my.data.alpha <- my.data[order(my.data[,"Name"]),]
> my.data.alpha
   ID Name gender theta  X  Y  Z
 3 3 Alice Female  7  18  14  64
 1 1 Bob Male 1  12  2  1
 7 7 Chuck Male 2  13  4  2
 2 2 Debby Female 3  14  6  4
 5 5 Eric Male 4  15  8  8
 6 6 Fred Male 5  16 10 16
 4 4 Gina Female 6  17 12 32

> my.data.theta <- my.data[order(my.data[,4]),]
> my.data.theta
   ID Name gender theta  X  Y  Z
 1 1 Bob Male 1 12  2  1
 7 7 Chuck Male 1 13  4  2
 2 2 Debby Female 3 14  6  4
 5 5 Eric Male 4 15  8  8
 6 6 Fred Male 5 16 10 16
 4 4 Gina Female 6 17 12 32

It was harder to see the perfect relationship between $\theta$ and X, Y, and Z with the original data.
Multiple estimates of the central tendency using the `apply` function

1. The basic mean is applied to columns 4 - 7
2. Then do this, but trim the top and bottom 20%
3. Now, don’t trim, but winsorize
4. Compare with the harmonic mean
5. Compare with geometric mean.
Effect of reciprocal transformation upon means

**Table:** Hypothetical study of arousal using an exciting movie. The post test shows greater arousal if measured using skin conductance (higher skin conductance means more arousal), but less arousal if measured using skin resistance (higher skin conductance means less arousal)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subject</th>
<th>Skin Conductance</th>
<th>Skin Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest (Control)</td>
<td>1</td>
<td>2</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>.50</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>2</td>
<td>.50</td>
</tr>
<tr>
<td>Posttest (Movie)</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>.25</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>2.5</td>
<td>.61</td>
</tr>
</tbody>
</table>
Non linearity can influence means if the variances differ.
What is the "average" class size?

Table: Average class size depends upon point of view. For the faculty members, the median of 10 is very appealing. From the Dean’s perspective, the faculty members teach an average of 50 students per calls. But what about the students?

<table>
<thead>
<tr>
<th>Faculty Member</th>
<th>Freshman/Sophomore</th>
<th>Junior</th>
<th>Senior</th>
<th>Graduate</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12.5</td>
<td>10</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12.5</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12.5</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>20</td>
<td>100</td>
<td>10</td>
<td>10</td>
<td>35.0</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>200</td>
<td>100</td>
<td>400</td>
<td>10</td>
<td>177.5</td>
<td>150</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.0</td>
<td>39</td>
</tr>
<tr>
<td>Mean</td>
<td>56</td>
<td>46</td>
<td>110</td>
<td>10</td>
<td>50.0</td>
<td>39</td>
</tr>
<tr>
<td>Median</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12.5</td>
<td>10</td>
</tr>
</tbody>
</table>
**Class size from the students’ point of view.**

Table: Class size from the students’ point of view. Most students are in large classes; the median class size is 200 with a mean of 223.

<table>
<thead>
<tr>
<th>Class size</th>
<th>Number of classes</th>
<th>number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>400</td>
<td>1</td>
<td>400</td>
</tr>
</tbody>
</table>
Time in therapy

A psychotherapist is asked what is the average length of time that a patient is in therapy. This seems to be an easy question, for of the 20 patients, 19 have been in therapy for between 6 and 18 months (with a median of 12) and one has just started. Thus, the median client is in therapy for 52 weeks with an average (in weeks) \(\frac{1 \times 1 + 19 \times 52}{20}\) or 49.4.

However, a more careful analysis examines the case load over a year and discovers that indeed, 19 patients have a median time in treatment of 52 weeks, but that each week the therapist is also seeing a new client for just one session. That is, over the year, the therapist sees 52 patients for 1 week and 19 for a median of 52 weeks. Thus, the median client is in therapy for 1 week and the average client is in therapy of \(\frac{52 \times 1 + 19 \times 52}{52 + 19}\) = 14.6 weeks.
Does teaching effect learning?

1. A leading research team in motivational and educational psychology was interested in the effect that different teaching techniques at various colleges and universities have upon their students. They were particularly interested in the effect upon writing performance of attending a very selective university, a less selective university, or a two year junior college.

2. A writing test was given to the entering students at three institutions in the Boston area. After one year, a similar writing test was given again. Although there was some attrition from each sample, the researchers report data only for those who finished one year. The pre and post test scores as well as the change scores were as shown below:
Types of teaching affect student outcomes?

Table: Three types of teaching and their effect on student outcomes

<table>
<thead>
<tr>
<th>School</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior College</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Non-selective university</td>
<td>5</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>Selective university</td>
<td>27</td>
<td>73</td>
<td>45</td>
</tr>
</tbody>
</table>

From these data, the researchers concluded that the quality of teaching at the selective university was much better than that of the less selective university or the junior college and that the students learned a great deal more. They proposed to study the techniques used there in order to apply them to the other institutions.
Teaching and math performance

Another research team in motivational and educational psychology was interested in the effect that different teaching at various colleges and universities affect math performance. They used the same schools as the previous example with the same design.

Table: Three types of teaching and their effect on student outcomes

<table>
<thead>
<tr>
<th>School</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior College</td>
<td>27</td>
<td>73</td>
<td>45</td>
</tr>
<tr>
<td>Non-selective university</td>
<td>73</td>
<td>95</td>
<td>22</td>
</tr>
<tr>
<td>Selective university</td>
<td>95</td>
<td>99</td>
<td>4</td>
</tr>
</tbody>
</table>

They concluded that the teaching at the junior college was far superior to that of the select university. What is wrong with this conclusion?
Effect of teaching, effect of students, or just scaling?

Writing

Math

Performance

Performance

Ivy

TC

JC

Ivy

TC

JC

Scales

Scaling

Levels

Dispersion

References
The effect of scaling upon the latent variable - observed variable relationship
The problem of scaling is ubiquitous

1. A leading cognitive developmentalist believed that there is a critical stage for learning spatial representations using maps. Children younger than this stage are not helped by maps, nor are children older than this stage.

2. He randomly assigned 3rd, 5th, and 7th grade students into two conditions (nested within grade), control, and map use. Performance was measures on a task of spatial recall (children were shown toys at particular locations in a set of rooms and then asked to find them again later.) Half the children were shown a map of the rooms before doing the task.

3. Their scores were

<table>
<thead>
<tr>
<th>Grade</th>
<th>No Map</th>
<th>Maps</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd</td>
<td>5</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>5th</td>
<td>27</td>
<td>73</td>
<td>46</td>
</tr>
<tr>
<td>7th</td>
<td>73</td>
<td>95</td>
<td>22</td>
</tr>
</tbody>
</table>

- 3rd grade: Too young
- 5th grade: Critical period
- 7th grade: Too old
Map use is most effective at a particular developmental stage

Recall varies by age and exposure to maps

Recall varies by age and exposure to maps
R code for the prior figure

```r
mapuse <- matrix(c( 3,5,27,5,27,73,7, 73,95),ncol=3,byrow=TRUE)
colnames(mapuse) <- c("grade","nomaps","maps")
rownames(mapuse) <- c("3rd","5th","7th")
maps.df <- data.frame(mapuse)

with(maps.df,plot(maps~grade,ylab="Recall",ylim=c(0,100),
typ="b", main="Recall varies by age and exposure to maps"))
with(maps.df,points(nomaps~grade,ylab="Recall",
    ylim=c(0,100),typ="b",lty="dashed"))
> text(5,80,"maps")  #add line labels
> text(5,15,"nomaps")

grade nomaps maps
3rd  3  5  27
5th  5 27  73
7th  7 73  95
```
Yet another developmentalist

Another cognitive developmentalist believed that there is a critical stage but that it appears earlier than previously thought. Children younger than this stage are not helped by maps, nor are children older than this stage. He randomly assigned 1st, 3rd, 5th and 7th grade students into two conditions (nested within grade), control and mapa use. Performance was measured on a task of spatial recall (children were shown toys at particular locations in a set of rooms and then asked to find them again later. Half the children were shown a map of the room before doing the task. The scores were

<table>
<thead>
<tr>
<th>Grade</th>
<th>No Map</th>
<th>Maps</th>
<th>Effect</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st grade</td>
<td>2</td>
<td>12</td>
<td>10</td>
<td>Too young</td>
</tr>
<tr>
<td>3rd grade</td>
<td>12</td>
<td>50</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>5th grade</td>
<td>50</td>
<td>88</td>
<td>38</td>
<td>Critical period</td>
</tr>
<tr>
<td>7th grade</td>
<td>88</td>
<td>98</td>
<td>10</td>
<td>Too old</td>
</tr>
</tbody>
</table>
A critical period in developmental?

Recall varies by age and exposure to maps
R code for the prior figure

```R
mapuse <- matrix(c( 1,2,12,10,3,12,50,38,5,50,88,38,7,88,98,10),ncol=4,byrow=TRUE)
colnames(mapuse) <- c("grade","nomaps","maps","Diff")
rownames(mapuse) <- c("1st","3rd","5th","7th")
maps.df <- data.frame(mapuse)
maps.df
with(maps.df,plot(maps~grade,ylab="Recall",ylim=c(0,100),
typ="b", main="Recall varies by age and exposure to maps"))
with(maps.df,points(nomaps~grade,ylab="Recall",
    ylim=c(0,100),typ="b",lty="dashed"))
text(4,75,"maps")  #add line labels
text(4,20,"nomaps")
grange(mapuse[,1:3])
```

<table>
<thead>
<tr>
<th>grade</th>
<th>nomaps</th>
<th>maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>5th</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>7th</td>
<td>7</td>
<td>73</td>
</tr>
</tbody>
</table>
Traditional levels of measurement

Nominal  Categories: X, Y, W, V

Ordinal  Ranks ($X > Y > W > V$)

Interval  Equal Differences ($X - Y > W - V$)

Ratio  Equal intervals with a zero point ($X/Y > W/V$)
Types of scales and types of inference

1. Nominal allow us to say whether groups differ in frequency

2. Ordinal allows to compare rank orders of the data, is one score greater than another score. Any monotonic transformation will preserve rank order.

3. Interval is the claim that we can compare the magnitude of intervals. Only linear transformations will preserve interval information (i.e. we can add and subtract the numbers and preserve interval information. item Ratio scales preserve absolute magnitude differences.
1. Any monotonic transformation will preserve order
2. Inferences from observed to latent variable are restricted to rank orders
3. Statistics: Medians, Quartiles, Percentiles
Interval scales

1. Possible to infer the magnitude of differences between points on the latent variable given differences on the observed variable? $X$ is as much greater than $Y$ as $Z$ is from $W$
2. Linear transformations preserve interval information
3. Allowable statistics: Means, Variances
4. Although our data are actually probably just ordinal, we tend to use interval assumptions.
Ratio Scales

1. Interval scales with a zero point
2. Possible to compare ratios of magnitudes (X is twice as long as Y)
3. Are there any psychological examples?
The search for an appropriate scale

1. Is today colder than yesterday? (ranks) Is the amount that today is colder than yesterday more than the amount that yesterday was colder than the day before? (intervals)
   - $50F - 39F < 68F - 50F$
   - $10C - 4C < 20C - 10C$
   - $283K - 277K < 293K - 283K$

2. How much colder is today than yesterday?
   - (Degree days as measure of energy use) is almost ratio
   - K as measure of molecular energy
Measurement confusions – arousal

1. Arousal is a fundamental concept in many psychological theories. It is thought to reflect basic levels of alertness and preparedness. Typical indices of arousal are measures of the amount of palmer sweating.

2. This may be indexed by the amount of electricity that is conducted by the fingertips.

3. Alternatively, it may be indexed (negatively) by the amount of skin resistance of the finger tips. The Galvanic Skin Response (GSR) reflects moment to moment changes, SC and SR reflect longer term, basal levels.

4. High skin conductance (low skin resistance) is thought to reflect high arousal.
Arousal and anxiety

1. Anxiety is thought to be related to arousal. The following data were collected by two different experimenters. One collected resistance, conductance data.

   low anxiety  1, 5        1, .2
   high anxiety 2, 2        .5, .5

   The means were therefore:

   Resistance, conductance data.

   low anxiety  3        .6
   high anxiety 2        .5,

2. That is, the low anxiety participants had higher skin resistance and thus were more relaxed, but they also had higher skin conductance, and thus were more aroused.

3. How can this be?
Multiple measures of dispersion

Range  (highest - lowest) is sensitive to the number of observations, but is a very good way to detect errors in data entry.

MAD (Median Absolute Deviation from the Median) applied ordinal statistics to interval measures

Variance  \((\sigma^2)\) is the Mean Square deviation (implies interval data)

Standard Deviation  \((\sigma)\) is the Root Mean Square deviation.

Coefficient of Variation  \(\frac{\sigma_x}{\mu_x}\)

Average difference  \(\sigma_x \sqrt{2}\)
Normal and non-normal curves

Normal and non-normal

- N(0,1)
- Cauchy
- logistic
- N(0,2)
Three normal curves
Seriously contaminated data

Normal and contaminated data

- Normal(0,1)
- Contaminated Normal(.3,3.3)

Probability of X

X
The normal curve and its frequent transforms

Alternative scalings of the normal curve

- Probability of z
- IQ
- SAT
- stanine
- Galton/Tukey
- lower fence
- Q1
- Median
- Q3
- Upper Fence
Decision making and the benefit of extreme selection ratios

1. Typical traits are approximated by a normal distribution.
2. Small differences in means or variances can lead to large differences in relative odds at the tails.
3. Accuracy of decision/prediction is higher for extreme values.
4. Do we infer trait mean differences from observing differences of extreme values?
The effect of small mean differences at the tails of a distribution

difference = .25

difference = .5

Probability x1 > x
The effect of small differences in variance at the tails of a distribution

- **Sigma = 1.1**

- **Sigma = 1.2**
**Table:** Tukey’s ladder of transformations. One goes up and down the ladder until the relationships desired are roughly linear or the distribution is less skewed. The effect of taking powers of the numbers is to emphasize the larger numbers, the effect of taking roots, logs, or reciprocals is to emphasize the smaller numbers.

<table>
<thead>
<tr>
<th>Transformation</th>
<th>effect</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^3$</td>
<td>emphasize large numbers</td>
<td>reduce negative skew</td>
</tr>
<tr>
<td>$x^2$</td>
<td>emphasize large numbers</td>
<td>reduce negative skew</td>
</tr>
<tr>
<td>$\sqrt{x}$</td>
<td>emphasize smaller numbers</td>
<td>reduce positive skew</td>
</tr>
<tr>
<td>$-1/x$</td>
<td>emphasize smaller numbers</td>
<td>reduce positive skew</td>
</tr>
<tr>
<td>$\log(x)$</td>
<td>emphasize smaller numbers</td>
<td>reduce positive skew</td>
</tr>
<tr>
<td>$-1/x^2$</td>
<td>emphasize smaller numbers</td>
<td>reduce positive skew</td>
</tr>
<tr>
<td>$-1/x^3$</td>
<td>emphasize smaller numbers</td>
<td>reduce positive skew</td>
</tr>
</tbody>
</table>
Figure: Tukey (1977) suggested a number of transformations of data that allow relationships to be seen more easily. Ranging from the cube to the reciprocal of the cube, these transformations emphasize different parts of the distribution.
The best scale is the one that works best

1. Money is linear but negatively accelerated with utility.
2. Perceived intensity is a log function of physical intensity.
3. Probability of being correct is a logistic or cumulative normal function of ability.
4. Energy used to heat a house is linear function of outdoor temperature.
5. Time to fall a particular distance varies as the square root of the distance \( s = at^2 \iff t = \sqrt{\frac{s}{a}} \)
6. Gravitational attraction varies as \( 1/distance^2 \) \( F = G \frac{m_1 m_2}{d^2} \)
7. Hull speed of sailboat varies as square root of length of boat.
8. Sound intensity in db is \( \log(\text{observed}/\text{reference}) \)
9. pH of solutions is \( -\log(\text{concentration of hydrogen ions}) \)
